cogitatio

SOCIAL INCLUSION

Fostering the Socially and Ecologically Sustainable Digitalisation of Welfare States

Edited by Paula Saikkonen and Marta Choroszewicz

Volume 13

2025

Open Access Journal

ISSN: 2183-2803

Social Inclusion, 2025, Volume 13
Fostering the Socially and Ecologically Sustainable Digitalisation of Welfare States

Published by Cogitatio Press
Rua Fialho de Almeida 14, 2º Esq.,
1070–129 Lisbon
Portugal

Design by Typografia®
http://www.typografia.pt/en/

Cover image: © Peach_iStock from iStock

Academic Editors
Paula Saikkonen (Finnish Institute for Health and Welfare)
Marta Choroszewicz (University of Eastern Finland)

Available online at: www.cogitatiopress.com/socialinclusion

Table of Contents

Social Inclusion
2025 • Volume 13 • Article 10937
https://doi.org/10.17645/si.10937

cogitatio

EDITORIAL

Open Access Journal

Fostering Socially and Ecologically Sustainable Digitalisation of Welfare States

Paula Saikkonen [1] **and Marta Choroszewicz** [2]

[1] Finnish Institute for Health and Welfare (THL), Finland
[2] University of Eastern Finland, Finland

Correspondence: Paula Saikkonen (paula.saikkonen@thl.fi)

Submitted: 11 July 2025 **Published:** 2 September 2025

Issue: This editorial is part of the issue "Fostering the Socially and Ecologically Sustainable Digitalisation of Welfare States" edited by Paula Saikkonen (Finnish Institute for Health and Welfare) and Marta Choroszewicz (University of Eastern Finland), fully open access at https://doi.org/10.17645/si.i514

Abstract

This thematic issue discusses how AI and digital technologies often overlook vulnerable citizens and reinforce inequalities. Articles highlight challenges like digital exclusion, excessive reliance on individual agency, and declining institutional trust. Socially sustainable digitalisation must ensure equitable access, offer in-person alternatives when needed, and uphold legitimacy. Often neglected ecological concerns must also be addressed through responsible data use and energy-conscious ICT systems. Achieving sustainability requires coherent policies that embed sustainability principles into welfare technology. Interdisciplinary research, citizen involvement, and empowering citizens are essential for a truly sustainable welfare state.

Keywords

artificial intelligence; digitalisation; social sustainability; trust; welfare system

1. Introduction

Digitalisation of societies and welfare systems is often touted as a driver of increased efficiency and service quality, enabling flexibility for service users and a possibility to save on costs. Big data, data analysis tools, and artificial intelligence (AI) are argued to bring opportunities for managers and decision-makers to lead better with knowledge derived from so-called real-time data. Yet, in practice, data pose numerous challenges to interpretation and simultaneous utilisation for multiple purposes (e.g., Hoeyer, 2023). Numerous studies have highlighted a gap between the expectations and the actual outcomes of new technologies. In welfare states, where services are often shaped by strict budgetary constraints, it is essential that reforms lead to genuine improvements.

The starting point for this thematic issue was our observation that sustainability and digitalisation are frequently addressed as separate concerns, digitalisation often being regarded primarily as a technological matter that automatically enhances access to and efficiency in service delivery. Only a limited number of reports have emphasised the ecologically unsustainable aspects of digitalisation or AI. Social sustainability has been predominantly acknowledged within the discussions on "decent work," the platform economy, datafication, and surveillance, rather than focusing on the social sustainability of welfare systems. Digitalisation, datafication and the effects of AI should also be examined as part of the development of societies. New technologies are shaping the conditions for well-being for current and future generations by influencing both social and ecological sustainability. With this thematic issue, we draw attention to the ways in which digitalisation either supports or undermines social sustainability, as policies often overlook its broader implications for social justice.

2. Social Sustainability at the Core of the Welfare System

2.1. Problematising the Current Expectation of Citizens' Strong Agency

The digitalisation and datafication of the welfare systems do not happen in a vacuum. The existing systems and their earlier development have an impact on these processes, political decision-making steers the aims of digitalisation (e.g., cost-efficiency, public or private system supplier, data collection, use of data), and citizens' digital literacy—with other skills and devices—influence their opportunities to utilise digital welfare systems. The welfare systems interact with citizens, especially in changing life circumstances like sickness, parenthood, unemployment, retirement, or rehabilitation. Interaction formulates citizens' perceptions of authorities and public policies may influence people's subjective beliefs on the welfare state and their role in it. For instance, are citizens only customers receiving services or active citizens participating in society?

Digitalised welfare systems, with their novel technologies, reformulate the relationships between citizens and institutions by strengthening social inclusion for some citizens and amplifying old—or creating new—disadvantages for others (Choroszewicz & Mäihäniemi, 2020). For social justice, it is essential that all citizens have access to necessary benefits and services, regardless of their varying skill levels or ability to use digital services (Saikkonen & Ylikännö, 2020). Previous studies have shown the pivotal role of public policies for participation, especially for those who have fewer resources (Shore, 2020). However, when reading policy documents about e-government or promises of AI in the public sector, it is striking how little attention is paid to those who are unable to use digital services.

The topic of social justice was highlighted in several articles in this issue. The articles address both the barriers to digital inclusion experienced by vulnerable groups and the strong normative expectation that citizens must actively exercise their agency in order to secure access to services. Dyrlev (2025) shows that the use and experience of mental health digital solutions (MHDS) in Denmark are socially stratified. Digital and social inequalities intersect, suggesting that MHDS should complement, not replace, in-person mental health services to ensure social sustainability.

Tarkkala et al. (2025) discuss the trend towards digital-first administration. They investigate the interaction between citizens and the state through the analysis of digital pre-screening tests as a new form of "screen-level bureaucracy." Their research highlights the ambivalence of these tests: while these tests can be

useful as communication channels and advice-giving resources, they can also create new barriers to accessing public services.

Mielismäki and Husso (2025) contribute to the ongoing discussion on the challenges, opportunities, and ethical considerations of the integration of AI-driven chatbots into domestic violence support systems. Their study highlights the issue of responsibility in the implementation and use of these chatbots. They also show that while these chatbots have the potential to lower barriers to help-seeking, their limitations in situational assessment may put extensive demand for agency on victims or survivors.

Rebergen et al. (2025) highlight a paradox in digitalized welfare systems: While they work well for well-off citizens with simple needs, those facing complex life situations often struggle. In the Netherlands, libraries provided IT support, but librarians lacked training in welfare systems. This gap frustrated citizens seeking substantive help and placed emotional strain on librarians who felt unable to meet those needs.

Smit et al. (2025) explore digital inclusion and exclusion among low-literate Dutch citizens within a regime of self-optimisation that frames digital participation as both a moral duty and a practical necessity. They show that disadvantaged individuals often rely on trusted family members or neighbors to navigate digital services. Their study emphasises that sustainable digital inclusion requires not only promoting individual self-reliance but also actively incorporating collective learning and supportive community practices.

Palukka et al. (2025) focus on the experiences of mental health rehabilitees in digital encounters. Their study shows how administrative burdens and their costs produce distrust in public administration and socially exclusive identities for marginalised citizens, especially when claiming welfare benefits, such as positions of feeling dispossessed, unreliable, insignificant, and inferior. The study highlights that this positioning of mental health rehabilitees prevents full realization of their citizenship.

2.2. Trust and Equality of the System

The perceived fairness in the service processes is extremely important for trust in the welfare systems, often even more important than realised outcomes (Van Ryzin, 2011). In the Nordic welfare states, institutional and generalised trust have generally been at high levels. Yet, this is not always the case among vulnerable citizens (Palukka et al., 2025; Tetri et al., 2024), which is why more attention should be paid to how they perceive the welfare system. Distrust in welfare systems hampers their legitimacy. The topic of trust was addressed in some articles within this issue. These contributions deepen our understanding of why maintaining and strengthening trust in welfare systems is crucial, especially when implementing new digital tools, services, and AI systems.

Petroons et al. (2025) examine whether algorithmic decision-making (ADM) contributes to harmonized welfare decisions across local agencies in Belgium. Focusing on the REDI system (a digital rule-based algorithmic system designed to assess families' financial needs), findings show that while ADM promotes uniformity in the form and amount of monthly financial support, significant local variation remains. The results point to only partial harmonization, shaped by financial and normative factors at the organizational level.

Rannisto and Vainionpää (2025) discuss in their article what it would take to maintain the high level of trust in the Nordic welfare state when algorithms and AI are part of the welfare systems. They elaborate on the

dichotomy between technological and social studies in trust research and suggest an approach to overcome that.

Palsa et al. (2025) investigate the future imaginaries of datafied education to identify the central features of the Finnish education system, such as the culture of trust and the strong role of the public sector. They discuss how these features might be subject to negotiations in the datafication processes. Their research underscores the importance of inclusive debates on the future of education to ensure that digitalization supports sustainable development in education.

3. The Missing Aspects of Sustainability

By editing this thematic issue, we have gained a better understanding that ecological and social sustainability are still insufficiently integrated into the digitalisation of welfare systems. For social sustainability, it would be important to better understand the relationship between online and on-site services and how they could be combined fruitfully. The welfare systems deliver benefits and services, but they also impact citizens' perceptions of public policies. Furthermore, the welfare systems have a direct impact on ecological sustainability (e.g., energy efficiency, the ecological footprint of the ICT system, GHG emissions) which cannot be ignored.

Ecological and social sustainability should be acknowledged when planning and purchasing the new ICT systems for welfare services. To strengthen sustainability, greater policy coherence is needed. In the digitalisation of welfare systems, this coherence can be achieved through discussions on the foundational principles of new ICT systems' development (see also Rannisto & Vainionpää, 2025). The socially sustainable digitalization requires the involvement of all stakeholders (e.g., citizens, frontline workers, managers, decision-makers). Practices should be planned with careful consideration of where and how digitalisation, semi- and automatic decision-making, or AI bring betterments to all groups of citizens and their well-being (Petroons et al., 2025).

During the joint journey of editing this issue, many activities took place. Apart from having launched a broad call for papers to bring this thematic issue to life, we also organised several panels and workshops on the topic at European scientific conferences. It was fascinating to see the variety of articles and presentations that emerged in response to our call. As readers may notice, the range of methods and topics is extensive in the issue. Yet, despite the broad call, we received only a few articles discussing ecological sustainability in relation to digitalisation, datafication, or AI in the welfare systems. This may reflect our own background as social scientists, but it might also indicate a concerning lack of multidisciplinary research on this topic.

Researchers have outlined some characteristics for sustainable welfare systems (see Bridgen & Saikkonen, 2025), which can be applied to the digitalised welfare system. Firstly, the nine planetary boundaries formulated by the Stockholm Resilience Centre (Richardson et al., 2023) should be considered in the context of digitalised welfare systems. Digitalisation, datafication, and the use of AI consume significant amounts of energy. While energy efficiency and the sources of energy are crucial considerations, it is important to recognise that greenhouse gas emissions and energy consumption also occur in traditional face-to-face service delivery.

Secondly, the welfare systems should become less dependent on economic growth, particularly given concerns about the feasibility of rapidly decoupling economic expansion from greenhouse gas emissions (Büchs, 2021). In digitalised welfare systems, this necessitates careful consideration of how data is used and what kinds of data are produced in the first place. Increasing datafication may also result in unexpected consequences. For instance, in the education system, it may reduce teachers' actual educational leeway and hinder students' opportunities for social mobility (Palsa et al., 2025). Large administrative datasets are also of significant interest to commercial actors, but the implications vary greatly depending on whether data is harnessed for profit-driven purposes or for research aimed at addressing sustainability challenges.

Thirdly, given material and financial constraints, welfare systems should focus more strongly on need-based provisions. In current systems, "need" is often poorly operationalised (Laruffa, 2022). Post-war welfare states have largely relied on expert-defined needs, primarily to determine eligibility and legitimise professionalised services. In digitalised welfare systems, it becomes essential to recognise when face-to-face interaction is necessary for individual assessments (see Dyrlev, 2025). Digital tools and systems can be used for pre-screening in ways that risk excluding individuals (Mielismäki & Husso, 2025; Palukka et al., 2025; Smit et al., 2025; Tarkkala et al., 2025), or they can complement service delivery, such as automated pharmacy vouchers in social assistance or capped out-of-pocket health expenses, as recommended by the WHO.

Fourthly, one characteristic is fair distribution (Büchs, 2021). In digitalised welfare systems, this means equal access and organizing supportive services in a way that adequate services are available for everyone independently from digital skills or devices in use. Furthermore, while access to services and benefits should be easy (digitalised or not), the easy access or automated decision-making should not blur political decisions behind the systems (e.g., eligibility, adequacy of benefits). After all, the welfare system is steered by the policymakers whom citizens have elected.

Fifthly, welfare systems should play an active role in sustainability transitions (Saikkonen & Ilmakunnas, 2023). They can provide support and incentives (Bohnenberger, 2020), for example, by encouraging lifestyle changes and cultural shifts aligned with less resource-intensive, post-consumerist notions of well-being. In digitalised welfare systems, this could involve using broader datasets and AI to improve the relevance of job offers provided by employment services. The system could also generate forward-looking suggestions, such as training or work opportunities with sustainability potential, rather than focusing solely on the current labour market and available vacancies.

Sixthly, citizen involvement is essential. Attention has been directed both to macro-level approaches concerning democratic governance (Büchs, 2021) and to micro-level aspects of policy design (Bohnenberger, 2020), particularly how recipients can influence the support they receive. In both cases, the key concern is to avoid paternalism and to uphold legitimacy (Büchs, 2021). In the context of digitalised welfare systems, this means designing systems that respond to citizens' capabilities and needs, rather than being driven solely by what is available on the market.

We argue that neglecting ecological and social sustainability is short-sighted, particularly given the urgent need for welfare states to tackle environmental crises. Although research on digital welfare systems is growing, it does not appear to accumulate in ways that effectively support the sustainable digitalisation of welfare systems. Academic research often focuses on narrowly defined problems, raising the question of whether we

have adequate tools to synthesise research findings in a way that informs sound political decision-making that would facilitate the transition of welfare states toward ecological sustainability.

Funding

The writing of this editorial, as well as the editing of the thematic issue, were done as part of the projects SustAgeable (Finnish research Council nro 364493) and Perpetual Piloting and Invisible Work of Automating Public Services in Finland (Finnish Cultural Foundation, Teresia ja Rafael Lönnström Fund).

Conflict of Interests

The authors declare no conflict of interests.

Data Availability

Due to the nature of the research, data sharing is not applicable to this article.

References

Bohnenberger, K. (2020). Money, vouchers, public infrastructures? A framework for sustainable welfare benefits. *Sustainability*, *12*(2), Article 596. https://doi.org/10.3390/su12020596

Bridgen, P., & Saikkonen, P. (2025). Integrated sustainability: What is it and what are its implications for social care in ageing welfare states? In M. Vaalavuo, K. Nelson, & K. Kuitto (Eds.), *Social sustainability in ageing welfare states* (pp.229–250). Edward Elgar Publishing.

Büchs, M. (2021). Sustainable welfare: How do universal basic income and universal basic services compare? *Ecological Economics*, *189*, Article 107152. https://doi.org/10.1016/j.ecolecon.2021.107152

Choroszewicz, M., & Mäihäniemi, B. (2020). Developing a digital welfare state: Data protection legislation and the use of automated decision-making across six EU countries. *Global Perspectives*, *1*(1), Article 12910. https://doi.org/10.1525/gp.2020.12910

Dyrlev, E. K. (2025). Potentials and pitfalls of self-help tools: A survey study of digital psychiatry in Denmark. *Social Inclusion*, *13*, Article 9990. https://doi.org/10.17645/si.9990

Hoeyer, K. (2023). *Data paradoxes: The politics of intensified data sourcing in contemporary healthcare*. MIT Press.

Laruffa, F. (2022). The dilemma of "sustainable welfare" and the problem of the future in capacitating social policy. *Sustainability: Science, Practice and Policy*, *18*(1), 822–836. https://doi.org/10.1080/15487733.2022.2143206

Mielismäki, K. H., & Husso, M. (2025). Ethical implications of AI-driven chatbots in domestic violence support. *Social Inclusion*, *13*, Article 9998. https://doi.org/10.17645/si.9998

Palsa, L., Fagerlund, J., & Mertala, P. (2025). Exploring the futures of datafied welfare state education: Thematic analysis of sociotechnical imaginaries. *Social Inclusion*, *13*, Article 10087. https://doi.org/10.17645/si.10087

Palukka, H., Koski, A., Parviainen, J., & Eilola, L. (2025). Experiencing social exclusion and distrust: Mental health rehabilitees struggling with digital administrative burdens. *Social Inclusion*, *13*, Article 9950. https://doi.org/10.17645/si.9950

Petroons, J., Brotcorne, P., Wagener, M., Hermans, K., & Van Lancker, W. (2025). Algorithmic decision-making and harmonization in multi-level governance welfare practices: Empirical evidence from Belgium. *Social Inclusion*, *13*, Article. https://doi.org/10.17645/si.10289

Rannisto, A., & Vainionpää, F. (2025). Trust-affording action: Citizens' everyday relations with algorithmized public services. *Social Inclusion*, 13, Article 10001. https://doi.org/10.17645/si.10001

Rebergen, M., Swart, J., & Broesma, M. (2025). Directing digital citizenship: How librarians mediate the Dutch digital welfare state. *Social Inclusion*, *13*, Article 9949. https://doi.org/10.17645/si.9949

Richardson, K., Steffen, W., Bendtsen, J., Cornell, S. E., Donges, J. F., Drüke, M., Fetzer, I., Bala, G., Von Bloch, W., Feulner, G., Fiedler, S., Gerten, D., Gleeson, T., Hoffmann, M., Kummu, M., Mohan, C., Nogués-Bravo, D., Petri, S., . . . Rockström, J. (2023). Earth beyond six of nine planetary boundaries. *Science Advance*, *9*(37). https://doi.org/10.1126/sciadv.adh2458

Saikkonen, P., & Ilmakunnas, I. (2023). Reconciling welfare policy and sustainability transition—A case study of the Finnish welfare state. *Environmental Policy and Governance*, *34*(1), 53–64. https://doi.org/10.1002/eet.2055

Saikkonen, P., & Ylikännö, M. (2020). Is there room for targeting within universalism? Finnish social assistance recipients as social citizens. *Social Inclusion*, *8*(1), 145–154. https://doi.org/10.17645/si.v8i1.2521

Shore, J. (2020). How social policy impacts inequalities in political efficacy. *Sociology Compass*, *14*(5). https://doi.org/10.1111/soc4.12784

Smit, A., Swart, J., & Broersma, M. (2025). The regime of self-optimization: Lived experiences of enforced digital inclusion by low-literate citizens. *Social Inclusion*, *13*, Article 10052. https://www.cogitatiopress.com/socialinclusion/article/view/10052

Tarkkala, H., Koulu, R., Snell, K., & Soininvaara, H. (2025). "Complete the test first:" Prescreening tests at the margins of digital public administration. *Social Inclusion*, *13*. https://doi.org/10.17645/si.9971

Tetri, B., Rantanen, T., & Kouvonen, A. (2024). Digital skills and intention to use digital health care and social welfare services among socially marginalized individuals in Finland: A cross-sectional study. *Finnish Journal of EHealth and EWelfare*, *16*(2), 117–130. https://doi.org/10.23996/fjhw.143006

Van Ryzin, G. G. (2011). Outcomes, process, and trust of civil servants. *Journal of Public Administration Research and Theory*, *21*(4), 745–760. https://doi.org/10.1093/jopart/muq092

About the Authors

Paula Saikkonen works as a research manager at the Finnish Institute for Health and Welfare (THL). She is specialised in research on social welfare, well-being, and eco-social welfare state. Her current interests are related to social security, sustainable welfare, ecological and social sustainability, and welfare policies in sustainability transformation.

Marta Choroszewicz is a senior researcher in sociology at the University of Eastern Finland, Finland. Her research includes professional work and careers, social and digital inequalities, AI automation of public administration and services in Finland, and social sustainability in the context of digitalisation and automation of public administration and services.

Social Inclusion
2025 • Volume 13 • Article 9990
https://doi.org/10.17645/si.9990

cogitatio

ARTICLE

Open Access Journal

Potentials and Pitfalls of Self-Help Tools: A Survey Study of Digital Psychiatry in Denmark

Emilie Kristine Dyrlev

Department of Political Science, Aarhus University, Denmark

Correspondence: Emilie Kristine Dyrlev (emdy@ps.au.dk)

Submitted: 30 January 2025 **Accepted:** 12 May 2025 **Published:** 28 July 2025

Issue: This article is part of the issue "Fostering the Socially and Ecologically Sustainable Digitalisation of Welfare States" edited by Paula Saikkonen (Finnish Institute for Health and Welfare) and Marta Choroszewicz (University of Eastern Finland), fully open access at https://doi.org/10.17645/si.i514

Abstract

Welfare systems around the world are struggling to cope with the growing number of people needing psychiatric care. Consequently, digitalization has become a beacon of hope, making treatment more accessible and bolstering patient empowerment. However, scholars have shown that digital divides prevail. This study examines the social sustainability of digital psychiatry by illuminating patient perspectives on mental health digital solutions (MHDS) in Denmark. This is done via a unique survey sample from register data of 1,478 adults in psychiatric treatment in 2023. First, the study examines the association between seven predictors—socioeconomic position, severity of mental illness, age, gender, geographic location, migrant status, and social support—and MHDS usage via binary logistic regression analysis. The analysis reveals a social stratification behind the usage of MHDS. Second, the study conducts two latent class analyses—one for MHDS users and one for non-users—to identify underlying groups that characterize patient perspectives on MHDS. For the users, the analysis reveals latent classes characterized by experiences of participation as well as isolation. For the non-users, the analysis highlights latent classes characterized by few barriers to using MHDS as well as by multiple barriers related to the limited affordances of MHDS. Lastly, the study uses multinomial logistic regression analysis to examine the association between the predictors and the latent classes, showing that latent class membership has a social component. Taken together, the findings indicate that social and digital inequalities are intertwined. To become socially sustainable, digital initiatives should complement, and not replace, in-person treatment.

Keywords

digital divide; digital inclusion; digitalization; empowerment; psychiatry; participation; welfare state

1. Introduction

Mental health has been declared a pressing and ubiquitous issue worldwide. One key solution advocated by the World Health Organization (2021), as well as national and local services, is the use of mental health digital solutions (MHDS). These solutions are supposed to create more cost-efficient, accessible, and personalized treatment for psychiatric patients while also reducing pressure on welfare professionals. We have thus seen a proliferation of MHDS in recent years.

On the one hand, MHDS can be considered a lever for patient empowerment, because they allow patients to access information about their treatment, monitor their symptoms, and communicate with welfare professionals across time and space (Berrouiguet et al., 2018; Carpenter-Song et al., 2022; Cheng et al., 2021; Vitger et al., 2021). On the other hand, existing research shows that digital divides and inequalities prevail (Buchert et al., 2022; Eubanks, 2007; Reisdorf & Rhinesmith, 2020; Schradie, 2020), and that individuals with mental illness are more likely to have limited access, skills, and desire when it comes to using technology (Borzekowski et al., 2009; Comber et al., 1997; Dobransky & Hargittai, 2016; Ennis et al., 2012; Li & Kirkup, 2007; Tobitt & Percival, 2019; Wang et al., 2011; Wong et al., 2020). Scholars suggest that MHDS can exclude patients who do not have the means to use technology or need intensive psychiatric care (Greer et al., 2019; Robotham et al., 2016). Consequently, MHDS can be seen as a double-edged sword, carving channels for both social inclusion and exclusion.

Seen through a critical lens, MHDS are the result of a neoliberalization of welfare, where participation and responsibility are two sides of the same coin. In this way, MHDS evoke self-help and rest on the notion of an actively participating patient (Lupton, 2013, 2014). This study examines how this notion of participation can be empowering for some groups of patients while potentially excluding others. Thus, the study is governed by a two-fold research question: First, what characterizes psychiatric patient groups in terms of their experiences with and attitudes toward MHDS? Second, how are these experiences and attitudes related to social inequalities?

The first part of the research question is examined via two latent class analyses (LCAs). One of the LCAs uncovers patient groups in terms of experiences with using MHDS, while the other LCA sheds light on patient groups in terms of reasons for not using MHDS. Empirically, the LCAs are based on a large survey among psychiatric patients in Denmark investigating patient experiences with and attitudes towards MHDS. The survey is based on access to unique Danish register data, providing a representative gross survey sample. This access to register data is a remarkable strength of the study, bolstering the external validity of the findings.

The second part of the research question is investigated via regression analyses. As a preliminary analysis, the article examines the association between seven predictors—socioeconomic position, severity of mental illness, age, gender, geographic location, migrant status, and social support—and usage of MHDS. This provides insights into the differentiating mechanisms behind MHDS usage. Subsequently, the article conducts two multinomial logistic regression analyses examining whether the predictors are associated with latent classes from the LCAs and thus create unequal conditions in experiences and attitudes.

In contrast to much of the existing literature in clinical psychology and psychiatry (see e.g., Andersson et al., 2019; Barak et al., 2008; Batra et al., 2017; García-Lizana & Muñoz-Mayorga, 2010; Hayes et al., 2016; Hubley

et al., 2016; Yellowlees et al., 2021), this study does not focus on the treatment effects of MHDS. Instead, it complements the existing literature by providing a sociological perspective on the ambiguity of participation and responsibilization that MHDS instigate.

2. Scholarship on Empowerment

This study uses empowerment as an overall analytical perspective. Empowerment plays a central role in welfare literature and has a multifaceted and ambiguous meaning (Frank & Bjerge, 2011). Some seminal definitions suggest that empowerment is related to feeling competent and willing to take action (Zimmerman & Rappaport, 1988). Similarly, Arnstein's (2019) ladder of participation depicts empowerment as an ascending journey from dependence and coercion to agency and autonomy.

Scholars have examined how the call for user empowerment and participation alters the conventional hierarchy between welfare professionals and welfare users. It is suggested that the call for user participation promotes user expertise and challenges professional authority (Järvinen & Kessing, 2021; Schneider-Kamp & Askegaard, 2020). Welfare professionals are no longer called to prescribe the course of action, but should instead guide welfare users in self-management (Greenhalgh, 2009). This proves a difficult balance for welfare professionals: On the one hand, they must endorse the individual autonomy of welfare users; on the other hand, they must provide support that is consistent with auditing goals (Järvinen, 2016).

The multifaceted nature of empowerment means that it can become a vessel for different, sometimes even contradictory, political agendas (Andersen & Elm Larsen, 2011; Mayo et al., 2004). In this way, empowerment is a normative concept. Some critical scholars suggest that empowerment understood as self-help is a neoliberal way of individualizing social problems (Bjerge, 2007; Bjerge & Nielsen, 2014; Oute et al., 2015). Paradoxically, empowerment can then lead to exclusion. This happens when individuals who fail to live up to the imperative of helping themselves are excluded from welfare services because their problems are considered self-inflicted (Howell & Voronka, 2012; Mik-Meyer, 2017; Ringø et al., 2017). In the area of psychiatry, scholars criticize the prevailing recovery model for responsibilizing psychiatric users to get better and obey prescriptive standards of normality (Jørgensen et al., 2020; Oute & Ringer, 2014; Rizq & Jackson, 2019; Speed, 2011).

At the same time, empowerment can be seen as a way of addressing social injustice by giving voice to the users of psychiatric services. As highlighted by prominent critic of the psy-disciplines Nikolas Rose (N. Rose, 2018), psychiatry allows coercion and force if psychiatric patients are deemed at risk of harming themselves or others. Historically, mental illness has been used to denounce and silence the voices of patients. To counter this "epistemic injustice," Diana Rose and Nikolas Rose argue that we must attribute intrinsic value to the lived experiences and embodied expertise of psychiatric patients (D. Rose & Rose, 2023, p. 46).

"Nothing about us without us" is the slogan of mental health and disability empowerment activism (Charlton, 1998, p. 4). This call for participation and involvement can be seen through at least two lenses. On the one hand, calling for participation may be no different from calling for self-help, neglecting the power relations and coercion that hinder participation. On the other hand, participation can be a lever for freedom, autonomy, and agency. This study examines the heterogeneous effects of empowerment initiatives like MHDS across different patient groups. In this way, it seeks to provide a nuanced perspective on empowerment and MHDS, highlighting impediments to socially sustainable psychiatry.

3. Psychiatric Welfare Services in Denmark

Denmark has a universal welfare system that provides free access to social and health care services. However, the Danish welfare system is increasingly marked by inequalities related to the liberalization of welfare services. For instance, there is a rapid growth in supplementary private healthcare insurance, constituting a reduction in the universality of healthcare services (Greve, 2020). Individuals with private insurance have faster and better access to healthcare compared to individuals without. This access inequality is particularly salient in terms of psychological counseling (Greve, 2020, p. 140). Individuals without private insurance must either pay a substantial amount of money for psychological counseling or present with an acute, severe condition that qualifies for public psychiatric treatment. Even though public psychiatric services in Denmark solely treat severe mental illnesses, they nevertheless suffer from long waiting lists and scarce resources.

Scholars have shown a growing discrepancy in life expectancy across different socioeconomic and demographic groups (Brønnum-Hansen & Baadsgaard, 2012) and a corresponding inequality in terms of public health outcomes (Dybbroe, 2020). This is particularly salient in terms of psychiatric patients, who are statistically considered to be some of the most vulnerable groups in society. Studies have shown that psychiatric patients have excess mortality (Kugathasan et al., 2019; Nordentoft et al., 2013; Plana-Ripoll et al., 2020), possibly related to the increased risk of developing further diseases (Ribe et al., 2014) and the risk of worse healthcare treatment than the average population (Davydow et al., 2015, 2016; Graversen et al., 2021). In other words, Denmark's *de jure* equal access to healthcare is far from *de facto* equal access, and the association between digitalization and healthcare inequalities examined in this study is therefore relevant for other countries.

Denmark proves an interesting and relevant case for examining inequalities in relation to digitalization because the Danish healthcare sector is highly digitalized in general (Schmidt et al., 2019). The Danish psychiatric sector is currently undergoing an expansive reform as a response to the growing need for psychiatric care. This includes budget increases and changes to the way the sector is organized. Here, MHDS are an integral part of the reform, suggesting that MHDS will become a pivotal part of Danish psychiatry in the future (Indenrigs- og Sundhedsministeriet, 2023). Thus, it is salient to examine the potentials and pitfalls of MHDS that are expanding in Denmark in a time of growing global interest in digitalized welfare services.

4. Empirical Material and Data Collection

4.1. Questionnaire

The study was approved by the Institutional Review Board at Aarhus University (approval code: BSS-2023-084). It was preregistered before data collection on October 12, 2023, using OSF registers (https://osf.io/a8xyu/?view_only=e3b4ef242f8b4ce59bf4483c981a3401). Please note that there have been minor, non-substantive alterations. The concepts of responsibilization and self-technologies are not directly used as dependent variables in the LCA. The operationalizations of the predictors severity of mental illness and geographical location are also altered slightly, as shown in Supplementary File 1. In the test for representativeness, the study solely uses χ^2 goodness of fit tests and follows the conventional $p < 0.05$. Lastly, the study does not include an interaction term between socioeconomic position and severity of mental illness because of multicollinearity.

The study is based on a survey among patients who in 2023 were in either inpatient or outpatient psychiatric treatment in Denmark. The questionnaire inquires about the overall use of technology and perspectives on MHDS. Then, the questionnaire asks whether the participants have used MHDS. Those who have used MHDS receive questions about their experience. Those who have not used them are asked why not and about motivational factors. The full questionnaire is available in Supplementary File 2.

Some of the questionnaire items are inspired by international survey studies with a similar target group (Bonet et al., 2018; National Alliance on Mental Illness, 2014) and were translated into Danish by the author. The remaining questionnaire items are designed specifically for this study. The questionnaire has been validated by professionals at the Center for Digital Psychiatry in the Region of Southern Denmark, who are experts in communication with psychiatric patients. They provided crucial feedback on sentence structure, word choice, and other linguistic aspects. As another way of ensuring measurement validity, the study began with a pilot testing phase. The author and a research assistant conducted cognitive interviews with nine individuals who recently had been or were currently in psychiatric treatment. Since the questionnaire varies depending on whether the respondents have used MHDS, the questionnaire was tested on both groups. Participants listened to each survey item read out loud and were then asked to "think out loud," reasoning their response. The questionnaire was revised after the pilot phase, correcting ambiguous or misleading formulations.

4.2. Population and Sample

Based on simple random sampling selection, the survey was distributed to 7,000 patients. The gross sample came from the Danish Health Data Authorities. As an inclusion criterion, participants were required to have had at least one contact with Danish psychiatry in the period of January 1, 2023–November 11, 2023, with a psychiatric working diagnosis according to the 2019 version of the ICD-10 diagnosis classification manual (World Health Organization, 2019). Participants were also required to be 18 years or older. Excluded from the sample were participants with diagnoses classified as mental retardation, F71*–F73* according to the 2019 version of the ICD-10 diagnosis classification manual (World Health Organization, 2019), as well as legally incompetent participants, individuals living outside of Denmark at the data extraction time, and individuals under name and address protection.

Uniquely, the study has access to information about the whole population of psychiatry users. This register data from the Danish Health Data Authorities enables us to render the gross sample representative of the population of psychiatry users in Denmark. Additionally, the study employs register data from Statistics Denmark, providing highly reliable socioeconomic and health information about the population of psychiatry users. We are thus able to avoid the problems of missing data and fallacies of self-reporting that most other survey studies face.

4.3. Data Collection Approach

Psychiatric patients are a hard-to-reach population. Therefore, rather than relying solely on Computer-Assisted Web Interviewing (CAWI), the study used a host of different survey approaches to increase the response rate. The survey was primarily sent out as an official digital mail, supplemented with Paper-and-Pencil Interviewing (PAPI) for the 11% of the gross sample who were exempt from receiving

digital mail. Further, patients received SMS reminders and digital mail reminders. Lastly, the study used Computer-Assisted Telephone Interviewing (CATI) with patients who had an available telephone number. PAPI and CATI are known to be costly and highly laborious in comparison to CAWI, but they are also known to be the most effective ways of contacting hard-to-reach populations (Kagerbauer et al., 2013). Additionally, PAPI and CATI ensured that patients facing digital difficulties had the opportunity to participate. The data were collected over a two-month period between December 28, 2023–February 28, 2024.

The survey obtained 1,478 full responses and had a response rate of 21.1%. Because this population is considered hard to reach, this is deemed a favorable response rate. Importantly, 103 partial responses, or 1.5% of the gross sample, were excluded from the analysis. Most of the partial responses had only responded to the initial items without completing the rest of the questionnaire.

5. Method

The statistical analysis consists of three parts. First, binary logistic regression analysis is used to examine the association between MHDS usage and key predictors. Second, LCA is used to identify underlying groups that characterize participants' experiences of and attitudes toward MHDS. Third, the study employs multinomial logistic regression to examine the association between the latent classes and key predictors.

LCA is a statistical method used to identify underlying subgroups within a set of observed categorical variables. It assumes that these subgroups, or latent classes, explain the variation in the observed data. LCA is a person-oriented approach examining qualitative characteristics on a large scale (Collins & Lanza, 2010) and is a particularly useful method for uncovering heterogenous vulnerabilities (T. Rose et al., 2017; Scotto Rosato & Baer, 2012) and subgroups that would benefit from a particular intervention (Weller et al., 2012). Thus, LCA is a well-suited method for this study, because it enables a nuanced account of different groups of patients and their perspectives on MHDS. In this case, the assumption is that the latent classes explain the variation in participants' experiences and attitudes. Hence, the analysis consists of two LCAs: one for participants who have used MHDS and one for participants who have not. Both LCAs are carried out via the statistical software program Stata using the LCA Stata plugin (Lanza et al., 2018).

There are different information criteria (IC) to evaluate the model fit and thus determine the number of latent classes. In this study, the model fit is assessed via the Bayesian information criterion (BIC), the sample-size-adjusted Bayesian information criterion (*a*BIC), and the Akaike information criterion (AIC). Lower values entail better model fit. However, if the IC point in different directions, the best model fit is determined based on the BIC, which is often deemed superior to the other IC (Collins & Lanza, 2010; Nylund et al., 2007; Vermunt, 2002). In addition, this study reports entropy, which can be used as a diagnostic criterion showing how well the model characterizes the latent classes. Values close to 1 are ideal (Celeux & Soromenho, 1996), but there is no official cutoff point.

LCA conventionally employs maximum likelihood estimation to estimate model parameters. Key parameters are the latent class prevalences and the item response probabilities. The latent class prevalences indicate the size of each latent class in the population. The latent classes are mutually exclusive and exhaustive, meaning that everyone in the population will be assigned to a class based on probabilistic assessment. The item response probabilities indicate the conditional probability of observing a particular response pattern given

latent class membership. A central part of the analysis is to name and characterize the latent classes based on the item response probabilities (Collins & Lanza, 2010).

In the subsequent multinomial logistic regression analyses, modal assignment is used to determine the best class match for each individual based on the posterior probabilities using the LCA Stata plugin (Lanza et al., 2018). For all the regression analyses, statistical significance is determined via $p < 0.05$.

6. Concepts and Measurement

6.1. Digital Psychiatry

MHDS are part of the movement towards digital psychiatry. However, there is no official definition of digital psychiatry, and the concept has been defined in a variety of ways (Bucci et al., 2019; Golinelli et al., 2020; Stern et al., 2023). In this study, digital psychiatry is seen as a broad concept spanning telemedicine (technology-mediated consultations), eHealth (online health promotion and treatment), mHealth (mobile health technology), and algorithm-based medicine (use of algorithms and big data for health promotion and treatment; Marent & Henwood, 2022).

This study draws on the definition offered by Legind et al. (2022), who adapt Topol's (2019) definition of digital psychiatry to a Danish context. Specifically, digital psychiatry includes video consultations, apps, sensor technology, online therapy, artificial intelligence, and virtual reality. Based on the World Health Organization's (2018) taxonomy of digital health technologies, digital psychiatry is understood as a form of intervention—that is, it intervenes in the patient's life to instigate change. Therefore, activities like ordinary internet searches, podcasts, and videos on YouTube are not part of the concept of digital psychiatry.

6.2. Empowerment

Measuring empowerment in relation to MHDS, this study builds on an empowerment survey consumer-constructed scale, specifically focusing on the dimensions of self-efficacy/self-esteem, power/powerlessness, community activism/autonomy, and optimism/control over the future (Cottrell & Langzettel, 2005; Rogers et al., 1997). In a similar vein, the study draws on the first subscale of the Health Care Empowerment Scale, which consists of being informed, committed, collaborative, and engaged (Johnson et al., 2012). However, the specific operationalizations of empowerment have been specifically designed for this study, because no other studies have operationalized empowerment in relation to MHDS. For the MHDS users, empowerment is measured via nine Likert scale items:

- I have felt left to myself when using digital tools.
- Digital tools have given me a better understanding of my mental health.
- Overall, I have benefited from using digital tools.
- Digital tools have made it easier for me to manage on my own in my daily life.
- Digital tools have made it easier to determine where and when my treatment should take place.
- I have only used digital tools because there were no other treatment options.
- Digital tools have made it easier to get information about my treatment options.
- Because of digital tools, I need less contact with my welfare professional.

- Digital tools have made me more aware of my symptoms.

For the non-users, the focus is on the perceived barriers to using MHDS based on six dummy scale items:

- I do not have access to necessary IT devices (e.g., computer or phone).
- I do not want to use more IT devices in my daily life (e.g., computer or phone).
- I do not think it is as beneficial as meeting in-person with my welfare professional.
- I do not think I have the right technical skills.
- I do not trust the data security.
- I do not know which tools are available.

6.3. Socioeconomic Position

The study is based on the materialistic dimension of socioeconomic position. Consequently, it does not capture the symbolic dimension (Krieger et al., 1997) or the subjective dimension (Diemer et al., 2013) of socioeconomic position. The study draws inspiration from Duncan's index of socioeconomic inequality (Stevens & Featherman, 1981), albeit without the income dimension since this indicator is presumed to correlate strongly with education and employment status. Socioeconomic position is measured via register data from Statistics Denmark about labor market position and highest educational attainment. Supplementary File 1 shows an elaborate description of each register and each register variable used in the study. Socioeconomic position is calculated as an index between labor market status and highest educational attainment, where the two dimensions are given equal weight. Labor market position is based on data from December 31, 2022, and educational attainment is based on data from September 30, 2022, due to register data availability. However, socioeconomic position is considered a relatively stable social phenomenon. Therefore, the time difference should not significantly influence the results.

6.4. Severity of Mental Illness

Severity of mental illness describes how pervasive and potentially disabling an individual's mental illness is. In line with other studies (Ruggeri et al., 2000), this study defines severe mental illness as diagnoses with psychotic symptoms, that is F20–F22*, F24*, F25*, F28–F31*, F32.3*, and F33.3*, according to the 2019 version of the ICD-10 diagnosis classification manual (World Health Organization, 2019). Severity of mental illness is a multifaceted concept, and a limitation of the present study is that it does not focus on the duration of illness or functional impairment (National Institute of Mental Health, 1987). Severity of mental illness is measured via register data from the Danish Health Data Authorities. A participant is considered to have severe mental illness if they have at least one registration of the mentioned diagnoses in the period January 1, 2023–December 15, 2023.

6.5. Social Support

Social support is defined as the social resources that an individual has available or assumes they have available in formal or informal contexts that are not provided by welfare professionals (Cohen et al., 2001). In this way, it relates to cognitive and bonding social capital (Harpham et al., 2002). Barrera (1986) introduces three distinct dimensions of social support, namely embedded, perceived, and enacted social support. The first is

related to the magnitude and quality of an individual's social network; perceived social support refers to how much support an individual presumes to have available in their network; and enacted social support refers to how much support the individual receives from their network (Barrera, 1986). In this study, embedded and perceived social support are examined building on the work of Krause (1999). Enacted social support is investigated by drawing on the Inventory of Socially Supportive Behaviors (Barrera & Baca, 1990). Social support is measured via the survey questionnaire as an index where each dimension is given equal weight.

The remaining predictors—gender, geographical location, migrant status, and age—are from June 30, 2023. The former three consist of binary outcomes, while the latter represents the participant's actual age.

7. Contextual Analysis of MHDS in Psychiatry

7.1. Distributions and Representativeness

Table 1 shows the distributions of each register predictor in the population as well as in the gross and net samples. As previously mentioned, the gross sample was statistically representative of the population using simple random sampling selection. The net sample is not representative of all parameters because of low response rates for certain groups.

Table 1. Distributions in population, sample, and test for representativeness.

Predictors	Population	Gross sample	Net sample
Age			
18–27 years	28.6%	28.7%	23.8%
28–37 years	23.3%	23.2%	19.4%
38–47 years	15.1%	15.0%	16.3%
48–57 years	13.2%	13.5%	18.7%
58+ years	19.9%	19.6%	21.7%
Total	100% (89,266)	100% (6,995)	100% (1,477)
Gender			
Male	42.8%	42.4%	37.9%
Female	57.2%	57.6%	62.1%
Total	100% (89,266)	100% (6,995)	100% (1,477)
Migrant status			
Non-Danish	14.1%	14.1%	9%
Danish	85.9%	85.9%	91%
Total	100% (89,266)	100% (6,995)	100% (1,477)
Geographical location			
Small city	55%	54.5%	57.3%
Large city	45%	45.5%	42.7%
Total	100% (89,266)	100% (6,995)	100% (1,477)

Table 1. (Cont.) Distributions in population, sample, and test for representativeness.

Predictors	Population	Gross sample	Net sample
Severity of mental illness			
Non-severe	66.3%	65.5%	67.7%
Severe	33.7%	34.5%	32.3%
Total	100% (89,371)	100% (7,000)	100% (1,478)
Socioeconomic position			
Low	34.8%	34.8%	23.9%
Average	42.9%	43.7%	48.2%
High	22.3%	21.5%	27.9%
Total	100% (87,495)	100% (6,870)	100% (1,467)

Notes: Different totals are due to a lack of information at the register status time; for statistical purposes, the continuous variables socioeconomic position and age have been recoded as categorical variables; statistically significant p-values = the sample is significantly different from the population; age: $\chi^2 = 59.48^{**}$; socioeconomic position: $\chi^2 = 80.59^{**}$; gender: $\chi^2 = 14.58^{**}$; migrant status: $\chi^2 = 31.17^{**}$; geographical location: $\chi^2 = 3.23$; severity of mental illness: $\chi^2 = 1.16$; * $p < 0.05$, ** $p < 0.001$.

Table 1 shows that the net sample has an underrepresentation of young individuals, men, individuals with a low socioeconomic position, and individuals with a migrant background. However, the sample is representative in terms of geographical location and severity of mental illness. This means that the sample reflects the overall population in terms of the proportion of people with disorders featuring psychotic symptoms. The participants without severe mental illness are those with any other psychiatric disorder that, importantly, can still be experienced as strongly affecting their everyday lives.

7.2. Descriptive Findings

We will first look at the descriptive features of the sample relevant for the statistical analyses. Table 2 shows the item responses for patients' overall usage of devices and their general opinion about MHDS. Table 2 shows that the participants seem to use technology in their everyday lives: 92.8% use a smartphone and 56.2% use a laptop computer. This is similar to the overall Danish population (Danmarks Statistik, 2020; Epinion, 2023). However, the participants do not seem to prioritize having the latest digital devices. Among some of the participants, there seems to be an understanding that MHDS are less desirable than meeting face-to-face with a welfare professional. Approximately half of the participants indicated a fear that MHDS will eventually come to replace face-to-face interactions with welfare professionals. Related to this, 59.5% of the participants state that the biggest downside of MHDS is that you are left to yourself to a greater degree. In turn, the biggest upsides to MHDS concern flexibility and outreach to people who live far away from psychiatric services. The reported upsides and downsides of MHDS can thus partly be seen as a reflection of the tension between individualized responsibility and freedom inherent in empowerment initiatives.

Crucially, Table 2 shows that MHDS are far from widespread among the participants: 60.8% of the participants have not used them, and only 33.2% have. As a precursor for the LCAs, it is salient to investigate whether there are statistically significant differences between the users and the non-users. This is done via binary logistic regression analysis assessing whether the predictors age, socioeconomic position, severity of mental illness, social support, geographical location, migrant status, and gender are associated with MHDS usage. Seeing it as a funnel, this preliminary analysis reveals differentiating mechanisms underlying the LCAs.

Table 2. Descriptive findings from the survey questionnaire.

Items	Response categories	Percentages (N = 1,478)
Have used MHDS	Yes	33.2%
	No	60.8%
	Do not know/Prefer not to answer	6%
Daily use of IT devices (in general)	Smartphone	92.8%
	Laptop	56.2%
	Headphones	55.6%
	Tablet	30.3%
	Desktop computer	23.6%
	Smartwatch	16.2%
	Other	7.2%
	Landline phone	4.1%
	Not using devices	1.3%
	Do not know/Prefer not to answer	0.5%
Prioritize having the latest IT devices	Agree or strongly agree	11.2%
	Neither agree nor disagree	22.8%
	Disagree or strongly disagree	63.9%
	Do not know/Prefer not to answer	2.2%
Afraid that MHDS will replace in-person meetings with welfare professional	Agree or strongly agree	50.5%
	Neither agree nor disagree	18.8%
	Disagree or strongly disagree	24.2%
	Do not know/Prefer not to answer	6.4%
Biggest downside to MHDS	Left to yourself	59.5%
	Requires access to devices	13.7%
	Data security risks	7.2%
	No downsides	8.8%
	Do not know/Prefer not to answer	10.8%
The biggest upside to MHDS	Flexible treatment	30.9%
	Better outreach	30.6%
	More autonomy in daily life	13.1%
	No upsides	11.3%
	Do not know/Prefer not to answer	14.1%

Notes: The participants could choose multiple options in the item "Daily use of IT devices (in general)," which is why the percentages do not sum to 100%; the items "Prioritize having the latest IT devices" and "Afraid that MHDS will replace in-person meetings with welfare professional" have been recoded from a five-category Likert scale to three categories; "Do not know" and "Prefer not to answer" were separate categories but are here coded together because the latter contains few responses (1–2% on each item).

7.3. Regression Analysis of MHDS Usage

Table 3 shows the results of the binary logistic regression analysis of MHDS usage and the predictors. The number of observations differs from those presented in Table 2 because observations with missing

values on any of the regression variables are excluded from the analysis. All continuous variables have been standardized. Non-use serves as the reference category.

Table 3. Binary logistic regression analysis with odds ratios for MHDS usage.

Predictors	MHDS usage
Age	0.71**
Socioeconomic position	1.14*
Geographical location (ref. small city)	0.77*
Gender (ref. male)	1.38*
Social support	1.19*
Migration status (ref. non-Danish)	1.52
Severity of mental illness (ref. non-severe)	1.03

Notes: $N = 1{,}264$; Likelihood Ratio $\chi^2 = 59.63$**; log-likelihood = −794.34; * $p < 0.05$, ** $p < 0.001$.

The full model in Table 3 is significant ($p < 0.001$). The table shows that age is negatively associated with MHDS usage, meaning that younger people are more likely to use MHDS than older people. Interestingly, there is also a negative association between geographical location and MHDS usage, indicating that people from small cities are more likely to use MHDS than people from big cities. This could be related to the fact that MHDS include video consultations, which could be attractive for people who live far away from treatment facilities. As such, MHDS makes treatment more accessible by surmounting physical distances. Table 3 shows a positive association between socioeconomic position, gender, and social support, meaning that people with high socioeconomic position are more likely to use MHDS than people with low socioeconomic position, females are more likely to use MHDS than males, and people with high levels of social support are more likely to use MHDS than people with low levels of social support. MHDS usage thus seems to have a social gradient.

For the MHDS users, we do not know the details of their MHDS usage and in-person treatment because of the nature of the short survey. However, we do know that the most common tools are video consultations (59.7% of the users) and apps (45.6% of the users). Relatively few participants have been in online therapy (16.7% of the users) or used artificial intelligence (11.6% of the users), sensor technology (10.4% of the users), or virtual reality (3.7% of the users). Video consultations were most often suggested by a welfare professional, whereas apps were most often found by the users themselves. The next section looks more closely at the LCA for MHDS users.

8. Analyses of (Differentiated) Experiences and Attitudes Towards MHDS

8.1. LCA for MHDS Users

Table 4 shows the IC for eight latent class models. The IC favor different models. As mentioned previously, the BIC is given superiority and a three-class model is chosen. In comparison to the other IC, the BIC prioritizes parsimony and reduces model complexity and the risk of overfitting (Weller et al., 2020). Indeed, the AIC and *a*BIC would induce too much model complexity, both suggesting an eight-class model. In this case, it seems fruitful to prioritize parsimony over complexity.

Table 4. IC and entropy for eight latent class models, MHDS users.

N classes	BIC	AIC	*a*BIC	Entropy
1	3152.81	3077.45	3095.67	1.00
2	2657.98	2503.09	2540.54	0.80
3	2549.43	2315.01	2371.69	0.78
4	2595.77	2281.80	2357.72	0.76
5	2629.57	2236.07	2331.22	0.77
6	2696.97	2223.93	2338.31	0.78
7	2750.05	2197.47	2331.09	0.79
8	2789.72	2157.60	2310.45	0.82

Note: Highlighted cells indicate the best model solution.

Figure 1 shows the average item response probabilities for each latent class and the latent class prevalences. The item response probabilities concern the likelihood of observing a particular response, for instance "agree or strongly agree," given class membership. In Figure 1, the item response probabilities are shown as average values across the items mentioned in Secition 6.2 for the MHDS users. Supplementary File 3 shows all the raw item response probabilities for both LCAs. The latent classes are characterized and named via the item response probabilities. For instance, the latent class "Self-help" is so named because members of this class are likely to agree with positive statements about MHDS. The latent class "Neither good nor bad" is named in this way because the item response probabilities show that members of this latent class are likely to neither agree nor disagree with the statements about MHDS, indicating a somewhat ambivalent position. A possible interpretation could be that these individuals do not think that MHDS have had a notable influence on their mental health. Item response probabilities for the latent class "Not useful" reveal that members of this latent class have a negative outlook on MHDS, being likely to disagree with positive statements about MHDS.

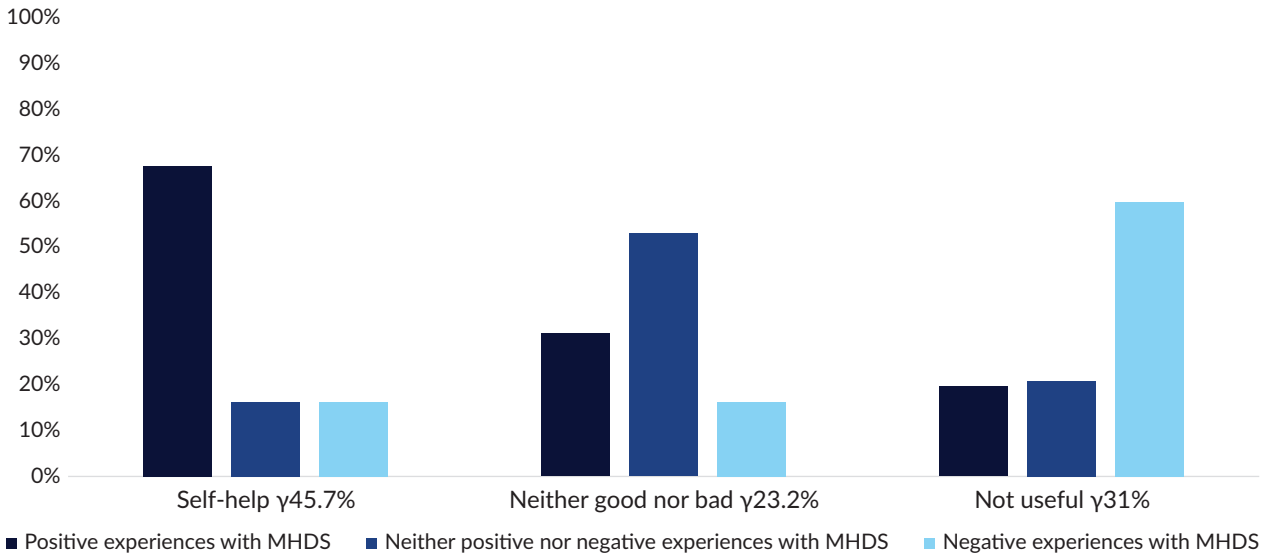

Figure 1. Average item response probabilities for each latent class and latent class prevalences (γ). Notes: *N* = 486; the response categories "Do not know" and "Prefer not to answer" are coded as missing values; 5 responses with missing values on all items were excluded from the analysis; recoded item categories: "Agree" and "Strongly agree" = Positive, "Neither agree nor disagree" = Neither positive nor negative, "Disagree" and "Strongly disagree" = Negative; two items were reverse-coded.

As the latent class prevalences in Figure 1 show, the positive "Self-help" class is the largest class in the population, covering 45.7%. However, the second-largest class is the critical "Not useful," covering 31% of the population. Lastly, the ambivalent "Neither good nor bad" class is the smallest, covering 23.3%. We will next take a step further and examine whether the predictors are associated with the latent classes for MHDS users.

8.2. Regression Analysis for MHDS Users

Table 5 shows the results of the multinomial logistic regression analysis of the three latent classes and the predictors age, socioeconomic position, severity of mental illness, social support, geographical location, migrant status, and gender. The regression analysis is based on the sample of the LCA for MHDS users. However, the number of observations differs because participants with missing data on any of the predictors are excluded from the analysis. All continuous variables have been standardized. Among the latent classes, the "Not useful" class was the reference category.

Table 5. Multinomial logistic regression analysis for MHDS users with odds ratios for latent class membership.

Predictor	Neither good nor bad	Self-help
Age	1.07	1.42*
Socioeconomic position	0.95	0.88
Geographical location (ref. small city)	0.90	1.00
Gender (ref. male)	0.97	1.05
Social support	0.90	1.34*
Migration status (ref. non-Danish)	2.34	0.85
Severity of mental illness (ref. non-severe)	1.18	1.40

Notes: $N = 448$; Likelihood Ratio $\chi^2 = 25.71$*; log-likelihood = −454.87; * $p < 0.05$, ** $p < 0.001$.

The full model in Table 5 is significant ($p < 0.05$). Table 5 shows that there are no statistically significant differences between the "Neither good nor bad" latent class and the reference category "Not useful." However, Table 5 reveals a positive association between age, social support, and membership of the "Self-help" latent class relative to the "Not useful" latent class. The fact that the other predictors are not significant could be related to low statistical power and be a limitation of the analysis. It could also be related to a selection effect, as demonstrated in the preliminary analysis of MHDS usage. The sample of MHDS users has substantially different characteristics than the non-users in terms of geographical location, gender, socioeconomic position, social support, and age. Figuratively speaking, this sample seems to consist of those who made it through the funnel.

Surprisingly, the association between age and the latent "Self-help" class is in the opposite direction than anticipated. Remembering that this sample is more likely to be comprised of younger participants compared to the non-users, it could be that those with a higher age in this sample have distinct qualities. It could be that they find MHDS more novel and impactful than those who are younger. Moreover, young individuals are more often exposed to digital tools, e.g., in educational settings or via social media, potentially making them less excited about MHDS. This would resonate with scholars who suggest that social media potentially inflicts young people with digital fatigue (Liu & He, 2021).

Further, Table 5 shows a positive association between social support and the "Self-help" latent class relative to the "Not useful" latent class. This lends a certain irony to the naming of the "Self-help" latent class. Members of this latent class are likely to feel that MHDS have, e.g., made it easier to manage their daily lives, and that they need less contact with their welfare professional. At the same time, they are likely to have higher levels of social support relative to the "Not useful" latent class, suggesting that they are not only more inclined towards self-help but also receive more help from others. Moreover, this challenges the individualistic connotation of self-help, underscoring the importance of having a network to draw on when managing mental illness.

8.3. LCA for MHDS Non-Users

This section follows the same LCA approach for the MHDS non-users. Table 6 shows the IC for eight latent class models. Again, the IC favor different models, and the BIC points towards a three-class solution, which is the model chosen. Importantly, entropy is the lowest for the three-class solutions, possibly imposing statistical uncertainty in the modal assignment used for the regression analysis. The entropy is not alarmingly low, but could serve as a limitation to the modal assignment.

Table 6. IC and entropy for eight latent class models, MHDS non-users.

N classes	BIC	AIC	*a*BIC	Entropy
1	534.20	506.16	515.14	1.00
2	363.48	302.74	322.20	0.75
3	285.15	191.71	221.64	0.67
4	302.00	175.85	216.26	0.78
5	305.36	146.51	197.40	0.80
6	339.39	147.84	209.20	0.84
7	372.10	147.84	219.67	0.87
8	378.89	121.93	204.23	0.90

Note: Highlighted cells indicate the best model solution.

Figure 2 shows the item response probabilities for responding "Yes" to the barrier items mentioned in Section 6.2 for each latent class and the latent class prevalences. The latent class "Lacking knowledge" receives this name because the item response probability for not knowing which tools are available is very high in this latent class. Conversely, the item response probabilities are low for the remaining items, indicating that these latent class members are not likely to face other barriers. Potentially, members of this latent class could have a positive outlook on MHDS if they knew which ones were available.

In contrast, members of the "Wanting in-person treatment" latent class seem to not use MHDS because they think they are less beneficial than meeting in-person with a welfare professional. Members of this class do not seem to have an aversion toward technology, given the low item response probabilities for the other items, but believe that digital therapy is of lower quality than in-person treatment. Again, this highlights the great importance attributed to welfare professionals.

Members of the "Multiple reasons for non-use" latent class seem to have medium-high item response probabilities on most of the items, suggesting that this class has multiple reasons for not using MHDS

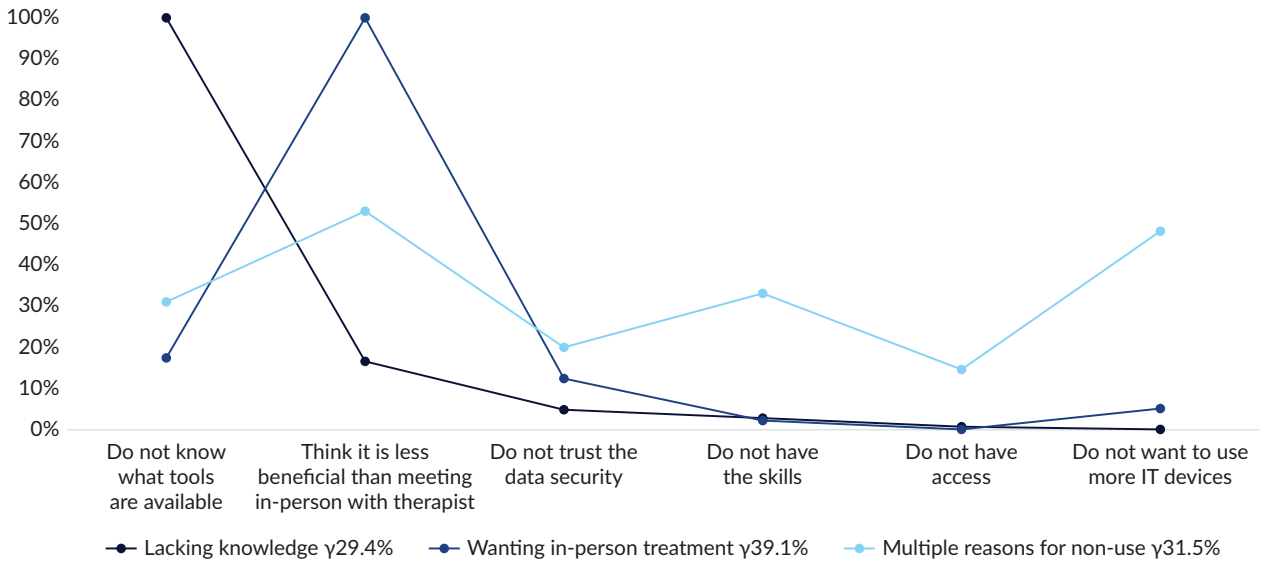

Figure 2. Plot of item response probabilities for each latent class and latent class prevalences (γ). Notes: *N* = 790; the response categories "I do not know" and "I prefer not to answer" are coded as missing values; 108 responses with missing values on all items were excluded before the analysis.

without any of the reasons being decisive. Moreover, many of the item response probabilities are homogenous for this class, making it difficult to articulate a clear characterization. This relates to the issue of latent class separation (Collins & Lanza, 2010), whereby it becomes difficult to separate the "Multiple reasons for non-use" latent class from the other latent classes. The most pronounced reasons are thinking that MHDS are less beneficial than meeting in-person with a welfare professional, not wanting to use more IT devices, and a lack of skills. Interestingly, none of the latent classes have high item response probabilities for a lack of access or trust in data security. However, members of the "Multiple reasons for non-use" latent class are more likely to face these barriers than members of the other latent classes. Thus, this latent class seems to be the most vulnerable in terms of MHDS.

Figure 2 shows that the three latent classes are somewhat equally distributed in the population. The hesitant latent class "Wanting in-person treatment" is the largest, covering 39.1%, while the critical latent class "Multiple reasons for non-use" covers 31.5%. In turn, 29.4% of the population belongs to the possibly open-minded latent class "Lacking knowledge."

8.4. Regression Analysis for MHDS Non-Users

Table 7 shows the results of the multinomial regression analysis of the three latent classes and the predictors. As with the previous regression analysis, the number of observations differs from the LCA because responses with missing data on any of the predictor variables are excluded. The continuous variables were standardized and the "Multiple reasons for non-use" latent class serves as the reference category.

The full model in Table 7 is significant ($p < 0.001$). Table 7 shows a negative association between age and membership of the "Wanting in-person treatment" latent class and the "Lacking knowledge" latent class relative to the "Multiple reasons for non-use" latent class. Thus, older people are more likely to be members of the critical "Multiple reasons for non-use" latent class than the two other latent classes. At the same time,

Table 7. Multinomial logistic regression analysis for MHDS non-users with odds ratios for latent class membership.

Predictor	Wanting in-person treatment	Lacking knowledge
Age	0.69**	0.64**
Socioeconomic position	1.24*	1.32*
Geographical location (ref. small city)	1.42	1.15
Gender (ref. male)	0.84	0.81
Social support	1.03	1.05
Migration status (ref. non-Danish)	1.32	1.09
Severity of mental illness (ref. non-severe)	1.25	1.03

Notes: $N = 720$; Likelihood Ratio $\chi^2 = 39.62$**; log-likelihood = −748.12; * $p < 0.05$, ** $p < 0.001$.

Table 7 shows a positive association for socioeconomic position, indicating that individuals with higher socioeconomic position are more likely to be members of the "Wanting in-person treatment" latent class and the "Lacking knowledge" latent class relative to the "Multiple reasons for non-use" latent class.

The "Multiple reasons for non-use" latent class seems to be characterized by older age and lower socioeconomic position compared to the other latent classes. Importantly, this sample is already statistically associated with older age and low socioeconomic position compared to the MHDS users, as demonstrated in the preliminary analysis. The key features of this latent class are that its members face multiple barriers to using MHDS compared to the other latent classes and are more likely to face barriers related to skills and technology aversion compared to the other latent classes. This indicates that those who are the most critical towards MHDS among the non-users are also the most socially vulnerable—here, in the form of older age and lower socioeconomic position.

9. Discussion

The descriptive analysis of the usage of digital technology and the LCA for MHDS non-users indicates that, while not without importance, access does not seem to be a key barrier to using MHDS among psychiatric patients in Denmark. This is supported by Bonet et al. (2018), who find that psychiatric patients in Spain have similar access to technology compared to the rest of the Spanish population. In an American context, scholars similarly suggest that psychiatric patients have a comparable usage of technology to the rest of the population (Gay et al., 2016; Gitlow et al., 2017; Naslund et al., 2016).

There are other studies that suggest differently: Wong et al. (2020) find that people with schizophrenia in Australia have limited access to technology. However, the present study does not find associations between the severity of mental illness and the usage of MHDS or latent class membership. This could be related to sample differences; whereas Wong et al. (2020) focus on a small sample of individuals with a schizophrenia diagnosis, this study draws on a large sample where severe mental illness is understood more broadly as diagnoses with psychotic symptoms. Similar to Wong et al., Tobitt and Percival (2019) find that psychiatric patients in the UK have a substantially lower usage of technology than the rest of the UK population. The authors write that patients in the psychiatric facility under study often have long-term psychosis and "can be socially excluded" (Tobitt & Percival, 2019, p. 4). Thus, this sample may also vary substantially from

that of the present study. As previously noted, severity of mental illness is a complicated matter, in which dimensions beyond the scope of this article such as duration of illness, need for in-patient treatment, and functional impairment could play a significant role in technology engagement.

Moreover, the importance of technology usage and access is disputed and could be sensitive to national contexts as well as sampling specificities. Still, this study supports the argument made by Torous et al. (2021) that time is a crucial factor and that technology access disparities are becoming less important. Instead, a lack of knowledge or motivation as well as the belief that MHDS are a poor replacement for in-person treatment could be key components of potential (self) exclusion.

As an empowerment initiative, MHDS pertain to broader modern sociological movements such as individualization and reflexivity (Giddens, 2009), do-it-yourself biography (Beck, 1993), self-technologies, and governmentality (Foucault, 1997). In different ways, these thinkers emphasize a turn away from traditional institutions towards the individual subject, which ought to be guided by self-optimization and self-determination. This entails critical reflexivity and the loss of irrevocable professional authority (Turner & Samson, 2007). Thus, scholars have illuminated how the empowerment and treatment inclusivity offered by digital solutions challenge the professional authority of welfare professionals (Denneson et al., 2017; Farnood et al., 2020; Fiske et al., 2020; Gabriels & Moerenhout, 2018). In line with the findings of Pilnick and Dingwall (2011), this study does not necessarily support this claim. Rather, participants seem to immensely value in-person meetings with welfare professionals.

The regression analysis for the MHDS users showed that members of the "Self-help" latent class are likely to have higher levels of social support relative to the "Not useful" latent class. The paradox of inclusion and assigning responsibility inherent in empowerment initiatives can also be related to social support. On the one hand, network involvement can be seen as a way of subverting authoritative hierarchies by enabling the network to understand and help the person in treatment. This could ensure sustainable welfare in the sense that one's network can provide long-term help and diminish dependency on welfare professionals. On the other hand, network involvement may be tied to the neoliberal focus on reducing welfare costs by diffusing responsibility to patients' networks. Seen from this point of view, the utilization of personal networks would inhibit socially sustainable welfare by leaving individuals without strong networks in a worse position than those who do have strong networks. This is not to say that network involvement is undesirable, nor that network involvement is borne out of neoliberalism. As with empowerment, the point is to flesh out a genuine dilemma, where inclusion and participation can go hand-in-hand with increased responsibility.

The regression analysis for the MHDS non-users showed that digital vulnerabilities are linked to social vulnerabilities. A critical interpretation could be that MHDS require patients to be resourceful in order to become (more) resourceful. This can be seen as an empowerment Matthew effect, exacerbating social inequalities if MHDS become the only treatment option. From a policy perspective, it is important to remember these structures. In the worst case, the implementation of MHDS could worsen the position of the most vulnerable groups of patients, who are not equipped for digital therapy. Additionally, given the overall importance attributed to welfare professionals, a socially sustainable digitalization of psychiatry would ensure that digital treatment does not supersede in-person treatment.

10. Conclusion

Paying attention to differentiating mechanisms underlying MHDS usage, this article presented a binary logistic regression showing that the predictors geographical location, gender, socioeconomic position, social support, and age are associated with MHDS usage. Second, the article presented two LCAs to characterize underlying groups among MHDS users and the non-users. The LCAs show that one size does not fit all. As shown with the latent class "Self-help," MHDS can be empowering for some patient groups. At the same time, others may feel left on their own as shown with the latent class "Not useful." Among non-users, MHDS can be excluding if they are experienced as the only accessible treatment option. This is particularly salient with members of the latent class "Multiple reasons for non-use." Finally, the article presented multinomial logistic regression analyses to examine the association between the predictors and the latent classes. These analyses reveal associations between age, social support, socioeconomic position, and the latent classes.

The article has used the concept of empowerment as an overall framework. Thus, the ambition of the article was to demonstrate how MHDS can be interlinked with both participation and increased responsibility. Importantly, empowerment is an ambiguous concept without a clear empirical foundation. This study has drawn on existing operationalizations of empowerment. However, no studies have constructed a validated measure for empowerment in relation to MHDS, and this could be an avenue for future research.

Furthermore, the findings do not seem to support the claim that MHDS diminish the importance of welfare professionals. Future research could examine the extent to which MHDS challenge, or perhaps reinforce, professional authority, as well as the ways in which social and digital inequalities may be mutually reinforced.

Acknowledgments

I would like to express gratitude to research assistants Malene Hoffmann Buskbjerg, Sarah Vestergaard Lydiksen, Andrea Bæk Noesgaard, and Mille Thuesen Hirsch Sørensen. They have provided indispensable aid with research design, survey programming, and data collection. I would like to thank my mentors Lars Thorup Larsen and Bagga Bjerge for providing valuable input and support. I would also like to thank the taskforce for digital psychiatry for crucial insights from psychiatry practice. I am also grateful for the support of communication experts at the Center for Digital Psychiatry, who validated the questionnaire. In addition, I would like to thank the reviewers for their thorough comments that have greatly improved the manuscript. Lastly, I would like to express gratitude to the study participants.

Funding

This work has received one grant of 485,000 DKK from the Danish Regions and another grant of 30,000 DKK from internal department funding.

Conflict of Interests

The author declares no conflict of interests.

Data Availability

The data are confidential and not publicly available. Please contact the author for further information.

Supplementary Material

Supplementary material for this article is available online in the format provided by the author (unedited).

References

Andersen, J., & Elm Larsen, J. (2011). Socialpolitik set i empowermentperspektivet. In I. Hornemann Møller & J. Elm Larsen (Eds.), *Socialpolitik* (pp. 475–489). Hans Reitzel.

Andersson, G., Carlbring, P., & Rozental, A. (2019). Response and remission rates in internet-based cognitive behavior therapy: An individual patient data meta-analysis. *Frontiers in Psychiatry*, *10*, Article 749. https://doi.org/10.3389/fpsyt.2019.00749

Arnstein, S. R. (2019). A ladder of citizen participation. *Journal of the American Planning Association*, *85*(1), 24–34. https://doi.org/10.1080/01944363.2018.1559388

Barak, A., Hen, L., Boniel-Nissim, M., & Shapira, N. (2008). A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. *Journal of Technology in Human Services*, *26*(2/3/4), 109–160. https://doi.org/10.1080/15228830802094429

Barrera, M. (1986). Distinctions between social support concepts, measures, and models. *American Journal of Community Psychology*, *14*(4), 413–445. https://doi.org/10.1007/BF00922627

Barrera, M., & Baca, L. M. (1990). Recipient reactions to social support: Contributions of enacted support, conflicted support and network orientation. *Journal of Social and Personal Relationships*, *7*(4), 541–551. https://doi.org/10.1177/0265407590074010

Batra, S., Baker, R. A., Wang, T., Forma, F., DiBiasi, F., & Peters-Strickland, T. (2017). Digital health technology for use in patients with serious mental illness: A systematic review of the literature. *Medical Devices: Evidence and Research*, *10*, 237–251. https://doi.org/10.2147/MDER.S144158

Beck, U. (1993). *Risk society: Towards a new modernity*. Sage.

Berrouiguet, S., Perez-Rodriguez, M. M., Larsen, M., Baca-Garcia, E., Courtet, P., & Oquendo, M. (2018). From eHealth to iHealth: Transition to participatory and personalized medicine in mental health. *Journal of Medical Internet Research*, *20*(1), Article e2. https://doi.org/10.2196/jmir.7412

Bjerge, B. (2007). Empowerment—Lettere sagt end gjort. *STOF – Viden om rusmidler og samfund*, *9*, 37–42.

Bjerge, B., & Nielsen, B. (2014). Empowered and self-managing users in methadone treatment? *European Journal of Social Work*, *17*(1), 74–87. https://doi.org/10.1080/13691457.2012.739560

Bonet, L., Llácer, B., Hernandez-Viadel, M., Arce, D., Blanquer, I., Cañete, C., Escartí, M., González-Pinto, A. M., & Sanjuán, J. (2018). Differences in the use and opinions about new eHealth technologies among patients with psychosis: Structured questionnaire. *JMIR Mental Health*, *5*(3), Article e51. https://doi.org/10.2196/mental.9950

Borzekowski, D. L. G., Leith, J., Medoff, D. R., Potts, W., Dixon, L. B., Balis, T., Hackman, A. L., & Himelhoch, S. (2009). Use of the internet and other media for health information among clinic outpatients with serious mental illness. *Psychiatric Services*, *60*(9), 1265–1268. https://doi.org/10.1176/ps.2009.60.9.1265

Brønnum-Hansen, H., & Baadsgaard, M. (2012). Widening social inequality in life expectancy in Denmark. A register-based study on social composition and mortality trends for the Danish population. *BMC Public Health*, *12*(1), Article 994. https://doi.org/10.1186/1471-2458-12-994

Bucci, S., Schwannauer, M., & Berry, N. (2019). The digital revolution and its impact on mental health care. *Psychology and Psychotherapy*, *92*(2), 277–297. https://doi.org/10.1111/papt.12222

Buchert, U., Kemppainen, L., Olakivi, A., Wrede, S., & Kouvonen, A. (2022). Is digitalisation of public health and social welfare services reinforcing social exclusion? The case of Russian-speaking older migrants in Finland. *Critical Social Policy*, *43*(3), 375–400. https://doi.org/10.1177/02610183221105035

Carpenter-Song, E., Acquilano, S. C., Noel, V., Al-Abdulmunem, M., Torous, J., & Drake, R. E. (2022). Individualized intervention to support mental health recovery through implementation of digital tools into clinical care: Feasibility study. *Community Mental Health Journal*, *58*(1), 99–110. https://doi.org/10.1007/s10597-021-00798-6

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, *13*(2), 195–212. https://doi.org/10.1007/BF01246098

Charlton, J. I. (1998). *Nothing about us without us: Disability oppression and empowerment* (1st ed.). University of California Press. https://doi.org/10.1525/9780520925441

Cheng, V. W. S., Piper, S. E., Ottavio, A., Davenport, T. A., & Hickie, I. B. (2021). Recommendations for designing health information technologies for mental health drawn from self-determination theory and co-design with culturally diverse populations: Template analysis. *Journal of Medical Internet Research*, *23*(2), Article e23502. https://doi.org/10.2196/23502

Cohen, S., Gottlieb, B. H., & Underwood, L. G. (2001). Social relationships and health: Challenges for measurement and intervention. *Advances in Mind-Body Medicine*, *17*(2), 129–141.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis*. Wiley.

Comber, C., Colley, A., Hargreaves, D. J., & Dorn, L. (1997). The effects of age, gender and computer experience upon computer attitudes. *Educational Research*, *39*(2), 123–133. https://doi.org/10.1080/0013188970390201

Cottrell, R. P. F., & Langzettel, M. (2005). Identifying dimensions of empowerment in consumers of psychiatric services. *Occupational Therapy in Mental Health*, *21*(1), 13–38. https://doi.org/10.1300/J004v21n01_02

Danmarks Statistik. (2020). *It-anvendelse i befolkningen*.

Davydow, D. S., Fenger-Grøn, M., Ribe, A. R., Pedersen, H. S., Prior, A., Vedsted, P., Unützer, J., & Vestergaard, M. (2015). Depression and risk of hospitalisations and rehospitalisations for ambulatory care-sensitive conditions in Denmark: A population-based cohort study. *BMJ Open*, *5*(12), Article e009878. https://doi.org/10.1136/bmjopen-2015-009878

Davydow, D. S., Ribe, A. R., Pedersen, H. S., Fenger-Grøn, M., Cerimele, J. M., Vedsted, P., & Vestergaard, M. (2016). Serious mental illness and risk for hospitalizations and rehospitalizations for ambulatory care-sensitive conditions in Denmark: A nationwide population-based cohort study. *Medical Care*, *54*(1), 90–97. https://doi.org/10.1097/MLR.0000000000000448

Denneson, L. M., Cromer, R., Williams, H. B., Pisciotta, M., & Dobscha, S. K. (2017). A qualitative analysis of how online access to mental health notes is changing clinician perceptions of power and the therapeutic relationship. *Journal of Medical Internet Research*, *19*(6), Article e208. https://doi.org/10.2196/jmir.6915

Diemer, M. A., Mistry, R. S., Wadsworth, M. E., López, I., & Reimers, F. (2013). Best practices in conceptualizing and measuring social class in psychological research. *Analyses of Social Issues and Public Policy*, *13*(1), 77–113. https://doi.org/10.1111/asap.12001

Dobransky, K., & Hargittai, E. (2016). Unrealized potential: Exploring the digital disability divide. *Poetics*, *58*, 18–28. https://doi.org/10.1016/j.poetic.2016.08.003

Dybbroe, B. (2020). Health promotion in Denmark: From critical potential to individualisation and marginalisation. *Socialmedicinsk Tidskrift*, *97*(3), 417–432.

Ennis, L., Rose, D., Denis, M., Pandit, N., & Wykes, T. (2012). Can't surf, won't surf: The digital divide in mental health. *Journal of Mental Health*, *21*(4), 395–403. https://doi.org/10.3109/09638237.2012.689437

Epinion. (2023). *Befolkningens oplevelser og udfordringer i et digitalt samfund—Med fokus på ældre*.

Eubanks, V. E. (2007). Trapped in the digital divide: The distributive paradigm in community informatics. *The Journal of Community Informatics*, *3*(2). https://doi.org/10.15353/joci.v3i2.2373

Farnood, A., Johnston, B., & Mair, F. S. (2020). A mixed methods systematic review of the effects of patient online self-diagnosing in the 'smart-phone society' on the healthcare professional–patient relationship and medical authority. *BMC Medical Informatics and Decision Making*, *20*(1), Article 253. https://doi.org/10.1186/s12911-020-01243-6

Fiske, A., Buyx, A., & Prainsack, B. (2020). The double-edged sword of digital self-care: Physician perspectives from Northern Germany. *Social Science & Medicine*, *260*, Article 113174. https://doi.org/10.1016/j.socscimed.2020.113174

Foucault, M. (1997). Technologies of the self. In P. Rabinow (Ed.), *Ethics: Subjectivity and truth—Essential works of Michel Foucault 1954–1984* (pp. 223–251). The New Press.

Frank, V. A., & Bjerge, B. (2011). Empowerment in drug treatment: Dilemmas in implementing policy in welfare institutions. *Social Science & Medicine*, *73*(2), 201–208. https://doi.org/10.1016/j.socscimed.2011.04.026

Gabriels, K., & Moerenhout, T. (2018). Exploring entertainment medicine and professionalization of self-care: Interview study among doctors on the potential effects of digital self-tracking. *Journal of Medical Internet Research*, *20*(1), Article e10. https://doi.org/10.2196/jmir.8040

García-Lizana, F., & Muñoz-Mayorga, I. (2010). What about telepsychiatry? A systematic review. *Primary Care Companion for CNS Disorders*, *12*(2), e1–e5. https://doi.org/10.4088/PCC.09m00831whi

Gay, K., Torous, J., Joseph, A., Pandya, A., & Duckworth, K. (2016). Digital technology use among individuals with schizophrenia: Results of an online survey. *JMIR Mental Health*, *3*(2), Article e15. https://doi.org/10.2196/mental.5379

Giddens, A. (2009). *The consequences of modernity*. Polity Press.

Gitlow, L., Abdelaal, F., Etienne, A., Hensley, J., Krukowski, E., & Toner, M. (2017). Exploring the current usage and preferences for everyday technology among people with serious mental illnesses. *Occupational Therapy in Mental Health*, *33*(1), 1–14. https://doi.org/10.1080/0164212X.2016.1211061

Golinelli, D., Boetto, E., Carullo, G., Nuzzolese, A. G., Landini, M. P., & Fantini, M. P. (2020). Adoption of digital technologies in health care during the Covid-19 pandemic: Systematic review of early scientific literature. *Journal of Medical Internet Research*, *22*(11), Article e22280. https://doi.org/10.2196/22280

Graversen, S. B., Pedersen, H. S., Ribe, A. R., Foss, C. H., & Sandbaek, A. (2021). The significance of depression for short-term readmission and mortality after a pneumonia admission. *Medical Care*, *59*(10), 872–880. https://doi.org/10.1097/MLR.0000000000001626

Greenhalgh, T. (2009). Patient and public involvement in chronic illness: Beyond the expert patient. *BMJ*, *338*(7691), Article b49. https://doi.org/10.1136/bmj.b49

Greer, B., Robotham, D., Simblett, S., Curtis, H., Griffiths, H., & Wykes, T. (2019). Digital exclusion among mental health service users: Qualitative investigation. *Journal of Medical Internet Research*, *21*(1), Article e11696. https://doi.org/10.2196/11696

Greve, B. (2020). Denmark—A universal welfare system with restricted austerity. In S. Blum, J. Kuhlmann, & K. Schuber (Eds.), *Routledge handbook of European welfare systems* (2nd ed., pp. 129–144). Routledge. https://doi.org/10.4324/9780429290510-8

Harpham, T., Grant, E., & Thomas, E. (2002). Measuring social capital within health surveys: Key issues. *Health Policy and Planning*, *17*(1), 106–111. https://doi.org/10.1093/heapol/17.1.106

Hayes, J. F., Maughan, D. L., & Grant-Peterkin, H. (2016). Interconnected or disconnected? Promotion of mental health and prevention of mental disorder in the digital age. *British Journal of Psychiatry*, *208*(3), 205–207. https://doi.org/10.1192/bjp.bp.114.161067

Howell, A., & Voronka, J. (2012). Introduction: The politics of resilience and recovery in mental health care. *Studies in Social Justice*, *6*(1), 1–7. https://doi.org/10.26522/ssj.v6i1.1065

Hubley, S., Lynch, S. B., Schneck, C., Thomas, M., & Shore, J. (2016). Review of key telepsychiatry outcomes. *World Journal of Psychiatry*, *6*(2), 269–282. https://doi.org/10.5498/wjp.v6.i2.269

Indenrigs- og Sundhedsministeriet. (2023). *En bedre psykiatri—Hurtigere behandling til børn og unge, bedre akut hjælp og større tryghed for medarbejderne*.

Järvinen, M. (2016). Expertise and ambivalence in user-focused human service work. In J. Gubrium, T. Andreassen, & P. Solvang (Eds.), *Reimagining the human service relationship* (pp. 59–78). Columbia University Press. https://doi.org/10.7312/gubr17152-004

Järvinen, M., & Kessing, M. L. (2021). Self-casting and alter-casting: Healthcare professionals' boundary work in response to peer workers. *Current Sociology*, *71*(3), 414–431. https://doi.org/10.1177/00113921211048532

Johnson, M. O., Rose, C. D., Dilworth, S. E., & Neilands, T. B. (2012). Advances in the conceptualization and measurement of health care empowerment: Development and validation of the health care empowerment inventory. *PLoS ONE*, *7*(9), Article e45692. https://doi.org/10.1371/journal.pone.0045692

Jørgensen, K., Præstegaard, J., & Holen, M. (2020). The conditions of possibilities for recovery: A critical discourse analysis in a Danish psychiatric context. *Journal of Clinical Nursing*, *29*(15/16), 3012–3024. https://doi.org/10.1111/jocn.15311

Kagerbauer, M., Manz, W., Zumkeller, D., Carrasco, J. A., Zmud, J., Munizaga, M., & Lee-Gosselin, M. (2013). Analysis of PAPI, CATI, and CAWI methods for a multiday household travel survey. In J. Zmud, M. Lee-Gosselin, J. A. Carrasco, & M. A. Munizaga (Eds.), *Transport survey methods* (pp. 289–304). Emerald Group Publishing. https://doi.org/10.1108/9781781902882-015

Krause, N. (1999). Assessing change in social support during late life. *Research on Aging*, *21*(4), 539–569. https://doi.org/10.1177/0164027599214002

Krieger, N., Williams, D. R., & Moss, N. E. (1997). Measuring social class in US public health research: Concepts, methodologies, and guidelines. *Annual Review of Public Health*, *18*(1), 341–378. https://doi.org/10.1146/annurev.publhealth.18.1.341

Kugathasan, P., Stubbs, B., Aagaard, J., Jensen, S. E., Munk Laursen, T., & Nielsen, R. E. (2019). Increased mortality from somatic multimorbidity in patients with schizophrenia: A Danish nationwide cohort study. *Acta Psychiatrica Scandinavica*, *140*(4), 340–348. https://doi.org/10.1111/acps.13076

Lanza, S. T., Dziak, J. J., Huang, L., Wagner, A. T., & Collins, L. M. (2018). *LCA Stata plugin users' guide* (Version 1.2.1). The Methodology Center, Penn State. Microsoft Word - Stata LCA Plugin-v1.2.1

Legind, C. S., Faurholt-Jepsen, M., Ebdrup, B. H., Anhøj, S., & Benros, M. E. (2022). Digital innovation giver nye perspektiver i psykiatrien. *Ugeskrift for Laeger*, *184*(9), Article V08210649.

Li, N., & Kirkup, G. (2007). Gender and cultural differences in internet use: A study of China and the UK. *Computers and Education*, *48*(2), 301–317. https://doi.org/10.1016/j.compedu.2005.01.007

Liu, Y., & He, J. (2021). "Why are you running away from social media?" Analysis of the factors influencing social media fatigue: An empirical data study based on Chinese youth. *Frontiers in Psychology*, *12*, Article 674641. https://doi.org/10.3389/fpsyg.2021.674641

Lupton, D. (2013). The digitally engaged patient: Self-monitoring and self-care in the digital health era. *Social Theory & Health*, *11*(3), 256–270. https://doi.org/10.1057/sth.2013.10

Lupton, D. (2014). Apps as artefacts: Towards a critical perspective on mobile health and medical apps. *Societies*, *4*(4), 606–622. https://doi.org/10.3390/soc4040606

Marent, B., & Henwood, F. (2022). Digital health. In K. Chamberlain & A. C. Lyons (Eds.), *Routledge international handbook of critical issues in health and illness* (pp. 261-275). Routledge.

Mayo, M., Siim, B., Andersen, J. (2004). Exclusion, inclusion and empowerment: Community empowerment?

Reflecting on the lessons of strategies to promote empowerment. In J. Andersen & B. Siim (Eds.), *The politics of inclusion and empowerment: Gender, class and citizenship* (pp. 139–158). Palgrave Macmillan. https://doi.org/10.1057/9781403990013_9

Mik-Meyer, N. (2017). *The power of citizens and professionals in welfare encounters: The influence of bureaucracy, market and psychology*. Manchester University Press.

Naslund, J. A., Aschbrenner, K. A., & Bartels, S. J. (2016). How people with serious mental illness use smartphones, mobile apps, and social media. *Psychiatric Rehabilitation Journal*, *39*(4), 364–367. https://doi.org/10.1037/prj0000207

National Alliance on Mental Illness. (2014). *Health and technology study 2014*.

National Institute of Mental Health. (1987). *Towards a model for a comprehensive community-based mental health system*.

Nordentoft, M., Wahlbeck, K., Hallgren, J., Westman, J., Ösby, U., Alinaghizadeh, H., Gissler, M., & Laursen, T. M. (2013). Excess mortality, causes of death and life expectancy in 270,770 patients with recent onset of mental disorders in Denmark, Finland and Sweden. *PLoS ONE*, *8*(1), Article e55176. https://doi.org/10.1371/journal.pone.0055176

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Oute, J., Huniche, L., Nielsen, C. T., & Petersen, A. (2015). The politics of mental illness and involvement: A discourse analysis of Danish anti-stigma and social inclusion campaigns. *Advances in Applied Sociology*, *5*(11), 273–285. https://doi.org/10.4236/aasoci.2015.511026

Oute, J., & Ringer, A. (2014). Psychiatry in crisis—A battle over definitions. *Sygeplejersken*, *114*(7), 78–82.

Pilnick, A., & Dingwall, R. (2011). On the remarkable persistence of asymmetry in doctor/patient interaction: A critical review. *Social Science & Medicine*, *72*(8), 1374–1382. https://doi.org/10.1016/j.socscimed.2011.02.033

Plana-Ripoll, O., Musliner, K. L., Dalsgaard, S., Momen, N. C., Weye, N., Christensen, M. K., Agerbo, E., Iburg, K. M., Laursen, T. M., Mortensen, P. B., Pedersen, C. B., Petersen, L. V., Santomauro, D. F., Vilhjálmsson, B. J., Whiteford, H. A., & McGrath, J. J. (2020). Nature and prevalence of combinations of mental disorders and their association with excess mortality in a population-based cohort study. *World Psychiatry*, *19*(3), 339–349. https://doi.org/10.1002/wps.20802

Reisdorf, B., & Rhinesmith, C. (2020). Digital inclusion as a core component of social inclusion. *Social Inclusion*, *8*(2), 132–137. https://doi.org/10.17645/si.v8i2.3184

Ribe, A. R., Laursen, T. M., Sandbaek, A., Charles, M., Nordentoft, M., & Vestergaard, M. (2014). Long-term mortality of persons with severe mental illness and diabetes: A population-based cohort study in Denmark. *Psychological Medicine*, *44*(14), 3097–3107. https://doi.org/10.1017/S0033291714000634

Ringø, P., Nissen, M. A., Fallov, M. A., Birk, R. H., & Kjærulff, J. (2017). Behind political ideas of welfare and productivity—Exploring ontological models and forms of exclusion. *Social Work & Society*, *15*(2), 1–16. https://ejournals.bib.uni-wuppertal.de/index.php/sws/article/view/526

Rizq, R., & Jackson, C. (2019). *The industrialisation of care: Counselling, psychotherapy and the impact of IAPT*. PCCS Books.

Robotham, D., Satkunanathan, S., Doughty, L., & Wykes, T. (2016). Do we still have a digital divide in mental health? A five-year survey follow-up. *Journal of Medical Internet Research*, *18*(11), Article e309. https://doi.org/10.2196/jmir.6511

Rogers, E. S., Chamberlin, J., Ellison, M. L., & Crean, T. (1997). A consumer-constructed scale to measure

empowerment among users of mental health services. *Psychiatric Services*, *48*(8), 1042–1047. https://doi.org/10.1176/ps.48.8.1042

Rose, D., & Rose, N. (2023). Is 'another' psychiatry possible? *Psychological Medicine*, *53*(1), 46–54. https://doi.org/10.1017/S003329172200383X

Rose, N. (2018). *Our psychiatric future*. Wiley.

Rose, T., Lindsey, M. A., Xiao, Y., Finigan-Carr, N. M., & Joe, S. (2017). Mental health and educational experiences among Black youth: A latent class analysis. *Journal of Youth and Adolescence*, *46*(11), 2321–2340. https://doi.org/10.1007/s10964-017-0723-3

Ruggeri, M., Leese, M., Thornicroft, G., Bisoffi, G., & Tansella, M. (2000). Definition and prevalence of severe and persistent mental illness. *British Journal of Psychiatry*, *177*(2), 149–155. https://doi.org/10.1192/bjp.177.2.149

Schmidt, M., Schmidt, S. A. J., Adelborg, K., Sundbøll, J., Laugesen, K., Ehrenstein, V., & Sørensen, H. T. (2019). The Danish health care system and epidemiological research: From health care contacts to database records. *Clinical Epidemiology*, *11*, 563–591. https://doi.org/10.2147/CLEP.S179083

Schneider-Kamp, A., & Askegaard, S. (2020). Putting patients into the centre: Patient empowerment in everyday health practices. *Health*, *24*(6), 625–645. https://doi.org/10.1177/1363459319831343

Schradie, J. (2020). The great equalizer reproduces inequality: How the digital divide is a class power divide. In B. Eidlin & M. A. McCarthy (Eds.), *Rethinking class and social difference* (pp. 81–101). Emerald Publishing. https://doi.org/10.1108/S0198-871920200000037005

Scotto Rosato, N., & Baer, J. C. (2012). Latent class analysis: A method for capturing heterogeneity. *Social Work Research*, *36*(1), 61–69. https://doi.org/10.1093/swr/svs006

Speed, E. (2011). Discourses of acceptance and resistance: Speaking out about psychiatry. In M. Rapley, J. Moncrieff, & J. Dillon (Eds.), *De-medicalizing misery: Psychiatry, psychology and the human condition* (pp. 123–140). Palgrave Macmillan. https://doi.org/10.1057/9780230342507_10

Stern, E., Micoulaud Franchi, J.-A., Dumas, G., Moreira, J., Mouchabac, S., Maruani, J., Philip, P., Lejoyeux, M., & Geoffroy, P. A. (2023). How can digital mental health enhance psychiatry? *The Neuroscientist*, *29*(6), 681–693. https://doi.org/10.1177/10738584221098603

Stevens, G., & Featherman, D. L. (1981). A revised socioeconomic index of occupational status. *Social Science Research*, *10*(4), 364–395. https://doi.org/10.1016/0049-089X(81)90011-9

Tobitt, S., & Percival, R. (2019). Switched on or switched off? A survey of mobile, computer and internet use in a community mental health rehabilitation sample. *Journal of Mental Health*, *28*(1), 4–10. https://doi.org/10.1080/09638237.2017.1340623

Topol, E. (2019). *The Topol Review—Preparing the healthcare workforce to deliver the digital future*. The National Health Service in England.

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*, *20*(3), 318–335. https://doi.org/10.1002/wps.20883

Turner, B. S., & Samson, C. (2007). Professions, Knowledge and power. In B. S. Turner & C. Samson (Eds.), *Medical power and social knowledge* (pp. 130–152). Sage. https://doi.org/10.4135/9781446250426.n7

Vermunt, J. K. (2002). [Latent class analysis of complex sample survey data: Application to dietary data]: Comment. *Journal of the American Statistical Association*, *97*(459), 736–737. http://www.jstor.org/stable/3085712

Vitger, T., Korsbek, L., Austin, S. F., Petersen, L., Nordentoft, M., & Hjorthoj, C. (2021). Digital shared

decision-making interventions in mental healthcare: A systematic review and meta-analysis. *Frontiers in Psychiatry*, *12*, Article 691251. https://doi.org/10.3389/fpsyt.2021.691251

Wang, J.-Y., Bennett, K., & Probst, J. (2011). Subdividing the digital divide: Differences in internet access and use among rural residents with medical limitations. *Journal of Medical Internet Research*, *13*(1), Article e25. https://doi.org/10.2196/jmir.1534

Weller, B. E., Bowen, N. K., & Bowen, G. L. (2012). Linking students to appropriate interventions: A typology for social workers based on general strain theory. *Journal of Social Work*, *13*(4), 361–381. https://doi.org/10.1177/1468017311435446

Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent class analysis: A guide to best practice. *Journal of Black Psychology*, *46*(4), 287–311. https://doi.org/10.1177/0095798420930932

Wong, K. T. G., Liu, D., Balzan, R., King, D., & Galletly, C. (2020). Smartphone and internet access and utilization by people with schizophrenia in South Australia: Quantitative survey study. *JMIR Mental Health*, *7*(1), Article e11551. https://doi.org/10.2196/11551

World Health Organization. (2018). *Classification of digital health interventions v1.0*.

World Health Organization. (2019). *International statistical classification of diseases and related health problems: 10th revision (ICD-10)*. https://icd.who.int/browse10/2019/en

World Health Organization. (2021). *Mental Health Action Plan 2013–2030*.

Yellowlees, P. M., Parish, M. B., Gonzalez, A. D., Chan, S. R., Hilty, D. M., Yoo, B. K., Leigh, J. P., McCarron, R. M., Scher, L. M., Sciolla, A. F., Shore, J., Xiong, G., Soltero, K. M., Fisher, A., Fine, J. R., Bannister, J., & Iosif, A. M. (2021). Clinical outcomes of asynchronous versus synchronous telepsychiatry in primary care: Randomized controlled trial. *Journal of Medical Internet Research*, *23*(7), Article e24047. https://doi.org/10.2196/24047

Zimmerman, M. A., & Rappaport, J. (1988). Citizen participation, perceived control, and psychological empowerment. *American Journal of Community Psychology*, *16*(5), 725–750. https://doi.org/10.1007/BF00930023

About the Author

Emilie Kristine Dyrlev is a PhD fellow at the Department of Political Science, Aarhus University. She is a trained sociologist particularly interested in social inequalities, health policies, and forms of expertise. She is currently working on a project examining how digitalization alters psychiatric care practices and, consequently, the patient–provider relationship.

Social Inclusion
2025 • Volume 13 • Article 9971
https://doi.org/10.17645/si.9971

cogitatio

ARTICLE

Open Access Journal

"Complete the Test First": Prescreening Tests at the Margins of Digital Public Administration

Heta Tarkkala [1], Riikka Koulu [2], Karoliina Snell [1], and Helmi Soininvaara [2]

[1] Faculty of Social Sciences, University of Helsinki, Finland
[2] Faculty of Law, University of Helsinki, Finland

Correspondence: Heta Tarkkala (heta.tarkkala@helsinki.fi)

Submitted: 30 January 2025 **Accepted:** 12 May 2025 **Published:** 12 June 2025

Issue: This article is part of the issue "Fostering the Socially and Ecologically Sustainable Digitalisation of Welfare States" edited by Paula Saikkonen (Finnish Institute for Health and Welfare) and Marta Choroszewicz (University of Eastern Finland), fully open access at https://doi.org/10.17645/si.i514

Abstract

Western welfare states are rushing to digitalise access to public services and citizens' interactions with public institutions. Finland exemplifies this trend towards digital-first administration. In this article, we discuss digital public administration and the interaction between citizens and the state through the example of digital prescreening tests. Digital prescreening tests are interactive webpages or mobile apps in which a citizen-user is required to answer questions about their situation before they are granted contact with a civil servant, access to a public service, or further advice. We analyse these tests theoretically and empirically, asking what their legal status is, how their developers perceive them, and what impacts they have on citizens' legal standing and rights. Prescreening tests may be useful tools for citizens, but may also lead to harmful gatekeeping and socially unsustainable digital public administration. We argue that prescreening is an important, yet neglected phenomenon. Regulatory policy and academic research have focused on issues regarding automating decision-making processes, while advice-giving and first points-of-contact have received little attention. Drawing on discussions on street-level bureaucracy and its connections with digital systems, we analyse prescreening through interviews and legal sources. Our analysis shows how this new "screen-level bureaucracy" of prescreening tests influences both service provision and access as well as the rationalities and modalities of digital public administration.

Keywords

automated decision-making; digital administrative burden; digital public services; healthcare; medical devices; prescreening tests; screen-level bureaucracy; self-service; social welfare

1. Introduction

Contemporary western welfare states, particularly the Nordic welfare states, are rapidly digitalising access to public services and citizens' interactions with public institutions (Dencik & Kaun, 2020; Jørgensen, 2023; Lindgren et al., 2024). Finland is a prime example of this trend towards digital-first administration, characterised by the country's relatively non-critical and techno-optimistic culture and established tradition of providing public online resources and digital tools for citizens. In this article, we discuss a specific form of digital-first interaction between citizens and the state—digital prescreening tests (DPTs).

DPTs are interactive user interface tools, such as interactive webpages, mobile apps, or patient portals, deployed by public institutions and administrative bodies to automate advice-giving and people's access to public services, often through private-public partnerships. We examine the objectives of DPTs and how they are envisioned by experts involved in their design and deployment. We elaborate on the role of tests in digital service provision, and their implications for the rights and obligations of citizens and institutions. DPTs shape the form and function of digital public administration and thus relate to socially sustainable digital transformation and whether digital technologies truly improve citizens' lives (Nosratabadi et al., 2023, p. 2).

We provide a situated case study of DPTs in healthcare and social service provision in Finland and ask: What are DPTs, in law and practice, in relation to citizen-state interactions in digital public administration? For whom are they and for what ends? We contextualise our analysis both in terms of legislative policy on digital public administration and in relation to discussions of street-level algorithms (or bureaucracy) and citizens as customers of self-service administration (Bovens & Zouridis, 2002; Lipsky, 2010; Melin et al., 2024). Our analysis demonstrates the diversity and ambivalence at the core of DPTs, which take various forms: They may represent harmful gatekeeping, diverting citizens from public services, but they can also be useful as communication channels, learning devices, and advice-giving resources (for a similar discussion on chatbots see Verne et al., 2022). This ambivalence is at odds with ideologies of digital-first initiatives that portray digitalisation as unavoidable and fundamental for future public administration, especially in the health and social welfare sector (e.g., MFF, 2024; MSH, 2023).

We argue that digital prescreening is an important yet understudied and underregulated phenomenon in digital public administration (see also Adelmant & Raso, 2025). Much regulatory policy and academic research has focused on automated decision-making (ADM) processes. Law produces and aggravates this decision-making focus, for example through the EU's General Data Protection Regulation (GDPR), leading, at the legislative level, to a disregard for the "less-intrusive" digitalisation of advice-giving and first points-of-contact. Prescreening tests, as well as the digitalisation of state–citizen interactions more broadly, remain an "under the radar" phenomenon in digital public administration, neglected by regulatory policy as well. DPTs play a role as gatekeepers in the service chain, but their legal and administrative status remains unclear. This issue is central to citizens' legal rights as well as the service provider's duty to provide care or service.

Hence, we seek to broaden the debate on digital public administration by providing a situated example of digital public administration that goes beyond decision-making. By analysing the various roles that DPTs play, we demonstrate how their ambivalence also creates uncertainties about the need for and feasibility of regulatory intervention, what to regulate, and how.

Our analysis demonstrates that although the explicit justification for prescreening is to provide new information resources and channels for citizens seeking access to social and healthcare services, the tests are also developed to support professionals by reducing their workload and by providing structured preliminary information before actual direct contact. The tests embed a specific notion of a citizen who can navigate the digital landscape effortlessly despite legal uncertainties and informational unclarities. This raises questions about their social sustainability: When the legal status of prescreening is unclear and even the developers struggle, at times, to categorise tests as advice or decision-making, or as something else, what chance does the citizen have of understanding them?

We structure our article as follows. First, in Section 2, we discuss prior research on digital public administration, street-level and screen-level bureaucracies, and the increasing responsibilisation of citizens that digitalisation often entails. In Section 3, we draw attention to the role of law in creating ambivalence about DPTs and in coproducing the decision-making focus. In Section 4, we describe our methods and our case study on Omaolo, the comprehensive national platform for healthcare and social services that incorporates many DPTs. In Section 5, we analyse the multiple purposes of DPTs, the blurry boundaries between advice, decision-making, and customership, the burden of recognising rights, entitlements, and errors, and the dynamic adaptability of DPTs. Finally, we provide concluding remarks, calling for more research on the modalities of prescreening in digital public administration.

2. Digitalisation of Public Services and Self-Servicing Citizens

Digitalisation of welfare services can mean digitalising existing services, completely new modes of offering and delivering services, as well as new ways of making decisions on and assessing the needs of the population (Dencik & Kaun, 2020; Haug et al., 2024). It spans from self-service solutions to ADM systems and the use of data analytics for monitoring, and even fraud detection (Peeters & Widlak, 2023). Recognising that the boundaries between various interface tools, such as chatbots, interactive guides, web resources, and various surveys and tests are often blurry and prone to change, prescreening tests seem to bring particular logics of testing, filtering, and structuring to digital public administration.

Digital prescreening encourages, or even requires, a citizen-user to take the test first, to answer questions about their situation before they are granted contact with a civil servant, access to a public service, or further advice about the workings of administration. The terminology concerning such tests is not yet established. Some refer to tests as "self-assessment tests," "self-screening tests," "symptom checker tools," or "digital self-triage tools" (Wallace et al., 2022; Ziebart et al., 2023), whereas in our empirical material they are also named "service-assessment tests." Despite terminological differences, all these concepts focus on the increasing reliance on digital, automated, online testing tools that give the appearance of interaction and individual evaluation of a situation, despite being automated. We opt for DPTs to highlight how they precede other interactions and how they are used for initial assessment.

Symptom checkers have been analysed for their medical accuracy and validity (Chambers et al., 2019; Wallace et al., 2022), the variability of different tests (Ziebart et al., 2023), as well as the development of such services (Trifuljesko & Ruckenstein, 2024). Even though chatbots in public administration have been discussed from the viewpoint of state–citizen relationships (Kaun & Männiste, 2025; Verne et al., 2022), discussions on DPTs in public services are still lacking. There is some overlap between chatbots and DPTs, for example in their

focus on providing information and guidance related to frequently asked questions, but nevertheless tests are provided to citizens specifically as tests and not as a form of question-answer interaction.

Digital public administration and welfare constitute the background to digital prescreening. Prior research on digital public administration suggests a long-term trend of increasing responsibilisation of the citizen vis-à-vis the state. This development is characterised by the diminishing role of face-to-face encounters. Lipsky (2010) famously discussed street-level bureaucrats as people who, in practice, implement public policies and use discretion in making judgments. Recently, this concept has been reframed to digitalisation through discussions on street-level algorithms, screen-level bureaucracies, and system—or infrastructure-level bureaucracies, which describe the increasing distance between the citizen and public power facilitated by digitalisation (e.g., Alkhatib & Bernstein, 2019; Bovens & Zouridis, 2002; Melin et al., 2024; Peeters & Widlak, 2023). Impersonalisation, low levels of discretion, and formalisation of service provision are defining characteristics of screen-level or infrastructure-level bureaucracies, concepts that suggest less face-to-face interaction compared to the time of Lipsky's analysis. Digitalisation has been demonstrated to shape interaction, and to change decision-making processes and the roles of public servants (e.g., de Boer & Raaphorst, 2021).

Melin et al. (2024) identify specific bureaucratic roles related to the management and functioning of digital public services. The roles are (a) automated bureaucrat; (b) self-servicing citizen; (c) front-office employee; (d) back-office employee; and (e) specialised bureaucrat. Thus, what Lipsky's street-level bureaucrats once did alone is now distributed among these five roles. The first two roles are especially important for prescreening tests done by citizens themselves. Automated bureaucrat refers to "automated processes which conduct specific, individual activities" such as "data exchange, case-handling, and [guiding of] citizens," which are enabled by "automated data exchange and decision-making systems (algorithms)" as well as "chatbots" (Melin et al., 2024, p. 106). The self-servicing citizen, then, refers to the citizen who "conducts various activities through digital self-service solutions" and thus both identifies services and applies for them, as well as provides "information and documentation" through "web-portals" and "digital self-service solutions" (Melin et al., 2024, p. 106). Based on these roles, we see characteristics of the self-servicing citizen and automated bureaucrat at play in our case.

Citizens' encounters with administration are prone to producing diversion (Määttä, 2012; Soininvaara et al., 2024) by adding administrative burdens on the self-servicing citizen, as described in Herd and Moynihan's (2018) work on the difficulties, exclusions, and bureaucratic work that citizens encounter in their interactions with the state. Digital administrative burdens, in turn, may result from automated or data-assisted decision-making or digital interactions. According to Peeters (2023, p. 9), digital administrative burdens "may merely be costly and tedious but may also prove insurmountable and lead to exclusion from rights, benefits, and services to which citizens are legally entitled." The prescreening tests we observe operate mainly in seeking information on services, symptoms, and personal situations. However, some of these may become parts of administrative processes if a test leads to direct contact with administrators, establishes a customer relationship and duty of care, or provides structured information for later decision-making.

Administrative burdens describe the shifting of workload to citizens themselves, which comes with learning, psychological, and compliance costs regarding interactions with the government (Moynihan et al., 2015), as

well as the need to learn new administrative competencies (Heggertveit et al., 2022; Madsen et al., 2022; Peeters, 2023). Additionally:

> The level of administrative burden placed on an individual, as well as the distribution of burden between the state and the individual, will often be a function of deliberate political choice rather than simply a product of historical accident or neglect. (Moynihan et al., 2015, p. 43)

Thus, the burdens are a form of policy-making by other means (Peeters & Widlak, 2023). We are less interested in whether they are deliberately constructed or unintended consequences (Herd & Moynihan, 2018; Madsen et al., 2022), but we acknowledge how institutional factors "shape the incentives and constraints for policy-makers and street-level bureaucrats" (Peeters & Widlak, 2023, p. 865).

Digital administrative burdens are also related to questions of good governance (Peeters, 2023, p. 119). According to Madsen et al. (2022, p. 3), interacting with government influences the citizens' "understanding and perceptions of the government" and "affect whether people will be able to exercise the fundamental rights of citizenship." Even though we identify the potential for burdens in the case of prescreening tests, we do also acknowledge that prescreening tests might have positive effects and are not always burdening or diverting. They can, for example, alleviate the learning costs of a benefit or service system as a test may offer targeted information based on the citizen's specific situation (Madsen et al., 2022).

3. Legal Frameworks and Prescreening Tests

The role of law in shaping and being shaped by digitalisation policy and in overseeing its implementation in administrative practice is often ignored in debates on the digitalisation of the welfare state. Prescreening is often overlooked in research and in regulation that focus on decision-making and its automation. However, it is important to understand how legal frameworks not only produce and amplify this decision-making focus but also contribute to marginalising prescreening in two ways: first, by juxtaposing decision-making that affects citizen's rights with less-consequential advice-giving that does not; second, through the unclear legal status and practice of DPTs. These explain how prescreening tests as a form of screen-level bureaucracy fall into the margins of regulatory attention. The legal framings co-produce certain centers of attention and silences at their margins, resulting in an interesting tension. However, the relative regulatory invisibility of prescreening tests is at odds with their visibility for citizens, for whom interacting with tests may be inseparable from (digital) public administration.

The form and process of administrative decision-making are densely regulated through both national administrative law and European technology regulation. For example, ADM is prohibited in the EU's GDPR (art. 22, 679/2016), and, while the general rule includes several exceptions and leeway for national legislation, it significantly limits potential automation—in relation to decision-making (Esko & Koulu, 2023). Unlike ADM, European technology regulation does not provide a definition of DPTs or suggest corresponding legal concepts.

However, as some DPTs are digital products used within healthcare, they fall under the EU's Medical Device Regulation (745/2017), which imposes product safety rules. From the legal perspective, these multiple frameworks raise questions about potential overlaps and uncertainties, as well as fragmentation. This means

that some DPTs are regulated differently from others, although they are provided for the citizens through the same platform, as described in the next section.

An important question is: Are DPTs legally about decision-making or advice-giving? Although DPTs and their ensuing rationalities largely remain outside legal attention, national administrative law would conceptualise them and automated citizen-state interactions through the service principle as giving advice and service to citizens instead of deciding on their rights and obligations (434/2003, Administrative Procedure Act). In giving advice, DPTs need to comply with the substantive obligations for "service automation" (Act on the Provision of Digital Services 306/2019, § 6a) for example, human users must be informed about the artificiality of a service, the language should be appropriate, and the user should be given the option to contact a human civil servant.

In understanding the legal status of DPTs, the distinction between decision-making and advice is central. Many of the existing legal safeguards, as well as the right to appeal, obligation to give grounds for a decision, and civil servants' personal liability and discretionary power, are attached to the decision-making process. If there is no decision, there is no recourse to an appeal body, although extrajudicial complaint processes may be available. Although DPTs do not generally produce decisions, they are dynamic and sometimes feed seamlessly into a decision-making process, blurring the boundary. Nevertheless, they may entail major access to justice issues, if for example prescreening tests provide false information or divert citizen-users away from a public service they would be entitled to.

Another crucial element in relation DPTs is the notion of customership, because sectoral legislation on social and healthcare conceptualises citizens with the multifaceted concept of "customer." The customer has rights and the service provider has a duty to provide care. Legal definitions of "customer" are contextual, and their exact meaning may become vague from the merging of various contexts and functions in digital service provision. For example, sometimes the "customer" is a resident of a wellbeing services county, for whom the county has a statutory obligation to provide services (612/2021, Act on the organisation of social and healthcare), but in other situations the "customer" is a person who applies for or uses social care (Act on the Processing of Customer Data in Social and Healthcare 703/2023) or a patient using or subjected to healthcare services (Act on the Status and Rights of Patients 785/1992). Furthermore, service providers are obligated to inform their customers about their rights and about the information systems related to the processing of their data as well as the general operating principles of such systems (703/2023, 68 §).

The legal unclarity of DPTs is associated with fuzziness that seems endemic to digital public services. Such unclarities have been also the object of legal oversight. The Chancellor of Justice (2022) in Finland condemned the national online platform Omaolo, through which the DPTS are made available to citizens, as the platform had not sufficiently informed its users on whether they were accessing a public or private service.

Finally, binding regulation is often complemented by soft-law instruments and policy documents that are not binding as such but have legal relevance. The digitalisation of public services is driven by an ethos of "digital by default" in many countries (Kaun & Männiste, 2025; Lindgren et al., 2024; Madsen et al., 2022; Schou & Pors, 2019), including Finland (Andreassen et al., 2021). Digitalisation is promoted to save costs and streamline administration and service provision, leading even to mandatory self-service (Andreassen et al., 2021; Schou & Pors, 2019, p. 466). This direction is also clear in Finland, as ongoing legislative reforms push for more comprehensive digitalisation of public administration. For example, the current government intends

to introduce a far-reaching digital-first initiative (MFF, 2024). This change would entail automatic digital service of documents to those deemed capable of digital interaction from 2026 onwards, without requiring consent. Moreover, in the strategies of 21 healthcare regions in Finland, digitalisation of social and healthcare services is portrayed as inevitable (Iisakka & Alastalo, 2024).

4. The Omaolo Case, Data, and Methodology

In this article, we focus on the most prominent service in the health and social sector in Finland, the Omaolo platform (see also Trifuljesko & Ruckenstein, 2024). This is a nationwide digital service operated by DigiFinland, a state-owned private company. Omaolo uniquely includes DPTs from both healthcare and social welfare sectors. It contains 17 prescreening tests for specific health issues and one general contact form for symptoms not covered by the tests. These DPTs are medical devices. Alongside symptom tests, the platform offers a digital health check as well as 12 different coaching programmes, for example, to quit smoking or eat more healthily. Omaolo also offers three tests on eligibility for specific social services. With the help of these tests, the customers should be able to evaluate whether they are eligible for personal assistance, mobility support, or informal care support for a family member. Unlike symptom tests, the service eligibility tests are not medical devices.

The 17 medical DPTs are developed based on the most often read articles in Terveyskirjasto, a health information webpage about illnesses, provided by The Finnish Medical Society Duodecim. Duodecim also provides the knowledge base for health-related DPTs on Omaolo. The user first chooses the most suitable test for their situation, such as back pain, diarrhea, or urinary tract infection. The test consists of background questions including sex and age, and specific questions about the symptoms. Based on these answers, the test recommends various actions: self-treatment instructions, or a recommendation to contact a health centre or the emergency services. The Omaolo symptom test does not usually produce a diagnosis but instead assesses the urgency of the user's need for healthcare. Citizen-customers can use Omaolo with or without logging in with strong electronic identification. This means that anyone can take the tests, but only if one logs in can the test be used as an access channel to service—the results of the test can be sent directly to a healthcare unit. Other communication channels, such as phone lines, are also available, but many municipalities encourage citizens to handle their health issues through Omaolo. The original aim was that all public social welfare service and healthcare providers—currently the 21 wellbeing service counties—would adopt the service, but some have not or have resigned from it due to low usage and high costs. This has created a situation where residents of the same regions can only take the tests without logging in but receive general advice based on their results.

This article builds on our previous work exploring the usability of various prescreening tests (Soininvaara et al., 2024), where we examined a series of DPTs in practice. We scrutinised four tests from the social and healthcare sector that represented varying designs and uses, two of which were on Omaolo. Some tests came very close to digital service by, e.g., simulating calculations for subsidies, while others assessed the applicant's chances of being eligible for a highly discretionary support. We found that tests can both guide individuals to services or turn them away through inadequate instructions, and they may also enable "gaming" the tests. As with automatic decision-making, pre-tests may overlook life situations that do not fit well with the system's logic. Another observation was that it was not always clear whether the tests could be understood as decisions for customers or advice to citizens.

We used an iterative and reflexive method (see Montgomerie, 2017; Srivastava & Hopwood, 2009), where we placed these initial, empirical findings in dialogue with the legal analysis. As described in Section 3, numerous regulations are associated with digital tests and their application areas: European data and technology regulation and national laws, both general administrative law and sector-specific laws on healthcare provision. The legal analysis was formed through systematic identification and examination of relevant laws. We aimed to identify the appropriate areas where legislation contributes to the DPTs and the challenges raised in our analysis and potential connections between legislation and the interviews.

After identifying gaps and challenges both empirically in the tests and the legal analysis, we refined our approach and questions iteratively (Srivastava & Hopwood, 2009). We conducted expert interviews among people currently or previously working with DPTs in the social and health sector, mainly in relation to Omaolo. The aim was to reflect our findings of the socio-legal analysis as well as to give the experts an opportunity to express their views on DPTs. To deepen our understanding, 9 people were interviewed in 2024 and early 2025 by one or two of the authors. Semi-structured interviews were conducted either online or in person. Two of the interviews included two informants. We interviewed people with the following expertise: national digital strategies and their implementation (2); public administration and service providers (5); planning prescreening tests (6); content creation (4); and technical development (4). One person might be included in several categories. Because of the small number of interviewees, we refer to them only by general descriptions.

Our aim was to understand the legal position and role of DPTs. Through the dynamic and iterative socio-legal analysis, we identified key challenges that the tests pose for public services, the rights of citizens, and the social sustainability of digital public service provision. Our data collection and analysis were also entwined, as initial findings based on experimenting with tests helped to identify gaps in legislation, which again fed into what we wanted to discuss in the interviews and needed to know more about. Based on this dynamic work, our refined focus on our topic guided how we organised our data. At least two of the authors read the interview data and sorted it according to the analytical sections we had formed based on our initial findings on previous practical experiments, legal analysis, and interviews. We then worked with these analytically sorted materials to further refine our argument through repeated focus on data, analysis, and writing (e.g., Kapiszewski et al., 2022).

5. Prescreening Tests in Public Administration

5.1. Multiple Purposes of Omaolo

Our interviewees reflected on the role of DPTs in many ways: as digital solutions providing gateways to services, as a vital part of healthcare and social services, and as part of a broader digital transformation of public services. An expert from the public sector explains the integral role of medical prescreening tests in optimising the care chain:

> We think of it always as comprehensive service, that the tests are part of the care chain. Tests are a good word, but we talk about preliminary information and surveys, and they are linked strongly to care pathways. With them, we can make the patient's access to care more agile, and at the same time get test results or preliminary information.

While DPTs can replace or complement phone calls, for example, many experts stressed that the goal is to use digital technology through its own affordances, not just as a parallel to human service. An expert on public sector digitalisation stated that fully digital services "have to be based on different symptom checkers or evaluations of the situation, be it symptoms or economic situation or anything." The logics of these automated bureaucrats (Melin et al., 2024) require ways to describe the customer or citizen in a manner that then produces relevant results without a human in the loop. Currently, most of the tests are access points to services performed by humans.

In the interviews, Omaolo symptom assessment tests were often framed as a means of alleviating the strain on the healthcare system and its communication channels. With their help, some cases could be treated automatically or initial steps such as recording patient information or symptoms could be automated. This, according to our interviews, may "save 5–10 minutes of nurses' time" if the patient is steered to an appointment. Tests were also seen to replace the assessments made by nurses regarding the urgency of treatment needs, thus supporting the human work and feeding into it. An expert stated that these systems have to be established, because "a chat with a doctor does not save any time" but a scalable digital service without a human could. It was also acknowledged that if the goal is to save money instead of providing better service, it does not often result in great digital services. Thus, the goals of savings and good service need to be balanced.

In the interviews, the goal of alleviating the strain on the healthcare system was complemented with delivering advice to citizen-customers. Providing "good advice" was seen by one interviewee as a way to keep "unnecessary people" out of the (physical) services, while simultaneously helping the person to take care of themselves if their condition could be treated at home. Another expert said that 20–30% of the people using tests receive advice and do not have further contact with healthcare professionals. This alleviates the burden on public services but can potentially divert people from services. However, a test can also be an easier channel for gaining access. An expert involved in the development described how the threshold to contact a service may be lower for people who are not comfortable with making phone calls. Also, certain symptom checkers such as the one for sexually transmitted diseases may be an easier venue for first contact. In some cases, the entire treatment pipeline is largely automated, and people can receive a test kit at home. Thus a test may allow seeking for help easily regardless of what the problem is. Additionally, for safety reasons, and as Omaolo symptom tests are medical devices, the threshold for a doctor's appointment is set relatively low for certain symptoms. The threshold may be lower in DPTs than in personal contact with a nurse who is able to spot important nuances related to, e.g., breathing.

In addition to health-related tests, Omaolo includes three service assessment tests that belong to the social welfare services. While health and social service prescreening tests—assessing symptoms and support needs—are of similar status in the platform design, they are inherently different. Whereas the health tests are medical devices and the results are subject to strict requirements, the service assessment tests reflect the highly discretionary process involved in support decisions. In the end, service assessment tests can only provide answers that the user "probably" is or is not entitled to the service. One expert described the difference between tests in social service and healthcare by stating that in healthcare the symptom assessment is a gateway to service: "You just send it, then the service starts, the matter is taken care of." In social services, there is a threshold that the prescreening tests cannot help users cross. The official decision must be made by a person doing a more extensive, individual assessment, with the ensuing rights and obligations.

The interviewees recognised this stark difference between Omaolo symptom assessments and social service assessments in practice and legally. They admitted that the social service tests provide "limited added value" and are rarely used. Having both healthcare and social welfare prescreening tests on the same platform reflected the goal of reforming and integrating the two spheres. In developing the Omaolo platform and its tests, however, the social service assessments were neglected, according to expert interviewees. An example follows:

> It must be partly confusing for the customer too that they are so different in the end. The symptom checkers in practice can lead to your matter being resolved. But then, as far as social welfare is concerned, it's a bit like a test to see whether you are a bird or a fish, slightly exaggerated….The service assessments have remained a bit underdeveloped and left out in the cold. But the reason why they are perhaps still kept there is, to some extent, in this state of hope or ambition that there would be some kind of integration between social welfare and healthcare.

The professionals interviewed agree that the social service DPTs are vague compared to the medical DPTs. Especially for discretionary social services such as support for informal care, the DPT cannot realistically give meaningful or truly tailored results. The tests are therefore of limited use. An expert from wellbeing services counties pointed out that, "according to the Social Welfare Act, actual social services cannot be provided through digital services." Thus, social services can provide digital solutions for automated bureaucracy (Melin et al., 2024) only for things that precede social welfare or service, which differs significantly from healthcare.

Social service pre-tests tend to emphasise how the test result is an assessment of one's own eligibility, producing results that a user "might" or "might not" be qualified for social service. The social services tests may be more misleading than symptom checkers. First, despite the careful wording, the interface visually corresponds with the health-related DPTs, giving the impression of similarity and interconnectedness. Second, healthcare-related tests result in instructions and potential contact with healthcare providers if the user is logged in. Social service tests, with their uncertain results, might end up diverting people from applying for the service altogether, if a negative assessment is interpreted as a response and evaluation from an authority. One of the interviewees said that it was discussed very early on that the test had to be formulated carefully so that no one could claim it had promised or denied eligibility. However, the vagueness requires the self-servicing citizen to understand the public service system, creating administrative burdens for them. The ethos of social care is to support people in their individual life situations, which fits poorly with prescreening. This problem was acknowledged in the interviews. It was stated that it might be wiser to think of a variety of options to help a person than to focus on "very individual things" as the tests do.

5.2. Advice, Decision-Making, and Customership

The outcome of an Omaolo DPT is usually either an assessment of the situation and guidance, or sometimes even a diagnosis. The test can be seen as a sort of "transmission mechanism" for the service providers. Sometimes a customer relationship begins, and the citizen is guided to actual services. However, how customership, decision-making, and guidance unfold or become entangled in practice, and how an individual's situation shifts from category to category, is unclear. An expert from a wellbeing services county reflected on how a person becomes a customer through strong online identification:

> I think that the treatment path cannot officially begin if the client has not been identified. We can provide general advice, but we cannot record anything in the patient information system without identification. In that case, it will just remain at general level.

Considering the legal distinction between advice and decision-making, identification becomes a decisive boundary from which rights and obligations follow. The interviewees regarded test results without identification as advice or guidance, not as an official decision associated with the legal safeguards of administrative decision-making. The users are considered customers only after identification or after the information has been sent to the service provider, again requiring identification. Even though the experts stated unanimously that information on an identified individual has to be sent over to the officials, the line between test users and customers and the status of the prescreening tests as either administrative advice or administrative decision-making remains complicated.

The purpose of the medical DPT is to assess the need for urgent care, not make diagnostic decisions as such. Despite this, there are also prescreening tests that perform diagnostics in practice, for example concerning urinary tract infection in fertile women. It can also be questioned whether instructions for self-care can be considered as decisions when, according to the test, no visit to a healthcare facility is needed. Some of the professionals noted that a prescreening test must more readily guide a person to contact the healthcare providers because a test does not have the same expertise in assessing conditions as, for example, experienced nurses.

The importance of customership and its legal underpinnings for the provider's duties was acknowledged and critically reflected by interviewees, but not elaborated to the citizen-user in the layout and design of Omaolo. An expert involved in developing Omaolo said that it required a lot of consideration to determine when responsibility for the treatment begins in these DPTs. In Omaolo, a customer relationship and therefore responsibility for care starts "at the point when the client sends the symptom assessment to the professional." Another expert said that at least when results of a prescreening test are sent to a professional, it creates an obligation for a wellbeing services county to respond, and "there is then a certain time, within which the case has to be taken care of, and the issue steered forward."

However, another informant stated that "being a customer is a difficult term here" since people living in a certain municipality are in principle always customers in relation to a certain wellbeing services county. There are also situations when this relationship is enacted and put practically in force, for example, when a person needs healthcare services or guidance and actively seeks help. An expert involved in development elaborated:

> There is also a customer relationship in situations where the Omaolo symptom assessment, unlike many other devices, is so safe to use that it can issue a self-care recommendation. This also indicates that you do not need to seek treatment. And this, compared to a few other instruments, is different. It has resolved the client's issue. There are, however, still two options. Was the client identified, or is there even an actual possibility of knowing who has been there, or whether the test was used without being identified? In the latter case, the customer relationship is basically to the service provider.

The service provider in this latter case would be DigiFinland, which owns Omaolo, not the wellbeing service county and the local healthcare provider. This can also happen if a specific region does not use Omaolo.

In these regions, there is no integration with local healthcare services, but Omaolo is openly available on the internet. An expert concluded that, even though "the concept of customer relationship has been discussed a lot," it is still not "entirely clear."

The Chancellor of Justice (2022)—the highest instance of legality control in Finland, who oversees digital public administration—also spotted inconsistencies in the legal status of Omaolo. The Chancellor made a decision in 2022 concerning the terms of use disclosed to users. According to the condemnatory decision, the Omaolo service entity falls somewhere between public and private social and health services, but despite its unregulated status should nevertheless follow the principles of good governance when using public power. This distinction between private and public services, and the users' legal rights and obligations, were not sufficiently disclosed to the user. Even though the terms of service now provide this information, the underlying problem remains.

There are multiple layers of uncertainty related to DPTs, including the public/private divide and how it affects becoming a customer or receiving guidance. Despite the same current care guidelines being used to guide medical treatments throughout the country, it is still a challenge to offer tests through a nationally united interface with integrated regional service provisions. How services are locally organised in terms of "how to get the customer to the right place at the right time and to the right address, either physically or by calling or sending electronically or something like that" were diverse and even described as "wild and free" by an expert we interviewed. To direct patients in healthcare services similarly throughout the country would have required a large-scale systemic renewal, which was not possible. Thus, according to an expert involved in development, "the shockingly confusing service production field" was turned into a digital service that, for example, in its early years directed people quite differently depending on "where the customer was, what their home municipality was, what day of the week it was, what time it was, whether it was a weekend or a public holiday." There are thus multiple layers in an Omaolo prescreening test. One is national and based on guidelines, and the second, local version, which is not used in all wellbeing services counties, supplements the national service by adding local instructions. In addition, there is an open third layer. In the wellbeing service counties that no longer employ Omaolo officially, residents can still take the tests. The webpages are open for anyone to use. But in such cases, even basic contact details of the particular region might be missing or out of date.

5.3. Burden of Recognising Rights, Entitlements, and Errors

The difference between advice and decisions is important in a regulatory sense and defines responsibilities and rights. For example, advice-giving does not involve instructions for appeal, which a user receives if a decision has been made. However, if the user does not seek access to services or benefits in the first place, because the test advised against it even though they would be eligible, the outcome cannot be appealed. An expert from another institution providing an eligibility test reflected that the option would then be to complain. However, extrajudicial complaints may be an even higher threshold than appeal for most citizens.

In addition, it seems that appeals are more significant for social services than healthcare. In healthcare, the treatment decisions of doctors, if officially complained or questioned, are handled at Valvira, the National Supervisory Authority for Health and Welfare. But "treatment decisions" might be closer to "incorrect administrative decisions" that one could "appeal to the provincial government or the court," a public sector

expert explained: for example, a "referral" that contains a "decision that you will now receive this specific service."

Because the law focuses so clearly on decision-making, the role of DPTs as advice is under-discussed and its ambivalent consequences remain largely hidden. The legal unclarity was critically reflected by an expert who specialises in public service digitalisation. The expert stated that in their operations authorities focus on whether administrative decisions have been made or not. Tests are in this sense often unclear because, as previously described, an administrative decision requires identification in a similar manner as is required for a customer relationship. The focus of legislation on decision-making thus complicates guidance, because the responsibility for advice and instructions is less clearly defined. One expert pondered whether guidance involves the same official responsibility that all public servants have when they give "advice by the authority," or if it is merely general advice that does not imply a customer relationship followed by official responsibility. The expert continued:

> The question is where the customer relationship begins, whether it starts when you get some answers to questions or even when you just read instructions on a website. Or whether it is already such a customer relationship, and someone is responsible that the instructions are correct….So this is a big problem, which is connected to all these electronic services in exactly the same way.

The fuzzy boundaries related to advice, decisions, and customership contribute to how self-servicing citizens may or may not recognise their rights. In digital public services, how law currently corresponds to the issues raised by DPTs seems to be partial and require a high level of understanding and knowledge of the welfare system from the citizen to recognise possible problems.

Moreover, even the authorities may not be aware of problems regarding these technologies, as they may not be visible. DPTs may produce what Adelmant and Raso (2025) have called "bureaucratic disempowerment," a situation where all actors struggle to recognise and address errors, sometimes because the involvement in decision-making is so unclear in the first place. In our case, this can be expanded to blurry relations between decision-making, general guidance, and advice by authority.

5.4. Adapting Tests, Adapting Users: Dynamics of Digital Public Administration

The landscape of prescreening tests is dynamic and evolving, not just as a whole but also within a single test. This can be exemplified with the test used during the Covid-19 pandemic. Starting from 2020, Omaolo offered a symptom checker for Covid-19, which was also used as a gatekeeper for Covid testing. To get tested, one had to take the DPT first. Throughout the lifespan of the test, the thresholds for services provided according to test results altered constantly with changes in national recommendations and knowledge base. Eventually, in 2023 the test was taken down and a more general "respiratory tract" symptom assessment was offered instead.

Nevertheless, during the pandemic, the test was in practice used to sort out access to laboratory tests and help relieve the pressure on healthcare services. According to Trifuljesko and Ruckenstein (2024, p. 11):

> The service engine materialised as a diagnostic tool to query the possibility of an infection, while the treatment facilitator opened a communicative channel with care professionals. This dual task—in

> decision-making and in enabling interaction—made the symptom checker an effective public health support at the heart of the pandemic.

This suggests that diversity of tests is not good or bad as such. Tests can also be "gamed" by the users. Users were able to take the test repeatedly and modify the symptoms reported according to the threshold needed to be directed to the official Covid test. Thus citizens also learned the logic of public service provision and how to gain access to services, which could also be considered empowering (see also Buffat, 2015). The practice of gaming the tests was not familiar to all the experts we interviewed, but some said that it was obvious that this would happen, because people might experience a need to contact the healthcare providers and would work for the access even if it meant having to "exaggerate [their] symptoms." Another expert reflected:

> At the end of the day, you can't directly prevent that scam. Sure, we can instruct that you should be answering truthfully, and by filling out the form you accept these terms of use,…but if a person wants to give different answers, you can't really prevent that.

It was also recognised that as the system is not currently fully digital, and the digital interface is only one route for contacting healthcare services, the customer might use multiple routes. Thus, they might game both the tests and the system as a whole, which actually increases the burden on the healthcare system. An expert described a potential case:

> If they are not getting the result they want this way [Omaolo], they will try to contact another way. So they don't believe the reply advising self-care….And they do not want to wait for the contact from the professional, as it is not in real time, but can take a couple of days….So, we might end up with three to four different contacts from the same person on the same issue.

But even when both tests and citizens adapt to the changes, there are also limits that have not yet been crossed. One concerns the languages in which tests are provided. Finland has two official languages, Finnish and Swedish, so public service is available in at least these two languages. In addition, Omaolo offers medical prescreening tests in English, but when it comes to social support needs there is an "other contact" form to be filled. But why are tests not available in other widely spoken languages such as Somali, Russian, or Arabic? An expert involved in the development work described the principle: "This is not just about filling the symptom assessment, but also about service in that language for the rest of the contact." That is, if an actual service contact followed, it would have to be provided in the same language as the test was taken.

Another problem for adaptation in terms of tests regards the specificity of problems that testing is suitable for. The prescreening tests are focused on one issue at a time, but people often have other symptoms or difficulties at the same time, as discussed in relation to social service assessments. The tests may fail to see the person in the relevant context, placing the burden on them concerning the realisation of their rights and entitlements (Herd & Moynihan, 2018). Prescreening tests may then distribute administrative burdens unevenly among the population, depending on the complexity of their situation, knowledge of the service system, and language skills.

6. Conclusion

From analysing DPTs and especially the Omaolo service, we draw three main conclusions. First, prescreening tests allow us to examine the digital welfare state in practice. Tests may shift the administrative burden from public institutions to individuals, leading to increasing responsibilisation of the citizen. These tests seem to align with the notion of the self-servicing citizen suggested in previous research and can contribute to creating digital administrative burdens—barriers and costs that citizens face when in need of public services (Madsen et al., 2022; Peeters, 2023). Identifying these burdens is complicated by the characteristics of different tests and contexts of use, a complexity further amplified by their unclear legal status and multiplicity of relevant legal frameworks. Some tests provide comprehensive assessments with clear outcomes and paths of action, whereas others result in an estimation of individuals' situations with no clear path forward. Some of these tests are highly regulated medical devices, whereas others are primarily classified as service automation with limited legal requirements, yet the citizen must navigate these. While the experts recognise many of the challenges—e.g., a blurry boundary between advice and decision, potential confusion of citizen-users, unclarity about rights and provider obligations—there are no easy solutions to these issues, and individual critique does not afford systemic reflection.

The operational logic of prescreening tests reflects and shapes both the appearance as well as the content of digital public administration and the drive towards automated bureaucracy. These tests risk turning users away, leading to scalable losses of rights, particularly affecting those users and user groups whose circumstances fall outside the typical cases coded into the tests. The tests also embody the logic of self-service administration, where citizen-users are made responsible for recognising and asserting their own rights and even challenging the outcomes (see also Adelmant & Raso, 2025).

Second, the demarcation between decision-making and advice-giving is not straightforward, even for the developers and deployers of prescreening tests. This appears particularly in the discussions on whether the completion of a test marks the beginning of a customer relationship between the citizen and the service provider, with ensuing legal obligations for the latter to provide further services. In our interview data, the complexities of establishing a customer relationship were left somewhat open, despite its significance for the legal status and rights of the citizen and the associated legal obligations for the welfare provider. Strong electronic identification was usually considered the prerequisite for customership but, at the same time, everyone is by default a customer of their own wellbeing services county through their residence. The vagueness of customership in practice corresponds with the vagueness and multiplicity of legal definitions of the customer.

As we have observed, the law contributes to marginalising prescreening tests and similar digital interaction tools. From the doctrinal-legal perspective, advice-giving and service to citizens are "softer" obligations—they are not as heavily regulated or sanctioned as rendering legally binding decisions with legal effects, giving public institutions more leeway to develop digital solutions. Yet DPTs are not separate phenomena, and they may also contribute to blurring boundaries between different administrative functions, changing the modalities of digital public administration. As the interface tools merge with one another, a test may become structured input data for a decision-making pipeline. Furthermore, the rationalities of prescreening tests are related not only to digital interaction but may also end up framing other forms of interaction, as human-led practices such as phone conversations can emulate digital-first practices with their logic of structured data and decision trees.

Third, the focus of current scholarship on the (semi-)automation of decision-making may lead to disregard for other aspects of the digital welfare state, such as prescreening tests, that may have significant implications for citizens' legal standing and fulfilment of their rights. Further research, both theoretical and empirical, is needed to broaden the existing debate on the social and legal implications of digital technologies as well as to fully grasp the rationalities associated with DPTs and other automation techniques beyond the decision-making focus. Ultimately, DPTs need to be assessed in terms of the social sustainability of the digital welfare state, since as an empirical case they encapsulate key dimensions of digital public services. Our analysis of DPTs contributes to an understanding of the effectiveness of existing laws in tackling digital public administration. The developers and deployers are aware of legal conceptualisations and these inform the DPT design, but the logic of digital prescreening materialises its own interpretation of legal rules.

In the future, it will be necessary to include the experiences and views of citizens in the study of prescreening tests and digital public services (also Peeters, 2023). This would help to capture how people experience prescreening tests as a part of service provision, advice-giving, and decision-making, and how tests contribute to diversion and administrative burdens. Problems of digitalisation are often framed in terms of special groups—how digitalisation impacts vulnerable groups or results in digital divides. However, prescreening tests show that the design of digital public services is a key question for everyone in terms of how entitlements and rights as citizens are realised. For some, tests might offer an easy and uncomplicated way to improve their situation, since they can offer information and guidance at a suitable moment for the citizen, even outside office hours.

It could also be asked what policies prescreening tests promote as a specific form of administration. This goes beyond "digital first" policies. Prescreening tests currently appear to be a grey area, both as public service and legally. Tests with quite different outcomes and further pathways offered side by side make the context of the phenomenon unclear—a medical device differs significantly from a simple, form-based decision tree meant to advise. It is generally not clearly defined when a person using a test is receiving service instead of general guidance, and at which point a person becomes a customer of public services. Thus, it is not easy to understand, legally and in terms of public service provision, what the tests are as a mode of governance nor in legal terms, and what the true benefits and costs are for both citizens and the state.

Acknowledgments

We would like to thank our interviewees for providing their time and expertise for our study, as well as two anonymous reviewers and the thematic issue editors, Paula Saikkonen and Marta Choroszewicz, for their feedback during the process.

Funding

This work is funded by the Strategic Research Council (SRC) established within the Research Council of Finland (358245; 353531; 353398), the Research Council of Finland (341434), and Svenska Litteraturskällskapet i Finland (4706474). Publication of this article in open access was made possible through the institutional membership agreement between the University of Helsinki and Cogitatio Press.

Conflict of Interests

The authors declare no conflict of interests.

Data Availability

The interview data is not stored in open data archives.

References

Adelmant, V., & Raso, J. (2025). Data entry and decision chains: Distributed responsibility and bureaucratic disempowerment in the UK's universal credit programme. *Oxford Journal of Legal Studies*. Advance online publication. https://doi.org/10.1093/ojls/gqaf006

Alkhatib, A., & Bernstein, M. (2019). Street-level algorithms: A theory at the gaps between policy and decisions. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–13). Association for Computing Machinery.

Andreassen, R., Kaun, A., & Nikunen, K. (2021). Fostering the data welfare state: A Nordic perspective on datafication. *Nordicom Review*, *42*(2), 207–223.

Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, *62*(2), 174–184.

Buffat, A. (2015). Street-level bureaucracy and e-government. *Public Management Review*, *17*(1), 149–161.

Chambers, D., Cantrell, A. J., Johnson, M., Preston, L., Baxter, S. K., Booth, A., & Turner, J. (2019). Digital and online symptom checkers and health assessment/triage services for urgent health problems: Systematic review. *BMJ Open*, *9*(8), Article 027743.

Chancellor of Justice. (2022). *Service terms of use and other information provided on a service* (Decision OKV/2674/10/2020-OKV-7). https://oikeuskansleri.fi/documents/1428954/164296211/published_decision_OKV_2674_10_2020.pdf/232410a2-2f4a-99b0-9753-8291f2e0cd90/published_decision_OKV_2674_10_2020.pdf?t=1686142689169

de Boer, N., & Raaphorst, N. (2021). Automation and discretion: Explaining the effect of automation on how street-level bureaucrats enforce. *Public Management Review*, *25*(1), 42–62.

Dencik, L., & Kaun, A. (2020). Datafication and the welfare state. *Global Perspectives*, *1*(1), Article 12912.

Esko, T., & Koulu, R. (2023). Imaginaries of better administration: Renegotiating the relationship between citizens and digital public power. *Big Data & Society*, *10*(1), https://doi.org/10.1177/20539517231164113

Haug, N., Dan, S., & Mergel, I. (2024). Digitally-induced change in the public sector: A systematic review and research agenda. *Public Management Review*, *26*(7), 1963–1987.

Heggertveit, I., Lindgren, I., Madsen, C. Ø., & Hofmann, S. (2022). Administrative burden in digital self-service: An empirical study about citizens in need of financial assistance. In R. Krimmer, M. R. Johannessen, T. Lampoltshammer, I. Lindgren, P. Parycek, G. Schwabe, & J. Ubacht (Eds.), *Electronic participation* (pp. 173–187). Springer.

Herd, P., & Moynihan, D. P. (2018). *Administrative burden: Policymaking by other means*. Russell Sage Foundation.

Iisakka, E., & Alastalo, M. (2024). Digitaalisen sosiaali-ja terveydenhuollon lupaukset: Kriittinen luenta hyvinvointialue-strategioiden sosioteknisestä mielikuvastosta. *Sosiologia*, *61*(3), 211–227.

Jørgensen, R. F. (2023). Data and rights in the digital welfare state: the case of Denmark. *Information, Communication & Society*, *26*(1), 123–138. https://doi.org/10.1080/1369118X.2021.1934069

Kapiszewski, D., MacLean, L. M., & Read, B. L. (2022). Dynamic research design: Iteration in field-based inquiry. *Comparative Politics*, *54*(4), 645–670.

Kaun, A., & Männiste, M. (2025). Public sector chatbots: AI frictions and data infrastructures at the interface of the digital welfare state. *New Media & Society*, *27*(4), 1962–1985. https://doi.org/10.1177/14614448251314394

Lindgren, I., Madsen, C. Ø., Rydén, H. H., Heggertveit, I., Hofmann, S., & Sæbø, Ø. (2024). What public services are suitable for digitalization? A classification of public service characteristics. In M. Janssen, J. Crompvoets, J. Ramon Gil-Garcia, H. Lee, I. Lindgren, A. Nikiforova, & G. Viale Pereira (Eds.), *Electronic government: EGOV 2024* (pp. 201–217). Springer.

Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public service* (30th ed.). Russel Sage Foundation.

Määttä, A. (2012). *Perusturva ja poiskäännyttäminen.* Diakonia-ammattikorkeakoulu.

Madsen, C. Ø., Lindgren, I., & Melin, U. (2022). The accidental caseworker—How digital self-service influences citizens' administrative burden. *Government Information Quarterly*, *39*(1), Article 101653.

Melin, U., Madsen, C. Ø., & Larsson, K. K. (2024). Five bureaucratic roles in the age of digital transformation—Insights from Scandinavian public organizations. In M. Janssen, J. Crompvoets, J. Ramon Gil-Garcia, H. Lee, I. Lindgren, A. Nikiforova, & G. Viale Pereira (Eds.), *Electronic government: EGOV 2024* (pp. 99–115). Springer.

MFF. (2024). *Digi ensin—Ohjelma*. https://vm.fi/hanke?tunnus=VM006:00/2024

Montgomerie, J. (2017). Iterative reflexive research strategy. In J. Montgomerie (Ed.), *Critical methods in political and cultural economy* (pp. 100–114). Routledge.

Moynihan, D., Herd, P., & Harvey, H. (2015). Administrative burden: Learning, psychological, and compliance costs in citizen-state interactions. *Journal of Public Administration Research and Theory*, *25*(1), 43–69.

MSH. (2023). *Digitaalisuus sosiaali—ja terveydenhuollon kivijalaksi: Sosiaali—ja terveydenhuollon digitalisaation ja tiedonhallinnan strategia 2023–2035*. https://julkaisut.valtioneuvosto.fi/handle/10024/165288

Nosratabadi, S., Atobishi, T., & Szilárd Heged, S. (2023). Social sustainability of digital transformation: Empirical evidence from EU-27 countries. *Administrative Sciences*, *13*, Article 126.

Peeters, R. (2023). Digital administrative burdens: An agenda for analyzing the citizen experience of digital bureaucratic encounters. *Perspectives on Public Management and Governance*, *6*(1), 7–13.

Peeters, R., & Widlak, A. C. (2023). Administrative exclusion in the infrastructure-level bureaucracy: The case of the Dutch daycare benefit scandal. *Public Administration Review*, *83*, 863–877.

Schou, J., & Pors, A. S. (2019). Digital by default? A qualitative study of exclusion in digitalised welfare. *Social Policy & Administration*, *53*(3), 464–477.

Soininvaara, H., Koulu, R., Snell, K., & Tarkkala, H. (2024). "Tee ensin testi": Digitaaliset esitestit julkishallinnon ilmiönä. *Yhteiskuntapolitiikka*, *89*(3), 307–313.

Srivastava, P., & Hopwood, N. (2009). A practical iterative framework for qualitative data analysis. *International Journal of Qualitative Methods*, *8*(1), 76–84.

Trifuljesko, S., & Ruckenstein, M. (2024). Algorithmic configurations in caring arrangements. *Big Data & Society*. Advance online publication. https://doi.org/10.1177/20539517241299726

Verne, G. B., Steinstø, T., Simonsen, L., & Bratteteig, T. (2022). How can i help you? A chatbot's answers to citizens' information needs. *Scandinavian Journal of Information Systems*, *34*(2), Article 7.

Wallace, W., Chan, C., Chidambaram, S., Hanna, L., Iqbal, F. M., Acharya, A., Normahani, P., Ashrafian, H., Markar, S. R., Sounderajah, V., & Darzi, A. (2022). The diagnostic and triage accuracy of digital and online symptom checker tools: A systematic review. *NPJ Digital Medicine*, *5*(1), Article 118.

Ziebart, C., Kfrerer, M. L., Stanley, M., & Austin, L. C. A. (2023). Digital-first health care approach to managing pandemics: Scoping review of pandemic self-triage tools. *Journal of Medical Internet Research*, *25*, Article 40983.

About the Authors

Heta Tarkkala is an academy research fellow at the University of Helsinki, Finland. She is a sociologist and STS scholar, who has studied the making of biomedical knowledge, datafication, and questions related to uses of health data, especially in the welfare state setting.

Riikka Koulu is an associate professor of social and legal implications of AI at the University of Helsinki, and the director of Legal Tech Lab, an interdisciplinary research hub at the intersections of law, technology, and society. She combines socio-legal studies with STS to examine law's interfaces and interactions.

Karoliina Snell is a programme director at the University of Helsinki. Her sociology of health and medicine and STS research focuses on acceptance, policies, and implementation of new technologies in health care. She has published on social, ethical, and legal aspects on genomics, biobanking, and digitalisation and datafication of health.

Helmi Soininvaara is a project researcher at the University of Helsinki Legal Tech Lab. Her current work focuses on technology policy and public sector digitalisation.

Social Inclusion
2025 • Volume 13 • Article 9998
https://doi.org/10.17645/si.9998

cogitatio

ARTICLE

Open Access Journal

Ethical Implications of AI-Driven Chatbots in Domestic Violence Support

Hanna Mielismäki and Marita Husso

Faculty of Social Sciences, Tampere University, Finland

Correspondence: Hanna Mielismäki (hanna.mielismaki@tuni.fi)

Submitted: 30 January 2025 **Accepted:** 26 June 2025 **Published:** 2 September 2025

Issue: This article is part of the issue "Fostering the Socially and Ecologically Sustainable Digitalisation of Welfare States" edited by Paula Saikkonen (Finnish Institute for Health and Welfare) and Marta Choroszewicz (University of Eastern Finland), fully open access at https://doi.org/10.17645/si.i514

Abstract

Our study explored the opportunities, challenges, and ethical considerations of using artificial intelligence (AI)-driven chatbots in domestic violence (DV) support. DV is a serious public health and social problem. Identifying it as early as possible is important in violence prevention. However, victim-survivors of DV are often reluctant to disclose violence, and service practitioners may lack the capacity or confidence to address violence-related issues. To tackle these challenges, the use of AI-driven chatbots presents opportunities to address DV by providing information and guiding users to appropriate services. However, interactions between humans and AI systems lie at the intersection of the human need for practical assistance and the risks inherent in digital communication—raising ethical considerations, particularly in the vulnerable context of DV. Semi-structured interviews with 25 victim-survivors, DV professionals, and criminal justice experts suggest that while the implementation of AI-driven chatbots can greatly enhance access to information, it also poses significant challenges related to safety and accountability. This is because interactions with chatbots lack essential elements for comprehensive situational assessment and documentation of DV cases, and for the establishment of a support network. These insights underscore the critical role of human interaction in addressing DV cases, while also highlighting the potential roles of chatbots as intermediate support systems for victim-survivors and as supplementary tools for welfare service practitioners in identifying different forms of DV and supporting the appropriate referral of cases. This study advances understanding of how AI-driven chatbots can be ethically and sustainably implemented in DV support systems.

Keywords

AI chatbots; digital ethics; domestic violence; gender-based violence; help-seeking; human-technology interaction; responsible AI; sustainability; technology-facilitated abuse; welfare services

1. Introduction: Current Challenges in Seeking Help for Domestic Violence

Domestic violence (DV) is a serious global public health and social problem. It may significantly damage the mental and physical health of victim-survivors of DV (Bellis et al., 2019; Ellsberg et al., 2015; Miller & McCaw, 2019; Stubbs & Szoeke, 2022). Furthermore, the widespread use of digital tools has introduced new forms of violence. For example, technology-facilitated abuse (TFA) involves controlling, monitoring, or harming victim-survivors through mobile phones and other digital devices (Boethius et al., 2023; Storer et al., 2023; Woodlock et al., 2020). This form of abuse enables DV perpetrators to continue exerting control even after separation, making them omnipresent in victim-survivors' lives. As a result, the abuse extends beyond physical encounters and intensifies the psychological distress experienced by victim-survivors (Afrouz, 2023; Harris & Woodlock, 2019). However, victim-survivors often hesitate to disclose their experiences to authorities—including social and healthcare practitioners, as well as the police—due to shame and the stigma associated with victimisation (Harris & Woodlock, 2019; Husso et al., 2021; Meyer, 2016; Piippo et al., 2021) or of being judged or not believed by service practitioners (Storer et al., 2022). In addition, victim-survivors are often unaware of the different forms of DV and which of those they are experiencing, their legal rights related to DV, and the procedures for disclosing DV to authorities (Decker et al., 2019; Saxton et al., 2021).

Research indicates that social welfare and healthcare service providers are in key positions to assess risks of violence, implement interventions, and organise support services (Husso et al., 2021; Miller & McCaw, 2019; Siltala et al., 2023). These practitioners assess individual circumstances, identify the issue's broader social context, and determine which other professionals or services are already involved (Trevithick, 2012). Collaborative, multidisciplinary approaches are particularly vital when working with clients facing complex challenges. The ultimate aim of these approaches is to empower and support the client—both materially and non-materially—in enhancing their own well-being (Trevithick, 2012). For example, victim-survivors typically need basic information about different forms of DV and guidance in accessing the necessary support services, including those related to housing and economic and mental recovery (Rogers et al., 2023).

Although DV, as a form of gender-based violence, has long been recognised as a human rights violation, there remains a pressing need for more effective interventions within the service system (Husso et al., 2021; Piippo et al., 2021; Turner et al., 2017; Wright et al., 2022). For example, in Finland, most victim-survivors remain unidentified by social and healthcare service providers despite using these services twice as often as individuals who have not experienced DV (Siltala et al., 2023). Practitioners typically consider DV-related problems either too challenging or too complicated to solve (Husso et al., 2021; Piippo et al., 2021). It is suggested that the insufficient actions of practitioners are due to their inadequate education on DV (Husso et al., 2021; Piippo et al., 2021; Siltala et al., 2023).

This gap in responding to DV appears to result from victim-survivors' reluctance to disclose their experiences of violence and service providers' limited capacity, confidence, or willingness to intervene in such situations. However, the physical, emotional, and social consequences of unaddressed and unrecognised DV may be serious for both victim-survivors and those witnessing the violence (Callaghan et al., 2018; Holt et al., 2008; Miller & McCaw, 2019; Stubbs & Szoeke, 2022). Early identification of DV is therefore essential for enhancing the safety of victim-survivors and preventing further harm (Siltala et al., 2023).

Artificial intelligence (AI) is increasingly being integrated into digitised welfare services, enabling novel approaches to addressing DV, such as the use of AI-driven chatbots to support cases that might otherwise remain unaddressed (Novitzky et al., 2023). These automated computer programs simulate human interaction through spoken, written, or visual communication and perform tasks that previously required a human agent (Henman, 2020). The interaction interfaces act as service assistants and are often integrated into service providers' websites or available as mobile applications. While general-purpose chatbots are designed to handle a wide range of topics, domain-specific chatbots are designed to meet specific needs, with their content typically curated by experts in the relevant field.

The benefits of these tools include handling routine activities, such as answering simple and repetitive questions and sharing information (Draughon Moret et al., 2022). The automation of routine tasks in welfare services has been justified by its potential to free up experts to focus on more demanding and critical responsibilities. Yet, distinguishing between tasks that require human judgment and those that can be automated remains challenging (Parviainen & Rantala, 2022). For example, Longoni et al. (2019) discovered that people perceive human practitioners as more capable of addressing their diverse needs and individual circumstances than machines. Nevertheless, the implementation of AI-driven chatbots has the potential to enhance service efficiency and end-user experience by providing instant responses to multiple user requests simultaneously (Nordheim et al., 2019). However, interactions between humans and AI systems lie at the intersection of the human need for practical assistance and the risks inherent in digital communication. This raises ethical considerations regarding the use of chatbots in the sensitive and vulnerable context of DV (Butterby & Lombard, 2025; Henry et al., 2024; McGreevey et al., 2020).

In this study, we explore the opportunities, challenges, and ethical considerations of using AI-driven chatbots in DV support from the perspectives of victim-survivors, DV professionals, and criminal justice experts. This research contributes to the ongoing discussion on the ethical and sustainable integration of chatbots into support systems for victim-survivors and to the broader agenda of sustainable development by addressing several of the United Nations (2015) SDGs. These goals include fostering societies free from fear and violence (SDG 16), advancing the use of technology to promote the empowerment of women (SDGs 5 and 9), enhancing the accessibility and reliability of welfare services (SDGs 10 and 16), and promoting health and well-being (SDG 3). Additionally, in 2024, the UN adopted a resolution to promote the safe, secure, and reliable development and use of AI systems that respect human rights and foster sustainable development, further emphasising the importance of ethical AI technology in achieving the SDGs (United Nations, 2024).

In studying AI-driven chatbots within the service system and DV support, an interdisciplinary approach is essential (Henry et al., 2024; Lindgren & Dignum, 2023). Our research draws on perspectives that are grounded in justice, equity, and the ethical use of technology in sensitive support contexts. We integrate insights from research on interpersonal violence and TFA with the normative ethical framework provided by the High-Level Expert Group on AI (AI HLEG, 2019), which outlines principles for the responsible development and use of AI. This combination enables a nuanced understanding of both the technological and human-centred dimensions of AI implementation in support services.

2. AI-Driven Chatbots as Tools for DV Support

Finnish people have the most positive attitudes towards AI among the European countries surveyed (Bergdahl et al., 2023). Finland pioneered digitalisation in Europe in the mid-1990s through governmental initiatives. Current government policy prioritises digital assessment of service needs and utilisation of AI in service guidance. Key public services are already fully available online, and further advancement of service digitalisation and automation is targeted (Finnish Government, 2024). Currently, 97.61% of Finnish internet users utilise eGovernment services, and 82% of the population have at least basic digital skills (European Commission, 2024a, 2024b).

As chatbots become more integrated into support services, it is important to understand how their design and use may influence users' experiences and well-being in both online and offline settings. Researchers have noted that trust in the offered digital service might be compromised if it fails in safety (Kretzschmar et al., 2019; McGreevey et al., 2020; Xu et al., 2021). Trust is considered a key element in the adaptation of new technology (de Visser et al., 2018; Jobin et al., 2019). For victim-survivors of DV who engage with chatbot services, perceptions of safety and reliability may affect their willingness to use such tools and the extent to which they find them helpful.

The European Union AI Act aims to regulate the development, deployment, and use of AI systems, with a particular focus on safety and reliability (European Union, 2024). In parallel, the AI HLEG (2019), established by the European Commission, developed the *Ethics Guidelines for Trustworthy AI*. These guidelines are grounded in fundamental rights and European values and define "trustworthy AI" as lawful, ethical, and robust throughout its lifecycle. However, the guidelines remain relatively general in nature, as the implementation of AI systems presents diverse challenges and risks across various fields. Therefore, technological solutions should be examined as part of a broader sociotechnical system (Abdelnour-Nocera & Clemmensen, 2019) to address the practical challenges and conflicts that may arise during the implementation of AI systems.

In this study, we examined the use of AI-driven chatbots in the context of DV and the support system for victim-survivors. Drawing on our findings, we considered how the ethical principles outlined by the AI HLEG (2019) could inform the implementation of these systems in supporting victim-survivors. Our research question was: What are the opportunities, challenges, and ethical considerations of using AI-driven chatbots in DV support? We focused on the four main principles proposed by the AI HLEG: prevention of harm, fairness, explicability, and respect for human autonomy. These guidelines emphasise both technical robustness and the user's safety, well-being, and human rights.

Regarding the first principle, the AI HLEG (2019) emphasises that technological innovations are expected to contribute to individual and societal well-being, while also minimising the risk of harm, whether intentional or unintentional (see also Braunschweig & Ghallab, 2021; Floridi et al., 2018). This is of particular concern in this study, as AI-driven digital tools for help-seeking should not further disadvantage those who are already in a vulnerable position. One major challenge is that the same digital tools that enable victim-survivors to seek help may also allow perpetrators to exert harmful power and control—for example, by hacking devices, installing spyware, or monitoring victim-survivors' online activity via mobile phones and computers (Afrouz, 2023; Boethius et al., 2023; Woodlock et al., 2020). Closely related to these risks, privacy and data security

are relevant concerns, as victim-survivors may disclose sensitive personal information when interacting with chatbots (Butterby & Lombard, 2025; Henry et al., 2024).

The second principle, fairness, is outlined by the AI HLEG (2019) as emphasising the need for AI systems to operate in ways that support just, fair, and unbiased interactions with users. According to the international ethical principles of social work (International Federation of Social Workers, 2018), fairness involves a commitment to equity and inclusion, including efforts to reduce barriers to access and to ensure that support reaches vulnerable or marginalised populations. In the context of DV, fairness refers not only to equal treatment but also to equity, which requires recognising and addressing structural inequalities that may affect different user groups. For example, the data used to train chatbots may reflect existing cultural, gendered, or societal biases, which can inadvertently reinforce stereotypes or marginalise victim-survivors (Bailey & Burkell, 2021). To address these concerns, Kretzschmar et al. (2019) note that the training data and outputs should be evidence-based, inclusive, and accountable. Furthermore, scholars have drawn attention to the importance of ensuring that individuals with varying levels of digital literacy, language proficiency, or disabilities can access and benefit from these tools (Mishna et al., 2021; Taylor, 2017). This perspective aligns with the AI HLEG's (2019) principle of fairness, which includes non-discrimination and equal participation. Accordingly, the implementation of chatbots can promote more equal access to information and services, as they have been found to be especially useful for people who might not otherwise seek help (Kabacińska et al., 2022; Montagni et al., 2020; Torous et al., 2021). This is particularly relevant for victim-survivors who are isolated due to their abusers' coercive control, which involves restricting their movements and limiting their access to resources and support networks (Al-Alosi, 2020; Novitzky et al., 2023).

Regarding the third principle, explicability, the AI HLEG (2019) has pointed out that the requirement for it in AI systems is context-dependent and becomes increasingly critical when the consequences of erroneous or misleading outputs are severe. In the context of DV, automated support tools, such as AI-driven chatbots, may pose risks to victim-survivors' decision-making processes, particularly if they place undue trust in the system's outputs or perceive it as capable of providing optimal advice or solutions (AI HLEG, 2019; Henman, 2020). Users should be able to distinguish whether they are interacting with a human or an AI system, as this distinction is found important for supporting informed consent and fostering trust (AI HLEG, 2019; Xu et al., 2021).

Finally, the fourth principle—respect for human autonomy—emphasises the importance of fostering equitable interaction and empowering victim-survivors to make informed decisions with the support of AI systems, particularly in contexts involving vulnerable groups (AI HLEG, 2019). For instance, the ability to seek help anonymously through chatbots can reduce victim-survivors' fear of humiliation or judgment by service practitioners when they disclose deeply personal and emotionally sensitive experiences (Storer et al., 2023). This sense of anonymity has been shown to empower victim-survivors and increase their confidence in taking further steps to access support services (Al-Alosi, 2020; Novitzky et al., 2023). However, manipulative forms of guidance by AI systems have been identified as a potential threat to autonomous decision-making, as they may limit users' ability to act freely and independently (Devillers et al., 2021).

3. Methods

This study was conducted as part of a Horizon Europe project titled Innovative Solutions to Eliminate Domestic Abuse (ISEDA), which involved partners from nine European countries: Bulgaria, Catalonia (Spain), Cyprus, Greece, Finland, France, Germany, Italy, and Scotland. One of the key objectives of this project was to develop and pilot a chatbot for women victim-survivors of DV. Seven countries involved in the project conducted focus groups and/or individual interviews to better understand the challenges in seeking help for DV, as well as the positive and negative perceptions of chatbots.

Focus groups were considered suitable for this purpose, as they enabled participants to articulate their views through interaction with others in a shared setting. To support open and productive discussion, the groups were composed to be relatively homogeneous in terms of participants' background and familiarity with the topic, which is known to facilitate communication (Bryman, 2004; Fern, 2001). At the same time, differences in individual experiences contributed to a range of perspectives on the phenomenon (Kitzinger, 1994). First, we asked the participants about their perceptions of the challenges in seeking help for DV from public services, including social and healthcare services and the police. Then, we asked them to reflect on the opportunities, challenges, and risks associated with the use of AI-driven chatbots to seek help for DV. We defined a chatbot for them as a computer program designed to simulate conversations with humans.

In this study in Finland, we conducted 13 semi-structured interviews in 2023. In total, 25 informants were interviewed, either individually or in groups of 2 to 3. The details are outlined in Table 1. The interview format was selected based on participants' preferences and practical factors such as availability and comfort in discussing sensitive topics (Lazar et al., 2017).

Table 1. Interview formats and informant categories.

	Focus group interviews		Individual interviews
	Number of groups	Informants per group	
Victim-survivors	1	2	3
DV professionals	5	2–3	1
Criminal justice experts	1	3	2

The victim-survivors were women over 18 years of age whom violence prevention organisations selected for the interviews. These women acted as peer support in these organisations. The organisations ensured that these women were not under threat of violence before inviting them for an interview. The professionals and experts represented civil society organisations, as well as national and regional governmental administrations from different parts of Finland. Many of them had extensive experience in the field of DV and maintained a broad network within the relevant service system, including social and healthcare workers and the police.

We provided all the informants with detailed information about the purpose of this study and obtained their verbal informed consent to participate prior to conducting the interviews. Each interview lasted 53 to 82 minutes. We recorded a total of 14 hours of audio material and transcribed it for analysis. We removed all identifying information before completing the analysis and identified the informants only by their interview numbers (e.g., G3 for Group 3) and participant numbers (e.g., R2 for Respondent 2). The original interview data are in the Finnish language and the interview extracts used in this article are English translations of them.

We analysed the data using theory-driven thematic content analysis (e.g., Braun & Clarke, 2006), not directly based on theory but guided by concepts and findings on the topic from previous research. This approach enabled us to identify and examine patterns in the informants' understanding and experience of the process of seeking help for DV, and their perceptions related to it. The goal was to identify relevant ideas that repeatedly arose from the data.

In the first phase of the analysis, we repeatedly read the entire dataset to understand all aspects of the data. In the second phase, we systematically coded the dataset by identifying and labelling facilitating and challenging elements related to help-seeking for DV, interactions with authorities, and the use of chatbots. In the third phase, we began organising these codes into potential themes and sub-themes, using tables to support the sorting and comparison process. In the fourth phase, we reviewed the coded interview extracts for each theme to ensure their coherence and consistency with the research question. In the fifth phase, we examined the themes based on both the literature and the informants' perspectives and then deepened the analysis to identify the opportunities and risks associated with the use of AI-driven chatbots. Finally, in the sixth phase, we reflected on the findings through the lens of the AI HLEG's (2019) four ethical principles: prevention of harm, fairness, explicability, and respect for human autonomy.

This entire study complied with the ethical guidelines of the Finnish National Board on Research Integrity TENK (2023). The Ethics Committee for Human Sciences in the Tampere Region issued a statement on this study's ethical acceptability. Additionally, the necessary research permits were obtained from the interviewees' respective organisations.

4. Results: Opportunities and Challenges of AI-Driven Chatbots in DV Support

This section presents the interviewees' perceptions of AI-driven chatbots in the context of DV support, highlighting both opportunities and challenges associated with their use. The identified opportunities are (a) providing information and guidance during the early stages of DV, (b) offering a non-judgmental space for sharing experiences, and (c) reducing the emotional strain of violence disclosure. The identified challenges are (a) absence of case documentation, (b) lack of holistic situational assessment, and (c) concerns related to accountability and safety.

4.1. Providing Information and Guidance During the Early Stages of DV

All the interviewees emphasised the importance of knowledge about the different forms of DV and its multifaceted impacts on victim-survivors, including its physical, psychological, and social consequences. However, they expressed disappointment in what they perceived as the authorities' insufficient knowledge of DV. In contrast, they viewed AI-driven chatbots as more consistent in providing standardised and timely information, which they considered more reliable in certain situations. Accordingly, participants viewed these systems as low-threshold support tools during the early stages of DV, when abuse is just beginning to occur, to help prevent violence from becoming normalised in victim-survivors' lives. The following criminal justice expert pointed out an important aspect of the challenges facing those who disclose violence:

> A computer-based channel like this, where a person can reflect on their own experiences safely and there is no threshold for whether they are believed or whether their case is too trivial, or whether they

> are burdening these professionals unnecessarily, will probably be able to dispel human doubts about whether the person's situation is serious enough to warrant help. (Criminal justice expert, G1R1)

Victim-survivors often consider themselves not "ideal" victims worthy of support (Meyer, 2016) and feel that they have the burden of proof of being targets of violence (Piippo et al., 2021). The interviewees also expressed that victim-survivors need information on support services, including legal, medical, and psychological support. Accordingly, they perceived that the use of AI-driven chatbots in DV support could lower the threshold for seeking help by providing accessible information and guidance. This, in turn, was seen as potentially encouraging victim-survivors to reflect on their situations and make more informed decisions about how to proceed. The victim-survivors expressed that knowledge about DV is essential for them to prevent violence from becoming normalised, which Piippo et al. (2021) found to be the case for many victim-survivors. For example, the following interviewee described how she needed guidance in these matters instead of being belittled by service practitioners:

> I feel that I was not taken seriously. I was hushed up and told that it would be alright. And now, years later…it's not OK. If someone had taken me seriously and told me about my rights and the rights of children, what's right and what's wrong, that would certainly have helped. (Victim-survivor, G10R1)

> We victims are terribly skilled at not seeing the problem as big as it actually is, especially in the early phase. At that phase, AI could be useful. (Victim-survivor, G10R2)

These quotes highlight an important aspect of experiencing violence: victim-survivors may question their experiences and thereby prolong their suffering. For example, abusers may use coercive control, distort their targets' perceptions of reality, and erode their autonomy (Stark, 2007). The institutional practices of bypassing and silencing problems related to DV that G10R1 described may enforce the normalisation of violence.

Accordingly, most of the participants mentioned that the use of chatbots could facilitate the disclosure of DV by guiding victim-survivors on how to proceed with their cases. In addition to supporting victim-survivors directly, both the DV professionals and the criminal justice experts considered chatbots as potential tools to assist the social and healthcare practitioners and the police in facilitating the disclosure of DV and proceeding systematically with DV cases. The following professional described the systemic gap in which victim-survivors hesitate to disclose violence, while service practitioners may not be able to properly identify a case of DV—highlighting how chatbots could also support professionals in recognising and responding to such cases:

> A person who repeatedly visits health care services with various symptoms but is never asked about violence and does not disclose it themselves—we have considered this a situation where AI could assist: to guide, explore, and ask about the theme of violence. (DV professional, G3R3)

> I believe AI can certainly ask specific questions, perhaps even better, and will remember to ask all of them, provided the settings ensure that all necessary questions are included. In this way, human omission can be avoided. (Criminal justice expert, G2R1)

G2R1 described the mechanical feature of the AI systems as a benefit. This insight illustrates the diverse needs of different users in different contexts. This expert referred to the preliminary questioning, where contextual

matters, such as the hectic hours of a weekend, may affect a police officer's ability to routinely conduct all necessary questions.

4.2. Offering a Non-judgmental Space for Sharing Experiences

The interviewees perceived AI-driven chatbots as providing a supportive space for victim-survivors to share their experiences without the fear of being judged by the other party. The victim-survivors noted that the other party's supportive interaction and acknowledgement of their feelings are essential in vulnerable situations, as these show that their experiences are taken seriously. They described many ways in which victim-survivors are often ignored by authorities, including social and healthcare workers and the police. These approaches were associated with negative interactions, such as belittling and victim-blaming, which were perceived as negatively affecting the likelihood of them getting the help they need. In contrast, the interviewees perceived AI systems as lacking such attitudinal attributes:

> There is, of course, a risk if AI also ends up belittling the seriousness of the problem. That is, what kinds of questions and response options are created for early-stage support—and what kinds of paths the AI then leads the user into—there are enormous risks in that. But somehow, I see that these risks are not as big…as if you call the family counselling centre or the deaconess of a church and then they say: "Nonsense, try to get along." So, in a way, the bot would never respond like that. It would aim to resolve the situation. (Victim-survivor, G10R1)

> It significantly lowers the threshold when you do not have to disclose violence to potentially contemptuous and arrogant male police officers. Instead, if this process could be neutral, most people would likely prefer to tell an AI what has happened rather than fear being belittled for their experiences, which creates a very high threshold. (DV professional, G7R1)

The fear of being belittled by service practitioners hinders victim-survivors from disclosing violence, as seen in these quotes. G10R1 implied the importance of actions and the need to solve problematic situations. Acknowledging victim-survivors' experiences is essential for empowering them and helping them make sense of their situations (Keeling & Fisher, 2015; Wood et al., 2022). Without validation of their feelings, they may question their experiences and become accustomed to living in violent situations (Piippo et al., 2021).

However, the victim-survivors also expressed the need for genuine human connections when facing life challenges and being in vulnerable situations. The following quotes highlight the contrast between programmed interaction and human interaction:

> The victim has a need to be heard, and with a bot, they do not necessarily feel that they are genuinely heard, as it is only a program that answers as it is programmed to. (Victim-survivor, G8R1)

> At that moment, the victims long especially for humanity from these social and healthcare services. The conversational bot is anything…but humane….Typically, the victim feels alone in their situation and as if they are fighting against windmills. Especially in these situations, they long for human contact and to be seen as human. (Victim-survivor, G8R1)

These quotes emphasise the importance of distinguishing between hearing and listening. Whereas listening is often perceived as an easy or passive task, it is, in fact, a complex skill that requires conscious effort and practice to build trust and understanding (Trevithick, 2012).

4.3. Reducing the Emotional Strain of Violence Disclosure

According to the interviewees, fear, as a human emotion, can act as a barrier to accessing support, as, in their experience, it was linked to practitioners' reluctance to engage with DV issues. In contrast, the interviewees believed that the use of AI-driven chatbots could facilitate help-seeking by automatically presenting the necessary questions about violence, raising hopes for a more systematic handling of cases. For example, Husso et al. (2012) and Piippo et al. (2021) found that social workers may fear asking clients about DV because they worry about their clients' reactions or the potential emotional burden it places on the practitioners themselves, such as anxiety, frustration, or sadness:

> People might fear hearing about violence, which can prevent them from wanting to get involved. And I think that's quite common, at least, based on my experience….AI doesn't have that need to protect itself, which can lead to better solutions. (Victim-survivor, G9R1)

> There could be something that removes the human element in such a way that it doesn't always depend on someone's opinion whether the matter progresses or not. For example, some kind of entirely computer-based, AI-driven form system that would automatically score risks. (Victim-survivor, G9R1)

When the victim-survivors discussed interactions with human practitioners and AI, they often pondered the meaning of humanity and the non-human attributes of AI in interactions. The presence or absence of humanity seemed to both facilitate and hinder access to support. While the human practitioners' empathetic approach was considered a fundamental requirement for successful encounters, human emotions were seen as a burden to the successful progression of processes. As seen in the quotes, the victim-survivors regarded AI as a tool that can be designed to minimise subjective variability in interactions and to handle their cases more systematically.

The interviewees acknowledged that DV is deeply institutionalised and reflected in the attitudes of service practitioners. However, the interviewees also acknowledged the risk of human biases being embedded in AI systems through coding and training processes, as one DV professional noted:

> You have to be very careful about what database a chatbot uses so that the general biases of people are not transferred to it. (DV professional, G6R1)

This quote implies the variety of chatbots and the challenge of knowing which of them is reliable (Kabacińska et al., 2022; Montagni et al., 2020). Draughon Moret et al. (2022) found that the quality of apps varies significantly. Nevertheless, some of the interviewees saw more potential for eliminating negative bias from AI systems than from human practitioners.

4.4. Absence of Case Documentation

The victim-survivors regarded anonymous interaction with chatbots as beneficial for those who are not yet ready to take further steps in their current situation. This allows them to gain knowledge anonymously and decide whether to proceed with their cases, thereby avoiding unwanted consequences from authorities, such as child welfare notifications or the child being taken into care by social services. The victim-survivors explained that they might fear disclosing violence due to concerns about case documentation, as one interviewee expressed:

> The church deacon is the only person whom you have the courage to approach and talk to,…not the authorities, because they always document and save everything. And the consequences are much worse than…if you simply do not seek help and just stay quiet….if only you can just…seek help if you are in a violent relationship without having to fear that if you…disclose violence, they will take your children away or you will end up in a terrible hassle. (Victim-survivor, G5R1)

This quote shows a tension between the need to seek help and the fear of consequences. It also illustrates the pressure victim-survivors face when they have to stay in a violent relationship due to the fear of disclosing violence. This shows that victim-survivors need more decision-making power regarding how their cases are handled by authorities. However, avoiding documentation contradicts the professionals' views, which highlight the importance of documentation, as described in the following quote:

> How does the information transfer and get recorded or documented if the conversation with the AI doesn't lead to anything at that moment, but it leads to something two months later, or if the same person seeks help multiple times?...One would always hope to know at what point a human should get involved….Banks already operate like this. Just yesterday, I dealt with a bank bot first, and the bot concluded that a human was needed. (DV professional, G12R2)

The professionals noted that repetitive incidents of violence would remain undocumented if victim-survivors interacted only with chatbots. DV is typically an ongoing process rather than a single incident (see Stark, 2007); however, individual contacts with chatbots remain separate and do not form a client case that allows for the tracking of history or progression. The professionals acknowledged this nature of the phenomenon of DV by pointing out the need to regard DV cases as a process. Furthermore, because violence affects the entire family (Callaghan et al., 2018; Holt et al., 2008), the professionals were especially concerned about potential emergency child protection needs and the limitations of chatbots in identifying the criticality of situations:

> In these relationships, there are, on average, at least two to three underage children, especially among those that have lasted a long time. So, how does the situation unfold, considering that the need for child protection can be immediate, especially in urgent situations like these? (DV professional, G12R1)

Failing to recognise immediate risks and the need for instant reaction could endanger the safety and well-being of victim-survivors and their children. According to national statistics in Finland, in 65% of cases where violence has been recurring, children have been repeatedly exposed to violence in their families (Attila et al., 2023). However, the need for child protection may remain undiscovered with the use of chatbots, as expressed by DV professionals.

4.5. Lack of Holistic Situational Assessment

DV professionals expressed scepticism regarding the capacity of chatbots to adequately respond to crises and potential traumas caused by violence, which they viewed as requiring a dynamic and multidimensional approach. For example, one professional referred to the seriousness of DV, with potential lethal consequences, and raised concerns about the extent to which chatbots can execute the required critical situational assessments, whereas human practitioners were considered to have gained deep empirical knowledge through long experience in assisting victim-survivors:

> The assessment of the risk of violence is very dynamic and highly multidimensional. It includes a strong qualitative element, and although there are certain systematic aspects that can be considered, there is also a significant amount of experiential knowledge that develops only through direct encounters and over the years. So, when we talk about threats to life and health, to what extent can a chatbot reliably make such an assessment? (DV professional, G12R2)

As shown in this quote, recognising and responding to violence-related problems requires profound interaction and holistic observation of situations. According to Trevithick (2012), the initial meeting with a client should emphasise the use of open-ended questions to gain a deeper understanding of the client's situation and circumstances, including the broader social context. While chatbots are considered beneficial in answering simple and repetitive questions (Draughon Moret et al., 2022), in DV cases, there are no simple answers in diverse situations, as the following professional noted:

> At least, from the perspective of harassment, these cases are so terribly complex, and I don't know if there are any straightforward answers in the end. The amount of data that would need to be input is quite considerable. Even though I have had such a long work history in this field, almost every day, I encounter questions that make me think: "Wow, I've never heard of that before" and "I don't really know what advice to give or what to do here; let's think about it together," as there aren't really any good solutions. (DV professional, G7R2)

This quote highlights another essential element of situational assessment: Practitioners also use creative problem-solving to assess various solution options in complex DV cases where potential risks need to be assessed from various angles. Considering the amount of data required to address all nuanced DV scenarios, doubts arose about the capability of AI-driven chatbot systems to perform these kinds of tasks. Furthermore, the DV professional referred to the overwhelmingly complex nature of stalking cases, as digital devices have enabled the continuation of violence even after separation.

Accordingly, a substantial element of holistic assessment includes observing signs of TFA, about which most of the interviewees raised concerns. The intimate matters and vulnerabilities of victim-survivors may be exposed to hacking, misuse of information, or the installation of spyware by a current or ex-partner:

> What evidence do women have of this? If it's not bruises on their bodies, then it's in their stories or, just like R2 said, they should be able to show if there are message threads, or if a tracking program has been installed on their phone, or cameras at home, like last time, when cameras were eventually found in every room upon investigation. But how does the bot verify these? (DV professional, G3R3)

This form of abuse can significantly impact the safety and privacy of victim-survivors, making it essential to recognise the signs of abuse also in the environment where victim-survivors live. Besides determining the criticality of victim-survivors' circumstances, the situational assessment must succeed in preventing the escalation of violence. The DV professionals highlighted the need for more holistic observations than a mechanical assessment provided by AI, as it may not capture important signs of abuse, including nonverbal cues:

> How can victim-survivors recognise their own situation and the related issues if they have lived in it for a long time? Of course, the professional also makes other observations besides just ticking boxes. All of that can easily be missed if it's entirely done by AI, replacing the human. I see more risks in that. (DV professional, G3R1)

G3R1 referred to the tendency of violence to become normalised in victim-survivors' lives (see Husso et al., 2012; Piippo et al., 2021). However, violence typically becomes more serious over time, making early identification of DV essential and potentially life-saving. Such identification often relies on holistic observations, including listening, body language, and gaze direction (Reith-Hall & Montgomery, 2022). Digital technologies are not considered capable of emulating human characteristics in this sense, as they cannot read body language or perceive hurt and pain (Storer et al., 2023).

While the interviewees highlighted that chatbots may lack the capacity to appropriately process the complex and nuanced situations of victim-survivors, this challenge was described as even more pronounced for victim-survivors from immigrant backgrounds, who often face linguistic barriers and cultural differences in the meanings of concepts. Concerns were raised about chatbots potentially misinterpreting victim-survivors or providing incorrect instructions, especially when the language used was not standard or contained errors:

> If the client writes something incorrectly, will they get the right kind of help or the right answers? Or will they get something completely different from what they need? Also, if it's the first contact and you send a message and it doesn't work and feels frustrating, will you then seek help from anywhere else, or will you think, "I just can't do this anymore—I'm done"? (DV professional, G4R2)

> Whenever I've tried using a chatbot, I've never gotten the right answer through it. Especially the fact that you can't have any typos if you're trying to write in Finnish. It doesn't understand if you write something incorrectly. (DV professional, G4R1)

These reflections illustrate how easily misunderstandings and misguidance can occur in chatbot interactions, particularly for users with limited language proficiency. The inability of chatbots to process non-standard language or accommodate linguistic variation was identified as a significant barrier to effective support. Thus, special attention should be paid to language, imagery, and wording in the chatbot development process (Butterby & Lombard, 2025). Ensuring that these elements are carefully considered can help mitigate some of the risks associated with chatbot interactions.

4.6. Concerns Related to Accountability and Safety

The DV professionals were concerned that the responsibility for initiating contact with support services might fall on the victim-survivors once they had been provided with contact information by a chatbot. This was of special importance because the participants acknowledged the vulnerable position of victim-survivors due to the threat of violence and their potentially lowered functional capacity due to violence. Hence, there is a contradiction between their vulnerability and the expected strong agency to take action to advance their case.

> It seems that when interacting with an AI chatbot, much of the responsibility—including assessing the situation and maintaining the connection—is left to the person seeking help. (DV professional, G3R2)

> After certain programmed answers or questions, the person should be directed forward very quickly—because currently, seeking help and reporting violence are different matters. (Criminal justice expert, G2R1)

The DV professional pointed out the safety aspect: The need for help may not become apparent to service providers if interactions are limited to a chatbot. The criminal justice expert stated that these tools should be designed to detect indicators of emergencies and initiate the necessary predefined steps to ensure that the situations are reported and victim-survivors are guided forward. Many DV professionals pointed out that, for example, social work practitioners have a responsibility to ensure that victim-survivors have a safety network and that their children are protected from violence. Hence, there seems to be a lack of agency in the process of seeking help via chatbots.

> I strongly advocate for a human touch alongside this, as receiving emotional and psychosocial support is equally important. How can we ensure that this information flows while also connecting the help to the specific person? (DV professional, G3R3)

The interviewees highlighted that victim-survivors typically need multiprofessional support to secure safe and lasting disengagement from their abusers. The DV professionals were concerned that no one would be held accountable for victim-survivors' safety. On the whole, the interviewees considered DV too severe and complex to be entrusted to chatbots and expressed the need for a comprehensive assessment of the criticality of situations and concrete measures to ensure the safety of victim-survivors and their children.

Long-standing structural barriers to addressing DV were highlighted in the following comment, in which the respondent questions whether chatbots in DV support could overcome the challenges embedded in social structures and institutional practices:

> I do not see any added value a chatbot would have compared to what already exists. This is such a serious matter, and I feel that it requires a human touch. This problem is so deeply rooted in Finnish society. (Victim-survivor, G10R2)

This comment underscores the importance of critically assessing which specific issues technology can effectively address. On a similar note, the comment below acknowledges that while chatbot technology may help victim-survivors recognise DV and access information about support services, the broader service

system continues to pose structural challenges—particularly related to resources. Disclosing violence requires that practitioners have both the capacity and resources to respond to DV, offer support services, and handle a potential increase in criminal reports from victim-survivors:

> At the moment, the resources are probably insufficient, and knowing how long the criminal proceedings currently are, if there's even a little more pressure, it just means the process will be extended for everyone involved. (Criminal justice expert, G11R2)

Criminal processes are already lengthy and mentally exhausting, often leaving victim-survivors without adequate support before, during, and after legal proceedings. If these processes are further prolonged, their safety may be compromised. Several victim-survivors pointed out that one reason for not reporting violence is the lack of safety guarantees—even when their primary hope is simply for the violence to stop.

5. Ethical and Sustainable Integration of AI-Driven Chatbots in Support Services

To understand the ethical implications of using chatbots in DV support services, the findings were reflected on considering the ethical principles of prevention of harm, fairness, explicability, and respect for human autonomy. The results indicate that the implementation and use of AI-driven chatbots both align with these ethical principles and face challenges regarding them, illustrating how abstract guidelines can encounter various practical challenges depending on the context in which these AI systems are deployed.

Considering the ethical principle of preventing harm, the results suggest that the implementation of AI-driven chatbots may be consistent with this principle in certain contexts, while also revealing inconsistencies or limitations in practice. The challenges lie particularly in the absence of a responsible party to take the necessary concrete actions for organising multidisciplinary support and establishing a safety network for victim-survivors and their children. The interviewees pointed out that current AI-driven chatbot systems have limitations in situational assessment, particularly in processing contextual information holistically and in identifying the criticality of situations or the potential for digital abuse. These concerns resonate with findings from previous studies. For example, Palanica et al. (2019), who explored the use of healthcare chatbots, referred to the need for personalised knowledge in complex cases—a need that was consistently highlighted across all focus groups in the present study. Similarly, Emezue et al. (2022) identified the problem of accountability in digital interventions. They highlighted the need to critically consider the demand for agency that is placed on victim-survivors, as they may be living under constant threat of violence. In such circumstances, the expectation of victim-survivors' agency may seem unreasonable, considering how significantly the violence may impair their ability to function (see Miller & McCaw, 2019; Stubbs & Szoeke, 2022).

Expanding the discussion on harm prevention, all focus groups acknowledged the risks of being exposed to TFA when using a chatbot to seek help, in line with recent research (Afrouz, 2023; Boethius et al., 2023; Storer et al., 2023; Woodlock et al., 2020). As Afrouz (2023) noted, victim-survivors may struggle to identify TFA, especially if they lack technical skills. However, as shown by Boethius et al. (2023) and Novitzky et al. (2023), service practitioners also face challenges in addressing TFA. Similarly, the present study raised concerns about the extent to which AI-driven chatbot systems can support the detection of potential misuse of the device on which they are deployed. Hence, the results indicate that the responsibility for assessing the safety of a

chatbot, verifying the validity of its responses, and taking further actions, including reaching out to support services, falls on victim-survivors.

These concerns underscore the importance of establishing clear criteria for assessing the reliability of digital tools (see Kabacińska et al., 2022; Montagni et al., 2020), given the variability in app quality that Draughon Moret et al. (2022) found in their study. As Boethius et al. (2023) noted, victim-survivors should not be forced to limit their online presence, as technology can also empower them in various ways, including by providing them with access to support and information. Therefore, chatbots could, for example, be integrated into existing welfare service websites, which citizens are familiar with and may rely on more than apps from different sources. This would also address the structural challenges of scattered services (see Husso et al., 2021) and prevent the recurrence of systemic deficiencies in the digital dimension.

While the ethical principle of preventing harm raised concerns in some respects, the integration of AI-driven chatbots may also support it by lowering barriers to help-seeking—a dynamic also observed in prior studies on digital support tools (Al-Alosi, 2020; Kabacińska et al., 2022; Montagni et al., 2020; Novitzky et al., 2023; Torous et al., 2021). The interviewees perceived the automated functions of AI systems—such as systematically asking about experiences of DV—as encouraging victim-survivors to speak about issues that they might otherwise consider too emotionally exhausting for service practitioners to address. This perception resonates with earlier findings (Husso et al., 2012; Piippo et al., 2021; Virkki et al., 2015), which suggest that practitioners' own emotions can hinder them from intervening in DV.

Reflecting this, many interviewees saw chatbots as potentially useful tools for supporting practitioners in their work with DV cases by systematically asking the necessary questions to enhance the disclosure of violence. Hence, this could partly address the systemic gap whereby victim-survivors are reluctant to disclose violence (see Harris & Woodlock, 2019; Husso et al., 2021; Meyer, 2016; Storer et al., 2022) and service practitioners may struggle to effectively identify and intervene in DV cases (see Siltala et al., 2023; Turner et al., 2017; Wright et al., 2022). However, practitioners' motivation to use chatbots as tools could be questioned, given the insufficient use of tools specifically developed to map experiences of violence. For example, in Finland, over 50% of police officers and social and healthcare service workers use no tools at all for mapping DV (Niklander et al., 2019).

Turning to the ethical principle of fairness, the results show that the implementation of AI-driven chatbots can enhance access to information, thereby increasing awareness of DV and the legal rights of victim-survivors. These findings align with recent studies that suggested chatbots' potential to enhance confidentiality and to encourage victim-survivors to make an initial outreach for help from support services (Butterby & Lombard, 2025; Novitzky et al., 2023; Storer et al., 2022). Receiving information on a low threshold is both fair and empowering, aside from respecting victim-survivors' autonomy to make informed decisions about their situations. However, the results also indicate unequal opportunities to access information, especially for victim-survivors from immigrant backgrounds who face linguistic barriers. Additionally, Emezue et al. (2022) highlighted that victim-survivors with diverse ethnic, cultural, and linguistic needs may face obstacles in using DV apps. This shows how vulnerabilities may intersect and accumulate, and how, to prevent intersectional inequalities, it is important to perceive the diversity of capabilities to utilise digital services (AI HLEG, 2019; Bailey & Burkell, 2021; Mishna et al., 2021).

With regard to the ethical principle of explicability, the findings highlight challenges in ensuring that AI systems are understandable and interpretable to all users. As the interviewees noted, misunderstandings and frustration caused by unsuccessful interactions can ultimately deter victim-survivors from seeking help. Earlier research suggests that trust and perceived safety are central to users' willingness to engage with digital services (de Visser et al., 2018; Jobin et al., 2019; Kretzschmar et al., 2019; McGreevey et al., 2020). If a chatbot fails to communicate effectively due to language or cultural barriers, its functioning may become opaque to the user, thereby undermining both the user's trust and their ability to make informed decisions—echoing findings by Mishna et al. (2021) and Xu et al. (2021). Furthermore, in the present study, DV professionals expressed significant concerns about the reliability of chatbot-generated responses in complex and nuanced situations, particularly in critical cases requiring immediate action to protect victim-survivors and their children, a concern also raised by Butterby and Lombard (2025).

Finally, reflecting on the ethical principle of respect for human autonomy, the results indicate challenges in balancing victim-survivors' self-determination and institutional responsibilities, particularly in relation to anonymous user interactions with chatbots. Based on the victim-survivor interviews, these interactions were perceived as retaining victim-survivors' decision-making power over the progression of their cases and as possibly helping them avoid unwanted consequences by authorities, such as child protection interventions. However, this presents difficulties in ensuring comprehensive support and accountability. While anonymous interaction may support autonomy and reduce fear of judgment (see Al-Alosi, 2020; Novitzky et al., 2023; Storer et al., 2023), DV professionals emphasised the importance of intervention and documentation of cases. They highlighted that emergency situations requiring child protection or cases of recurring violence would remain undocumented if only chatbots are interacted with. This concern is even greater when considering how underreported violence has been within the service system (see Husso et al., 2021; Piippo et al., 2021; Siltala et al., 2023), potentially creating tensions between respect for victim-survivors' autonomy and the legal obligation of authorities to protect children. This tension highlights that the balance between different ethical principles may vary depending on the end-users' needs.

6. Conclusion

The findings from the present study suggest that while the implementation of AI-driven chatbots can significantly enhance access to information on DV and the legal rights of victim-survivors, it also raises ethically significant concerns regarding safety and accountability. Interactions with chatbots lack essential elements for comprehensive situational assessment, documentation of violence, and the formation of a support network. Hence, the results highlight the importance of human interaction in DV cases, aligning with recent studies (Butterby & Lombard, 2025; Domingo-Cabarrubias et al., 2023; Henry et al., 2024). However, the need for human interaction also requires that service practitioners possess the essential professional skills and knowledge to address DV. This requirement remains a structural deficiency yet to be resolved (Husso et al., 2021; Niklander et al., 2019; Piippo et al., 2021).

The results further emphasise the need to consider the vulnerabilities of victim-survivors, ensuring their safety and acknowledging the severity of DV, while questioning which specific issues technology can effectively address. Attempting to address problems that are structural and deeply embedded in institutional practices with technological solutions may divert attention from efforts to strengthen existing service systems and practices (see Hunt et al., 2020; Lindgren & Dignum, 2023). Therefore, it is essential to ensure

that the implemented technology reinforces the service system while supporting and strengthening practitioners' expertise in addressing DV (see Emezue et al., 2022; Hunt et al., 2020; Inkster et al., 2018). This study thereby suggests that AI-driven chatbots could serve victim-survivors as an intermediate support system, aligning with findings by Henry et al. (2024). In addition, they may function as a supplementary tool for welfare service practitioners to identify DV and proceed with cases—an area that remains underexplored in existing research.

DV causes a wide range of social and health problems and significant human suffering. Developing effective intervention practices requires a broad understanding of the distinctions and interconnections between online and offline domains as well as the opportunities and limitations of AI-driven solutions. Meeting these challenges demands comprehensive ethical consideration of safety, accountability, and the potential impacts of automated interventions.

While this study contributes to the ethical and practical discourse on AI-driven interventions in DV support, it is not without limitations. First, the conceptual distinction between online and offline domains often fails to capture the experiences of victim-survivors, who navigate violence across both spheres, where digital and physical realities are deeply interwoven. This ambiguity presents challenges not only for victim-survivors but also for practitioners and researchers seeking adequate language to describe AI-mediated interactions. Second, additional research is needed to examine the user experiences of AI-driven chatbots specifically designed for victim-survivors. Third, while many interviewed DV professionals had relevant education and experience, social and healthcare service practitioners were not directly interviewed. Furthermore, this study identified several areas requiring further exploration, including the tension between respecting human autonomy and the duty of authorities to document DV cases, as well as the potential for practitioners to incorporate AI tools into violence prevention and intervention strategies.

Acknowledgments

We would like to thank our interviewees for sharing their time and expertise, and we would also like to thank the anonymous reviewers for their valuable feedback throughout the process.

Funding

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101073922 (ISEDA). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them. Publication of this article in open access was made possible through the institutional membership agreement between Tampere University and Cogitatio Press.

Conflict of Interests

The authors declare that there is no conflict of interest.

References

Abdelnour-Nocera, J., & Clemmensen, T. (2019). Theorizing about socio-technical approaches to HCI. In B. R. Barricelli, V. Roto, T. Clemmensen, P. Campos, A. Lopes, F. Gonçalves, & J. Abdelnour-Nocera (Eds.), *Human work interaction design. Designing engaging automation* (Vol. 544, pp. 242–262). Springer. https://doi.org/10.1007/978-3-030-05297-3_17

Afrouz, R. (2023). The nature, patterns and consequences of technology-facilitated domestic abuse: A scoping review. *Trauma, Violence, & Abuse*, *24*(2), 913–927. https://doi.org/10.1177/15248380211046752

Al-Alosi, H. (2020). Fighting fire with fire: Exploring the potential of technology to help victims combat intimate partner violence. *Aggression and Violent Behavior*, *52*, Article 101376. https://doi.org/10.1016/j.avb.2020.101376

Attila, H., Keski-Petäjä, M., Pietiläinen, M., Lipasti, L., Saari, J., & Haapakangas, K. (2023). *Sukupuolistunut väkivalta ja lähisuhdeväkivalta Suomessa 2021: Loppuraportti*. Statistics Finland. https://urn.fi/URN:ISBN:978-952-244-717-3

Bailey, J., & Burkell, J. (2021). Tech-facilitated violence: Thinking structurally and intersectionally. *Journal of Gender-Based Violence*, *5*(3), 531–542. https://doi.org/10.1332/239868021X16286662118554

Bellis, M. A., Hughes, K., Ford, K., Ramos Rodriguez, G., Sethi, D., & Passmore, J. (2019). Life course health consequences and associated annual costs of adverse childhood experiences across Europe and North America: A systematic review and meta-analysis. *The Lancet Public Health*, *4*(10), e517–e528. https://doi.org/10.1016/S2468-2667(19)30145-8

Bergdahl, J., Latikka, R., Celuch, M., Savolainen, I., Soares Mantere, E., Savela, N., & Oksanen, A. (2023). Self-determination and attitudes toward artificial intelligence: Cross-national and longitudinal perspectives. *Telematics and Informatics*, *82*, Article 102013. https://doi.org/10.1016/j.tele.2023.102013

Boethius, S., Åkerström, M., & Hydén, M. (2023). The double-edged sword—Abused women's experiences of digital technology. *European Journal of Social Work*, *26*(3), 506–518. https://doi.org/10.1080/13691457.2022.2040437

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braunschweig, B., & Ghallab, M. (2021). Reflections on AI for humanity: Introduction. In B. Braunschweig & M. Ghallab (Eds.), *Reflections on artificial intelligence for humanity* (pp. 1–12). Springer. https://doi.org/10.1007/978-3-030-69128-8_1

Bryman, A. (2004). *Social research methods*. Oxford University Press.

Butterby, K., & Lombard, N. (2025). Developing a chatbot to support victim-survivors who are subjected to domestic abuse: Considerations and ethical dilemmas. *Journal of Gender-Based Violence*, *9*(1), 153–161. https://doi.org/10.1332/23986808Y2024D000000038

Callaghan, J. E. M., Alexander, J. H., Sixsmith, J., & Fellin, L. C. (2018). Beyond "witnessing": Children's experiences of coercive control in domestic violence and abuse. *Journal of Interpersonal Violence*, *33*(10), 1551–1581. https://doi.org/10.1177/0886260515618946

Decker, M. R., Holliday, C. N., Hameeduddin, Z., Shah, R., Miller, J., Dantzler, J., & Goodmark, L. (2019). "You do not think of me as a human being": Race and gender inequities intersect to discourage police reporting of violence against women. *Journal of Urban Health*, *96*(5), 772–783. https://doi.org/10.1007/s11524-019-00359-z

Devillers, L., Fogelman-Soulié, F., & Baeza-Yates, R. (2021). AI & human values: Inequalities, biases, fairness, nudge, and feedback loops. In B. Braunschweig & M. Ghallab (Eds.), *Reflections on artificial intelligence for humanity* (pp. 76–89). Springer. https://doi.org/10.1007/978-3-030-69128-8_6

de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From "automation" to "autonomy": The importance of trust repair in human-machine interaction. *Ergonomics*, *61*(10), 1409–1427. https://doi.org/10.1080/00140139.2018.1457725

Domingo-Cabarrubias, L., Woodlock, D., Alexander, C., Sato, M., Grant, G., & Weinberg, J. (2023). The role of technology in improving access to justice for victims of family violence: Challenges and opportunities. *Law, Technology and Humans*, *5*(1), 1–10. https://doi.org/10.5204/lthj.2469

Draughon Moret, J., Todd, A., Rose, L., Pollitt, E., & Anderson, J. (2022). Mobile phone apps for intimate partner and sexual violence prevention and response: Systematic search on app stores. *JMIR Formative Research*, *6*(2), Article 28959. https://doi.org/10.2196/28959

Ellsberg, M., Arango, D. J., Morton, M., Gennari, F., Kiplesund, S., Contreras, M., & Watts, C. (2015). Prevention of violence against women and girls: What does the evidence say? *The Lancet*, *385*(9977), 1555–1566. https://doi.org/10.1016/S0140-*6736*(14)61703-7

Emezue, C., Chase, J. D., Udmuangpia, T., & Bloom, T. L. (2022). Technology-based and digital interventions for intimate partner violence: A systematic review and meta-analysis. *Campbell Systematic Review*, *18*(3), Article 1271. https://doi.org/10.1002/cl2.1271

European Commission. (2024a). *DESI indicators—eGovernment usage: Finland*. https://digital-decade-desi.digital-strategy.ec.europa.eu/datasets/desi/charts/desi-indicators?indicator=desi_egov_users_anys&indicatorGroup=desi2023-4&breakdown=ind_total&period=desi_2024&unit=pc_ind_ilt12

European Commission. (2024b). *Digital decade country report 2024: Finland*. https://digital-strategy.ec.europa.eu/en/factpages/finland-2024-digital-decade-country-report

European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Fern, E. F. (2001). *Advanced focus group research*. Sage. https://doi.org/10.4135/9781412990028

Finnish Government. (2024). *Finland's national roadmap: EU Digital Decade Policy Programme 2030* (Publication of the Finnish Government 2024:7). https://urn.fi/URN:ISBN:978-952-383-743-0

Finnish National Board on Research Integrity TENK. (2023). *The Finnish code of conduct for research integrity and procedures for handling alleged violations of research integrity in Finland 2023*. https://tenk.fi/sites/default/files/2023-05/RI_Guidelines_2023.pdf

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Harris, B. A., & Woodlock, D. (2019). Digital coercive control: Insights from two landmark domestic violence studies. *British Journal of Criminology*, *59*(3), 530–550. https://doi.org/10.1093/bjc/azy052

Henman, P. (2020). Improving public services using artificial intelligence: Possibilities, pitfalls, governance. *Asia Pacific Journal of Public Administration*, *42*(4), 209–221. https://doi.org/10.1080/23276665.2020.1816188

Henry, N., Witt, A., & Vasil, S. (2024). A "design justice" approach to developing digital tools for addressing gender-based violence: Exploring the possibilities and limits of feminist chatbots. *Information, Communication & Society*. Advance online publication. https://doi.org/10.1080/1369118X.2024.2363900

High-Level Expert Group on AI. (2019). *Ethics guidelines for trustworthy AI*. European Commission. https://doi.org/10.2759/346720

Holt, S., Buckley, H., & Whelan, S. (2008). The impact of exposure to domestic violence on children and young people: A review of the literature. *Child Abuse & Neglect*, *32*(8), 797–810. https://doi.org/10.1016/j.chiabu.2008.02.004

Hunt, X., Tomlinson, M., Sikander, S., Skeen, S., Marlow, M., du Toit, S., & Eisner, M. (2020). Artificial intelligence, big data, and mHealth: The frontiers of the prevention of violence against children. *Frontiers in Artificial Intelligence*, *3*, Article 543305. https://doi.org/10.3389/frai.2020.543305

Husso, M., Notko, M., Virkki, T., Holma, J., Laitila, A., & Siltala, H. (2021). Domestic violence interventions

in social and health care settings: Challenges of temporary projects and short-term solutions. *Journal of Interpersonal Violence*, *36*(23/24), Article 0886260519898438. https://doi.org/10.1177/0886260519898438

Husso, M., Virkki, T., Notko, M., Holma, J., Laitila, A., & Mäntysaari, M. (2012). Making sense of domestic violence intervention in professional health care. *Health and Social Care in the Community*, *20*(4), 347–355. https://doi.org/10.1111/j.1365-2524.2011.01034.x

Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR Mhealth and Uhealth*, *6*(11), Article 12106. https://doi.org/10.2196/12106

International Federation of Social Workers. (2018). *Global social work statement of ethical principles*. https://www.ifsw.org/global-social-work-statement-of-ethical-principles

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kabacińska, K., McLeod, K., MacKenzie, A., Vu, K., Cianfrone, M., Tugwell, A., & Robillard, J. M. (2022). What criteria are young people using to select mobile mental health applications? A nominal group study. *Digital Health*, *8*, Article 20552076221102775. https://doi.org/10.1177/20552076221102775

Keeling, J., & Fisher, C. (2015). Health professionals' responses to women's disclosure of domestic violence. *Journal of Interpersonal Violence*, *30*(13), 2363–2378. https://doi.org/10.1177/0886260514552449

Kitzinger, J. (1994). The methodology of focus groups: The importance of interaction between research participants. *Sociology of Health & Illness*, *16*(1), 103–121. https://doi.org/10.1111/1467-9566.ep11347023

Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., & Singh, I. (2019). Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical Informatics Insights*, *11*, Article 1178222619829083. https://doi.org/10.1177/1178222619829083

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction* (2nd ed.). Morgan Kaufmann.

Lindgren, S., & Dignum, V. (2023). Beyond AI solutionism: Toward a multi-disciplinary approach to artificial intelligence in society. In S. Lindgren (Ed.), *Handbook of critical studies of artificial intelligence* (pp. 163–172). Edward Elgar Publishing. https://doi.org/10.4337/9781803928562.00019

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *The Journal of Consumer Research*, *46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013

McGreevey, J. D., III, Hanson, C. W., III, & Koppel, R. (2020). Clinical, legal, and ethical aspects of artificial intelligence–assisted conversational agents in health care. *JAMA*, *324*(6), 552–553. https://doi.org/10.1001/jama.2020.2724

Meyer, S. (2016). Still blaming the victim of intimate partner violence? Women's narratives of victim desistance and redemption when seeking support. *Theoretical Criminology*, *20*(1), 75–90. https://doi.org/10.1177/1362480615585399

Miller, E., & McCaw, B. (2019). Intimate partner violence. *New England Journal of Medicine*, *380*(9), 850–857. https://doi.org/10.1056/NEJMra1807166

Mishna, F., Milne, E., Bogo, M., & Pereira, L. F. (2021). Responding to COVID-19: New trends in social workers' use of information and communication technology. *Clinical Social Work Journal*, *49*(4), 484–494. https://doi.org/10.1007/s10615-020-00780-x

Montagni, I., Tzourio, C., Cousin, T., Sagara, J. A., Bada-Alonzi, J., & Horgan, A. (2020). Mental health-related

digital use by university students: A systematic review. *Telemedicine and E-Health*, *26*(2), 131–146. https://doi.org/10.1089/tmj.2018.0316

Niklander, E., Notko, M., & Husso, M. (2019). *Intervening in domestic violence and training of professionals in social services and health care and the police: Evaluation of the EPRAS project*. National Institute for Health and Welfare (THL). https://urn.fi/URN:ISBN:978-952-343-413-4

Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An initial model of trust in chatbots for customer service—Findings from a questionnaire study. *Interacting With Computers*, *31*(3), 317–335. https://doi.org/10.1093/iwc/iwz022

Novitzky, P., Janssen, J., & Kokkeler, B. (2023). A systematic review of ethical challenges and opportunities of addressing domestic violence with AI-technologies and online tools. *Heliyon*, *9*(6), Article 17140. https://doi.org/10.1016/j.heliyon.2023.e17140

Palanica, A., Flaschner, P., Thommandram, A., Li, M., & Fossat, Y. (2019). Physicians' perceptions of chatbots in health care: Cross-sectional web-based survey. *Journal of Medical Internet Research*, *21*(4), Article 12887. https://doi.org/10.2196/12887

Parviainen, J., & Rantala, J. (2022). Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care. *Medicine, Health Care and Philosophy*, *25*(1), 61–71. https://doi.org/10.1007/s11019-021-10049-w

Piippo, S., Husso, M., Hirvonen, P., Notko, M., & Glumbíková, K. (2021). Institutional and affective practices of domestic violence interventions in social work: Malignant positioning of victims. In M. Husso, S. Karkulehto, T. Saresma, A. Laitila, J. Eilola, & H. Siltala (Eds.), *Violence, gender and affect: Interpersonal, institutional and ideological practices* (pp. 113–133). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-56930-3_6

Reith-Hall, E., & Montgomery, P. (2022). The teaching and learning of communication skills in social work education. *Research on Social Work Practice*, *32*(7), 793–813. https://doi.org/10.1177/10497315221088285

Rogers, M. M., Fisher, C., Ali, P., Allmark, P., & Fontes, L. (2023). Technology-facilitated abuse in intimate relationships: A scoping review. *Trauma, Violence, & Abuse*, *24*(4), 2210–2226. https://doi.org/10.1177/15248380221090218

Saxton, M. D., Olszowy, L., MacGregor, J. C. D., MacQuarrie, B. J., & Wathen, C. N. (2021). Experiences of intimate partner violence victims with police and the justice system in Canada. *Journal of Interpersonal Violence*, *36*(3/4), NP2029–NP2055. https://doi.org/10.1177/0886260518758330

Siltala, H., Hisasue, T., Hietamäki, J., Saari, J., Laajasalo, T., October, M., Laitinen, H., & Raitanen, J. (2023). *Domestic violence-related use of services and the resulting costs in health, social and legal services*. Prime Minister's Office. http://urn.fi/URN:ISBN:978-952-383-317-3

Stark, E. (2007). *Coercive control: The entrapment of women in personal life*. Oxford University Press.

Storer, H. L., Nyerges, E. X., & Hamby, S. (2022). Technology "feels less threatening": The processes by which digital technologies can facilitate youth's access to services at domestic violence and sexual assault organisations. *Children and Youth Services Review*, *139*, Article 106573. https://doi.org/10.1016/j.childyouth.2022.106573

Storer, H. L., Scott, C. F., Rodriguez, M., & Nyerges, E. X. (2023). Technology is a "blessing and a curse": The perceived risks and benefits of digital technology adoption at domestic violence organisations that serve teens. *Journal of Technology in Human Services*, *41*(1), 96–124. https://doi.org/10.1080/15228835.2023.2179158

Stubbs, A., & Szoeke, C. (2022). The effect of intimate partner violence on the physical health and health-related behaviors of women: A systematic review of the literature. *Trauma, Violence, & Abuse*, *23*(4), 1157–1172. https://doi.org/10.1177/1524838020985541

Taylor, A. (2017). Social work and digitalisation: Bridging the knowledge gaps. *Social Work Education*, *36*(8), 869–879. https://doi.org/10.1080/02615479.2017.1361924

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*, *20*(3), 318–335. https://doi.org/10.1002/wps.20883

Trevithick, P. (2012). *Social work skills and knowledge: A practice handbook* (3rd ed.). McGraw-Hill Education; Open University Press.

Turner, W., Hester, M., Broad, J., Szilassy, E., Feder, G., Drinkwater, J., Firth, A., & Stanley, N. (2017). Interventions to improve the response of professionals to children exposed to domestic violence and abuse: A systematic review. *Child Abuse Review*, *26*(1), 19–39. https://doi.org/10.1002/car.2385

United Nations. (2015, October 21). *Transforming our world: The 2030 agenda for sustainable development* (A/RES/70/1). https://digitallibrary.un.org/record/3923923

United Nations. (2024, March 21). *Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development* (A/RES/78/265). United Nations Digital Library. https://digitallibrary.un.org/record/4043244

Virkki, T., Husso, M., Notko, M., Holma, J., Laitila, A., & Mäntysaari, M. (2015). Possibilities for intervention in domestic violence: Frame analysis of health care professionals' attitudes. *Journal of Social Service Research*, *41*(1), 6–24. https://doi.org/10.1080/01488376.2014.917449

Wood, L., Hairston, D., Schrag, R. V., Clark, E., Parra-Cardona, R., & Temple, J. R. (2022). Creating a digital trauma informed space: Chat and text advocacy for survivors of violence. *Journal of Interpersonal Violence*, *37*(19/20), NP18960–NP18987. https://doi.org/10.1177/08862605211043573

Woodlock, D., McKenzie, M., Western, D., & Harris, B. (2020). Technology as a weapon in domestic violence: Responding to digital coercive control. *Australian Social Work*, *73*(3), 368–380. https://doi.org/10.1080/0312407X.2019.1607510

Wright, E. N., Anderson, J., Phillips, K., & Miyamoto, S. (2022). Help-seeking and barriers to care in intimate partner sexual violence: A systematic review. *Trauma, Violence, & Abuse*, *23*(5), 1510–1528. https://doi.org/10.1177/1524838021998305

Xu, L., Sanders, L., Li, K., & Chow, J. C. L. (2021). Chatbot for health care and oncology applications using artificial intelligence and machine learning: Systematic review. *JMIR Cancer*, *7*(4), Article 27850. https://doi.org/10.2196/27850

About the Authors

Hanna Mielismäki is a doctoral researcher in social sciences at Tampere University. Her research focuses on domestic violence and technology-facilitated abuse, responsible AI, digital ethics, and the digitalisation of welfare services. She explores the ethical and societal implications of AI-driven technologies in support services.

Marita Husso is a professor of social policy at Tampere University. Her research interests include gender-based violence, digitalisation, service systems, care work, and gendered agency. She is the first editor of *Interpersonal Violence—Differences and Connections* (Routledge, 2017) and *Violence, Gender and Affect—Interpersonal, Institutional and Ideological Practices* (Palgrave, 2021).

Social Inclusion
2025 • Volume 13 • Article 9949
https://doi.org/10.17645/si.9949

cogitatio

ARTICLE

Open Access Journal

Directing Digital Citizenship: How Librarians Mediate the Dutch Digital Welfare State

Maud Rebergen, Joëlle Swart, and Marcel Broersma

Centre for Media and Journalism Studies, University of Groningen, The Netherlands

Correspondence: Maud Rebergen (m.s.rebergen@rug.nl)

Submitted: 28 January 2025 **Accepted:** 23 June 2025 **Published:** 30 July 2025

Issue: This article is part of the issue "Fostering the Socially and Ecologically Sustainable Digitalisation of Welfare States" edited by Paula Saikkonen (Finnish Institute for Health and Welfare) and Marta Choroszewicz (University of Eastern Finland), fully open access at https://doi.org/10.17645/si.i514

Abstract

Digital citizenship has emerged as a prominent concept in policy and academic discourse, broadly referring to individuals' ability to access and use digital tools for public engagement. However, because its meaning varies between actors and across societal contexts, it is still an ambiguous term. This article considers how digital citizenship takes shape in practice by examining the everyday practices of librarians, the key mediators between citizens and the digital welfare state, and analysing how these contrast with conceptualizations of digital citizenship promoted by the Dutch national Digital Inclusion programme. Governments worldwide are adopting "digital-by-default" models, emphasizing ICT-driven public service solutions. While promising greater efficiency and accessibility, this transition exacerbates challenges for individuals lacking access to ICT resources or digital literacy, reinforcing social inequalities. Using the Netherlands as a case study, this research highlights the challenges of digital inclusion in highly digitalized societies. Despite high digitalization rankings, many Dutch citizens face difficulties using digital tools and accessing digital public services. To address this issue, the Dutch government launched the Digital Inclusion programme in 2019, establishing information points and digital skills courses in libraries. Based on ethnographic research in three public libraries, this study reveals a disconnect between policymakers' conceptualisations of digital citizenship, defining what competencies citizens should possess and what participatory practices are supported, and the lived reality of digital citizenship in public libraries. By highlighting the tensions and misalignments between policy and practice, this article aims to contribute to more inclusive conceptualizations of digital citizenship, to inform digital inclusion initiatives that foster equitable participation in digital societies.

Keywords

digital citizenship; digital government; digital inclusion; digital literacy; digital welfare state; public libraries; self-service solutions; social inequality

1. Introduction

In today's increasingly digital world, governments worldwide are adopting digital technologies to enhance governance and public service delivery (Liu & Yuan, 2015). The Netherlands is a pioneer in this transition (United Nations E-government Knowledgebase, n.d.) and has prioritised the introduction of digital government services and online identity verification (e.g., the DigiD identification tool) since the early 2000s (Digitale Overheid, 2020). Over the past two decades, numerous government websites and apps have been developed to enable citizens' access to a wide range of online public services, including tax filing, benefit applications, and study financing.

Like in other countries, digitalization here has involved the implementation of "self-service solutions" (Schou & Hjelholt, 2018), shifting the responsibility for accessing public services from public servants to citizens themselves. Henman (2010, pp. 216–217) argues that this reflects a neoliberal approach to state restructuring, emphasizing cost reduction and active citizen engagement over the passive dependency characteristic of paternalistic welfare states. Citizens are thus held accountable for acquiring the skills and resources needed to navigate online public services effectively.

However, this digital transformation comes with significant challenges, particularly for citizens who lack the necessary resources, skills, or confidence to navigate these systems effectively. While the Netherlands boasts relatively high internet coverage and literacy rates (Centraal Bureau voor de Statistiek [CBS], 2020), a significant portion of the population struggles to participate in the digital society (Digitale Overheid, 2020). An estimated 2.5 million of the 17 million people lack basic digital skills, and up to 5 million lack critical competencies. Additionally, approximately 4.5 million adults face challenges interacting with government services (Hoevenagel & Joossen, 2022).

To address these issues, in 2019, the Dutch government initiated the Digital Inclusion programme in collaboration with the National Library of the Netherlands and eight public service organizations, collectively referred to as the Manifestgroep ("Manifesto Group"). The programme aims to provide training and support for individuals lacking the skills needed to effectively use digital public services and participate in the digital society. As part of this initiative, 15 public libraries across the country expanded their existing offerings of free digital skills courses and courses for navigating digital public services. They also introduced information points ("IDOs") with walk-in consultation hours to provide information and guidance on the services of the participating parties (Staatscourant, 2019), making public libraries important sites for citizen-state interaction and librarians key intermediaries in the relationship between citizens and the digital welfare state.

This article explores this evolving role of librarians as digital care workers (Kaun & Forsman, 2022) by asking: How do the practices of librarians as key mediators between citizens and the digital welfare state contrast with conceptualizations of digital citizenship promoted by the Dutch national Digital Inclusion programme?

Our findings provide critical insights into the complexities of digital citizenship, highlighting persistent gaps between policy visions and the practical reality of digital inclusion policy. Moreover, it emphasizes the pivotal role of librarians in bridging divides in the digital welfare state, showing the barriers they face in supporting citizens' digital civic practices. Consequently, our study provides crucial insights to rethink digital inclusion initiatives, to more effectively address the structural challenges citizens face in increasingly digital societies.

2. Literature Review

2.1. Citizenship in the Digital Society

Since the mid-1990s, the concept of "digital citizenship" has emerged as a pivotal framework within academic and governance discourse, addressing the evolving relationship between digital technologies and civic participation (Isin & Ruppert, 2015; Katz, 1997; Mossberger et al., 2008). At its core, digitalization has reconfigured citizenship, providing new modalities for engagement in democratic and societal processes while challenging traditional boundaries of public and private spheres (Hintz et al., 2019). Despite its prevalence in public discussions, however, "digital citizenship" remains a contested and multifaceted concept, shaped by diverse interpretations and enacted differently across disciplines and practices, with no clear consensus or definition (cf. Mol, 2002).

Theorizing digital citizenship involves understanding it as a dynamic interplay of rights, responsibilities, and practices. Moving beyond a static legal status, citizenship emerges from participatory practices that construct agency within political, social, and cultural contexts (Lister & Campling, 1997). This performative dimension is further nuanced in digital environments, where actions such as "liking, coding, clicking, downloading, sorting, blocking, and querying" (Isin & Ruppert, 2015, p. 12) support novel participatory practices. Digital infrastructures afford empowerment and constraint simultaneously, where opportunities for agency coexist with regulatory mechanisms (Vivienne et al., 2016, p. 4).

Initial conceptualizations, such as Katz's (1997), imbued digital citizenship with a techno-optimistic ethos, envisioning it as a catalyst for democratization, community building, and individual empowerment. Subsequent critiques, however, revealed that digital citizenship is a more complex phenomenon. Scholars like Mossberger et al. (2007) emphasize the socio-economic stratifications that shape access and literacy, asserting that digital citizenship disproportionately benefits individuals with greater resources and competencies. This underscores the inequities inherent to digital systems and illustrates that targeted interventions are necessary to address systemic barriers.

In the Dutch political context, the promotion of digital citizenship has been a topic of discussion since the late 1990s, although a clear definition of the concept remained elusive for a long time. Wagenaar of the Labour Party (PvdA) was among the few to address this issue in depth during that period, connecting digital citizenship with equitable access to information during a parliamentary debate:

> The 21st century can develop into an information society or an infocracy, where information becomes a costly commodity accessible only to the economically strong. The PvdA chooses to support the development of an information society, where new media and related technologies are employed to promote accessibility to knowledge and information. (Tweede Kamer der Staten-Generaal, 1999)

During the next two decades, conceptualizations of digital citizenship widened to include a more comprehensive agenda. For example, a report of the Parliamentary System State Commission in 2018 emphasized civic participation and the mitigation of digital risks, such as disinformation and news avoidance (Staatscommissie Parlementair Stelsel, 2018). Suggestions for curricular reforms in primary and secondary education in 2019 further operationalized digital citizenship, focusing on two themes: the "digital citizen"

and "digital identity" (Ministerie van Onderwijs, Cultuur en Wetenschap, 2019). These "pillars" linked digital skills to awareness and critical reflection on the role of digital technologies in society, aiming to foster active and responsible participation.

By situating digital citizenship within a broader ecosystem of competencies, including self-regulation, critical thinking, and cultural awareness, contemporary Dutch policy advances a model that reconciles individual agency with collective welfare. This understanding of digital citizenship has further been implemented through the Digital Inclusion programme, which aims to empower citizens by teaching digital skills through courses and providing information on digital government services via the IDOs in public libraries. How librarians enact and understand this newly assigned role as key intermediaries in the increasingly digital-by-default citizen-state relationship, however, remains underexplored.

2.2. Transitioning to Digital-by-Default

During the past two decades, the Dutch government has shifted towards a digital-by-default model, prioritizing digital over traditional paper alternatives. Since the early 2000s, alongside the growing public use of the internet, the Netherlands has focused on developing digital "self-service solutions" (Schou & Hjelholt, 2018), where citizens with access to digital technologies manage their own affairs to reduce bureaucracy and costs. Recent amendments to Dutch legislation have even made it possible to make digital communication obligatory for certain matters (Kerssies & Wijngaards, 2022). This development is part of a broader global trend, where governments around the world have gradually implemented ICTs to digitize public services, with the aim of making administrative processes more efficient, cost-effective, and accessible (Liu & Yuan, 2015). This transition is generally encouraged by a strong belief in technological innovation and the desire to bring the government closer to citizens.

This "digital-by-default" strategy has received broad support in the Netherlands with a techno-solutionist consensus amongst the administrations of the past two decades, based on the belief that digital technologies improve state-citizen relations through more efficient, user-friendly, and reliable public services. However, while the digitalisation of the Dutch welfare state has indeed resulted in more efficient public service delivery for most, it has also created new challenges for disadvantaged groups. While digitally literate citizens benefit from the ease of use and speed of access to digital services, disadvantaged citizens face exclusion. Those without access to digital technologies or the necessary skills encounter barriers to participation, deepening existing inequalities and placing greater responsibility on individuals often without adequate support.

Scholars have highlighted that digital inequality is a complex issue intertwined with broader social, economic, and cultural inequalities (Hargittai, 2022; Helsper, 2021; Van Dijk, 2020). They point out that digital inequality and social inequality are interlinked, reinforcing one another (Helsper, 2021). A substantial body of research on digital inclusion in the Netherlands underscores this perspective, highlighting the connection between digital inequality and broader social disparities, such as socio-economic disadvantage (Goedhart et al., 2019), low literacy (Smit et al., 2023), and poverty (Boerkamp et al., 2025). To address these disparities, the responsibility of supporting disadvantaged groups has increasingly been delegated to public libraries. Librarians, thus, have come to play a pivotal role in providing citizens with access to and support with digital government services, especially as analogue means for communication with the government increasingly disappear.

2.3. The Rise of Digital Care Work

In Western democracies like the United States (Greene, 2021) and Finland (Kaun & Forsman, 2022), public libraries increasingly act as intermediaries between citizens and digital government services. In the Netherlands, this shift reflects broader efforts to redefine public libraries' role amid declining memberships, austerity measures, and reduced staffing since the 1980s (Huysmans & Oomes, 2018). These challenges prompted public libraries to reposition as "socio-educational living rooms" offering diverse services, including providing access to information, education, fostering debate, and cultural engagement (Huysmans & Oomes, 2018).

Dutch public libraries have embraced digital literacy as central to their transformation, paralleling global trends of "bootstrapping," where public libraries secure funding by offering access to technology and training services (Greene, 2021). In 2019, they integrated the national Digital Inclusion programme, hosting digital skills courses and IDOs to assist citizens with tasks like taxes, benefits, and digital identity management. The programme was financed through a multi-year subsidy with per capita subsidies increasing from EUR 0.03 in 2019 to EUR 0.98 in 2022. Additional funds, up to EUR 200,000, were allocated for training, knowledge sharing, and network building, contingent on the National Library's annual budget (Wetten.nl, 2022).

Since 2019, the programme's core focus has remained steady, but the number of IDOs has surged dramatically from 15 in 2019 to 741 by 2025 (Bibliotheeknetwerk, n.d.). Additionally, the Manifestgroep now consists of 17 government agencies and other participating organisations (Manifestgroep, n.d.). This growth reflects the objectives outlined in the 2020 Library Covenant, which emphasizes the vital role of libraries in fostering digital citizenship:

> The public library is a key partner alongside the education sector in promoting digital citizenship and literacy from an early age. Libraries empower individuals by offering courses and assisting citizens with questions about the (digital) government. They also serve as a referral point to local organizations that can address specific individual concerns. (Staatscourant, 2020)

The covenant emphasized their efforts in fostering self-reliance, offering courses, and guiding citizens in navigating digital government systems.

While public libraries have embraced this expanded role to enhance their societal relevance, Kaun and Forsman (2022) found that librarians are not always adequately equipped for their new position as "digital care worker[s]," pointing to shortcomings in current funding and training. They show how library staff are often required to provide support that exceeds their available resources, skills, and expertise, including offering technical assistance, addressing emotional barriers such as frustration, mistrust, and uncertainty, and referring individuals to social support organizations.

This article builds upon these findings by examining Dutch librarians' experiences of supporting digital citizenship, to highlight the tensions between conceptualizations of digital citizenship as embedded in digital inclusion initiatives versus librarians' everyday practices as digital care workers. Underscoring the need for more nuanced understandings of how digital citizenship policy is practiced, our findings illustrate how current digital inclusion policy falls short in addressing the digital divide effectively.

3. Methodology

To examine how librarians navigate the gap between top-down imposed governmental digital inclusion policies and citizens' everyday support needs and questions around digital citizenship, our study employed an institutional ethnographic approach (Smith, 1987). This method is particularly relevant for analysing librarians' mediating role in supporting digital citizenship, as it aims to explicate how people's everyday activities are organized and coordinated through institutionalized processes and practices, what Smith (2005, p. 10) calls "ruling relations." To capture both librarians' practices of supporting people's digital citizenship as well as how they discuss and understand these behaviours in light of the complex institutional context they are part of, we combined longitudinal, weekly participant observations of walk-in consultation hours and digital literacy courses (99 hours in total) with in-depth interviews with staff (N = 24), in libraries in three cities geographically spread across the Netherlands.

We selected libraries that participated as "forerunners" in the Digital Inclusion programme. Based in one of the largest cities in the Netherlands with an ethnically diverse population, Library A is a large multifunctional city library, with one major hub in the city centre and 19 smaller libraries in different neighbourhoods. In addition to lending books and offering digital support, it frequently hosts exhibitions, offers language and financial support, organizes workshops and courses, and provides a meeting space. Staff at mid-sized Library B explicitly described themselves as "community librarians," positioning their library as a warm environment offering a "socio-educational living room" to all citizens in the mid-sized town. In contrast, location C was a small library with a more traditional focus as an information provider and book lender, in a small town with a relatively low-income and low-educated population.

Observations and interviews took place from September to November 2020 during the second peak of the Covid-19 pandemic in the Netherlands, when vaccines were not yet available, and during the second half of the research period a partial lockdown was in place. Paradoxically, the growing need for digital support to stay engaged in social life coincided with a decline in library visits, especially by vulnerable groups, due to infection risks, even though libraries were among the few places that remained open. Due to the circumstances, our ability to study librarians' direct engagement with citizens was limited, resulting in fewer opportunities to observe interactions. Instead, we focused on understanding how library workers talked about their practices and their evolving role, from information providers to educators fostering digital inclusion, as well as their insights into the challenges citizens face. More importantly, library workers provide a valuable perspective for examining the institutional order, including its social relations and power dynamics (DeVault & McCoy, 2006; Greene, 2021; Smith, 2005), making their viewpoints a crucial lens for this study.

Team members each visited one of the libraries once or twice a week, on fixed mornings or afternoons, not just to attend the scheduled walk-in hours and digital literacy courses, but also to build rapport with library staff and volunteers who typically worked there part-time. We used an observation protocol to guide our observations and recorded them by writing fieldnotes, drawing floor plans, and taking pictures. First, we focused on mapping social, spatial, and temporal structures and dynamics, such as recurring activities, returning visitors to consultation hours, and the position of technological resources and materials. We then gradually shifted our attention to the interactions between library workers and visitors. The role of the observers, likewise, changed over time: Whereas during the earlier observations, researchers listened and watched carefully while taking detailed notes, with minimal engagement as to capture the research setting

without steering the activities going on, over the course of the study, researchers would approach librarians and visitors actively and ask detailed questions. While we were careful to introduce and identify ourselves as researchers, including wearing name badges, occasionally visitors would approach us for help with digital questions; we then always referred them to the library staff, having the librarian take the lead on providing support while we observed the interaction. Fieldnotes were structured chronologically and reflect researchers' personal perspectives on the observed events and their role when observing. Every two weeks, all members of the research team met to share and reflect on the observations.

After our first month of observations, we conducted semi-structured interviews with a selection of the librarians and volunteers at each location. This allowed us to probe for details about the patterns and events that we observed and to gain an in-depth understanding of staff's perceptions and experiences of the library as a digital literacy supporter. We selected staff ($N = 24$) across the three locations, using maximum variation sampling to account for possible diversity in viewpoints on the interplay between librarians, digital citizenship, and their institutional frameworks. Participants ranged in age from mid-20s to mid-60s. We interviewed librarians with different roles, years of experience, and levels of involvement in the library's digital inclusion-related tasks, including community librarians, domain specialists, project officers, and volunteers, until the point when additional interviews would no longer yield novel patterns or themes. Using an interview guide, the interviews discussed participants' tasks within the library, types of digital support requested, referral to other institutions offering support for digital or related issues, and the role of public libraries for digital inclusion. All interviews were audio recorded and then fully transcribed for qualitative thematic analysis. Pseudonyms were used for all participants, and informants gave consent for pseudonymized quotes to be included in the study.

All field notes and interview transcripts were analysed using Atlas.ti, applying a grounded theory-inspired, iterative approach (Corbin & Strauss, 2008). In a first round of descriptive open coding, the data were analysed line-by-line by the research team. We focused on library staff's practices, knowledge, skills, motivations, emotions, and perceived facilitators and barriers related to digital support and inclusion, in addition to mapping social, temporal, spatial, and technological structures. Second, aided by individual memo writing and regular discussion meetings within the research team, these open codes were compared in order for the research team to be able to merge them into overarching categories that identify similarities between situations and/or locations. Considering the specific purpose of this article, field notes excerpts and interview quotes related to the categories of "digital citizenship" have consequently been re-analysed in a final round of coding.

4. Results

4.1. Defining the Digital Citizen

Our observations and interviews reveal that the Digital Inclusion programme expects citizens to possess a combination of resources, skills, knowledge, and qualities to effectively use digital government services and their support systems. This is paradoxical, as those who can meet these expectations are typically already self-sufficient in the digital society. In contrast, we see that citizens and public service users who are already vulnerable to digital exclusion often struggle to meet the requirements embedded in the digital government and its support structures, putting them at risk of further marginalization and exclusion from digital public services.

4.1.1. The Requisites of Digital Citizenship: Access, Skills, and Literacy

Our observations and interviews indicate that citizens are assumed to have access to the necessary physical and technological resources to use digital public services. This includes equipment such as a computer, tablet, or smartphone, a stable and secure internet connection, and the ability to print forms or confirmations. In practice, we found that many people do not have access to these resources, or the ones they do have are difficult to use, broken, or outdated, which greatly limits their ability to utilize digital services. Sam, a 27-year-old community librarian at Library B, said:

> You do see significantly more people who still have a simple Nokia, or sometimes don't have a phone at all, which makes it very difficult because if you want to create a DigiD, you have to provide either an email address or a phone number, and some people don't have either.

All three libraries provided public computers with internet access, which were frequently used by citizens during our observations. At Library C, for instance, we observed two young men entering the library and requesting a computer, though communication was difficult due to language barriers. When asked, one of the men explained they were from Eritrea, speaking Tigrinya, and used the computer for work and to practice Dutch. As they did not have computers at home, they visited the library nearly every day to use the computer (fieldnotes, 23-10-2020, Library C). During the Covid-19 pandemic, however, the use of computers was limited as part of measures to curb the spread of the virus. This presented a major challenge for all three libraries, as their role as providers of digital resources (often the sole means of accessing public services during this period) became significantly more difficult due to these restrictions.

Beyond access, citizens are expected to possess an increasing level of digital competencies to effectively use digital government services. These include operating computers, tablets, or smartphones, navigating the internet to find government websites, and performing basic digital tasks such as logging in with e-identifiers, sending emails, or filling out online forms, while managing privacy and security risks such as phishing attempts or identity fraud. If any technical issues occur, citizens are expected to solve these independently or know where to seek help.

In practice, however, these demands sometimes appeared difficult to meet. At Library C, we observed a woman who struggled to submit a damage report. Alex, the 63-year-old facility and volunteer coordinator helping her, was unsure how to proceed and asked the researcher for help. The woman had completed the form digitally with the intention of printing and scanning it, unaware that she could have submitted it electronically right away. While trying to figure out how to print the form, she was logged out of her account for security reasons, losing her data in the process. Frustrated with the system she was using, she insisted she hadn't taken that long. The researcher therefore offered a slower-paced alternative that better matched the woman's digital skills: printing a blank form for her to fill out by hand, then assisting her in scanning and emailing it later (fieldnotes, 23-10-2020, Library C).

In addition to technological skills, using digital government services in the Netherlands requires a high level of Dutch language proficiency. Citizens are expected to understand official letters, forms, and instructions written in formal Dutch. In practice, however, even for native speakers, this can be a major barrier. Max, a 58-year-old community librarian at Library B, explained that many visitors had limited digital experience and faced additional difficulties when confronted with complex bureaucratic language:

> You pass them on to courses like Klik&Tik, because they really still have to learn the basics. And the people who then come to Digisterker can handle the computer, but they get stuck on the official language used by the government and forms that are sometimes complicated and that you have to do everything securely with DigiD and with passwords and things like that.

We observed how elderly citizens struggled to find information and navigate complex government websites due to a lack of digital literacy skills. According to our interviewees, this issue also affects other groups, including low-literate adults and young people. As librarian Robin (Library A) explained:

> Young people are often very skilled with buttons and technology and learn quickly, but they often overlook safety or the simplest ways to look things up. It's not self-evident that they know all of that.

According to our informants, the Digital Inclusion programme was inadequately designed and insufficiently promoted to address the needs of these groups, further restricting their access to digital public services.

4.1.2. Self-Reliance, the Golden Standard for Digital Citizenship

Alongside the emphasis on access, skills, and literacy as important requisites for digital inclusion, the Digital Inclusion programme places significant focus on pursuing self-sufficiency, linking individual self-reliance to collective prosperity. This approach shifts responsibility to citizens, requiring them to proactively address their needs and navigate support structures to effectively access public services rather than the system adapting to their circumstances.

For instance, citizens are expected to understand which government agencies, such as the Belastingdienst (Tax and Customs Administration), the UWV (Employee Insurance Agency), or the municipality, handle which matters. They should know the required documents for applications like allowances or benefits, and must be proficient with digital tools like DigiD and MijnOverheid (MyGovernment). According to Jessie, a librarian at Library A, many visitors were unable to meet such demands:

> People with lower incomes are the ones who need the most help. The people who often receive various benefits or extra welfare from the government. Because they already have so much contact with government services, they naturally require more assistance. I believe that especially people with lower incomes who struggle with the Dutch language would benefit the most, as they tend to leave important matters unattended. They may not be aware of their rights or the information they need. Typically, people come in with a specific question, but not many ask, "What can I get?"

Finally, citizens are assumed to know where to go for assistance when they get stuck. This includes knowledge of available support structures in public libraries, such as the IDOs and digital skills courses, and awareness of how, where, and when this help is provided. A certain level of willingness and mobility is also expected to physically visit a public library for assistance. Both in Library A and C, however, the information points were poorly sought out due to a combination of factors, including the ongoing pandemic and limited visibility of this new service. Jip, a 34-year-old domain specialist for "Language as a Basis" at Library C, said: "I have the feeling that they simply don't know. I think it's also a target group that doesn't come to the library on their own." They explained:

> I had a neighbour who was also a bit in that target group, and he heard that I worked at the library and literally asked, "Wow, does that still exist?" Well, I was actually quite shocked by that. You see, if you don't come to the library, you don't know that changes have taken place and what else you can do besides borrowing books.

As a result, "the people who need the help the most can find it the hardest," they noted. This underscores the paradox of the initiative, demonstrating how well-intentioned policies can inadvertently exacerbate existing societal disparities.

Thus, the set of qualities, skills and attitudes comprising the baseline for citizens' participation in the Digital Inclusion programme appeared to be far from self-evident, for disadvantaged groups. This not only hindered citizens' interaction with the digital government, but also resulted in feelings of fear, uncertainty, shame, frustration, and a lack of confidence in their own skills and in digital systems and the government more broadly, as highlighted by our observations and interviews. Jip, for instance, explained that people in the local community were often hesitant to seek help due to fear of stigmatization, especially due to the vulnerability associated with asking for assistance in a public setting. They noted:

> That's a certain culture: People don't easily show the back of their tongues. They are sometimes a bit reserved and narrow-minded. And also, a bit of the culture of "keep everything behind closed doors," especially when things aren't going well….That is a piece of pride that they, actually very unfortunate, do not get over….If you then have to go to the library to ask for help, yes, I think that can be a barrier for a lot of people.

Due to the inflexibility of current support structures, a divergence from the expected qualities and attitudes of citizenship resulted, at best, in long-term dependence on support systems such as one's social network, and at worst, in complete disengagement. Bo, a volunteer at Library C, explained:

> I think that is more of a fatalism, of "okay, this isn't it anymore, I give up." These are people who have experienced significant innovations in their lives….And now, suddenly, the bank buildings are closing, post offices are gone. All such things, it comes over them. And one learns to deal with it, and the other closes himself off from it.

These barriers prevented citizens from engaging with digital government services, illustrating how the programme's failure to accommodate a multimodal understanding of citizenship can result in exclusion.

4.2. Navigating Practices of Digital Citizenship

Our observations and interviews revealed the ongoing efforts of public libraries in assisting citizens with basic digital skills training and digital government services through the Digital Inclusion programme. While each library had its own approach, all three featured digital skills courses and IDOs where citizens could inquire about the participating digital government services. As public libraries have become a key mediator between citizens and digital government services, librarians play a crucial role in facilitating the interaction between citizens and the digital government.

According to our informants, many of the questions they address revolve around navigating government websites, such as logging in with DigiD, completing tax forms, or applying for benefits. Additionally, librarians assist with questions about specific government documents, including unclear letters, tax returns, and driver's licences, helping citizens understand and act on these matters. While this support is essential for enabling participation in the digital society, it addresses only a fraction of the broader demand for assistance. The information points were set up to provide basic information or guidance. When citizens face greater challenges or require more in-depth support, librarians are instructed to refer them to relevant courses or external partners.

In practice, librarians frequently faced situations that extended beyond the scope of their official duties. Questions from visitors did not only pertain to digital government services, but often addressed broader cultural, social, and economic needs. This underscores the complexities of offering comprehensive support for the multifaceted challenges of digital citizenship. As Max pointed out:

> It is a drop in the ocean, right?...The questions you get are often not from IDO partners. They are websites from a government, or a semi-government, or an organization....That patch is so much bigger than just the [participating] government agencies.

For example, cultural participation is an integral part of the digital society, enabling people to access digital resources for personal enrichment. Our informants regularly received questions about online library services, such as downloading e-books and audiobooks, using e-readers or platforms like YouTube, as well as using library computers to learn Dutch and gain a deeper understanding of the local culture. In Library B, for instance, we observed how a couple signed up for a library subscription and, with a staff member's guidance, learned how to download an app for reading e-books (fieldnotes, 15-10-2020, Library B).

In the context of social participation, librarians helped citizens stay connected through digital tools like email, social media, and video calling software. In Library C, Sacha, a 59-year-old volunteer, shared the story of an elderly woman who sought help after losing her husband, who had previously managed their digital needs. According to our informants, many citizens in similar situations rely on librarians to navigate technology for maintaining social connections. Furthermore, librarians promoted the social participation of citizens by offering digital skills courses, which typically had a strong social component. Jamie, a 45-year-old volunteer at Library C, said: "There are many people who visit us who are often alone and truly enjoy spending a morning at the library." Indy, a 44-year-old programme officer, noted: "It is the social [aspect] that is the starting point, not really because they want to become more digitally proficient."

Finally, economic participation in the digital welfare state increasingly relies on digital components. Librarians assist citizens with processes such as online banking, managing finances, creating CVs, and starting businesses. As Chris, a 39-year-old basic skills advisor at library A, noted:

> It is not just about language or digital skills; it is about empowering self-reliance and addressing specific questions, like, indeed, "how to become self-employed."

Our findings show that to effectively support the wide range of participatory practices displayed in the public library, librarians must go beyond simply providing information on government services and offer

practical help, such as navigating websites or correctly filling out forms. Such support is vital for citizens who cannot independently use digital services that are not covered by the Digital Inclusion programme. Based on our observations and interviews, librarians are dedicated to meeting the needs of citizens by providing support and guidance in navigating digital services. However, they also faced four major challenges that make fulfilling these responsibilities difficult: limited resources, gaps in required skills, regulatory barriers, and issues with referrals.

In all three public libraries, ongoing austerity policies have led to a shortage of resources, such as suitable equipment, spaces for courses and consultation, and staff shortages. This has resulted in the overburdening of permanent staff and an increased dependence on volunteers, which has affected the quality of service delivery. Billie, a 63-year-old domain specialist at Library C, explained:

> The cuts went so far that the only thing left to reduce was the permanent staff, and as a result, volunteers have been recruited who are now heavily involved in day-to-day operations. And you can really feel the effect now, as there are very few well-trained professionals on the floor to address these issues.

In the same library, we observed that the central desk was largely understaffed. Staff were only briefly present when they needed to attend to something at the desk, leaving visitors waiting for help. Many considered pressing the bell on the desk but hesitated and ultimately decided against it, likely due to the quiet atmosphere in the library (fieldnotes, 05-10-2020, Library C).

A lack of skills and expertise is another major challenge for public libraries, especially as the sector is aging and has undergone significant changes in recent years. Jip noted: "Many colleagues came to work at the library when we still had our old role, and [they] do not know very well whether they have the expertise for this new position." Dominique, a project officer at library A, said: "There is a very big shift involved…and that is scary for many people. Because now, after so many years, they are asked to do something quite different." Although training is provided through the Digital Inclusion programme to prepare librarians to support citizens, librarians note that they often lack digital literacy, technical knowledge, and practical experience, which prevents them from addressing citizens' needs. In Library C, for instance, one staff member mentioned that they sometimes receive questions when they are near the printers or computers. Recently, they were asked about donor registration, but they felt unqualified to help, as it was something they weren't familiar with. (fieldnotes, 05-10-2020, Library C)

Thirdly, privacy regulations, such as the GDPR, can severely limit direct assistance, as librarians are not authorized to handle inquiries involving personal information. Librarians told us they felt conflicted in cases where citizens needed help but were prohibited from providing assistance despite being able to do so. Dominique said:

> That is one of the main issues I'm still dealing with, actually, privacy. When people come in with a letter saying, "Look, I've received this," you're immediately faced with: "Sorry, I can't help you."

This leads to frustration, as Kim, a 41-year-old librarian from Library A pointed out: "We are not allowed to do anything at all. It is really hard to work like that."

Finally, although librarians were expected to refer citizens to other institutions when questions went beyond their formal duties, in practice, these referrals were challenging. Librarians mentioned that citizens often face urgent deadlines and need immediate assistance. As Remy explained:

> There is someone in front of you who is often completely desperate, and it is always something that they would prefer to solve today or yesterday, and otherwise they really need it tomorrow. So, there is always pressure behind it.

Billie (63, Library C) explained that local collaborators "fish in the same pond, as they must also guarantee their right to exist" to receive funding. According to the librarians, this competition, driven by limited government resources and overlapping service scopes, led to strained relationships between public librarians and local social domain collaborators. Staff also mentioned poor communication and ineffective collaboration between organizations as major reasons for inadequate referrals. Jip said that, as a result, people have "been sent from pillar to post," leading to frustration and declining trust among citizens. To prevent this, librarians often preferred to handle questions internally, as Noah, a librarian from Library A, explained: "You want to help from start to finish and answer the question completely."

5. Discussion

Conceptualisations of digital citizenship embedded in the Dutch national Digital Inclusion programme conflict with the practices of librarians, as key mediators between citizens and the digital welfare state. The programme envisions an idealized citizen who is capable, self-reliant, and fully equipped to navigate a digital society. This meritocratic vision links shared prosperity in society to individual self-sufficiency, leaving little room for those unable to meet these expectations due to a range of intersectional and structural factors that are often beyond the individual's control. By applying this normative, one-size-fits-all approach to citizenship, the programme risks excluding the people it aims to support. Moreover, the programme's limited focus on digital government services overlooks the broader spectrum of practices that support citizens' participatory capacities. This limits the programme's potential for citizens to enrich their digital citizenship through the Digital Inclusion programme. Consequently, the digital citizenship conceptualisations embedded in the programme run the risk of hindering its goal of bridging the digital divide, and in some cases, may even deepen societal marginalization for those who are already disadvantaged.

Librarians play a critical role in bridging gaps in digital citizenship in the Netherlands, offering support that goes beyond the transactional and instrumental focus of the programme. In essence, the involvement of public libraries in digital support is a fairly innovative idea, building on the renewed social relevance and contemporary purpose public libraries have been seeking due to a decline in traditional readership. Historically, libraries have been associated with access to information and literacy, making the shift to becoming a digital literacy supporter a logical extension of their role. However, it is important to note that although libraries are framed by the programme as low-threshold, trusted, and informal spaces associated with accessibility and public support, this new image has not yet fully taken hold in broader society. Moreover, as a public institution, the library is not necessarily well-equipped to handle sensitive information or to guarantee the protection of personal privacy.

We find that the efforts of the programme are particularly hampered by systemic constraints in its implementation, such as resource shortages, inadequate digital expertise, and limited collaboration with other organizations in the social domain. These limitations make it challenging to provide the comprehensive assistance necessary for meaningful participation in the digital society. The rigid expectations of the programme fail to account for the diverse social, cultural, and economic contexts in which citizens operate, highlighting the inadequacy of a one-dimensional and one-size-fits-all model of digital citizenship. Instead, digital citizenship must be reimagined as a dynamic, intersectional, inclusive concept, one that acknowledges and addresses the differences, inequities, and complex realities of citizens' lives and enables a more equitable and participatory digital society.

Paradoxically, current Dutch digital inclusion policy lacks explicit definitions of digital citizenship, yet embeds high expectations of an idealized citizen. Our observations show that this is problematic, as it fails to address and may even exacerbate marginalization in initiatives meant to promote inclusion and bridge societal divides. The experiences of our interviewees show that true self-reliance is a myth: All citizens depend, to varying degrees, on structures beyond themselves. This interdependence should be acknowledged and integrated into policies and practices. Consequently, our study demonstrates an urgent need for redefining digital citizenship in a way that adequately reflects the complexities, differences, and inequalities among citizens to enable effective interventions. We propose that a reconceptualization of digital citizenship must take an intersectional approach, designing interventions with the most vulnerable citizens (those facing multiple, compounding disadvantages) as the starting point to ensure that all citizens can benefit from digital inclusion initiatives. This shift requires viewing digital inclusion not as a technical or instrumental issue, but as a social challenge, where dismantling structural barriers requires addressing social inequalities.

The design and implementation of the Dutch digital inclusion policy serve as a cautionary example for digital inclusion initiatives worldwide. As one of the forerunners in digitalization across the globe, the Netherlands demonstrates a rising tension between technological advancement on one hand, and the challenge of social inclusion on the other. While leveraging and repurposing public libraries for digital support is a smart way of using existing infrastructures and repurposing them to fit a new role in an advancing welfare state, there leaves much to be desired in terms of implementation to ensure that public libraries are adequately equipped to take on their new role as intermediaries of the digital welfare state.

The observations across the three libraries were largely consistent, with the notable exception of the job title and role of the librarians at Library B. These librarians were designated as "community librarians," and their role was deliberately situated at the intersection of library work and social work. In contrast to the other two libraries, where staff adhered more closely to traditional forms of librarianship, the employees at Library B appeared more aware of the competencies required for their evolving role and were better equipped to fulfil it. While this study provides insights into the realities of digital citizenship policies, it was subject to several limitations due to the outbreak of the Covid-19 pandemic. Restrictions imposed during this period significantly reduced opportunities for direct observations of interactions between citizens and librarians, as fewer citizens accessed library services due to public health measures. This also hindered our efforts to gather input from citizens themselves, as researchers sought to avoid putting visitors at risk. Post-pandemic research can explore citizens' own conceptualizations of digital citizenship more directly.

6. Conclusion

This article contributes to the digital divide literature by advancing two key arguments. First, it underscores the pivotal role of librarians as intermediaries in increasingly digitalized welfare states. These digital care workers not only assist individuals in navigating digital services but also embody the human infrastructure required to make digital inclusion a reality. Second, the article critiques prevailing conceptions of digital citizenship, revealing their limitations in everyday practice. Current policies often presuppose an idealized, self-sufficient citizen, which risks excluding precisely those populations that digital inclusion initiatives are meant to support. A more effective approach requires reimagining digital citizenship in ways that acknowledge social interdependence, structural inequalities, and the varied realities of everyday digital engagement.

As welfare states become digital-by-default, digital citizenship plays an increasingly important role. The relationship between the digital welfare state and digital citizenship shapes access to rights and services, and increasingly defines what it means to participate fully in digital societies. It is vital that both academic research and policymaking pay close attention to the reciprocity of this relationship. On one hand, access to technology, digital literacy, and online public services has become essential for full participation in public life. On the other hand, active engagement in public life and access to public services can, in turn, strengthen an individual's development as a citizen and their inclusion in society. As a result, these two dimensions reinforce each other (both positively and negatively). Understanding how digital citizenship and welfare structures interact enables the development of initiatives and programmes that foster mutual reinforcement, ultimately promoting social inclusion. Enhancing digital support for citizens requires a comprehensive and equity-focused approach that addresses both structural and individual barriers. Based on the experiences of librarians as human interfaces of the digital welfare state, key areas for improvement include better access to technological resources, such as user-friendly public devices, private rooms for handling personal information, and well-equipped course spaces. Equally important is the ongoing development of digital competencies and language skills, supported by trained staff with both technical knowledge and the agency to guide citizens through challenging digital environments.

The authorization function of national identification and verification systems should also be redesigned to allow citizens to easily and securely authorize others to manage their administrative tasks, ideally through non-digital methods. This would enable digital care workers to provide practical, secure, and privacy-friendly support to individuals who struggle with digital services.

To make support truly inclusive, it must be flexible and extend beyond rigid frameworks, offering personalized assistance, support closer to home, and programmes tailored to specific users, with content, formats, and availability that match their needs. Alternative forms of support must be developed for citizens who are not mobile enough to visit public spaces, preferably through home visits to ensure a private and safe environment with all documents available. Information should be made available in multiple accessible formats, with simpler language and greater language support for individuals with low literacy or limited Dutch proficiency.

Crucially, digital inclusion policy must recognize that not everyone will achieve full digital independence; for some, ongoing support will be necessary. This support should build on strong social ties, trust, and positive reinforcement, and offer meaningful alternatives to purely digital participation. Moreover, digital inclusion

efforts should expand beyond access to digital government services to promote broader forms of digital citizenship, including social, cultural, economic, and political practices.

To realize these goals, sustained investment is needed in hiring trained professionals (besides volunteers), in purchasing appropriate equipment, and in building strong local networks and referral systems. Finally, to ensure these interventions are truly effective, more research is needed into citizens' own perspectives, experiences, and needs (especially those in disadvantaged positions) to inform just and responsive digital inclusion policy.

Acknowledgments

We would like to extend our gratitude to all the librarians who participated in our research, the participating libraries, as well as the National Library of the Netherlands for their continuous support in all phases of the research project. Furthermore, we would like to thank Miyabi Babasaki, Rhanna Haverkort, Romy Roomans, and Lucy Frowijn for doing parts of the ethnographic research for this study, and Liesbet van Zoonen, Jiska Engelbert, and Aodhán Kelly for the inspiring and fruitful theoretical and conceptual discussions as well as their contributions to the research design of this Team Science project.

Funding

This research was funded by a Team Science grant (2020) from COMMIT/ and the UNL Digital Society programme, as part of the project Digital Literacy in the Public Library: Fostering Digital Inclusion and Civic Participation. Publication of this article in open access was made possible through the institutional membership agreement between the University of Groningen and Cogitatio Press.

Conflict of Interests

The authors declare no conflict of interests.

Data Availability

The research data related to this article is not available in an open data repository.

References

Bibliotheeknetwerk. (n.d.). *Dashboard Informatiepunten Digitale Overheid (IDO)*. https://www.bibliotheeknetwerk.nl/dashboard/dashboard-informatiepunten-digitale-overheid

Boerkamp, L. G. P., van der Zeeuw, A., van Deursen, A. J. A. M., van Laar, E., & van der Graaf, S. (2025). Internet appropriation barriers in the lives of Dutch parents living in poverty: A qualitative study. *The Information Society*, *41*(3), 195–208. https://doi.org/10.1080/01972243.2025.2467031

Centraal Bureau voor de Statistiek. (2020). *Internet: toegang, gebruik en faciliteiten; 2012–2019*. https://cbs.nl/nl-nl/cijfers/detail/83429NED?dl=35852

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Sage. https://doi.org/10.4135/9781452230153

DeVault, M. L., & McCoy, L. (2006). Institutional ethnography: Using interviews to investigate ruling relations. In D. E. Smith (Ed.), *Institutional ethnography as practice* (pp. 15–44). Rowman & Littlefield. https://doi.org/10.4135/9781412973588.n43

Digitale Overheid. (2020). *NL DIGIbeter 2020: Agenda Digitale Overheid*. https://archief29.sitearchief.nl/archives/sitearchief/20221104100849/https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/nldigibeter

Goedhart, N. S., Broerse, J. E., Kattouw, R., & Dedding, C. (2019). "Just having a computer doesn't make sense": The digital divide from the perspective of mothers with a low socio-economic position. *New Media & Society*, *21*(11/12), 2347–2365. https://doi.org/10.1177/1461444819846059

Greene, D. (2021). *The promise of access: Technology, inequality, and the political economy of hope*. MIT Press. https://doi.org/10.7551/mitpress/11674.001.0001

Hargittai, E. (2022). *Connected in isolation: Digital privilege in unsettled times*. MIT Press. https://doi.org/10.7551/mitpress/14224.001.0001

Helsper, E. J. (2021). *The digital disconnect: The social causes and consequences of digital inequalities*. Sage. https://doi.org/10.4135/9781526492982

Henman, P. (2010). *Governing electronically: E-government and the reconfiguration of public administration, policy and power*. Palgrave Macmillan. https://doi.org/10.1057/9780230248496

Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2019). *Digital citizenship in a datafied society*. Polity Press.

Hoevenagel, R., & Joossen, J. (2022). *Wie bereikt de (Rijks)overheid niet?* Leids OnderzoeksKollectief. https://www.communicatierijk.nl/documenten/publicaties/2022/11/11/update-literatuurstudie-mensen-de-rijksoverheid-niet-bereikt

Huysmans, F., & Oomes, M. (2018). The people's palaces: Public libraries in the information society. In E. van Meerkerk & Q. L. van den Hoogen (Eds.), *Cultural policy in the polder: 25 years Dutch Cultural Policy Act* (pp. 219–242). Amsterdam University Press.

Isin, E., & Ruppert, E. (2015). *Being digital citizens*. Rowman & Littlefield.

Katz, J. (1997, January 12). The digital citizen. *Wired*. https://www.wired.com/1997/12/netizen-29

Kaun, A., & Forsman, M. (2022). Digital care work at public libraries: Making digital first possible. *New Media & Society*, *26*(7), 3751–3766. https://doi.org/10.1177/14614448221104234

Kerssies, B., & Wijngaards, M. (2022). Alleen nog digitaal communiceren met de overheid: Uitkomst of utopie? Een paar kanttekeningen. *Nederlands Juristenblad*, *97*(47), 3542–3547. https://www.recht.nl/vakliteratuur/algemeen/artikel/541416

Lister, R., & Campling, J. (1997). *Citizenship: Feminist perspectives*. Macmillan Press. https://doi.org/10.1007/978-1-349-26209-0

Liu, S. M., & Yuan, Q. (2015). The evolution of information and communication technology in public administration. *Public Administration and Development*, *35*(2), 140–151. https://doi.org/10.1002/pad.1717

Manifestgroep. (n.d.). *Welkom bij de Manifestgroep*. https://manifestgroep.nl

Ministerie van Onderwijs, Cultuur en Wetenschap. (2019, October 10). *Curriculumherziening primair en voortgezet onderwijs* (Parliamentary Paper 31293, No. 485). Tweede Kamer der Staten-Generaal. https://zoek.officielebekendmakingen.nl/kst-31293-485.html

Mol, A. (2002). *The body multiple: Ontology in medical practice*. Duke University Press. https://doi.org/10.2307/j.ctv1220nc1

Mossberger, K., Tolbert, C., & McNeal, R. S. (2007). *Digital citizenship: The Internet, society, and participation*. MIT Press. https://doi.org/10.7551/mitpress/7428.001.0001

Schou, J., & Hjelholt, M. (2018). *Digitalization and public sector transformations*. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-76291-3

Smit, A., Swart, J., & Broersma, M. (2023). Digital inclusion of low-literate adults: Challenging the sequential underpinnings of the digital divide. In *Proceedings of the Weizenbaum Conference 2022: Practicing Sovereignty* pp. 72–84). Weizenbaum Library. https://doi.org/10.34669/wi.cp/4.7

Smith, D. E. (2005). *Institutional ethnography: A sociology for people*. AltaMira.

Smith, D. E. (1987). *The everyday world as problematic: A feminist sociology*. Northeastern University Press.

Staatscommissie Parlementair Stelsel. (2018). *Lage drempels, hoge dijken. Democratie en rechtsstaat in balans* (Parliamentary Paper 34430, No. 9). Tweede Kamer der Staten-Generaal. https://zoek.officielebekendmakingen.nl/kst-34430-9.html

Staatscourant. (2019). *Tijdelijke subsidieregels Digitale Inclusie 2019–2021 Koninklijke Bibliotheek* (Government Gazette 2019, No. 41052). Koninklijke Bibliotheek. https://zoek.officielebekendmakingen.nl/stcrt-2019-41052.html

Staatscourant. (2020). *Bibliotheekconvenant 2020–2023: Convenant houdende afspraken over de bijdrage van de bibliotheekvoorziening in Nederland aan maatschappelijke opgaven* (Government Gazette 2020, No. 57290). Koninklijke Bibliotheek. https://zoek.officielebekendmakingen.nl/stcrt-2020-57290.html

Tweede Kamer der Staten-Generaal. (1999). *Wetgeving voor de elektronische snelweg* (Parliamentary Paper 25880, No. 5). https://zoek.officielebekendmakingen.nl/kst-25880-5.html

United Nations E-government Knowledgebase. (n.d.). *E-government development index (EGDI)*. https://publicadministration.un.org/egovkb/Data-Center

Van Dijk, J. (2020). *The digital divide*. Polity Press.

Vivienne, S., McCosker, A., & Johns, A. (2016). Digital citizenship as fluid interface. In A. McCosker, S. Vivienne, & A. Johns (Eds.), *Negotiating digital citizenship: control, contest and culture* (pp. 1–18). Rowman & Littlefield.

Wetten.nl. (2022). *Tijdelijke subsidieregels Digitale Inclusie 2019–2021 Koninklijke Bibliotheek—BWBR0042443*. https://wetten.overheid.nl/BWBR0042443/2022-12-17

About the Authors

Maud Rebergen is a PhD candidate at the Centre for Media and Journalism Studies at the University of Groningen. Her research focusses on how digital technologies in public services may deepen social inequalities by excluding those with limited access, and how these technologies reshape everyday relationships between citizens and the state.

Joëlle Swart is an assistant professor at the Centre for Media and Journalism Studies at the University of Groningen. Her research focuses on changing everyday news use and how users develop knowledge, practices, and habits around news and journalism. She is a member of the editorial board of Digital Journalism.

Marcel Broersma is a full professor and director of the Centre for Media and Journalism Studies and its Digital Inclusion Lab at the University of Groningen. His research focuses on the interface between the digital transformation of journalism, changing media use, and digital literacy and inclusion.

Social Inclusion
2025 • Volume 13 • Article 10052
https://doi.org/10.17645/si.10052

cogitatio

ARTICLE

Open Access Journal

The Regime of Self-Optimization: Lived Experiences of Enforced Digital Inclusion by Low-Literate Citizens

Alexander Smit, Joëlle Swart, and Marcel Broersma

Centre for Media and Journalism Studies, University of Groningen, The Netherlands

Correspondence: Alexander Smit (a.p.smit@rug.nl)

Submitted: 3 February 2025 **Accepted:** 12 May 2025 **Published:** 16 July 2025

Issue: This article is part of the issue "Fostering the Socially and Ecologically Sustainable Digitalisation of Welfare States" edited by Paula Saikkonen (Finnish Institute for Health and Welfare) and Marta Choroszewicz (University of Eastern Finland), fully open access at https://doi.org/10.17645/si.i514

Abstract

This article introduces the regime of self-optimization, a theoretical framework to understand how disadvantaged citizens are compelled to continually improve their digital skills and capacities to meet the demands of an increasingly digital welfare state. Consequently, we ask: How do low-literate Dutch citizens experience the regime of self-optimization in their everyday lives? Drawing on historical and discursive underpinnings of governmentality, responsibilization, and standardization, we propose a conceptual framework to examine how top-down digital norms impose moral and practical obligations for self-optimization. We distinguish two dimensions of self-optimization: a vertical and horizontal mode. The first emphasizes extrinsic norms, efficiency, and personal responsibility, effectively attributing digital exclusion to personal failure while overlooking more profound structural barriers. The latter centers around intrinsic motivations, social support structures, trust, and context-sensitive adaptation, thereby fostering inclusion and agency, although its effectiveness depends on social resources. Drawing on a longitudinal ethnographic study conducted in libraries, a community center, and a vocational school in the Netherlands (41 participant observations; 23 semi-structured interviews), our findings show how these competing dimensions collide in the everyday lives of low-literate Dutch citizens, revealing frictions that highlight broader socio-political tensions in digitized welfare systems. These tensions highlight how the mechanisms intended to foster digital inclusion, by encouraging citizens to optimize themselves, can instead reinforce marginalization. By centering the experiences of marginalized groups, researchers and policymakers can more effectively address socio-economic, linguistic, and cultural barriers to digital inclusion. This challenges the assumption that universalized digital inclusion initiatives are sufficient for all citizens. The regime of self-optimization, thus, provides insights for designing human-centered, context-sensitive digital inclusion interventions in rapidly digitizing societies.

Keywords
digital inclusion; digital exclusion; digital inequality; digital literacy; low literacy; disadvantaged citizens; self-optimization

1. Introduction

Societies worldwide are rapidly digitizing public services under the promise of greater efficiency, flexibility, and cost savings (European Union, 2024). In digitized welfare systems, this techno-solutionist logic can overshadow the complex ways digital infrastructures intersect with issues of inequality and sustainability (Ruiu & Ragnedda, 2024; Saikkonen & Ilmakunnas, 2024). Despite growing concerns about ever-expanding digital ecosystems, many national and supranational policies still endorse digital-first or digital-by-default models that ignore social-digital equity (European Union, 2024). For marginalized groups, such as low-literate adults, this approach can exacerbate existing socioeconomic vulnerabilities and lead to new forms of exclusion (Choroszewicz & Mäihäniemi, 2020; Goedhart et al., 2019). Instead of alleviating life circumstances, whether shaped by unemployment, low-literacy, or broader societal issues, digitized welfare systems can become mechanisms of inequality that require individuals to self-optimize, navigate complex interfaces, and adopt prescribed technological norms. This effectively shifts the burden of digital inclusion onto those least equipped to bear it (Goedhart, 2021; Helsper, 2021; Notley & Aziz, 2024). This is particularly true for disadvantaged or marginalized publics, for whom such individual self-optimization tends to be impossible due to socio-economic, linguistic, cultural, affective, and digital barriers (Buddeberg, 2019; Friemel et al., 2021). This article presents a conceptual framework for analyzing the lived experiences of top-down-imposed digital norms and assumptions through the lens of the regime of self-optimization.

To show the value of this conceptual framework, we analyze the case of low-literate citizens in the Netherlands, where despite a high level of digitalization almost one in six adults is considered to be low-literate, and one in four lack sufficient digital skills to safely participate in a digital society (Buisman et al., 2024; European Union, 2023). Consequently, a significant portion of the Dutch population faces fundamental linguistic and digital barriers to digital inclusion (Smit et al., 2024b). Meanwhile, little is known about how imposed digital inclusion is experienced from the perspective of these disadvantaged individuals with limited linguistic proficiency. Studying these experiences contributes to several persistent research gaps. First, while studies show that literacy is a foundational competence for participating in society (Carpentieri, 2015), there is still much unknown about the relationship between low literacy, low digital literacy, and how this affects digital inclusion and exclusion (Buddeberg, 2019; Grotlüschen et al., 2019). Although frameworks on the digital divide have made significant strides in highlighting disparities in access, skills, and outcomes (Helsper, 2021; van Deursen et al., 2017), they risk obscuring structural factors by treating individual development and improvement as the primary remedy to digital inequalities (Alper, 2017). Second, despite previous findings that literacy supporters (Buddeberg, 2019) and digital care workers (Kaun & Forsman, 2024) play a significant role in promoting digital inclusion, little is known about how social resources are enacted in the everyday lives of low-literate citizens to foster digital inclusion (Asmar et al., 2020). Third, self-optimization has received limited attention in digital inequality research (Nehring & Röcke, 2023) and even less attention concerning digital citizenship (Ceccarini, 2021). It is primarily described in relation to, for example, academic debates about the quantification of social life (Lupton, 2021) or self-help culture (Nehring, 2024).

Consequently, this study argues for a socio-political paradigm to understand self-optimization as a potential vehicle of marginalization. The term "regime" enables us to conceptualize self-optimization as "more or less stable socio-material assemblages that surface as coherent patterns of thinking and acting in the world" (Deleuze & Guattari, 1972/1977, p. 503). Thus, it enables us to highlight the top-down, hierarchical nature of development, sustainability, and inclusion, rooted in historical efforts to govern citizens through governmentality and responsibilization (Foucault, 1991; Juhila et al., 2021). Self-optimization is an assemblage of discourses and practices that encourage individuals to pursue the optimal, imaginable version of their identities, bodies, and citizenship in relation to their everyday lives (Nehring & Röcke, 2023). Moreover, this understanding highlights the material and discursive practices that shape knowledge, steer behavior, and impose normative logic. Examining how the regime of self-optimization unfolds in everyday practice reveals the frictions and resistances of marginalized groups, which they use to challenge and counteract its embedded strategies and norms. Rooted in neoliberal ideals of personal responsibility, autonomy, and self-reliance, the regime prioritizes standardized digital access and literacy frameworks, ultimately placing the responsibility of inclusion on those least equipped to bear it. Consequently, we ask: How do low-literate Dutch citizens experience the regime of self-optimization in their everyday lives?

Drawing on historical and discursive underpinnings of governmentality and responsibilization, we propose a theoretical framework to examine how top-down digital norms impose moral and practical obligations for self-optimization. Building on a longitudinal ethnographic study conducted in libraries, a community center, and a vocational school in the Netherlands (41 participant observations; 23 semi-structured interviews), we distinguish two dimensions of self-optimization: a vertical and horizontal mode. The first emphasizes extrinsic norms, efficiency, and personal responsibility, effectively attributing digital exclusion to personal failure while overlooking more profound structural barriers. The latter centers around intrinsic motivations, social support structures, trust, and context-sensitive adaptation, thereby fostering inclusion and agency, although its effectiveness depends on social resources. This study resonates with ongoing scholarly debates about digital inequalities and the digital divide, which often attribute individuals' responsibility for overcoming digital barriers solely to them (Goedhart et al., 2022). In doing so, they divert attention from collective and policy-level imperatives to enhancing individual skills. Failing to do this exacerbates socio-digital vulnerability and simultaneously widens social and digital inequalities, as enforced digitization renders especially disadvantaged citizens vulnerable to exclusion and further social and digital inequalities. Hence, the regime of self-optimization enables scrutiny of how top-down imposed norms, values, and assumptions foster socio-digital inequalities and vulnerability by neglecting structural barriers and making low-literate Dutch citizens responsible for their digital inclusion and exclusion.

2. The Regime of Self-Optimization

Based on neoliberal ideals of responsibilization and self-management (Juhila et al., 2021), and the underpinnings of Foucault's (2008) governmentality theory, we introduce the theoretical framework of "the regime of self-optimization." This framework is developed to understand how disadvantaged citizens must continually improve their digital skills and capacities to meet the demands of an increasingly digital welfare state. The regime of self-optimization can be theoretically situated within neoliberal governance and digital welfare studies. Under neoliberal governmentality, state policies increasingly cultivate self-governing subjects, citizens who regulate and improve themselves to participate and meet institutional aspirations (Nehring & Röcke, 2023). The regime epitomizes this: It imposes extrinsic norms of digital competence and

shifts responsibility for inclusion onto the individual (Juhila et al., 2021). In a digital context, the "ideal" citizen is an active, self-reliant user of e-services, requiring minimal state intervention (Henman, 2010). In this sense, neoliberal rationality produces citizens as rational actors and "entrepreneurs of the self," whose moral worth is measured by their capacity for self-care and self-management (Nehring, 2024). Drawing from this theoretical perception, the regime of self-optimization advances digital welfare and digital inequality studies to show how enforced digital inclusion manifests as a regime for disadvantaged citizens to forcefully optimize their digital competencies and compliance with e-governmental systems (Henman, 2010; Nehring & Röcke, 2023). In doing so, the regime of self-optimization highlights how what might be framed as empowerment or digital inclusion can, in practice, become a mandate to legitimize exclusion and potentially become a vehicle of marginalization.

Self-optimization has become a pervasive cultural narrative, emphasizing continuous self-improvement across various domains, such as health, productivity, and social relationships (Nehring, 2024; Nehring & Röcke, 2023). It promotes the belief that individuals must refine and enhance their capacities to meet societal demands, often framed as personal responsibility or self-reliance (Juhila et al., 2021). This emphasis intersects with digital inclusion efforts, particularly for disadvantaged, low-literate citizens, who are increasingly expected to adapt to digital environments and acquire skills to navigate digitized public services (Notley & Aziz, 2024). The concept of self-optimization, therefore, provides a critical lens for examining the socio-political dimensions of digital citizenship.

The neoliberal focus on self-optimization related to class, ethnicity, gender, education, and income (Helsper, 2021). Digital policies often mirror these expectations by placing the burden of digital inclusion on individuals, particularly through the promotion of skills training or access to digital tools. However, such frameworks often overlook the socio-economic, cultural, and infrastructural disparities that hinder genuine inclusion. The "belabored self" (McGee, 2005) exemplifies this contradiction, as individuals are encouraged to embrace digital skills not as a pathway to empowerment but as a necessity to remain competitive and employable within a digitized economy (Nehring & Röcke, 2023). Moreover, self-optimization operates within a 'therapeutic culture' that commodifies self-improvement, framing digital inclusion as both a moral obligation and a gateway to modern citizenship (Illouz, 2007; Rimke, 2000). This narrative often depoliticizes digital inclusion and exclusion, presenting them as failures of individual effort rather than addressing systemic barriers, such as insufficient infrastructure or inaccessible digital platforms. By examining self-optimization as a socio-political construct, this framework connects digital inclusion with broader debates on individualization, responsibilization, and the shifting nature of self-identity under neoliberal capitalism (Beck, 2000; Giddens, 1991).

The institutional embrace of the regime of self-optimization as a universal ideal often obscures the deeper issues of social justice. In neoliberal discourse, civic obligations to be "productive" and "self-sufficient" reinforce power differentials, making it difficult for low-literate and low-digital-literacy citizens to acquire skills on their terms (Juhila et al., 2021). This lies at the basis of the regime of self-optimization, rooted in a logic of meritocracy. It delineates an apparent dichotomy between the economically and politically dominant "haves" and the disenfranchised "have-nots." This suggests that battling inequality and exclusion by the latter requires adopting the wealth accumulation strategies of the former (van Assche & Hornidge, 2015).

From a political-economic perspective, the self-optimization of the citizenry emphasizes individuality, autonomy, self-reliance, and active participation (Newman & Tonkens, 2011). This logic suggests that people can empower themselves and others, add value to society, and enhance their societal position by developing themselves via practices of self-management and self-optimization (De Brabander, 2014). According to neoliberal discourse rooted in Western ideas of "modernization," such a development logic entails an ideology centered on social change, where new ideas are introduced into a social system to produce higher per capita incomes and living standards through more modern production methods and improved social organization. This paradigm consequently emphasizes that the primary manner in which "vulnerable," "disadvantaged," and "underdeveloped" people can improve their standard of living is by following the steps of the ideology of the homo-economicus (Weber et al., 2006), centralizing active socio-economic participation in an individualized configuration rooted in meritocracy. Digitization fosters an additional layer of expectations and norms, as citizens must appropriate their capabilities and preferences to fit within the "digital mold" (Broersma et al., 2024). However, this neoliberal paradigm often overlooks critical social dimensions, including human rights, environmental sustainability, and community empowerment (Newman & Tonkens, 2011).

2.1. Neoliberal Discourse as Hegemonic Practice

Discourses centering on autonomy, self-reliance, resilience, empowerment, self-management, and similar concepts have become part of institutions' hegemonic practices (Weber et al., 2006). This hegemonic practice discursively construes implicit neoliberal norms, civic duties, values, and interests as universal, projecting civic normativity (Smit et al., 2024a). Underneath this practice lies a regime of self-optimization that frames inclusion, participation, and development as universal, unambiguous, and inherently positive. The strength of this ideal lies in its ability to be discursively framed as inherently positive for democracies, underpinned by a range of political and economic agendas. Simultaneously, citizens are held responsible for their successes and failures (Juhila et al., 2021), and the ability to appropriate this regime of self-optimization to personal conditions is considered in terms of vulnerability, capability, and (dis-)advantage. Thus, self-optimization relates to both online and offline contexts, as it entails how digital access and usage are operationalized in terms of capabilities and limitations. This logic promotes neoliberal ideals for digital inclusion by emphasizing individual responsibility and self-improvement. Citizens are expected to internalize these ideals, adapting to fit the organizational norms without questioning the underlying structures (van Assche & Hornidge, 2015). This logic does not accommodate personal circumstances or foster mutual understanding; instead, it perpetuates a system in which individuals must continually optimize themselves to meet institutional demands.

The key aspect within these power mechanisms is the demand that individuals take action themselves through self-management services and, to a certain extent, govern themselves on the premise of having access to information on what is legitimate and their rights. Such operational aspects in digital public sector practices draw upon the preference for agentic, ideologically constructed concepts, such as "self-reliance," "self-optimization," and "self-efficacy," reflecting active citizenship ideals.

Although often framed as empowering citizens, such ideologies paradoxically emerge from positions of dominance and privilege. We understand this discursive dynamic as "bureaucratese" (Watson, 1997). Bureaucratese operates through analogical reasoning, meaning it remains bounded by internal logic and the

constraints of institutional political discourse. According to Hummel (1994), bureaucratese represents a distinct institutional language that imposes ideological frameworks onto citizens. Contrasting bureaucratese with everyday societal communication, Hummel (1994) highlights that institutional language is inherently one-dimensional, enforcing conformity rather than enabling reciprocal dialogue.

Central to bureaucratese is the institutional context itself, rather than the lived experiences of citizens or the broader social environment (Hummel, 1994). This language imposes an idealized institutional form on real-world interactions, embedding neoliberal ideologies, such as self-optimization, into daily communication practices (Watson, 1997). Unlike ordinary communication, where meaning emerges through shared contexts and mutual experiences, bureaucratese enforces its predefined context. Consequently, citizens interacting with institutions must adapt to predetermined norms and expectations. In essence, bureaucratese communicates *at* citizens rather than engaging *with* them in dialogue (Watson, 1997).

Political discourse, expressed through bureaucratic language, functions as a tool for disseminating and reinforcing neoliberal ideologies. Consequently, it reproduces societal and digital norms that presuppose citizens possess adequate socio-economic capital for digital access, basic functional and digital literacies for practical use, and physical and cognitive abilities to articulate and overcome barriers encountered in digital media. Disadvantaged citizens often find their circumstances misaligned with one or more of these expectations, creating tensions between institutional norms and their actual daily practices and capabilities. This discrepancy highlights the neoliberal emphasis on self-optimization, fostering a techno-solutionist logic wherein citizens are expected to manage and resolve civic matters independently through digital technologies perceived as efficient, effective, and accessible (Schou & Pors, 2018).

2.2. The Self-Service Turn

While self-optimization is commonly associated with the quantified self (Lupton, 2021), this article adopts a political-economic perspective, defining how individuals enhance their abilities to access, utilize, and benefit from digital media. This concept aligns with neoliberal ideologies emphasizing self-management, self-service, and individual responsibility (Juhila et al., 2021). Under such frameworks, disadvantaged citizens are often positioned as personally accountable when they fail to meet digital expectations, for instance, when they struggle to access public services such as social housing or welfare through e-government platforms. Consequently, their inability to navigate these infrastructures results in further marginalization and exclusion (Goedhart, 2021; Smit et al., 2023).

Critically examining the principles underpinning the regime of self-optimization sheds light on its implications for digital inclusion and civic participation (Ceccarini, 2021). These principles outline the skills, knowledge, behaviors, and conditions necessary for digital inclusion, embedding assumptions about the competencies required of citizens in a digital-by-default society (Mensonides et al., 2024). Additionally, this framework highlights the crucial roles of confidence, trust, and articulatory power in shaping experiences of digital marginalization. It emphasizes how disadvantaged individuals bear responsibility for their digital and social inclusion, reinforcing self-optimization as both a personal mandate and structural expectation (Juhila et al., 2021).

While examining the regime of self-optimization is especially pertinent today, as key societal functions increasingly adopt digital-by-default approaches, this concept itself is not entirely new. In the 1990s, the shift towards self-service highlighted elements of responsibilization and self-optimization. Henman (2010) describes this transition as part of a neoliberal approach that merges cost efficiency with customer-oriented service, fundamentally altering the relationship between citizens and welfare services. Unlike traditional welfare states, which position individuals as passive recipients, the self-service model expects citizens to actively engage with welfare institutions independently, ensuring they secure the benefits to which they are entitled.

More recently, this self-service trend has expanded significantly through digitalization initiatives, digital education, and digital policy frameworks, laying the groundwork for contemporary understandings of self-optimization. Central to this development is the concept of responsibilization, which underscores the transactional relationship between individuals and the state. Current policy documents, such as the Dutch government's "digital agenda" (State Government of the Netherlands, 2024), explicitly reinforce these neoliberal values by expecting citizens to proactively seek assistance through institutional channels, such as libraries or government service points: Citizens "must take the initiative themselves to visit a service provider" (State Government of the Netherlands, 2024, p. 18). This requirement poses additional challenges for disadvantaged or digitally low-literate individuals, as it burdens them with the responsibility of improving their digital competencies and navigating complex digital infrastructures independently (Bovens et al., 2017).

To examine how individuals cope with these demands and limitations imposed by the self-optimization regime, we utilize the concept of self-efficacy (Maddux, 2016). Self-efficacy refers to an individual's confidence in their ability to handle various situations effectively. This is crucial for understanding how individuals perceive and navigate digital barriers related to their capabilities and limitations, such as digital access and literacy (Kappeler, 2024). Self-efficacy correlates closely with perceived competence (Ryan & Deci, 2000) and perceived behavioral control (Ajzen, 2020), both of which significantly influence an individual's digital engagement (Fathali & Okada, 2018). By applying self-efficacy as a lens, we can better comprehend the tensions arising from imposing self-optimization strategies on disadvantaged individuals and how they experience and reflect upon these socio-digital power dynamics concerning their practical needs and capabilities.

2.3. Making Citizens Do More

Research since the 1990s has addressed issues of autonomy, participation, and the digital divide (Norris, 2000; van Dijk, 2020). However, less attention has been given to how disadvantaged citizens experience and perceive the digitization of essential democratic processes tied to digital self-optimization, despite evidence that digital means often perpetuate existing social exclusions and inequalities (Helsper, 2021). Imposing digitalization on citizens through frameworks of self-optimization can exacerbate feelings of alienation among disadvantaged groups, thereby reinforcing socio-digital vulnerability and contributing to the formation of a digital underclass (Helsper & Reisdorf, 2017). Helsper and Reisdorf (2017) highlighted that social inequalities have intensified over time, increasingly concentrating digital exclusion within vulnerable groups such as the elderly, individuals with lower education, and isolated populations. This trend has a significant impact on institutional roles, transforming social workers into digital care workers (Kaun & Forsman, 2024), where efficiency-driven austerity measures increasingly overshadow traditional social

support systems (Ceccarini, 2021). While efficiency arguments frame digitalization as a mechanism to support citizens' transition into a digital-by-default society, in practice, it simultaneously accelerates inequality and marginalization.

In the Dutch context and comparable European democracies, policies are increasingly focused on automating public administration and digitizing public services to empower citizens to use self-service (Margetts & Dunleavy, 2013). Digital and functional literacies have thus become essential prerequisites for citizen participation, interaction with public administration, and access to essential services, such as social welfare, online banking, and tax services, all of which are becoming predominantly digital (European Union, 2023; Frau-Meigs et al., 2017; Smit et al., 2025). Recent Dutch regulations permit institutions to conduct interactions exclusively through digital means, without providing non-digital alternatives (State Government of the Netherlands, 2024).On the one hand, the Netherlands portrays itself as a welfare state built on the principle of meritocracy, promoting autonomy and individual development. This national identity suggests that every citizen has the freedom and opportunity to realize their ambitions (De Brabander, 2014). On the other hand, the Dutch government actively promotes neoliberal norms and values, such as individualization and personal responsibility (Ballin, 2021), particularly among disadvantaged groups, creating tension between institutional ideals and practical realities (Juhila et al., 2021).

For example, the Dutch digital inclusion policy frequently presumes that providing digital access and training citizens in digital literacy will naturally reduce digital inequalities and promote inclusion (Ballin, 2021). This perspective overlooks the heterogeneity among digitally excluded groups, ignoring their varying degrees of access, skills, and socio-affective capabilities (Goedhart et al., 2022; Leurs, 2016). Moreover, it overlooks the significant impact of individuals' social environments on their use of ICT (Friemel et al., 2021) and the crucial role that social support structures play in promoting digital inclusion (Smit et al., 2024b). Third-party support actors are seldom incorporated into mainstream conceptualizations of digital divides and inclusion (Smit et al., 2023). Therefore, the regime of self-optimization encompasses political strategies aimed at integrating diverse social groups under the principles of autonomy, empowerment, self-reliance, and personal responsibility; yet, it often overlooks the practical complexities of citizens' lived realities.

3. Methodology

To understand how low-literate Dutch citizens navigate literacy barriers and experience digital citizenship in everyday life, we applied a grounded theory-inspired approach (see Charmaz, 2021). We conducted weekly participant observations in three libraries, a vocational school for adult education, and a community center across the Netherlands. From March 2022 to October 2023, we observed 41 participants and conducted 23 follow-up semi-structured interviews. These venues, located in two regions with higher poverty rates affecting between 10% and 24% of the population (Centraal Bureau voor de Statistiek, 2024), offered digital literacy courses and technical support for digital devices.

By combining observations with interviews, we witnessed digital interactions as they occurred, while also gathering participants' reflections, thereby enhancing the richness and reliability of our findings. Participants were recruited through a combination of snowball sampling, where participants recommended others facing similar language and digital challenges, and purposive sampling, targeting low-literate citizens via educational staff, library workers, social service personnel, and community center volunteers who had previously assessed

the literacy levels of potential participants. Assessments of literacy were based on national benchmarks for both native (NT1) and non-native (NT2) Dutch speakers, using specific linguistic proficiency levels (A2 for NT2 and 1F for NT1) as inclusion criteria (Stichting Lezen en Schrijven, 2017). Generally, NT2 participants had higher educational achievements and digital literacy but lower Dutch language proficiency, whereas NT1 participants exhibited the opposite.

During participant observations, we engaged with 41 individuals aged between 31 and 73 years ($M = 54$, $SD = 5.2$), comprising 56% women and 44% men from 10 different nationalities. These settings enabled us to examine how low-literate adults navigate digital technologies on a daily basis. From this group, 23 individuals were interviewed to gain a deeper understanding of their experiences with digital inclusion and exclusion. Participants needed to be adults (18+) to be included in the sample. Due to the limited availability of interview participants, they ranged in age from 37 to 71 years ($M = 53$, $SD = 5.1$), with 61% female and 39% male, representing three nationalities.

Participants were selected through a combination of snowball sampling, where respondents referred us to others who were willing to discuss challenges in language competence and digital skills, and purposive sampling, where organizations, partners, and individuals directed us to low-literate adults. Low-literate citizens were identified with the help of educational practitioners, library staff, social workers, and community center volunteers, who were already familiar with the individuals' literacy levels through previous assessments. These evaluations utilized the national NT2 reference framework for non-native speakers and the official language proficiency levels for native Dutch speakers (NT1; see Stichting Lezen en Schrijven, 2017). We applied the same linguistic proficiency levels (A2 and 1F, respectively; see Stichting Lezen en Schrijven, 2017) as a criterion for participant inclusion.

Interviews, lasting approximately 45 minutes, were conducted in various locations and explored daily digital practices and media usage, focusing on how participants adapt their digital literacies amidst linguistic challenges. Topics included digital media habits, device and app preferences, usage frequency, and motivations. The interviews were audio-recorded, transcribed verbatim, and subjected to thematic analysis using Atlas.ti, following Clarke and Braun's (2017) iterative coding methodology. This process involved generating initial codes, re-examining transcripts for refined coding, and iterative re-coding to explore the contexts in which socio-digital dispositions were analyzed. Through this analytical process, we identified the horizontal and vertical dimensions of self-optimization and how our participants experienced them.

All data were anonymized to ensure privacy, and pseudonyms were used for any names cited in our findings. Informed consent was consistently obtained, although challenges arose with participants with limited reading and writing skills, particularly those who did not speak Dutch or English. In such cases, oral consent was carefully obtained, often using translation software to ensure clarity. The lead researcher introduced himself, explained his role, and outlined the study's focus to ensure participants were fully informed. Individuals who could not comprehend the research objectives or provide informed consent were excluded from the study. The research received ethical clearance from the Ethics Board of the University of Groningen (CETO) on 17 May 2023, under reference number 572–93593093.

4. Findings

The vertical dimension within the regime of self-optimization is characterized by extrinsic, top-down institutional demands prioritizing standardized norms of digital competence, digital access, and digital literacy over situated individual circumstances. On the contrary, the horizontal dimension reflects socially situated and context-driven learning processes facilitated by social support networks. Frictions arise when these dimensions with their differing norms and values collide in the everyday practices of disadvantaged citizens, undermining intrinsic motivations, needs, and rights, and exacerbating digital exclusion and inequality. The dynamics in both dimensions are examined through the lived experiences of our participants. This highlights the tensions between institutional demands and personal needs, providing a valuable conceptual framework for studying vernacular instances of digital inclusion and exclusion.

4.1. Vertical Self-Optimization

The vertical dimension of self-optimization is deeply rooted in top-down expectations that frame digital inclusion as a universal necessity based on the assumption of a homogeneous baseline of skills, resources, and access. The vertical mode is essentially an instantiation of the neoliberal citizenship ideal in the digital arena: Individuals must continuously upgrade their skills, devices, and literacies to remain included and not be seen as burdensome to the state (Schou & Pors, 2019). These assumptions create significant barriers for disadvantaged individuals, who, in practice, face complex interfaces and inaccessible processes. At the same time, implicit norms burden the individual with digital participation. Importantly, the vertical dimension builds on existing digital inclusion and governance paradigms by emphasizing standardization and efficiency.

By centering the regime's demands on conformity and compliance, individuals are expected to align their capabilities and behaviors with standardized norms, irrespective of their socioeconomic, linguistic, or emotional realities. This approach often fails to account for the systemic and structural barriers that prevent disadvantaged groups from meeting these expectations. As a result, those unable to conform or develop themselves in alignment with these demands face exclusion and disempowerment. The implications of this exclusion are profound, as individuals who cannot meet these imposed requirements are often excluded from participation, thereby exacerbating their marginalization. The neoliberal ideology driving this vertical dimension of the regime of self-optimization reinforces the belief that everyone must contribute to society, perpetuating a system where failure to comply equates to failure to contribute. This framing neglects the structural inequalities and systemic barriers that disadvantaged groups face, making digital inclusion a burden rather than an opportunity.

Barbara (57, F) illustrates how vertical self-optimization can exacerbate exclusion by shifting the burden of digital inclusion onto individuals without adequate support. Her social worker advised her to request a free laptop and attend digital literacy courses to manage e-governmental services independently:

> My social worker advised me to ask for a free laptop from the municipality and follow educational courses to become [digitally] literate enough to make use of those [e-governmental] services and arrange my affairs from my own home [the participant is partially disabled]. I didn't really want to do this, but I accepted his proposition because I thought it would make things easier. However, it made it way more difficult. Everything went fine when a social worker came by once a week and helped me

> with those difficult digital things in person, and I also very much liked to talk with that person, as I'm not able to go outside much. Now I have to do it on my own, and if I need help, I need to call someone, which often makes things even more difficult.

This case highlights a key aspect of vertical self-optimization: the reliance on top-down directives prioritizing individual responsibility over tailored support. Barbara's limited mobility and lack of digital skills were not adequately considered, and the withdrawal of personal visits resulted in the loss of both practical assistance and meaningful social interaction. The institutional focus on cost-cutting measures and digital self-reliance imposed unrealistic expectations on Barbara, leading to dual exclusion: digitally, due to insufficient resources and education, and socially, as the reduced support further isolated her. Her experience highlights the limitations of vertical self-optimization in addressing structural barriers and individualized needs, ultimately reinforcing exclusion and inequality rather than fostering inclusion.

Carlos's (61, M) experience resembles Barbara's: In his case, structural and infrastructural spatial barriers contributed to the feeling of imposed digital inclusion and being a burden instead of an opportunity. Living in a rural area with unreliable internet connectivity, Carlos was unable to participate in online courses or access virtual communities. Unlike Barbara, whose burden stemmed from withdrawing personalized support, Carlos faced external barriers that rendered self-optimization impossible. These spatial challenges evoked feelings of helplessness and frustration, as Carlos lacked the agency to overcome the infrastructural limitations. His case highlights the systemic inequalities that are often overlooked when digital inclusion policies assume uniform access to and availability of resources.

In addition, Asaf's (43, M) experience echoes Barbara's and Carlos's perceptions in terms of emotional burden but differs in the specific affective consequences. Feeling coerced into acquiring digital skills for e-governmental services, Asaf struggled with the complexity of online courses and digital portals. While Asaf had access to the e-governmental services, he was unable to use them due to limited digital literacy, a lack of confidence in his capabilities, and fear that he would make a mistake. Hence, the enactment of e-governmental services is not solely determined by access or skills. His failure to meet institutional expectations led to shame, anxiety, and withdrawal from social participation. Unlike Barbara, whose burden was amplified by a lack of direct assistance, or Carlos, whose negative experience was affected by spatial issues, Asaf's challenges were compounded by internalizing his struggles as personal failures. This erosion of self-esteem and avoidance of further engagement illustrate how vertical self-optimization undermines self-efficacy and perpetuates exclusion.

In all three cases, institutional presuppositions of universal digital competence clashed with individual circumstances. Barbara's physical disability, Carlos's infrastructural limitations, and Asaf's low digital literacy were not accommodated, reflecting a failure to align policies with diverse user needs. Each case reveals the affective toll of enforced digital inclusion. Barbara experienced isolation from reduced social interaction, Carlos felt helpless due to spatial constraints, and Asaf internalized shame and frustration. These emotional responses underscore the limitations of policies that prioritize compliance over individual well-being.

Thus, the vertical dimension of the regime of self-optimization imposes rigid expectations on individuals, assuming they possess the necessary skills, devices, and agency to navigate increasingly digitized systems. This assumption systematically undermines the self-efficacy and agency of disadvantaged individuals,

creating a disconnect between institutional norms and personal realities. Bianca's (57, F) story exemplifies how this has a significant effect on participants' sense of competence:

> To register for a digital course, the municipality of [redacted] asked me to download a form from their website, fill in my personal information, and upload it. I tried it on my phone, but I couldn't open the file. I called the help desk, and the person advised me to open the file on a computer or access it through Gmail. However, I don't have a computer or a phone with Gmail. I did not tell this to the employee, as this isn't something I'm proud of. I just said thank you and hung up.

Ashamed to admit her limitations, Bianca ended the conversation by feigning gratitude, leaving the issue unresolved. Bianca's experience underscores how the vertical dimension reduces agency by framing failure to meet digital expectations as a personal inadequacy. Her inability to articulate her challenges reflects how the affective burden of shame inhibits participants from seeking further assistance. Rather than feeling empowered, Bianca's encounter left her feeling disengaged and unsupported, revealing the alienation caused by institutional frameworks that prioritize efficiency over empathy.

The internalization of struggles with imposed digitalization as personal failure harms citizens' self-esteem and contributes to a loss of trust in institutions, as seen in the case of Carla (53, F). Similar to Bianca, she felt the pressure of conforming to digital norms that she was ill-equipped to meet. For Carla, the vertical demands of self-optimization magnified her exclusion and created a psychological barrier to re-engaging with institutional systems. Hence, while the language of self-optimization may appear uplifting, it can mask systemic and affective barriers that disproportionately affect those already marginalized.

Feelings of shame are central to how vertical self-optimization undermines agency, as seen in Karim's (45, M) case, who was unable to navigate his child's school portal. This excluded him from participating in his child's education and amplified his sense of inadequacy as a parent. This highlights how the vertical dimension imposes normative expectations that individuals internalize, eroding their ability to challenge or negotiate these demands. This has profound consequences for affective dispositions, such as shame and distrust, fostering a sense of unbelonging and a lack of agency and control, which can potentially result in social exclusion.

Ella (63, F) experienced a similar loss of agency, as her inability to keep up with her friends' online interactions left her feeling isolated and guilty. Unlike Bianca, who faced direct material and systemic barriers, Ella's exclusion was relational, as her digital illiteracy severed her from social networks. Both cases illustrate the broader impact of vertical self-optimization, where systemic assumptions about digital competence lead to affective exclusion and a diminished sense of self-worth.

These examples collectively reveal how the vertical dimension of the regime of self-optimization disempowers individuals by imposing unrealistic expectations of independence and competence, where digital-by-default is the norm. The presumption of a homogenous baseline of digital skills and access ignores the structural and personal disparities marginalized groups face. For Bianca, Carla, Karim, and Ella, these expectations eroded their confidence, created feelings of inadequacy, and diminished their ability to assert agency within institutional systems. Addressing these issues requires a paradigm shift in digital inclusion policies, focusing on fostering self-efficacy and supporting diverse capabilities, rather than perpetuating a one-size-fits-all approach that marginalizes those unable to meet the regime's demands.

4.2. Horizontal Self-Optimization

By contrast, the horizontal dimension within the regime of self-optimization aligns with theoretical approaches that emphasize social context, collective support, and user-centric adaptation in digital inclusion (Goedhart, 2021). Rather than seeing the citizen as an isolated rational actor, the horizontal perspective sees them as embedded in a social fabric where learning and inclusion occur through relationships and trust. This resonates with socially situated understandings of learning and digital inclusion, for example, the notion that people learn more effectively in community-based practice contexts (Goedhart, 2021). The horizontal dimension builds on this by framing such peer or community support as an intrinsic mode of self-optimization: Individuals are motivated to improve their skills not just to meet an abstract standard, but to connect with others, to achieve personal goals, or to alleviate immediate fears (such as the fear of "feeling stupid" with technology) in a supportive environment.

Sylvia's (54, F) account illustrates how self-optimization in the horizontal dimension is deeply social and situated. She relies on her daughter's smartphone for essential digital interactions:

> I don't have any digital device, I'm afraid I'll do something wrong and lose my income or get scammed or something. Still, if I want to send a picture, call someone, or check the weather, I'll borrow my daughter's phone. First, she had to do all these things for me, but now that I've seen her do it hundreds of times, I know how to do it myself….But only the simple things. If you ask me to do those things with taxes or online banking, that's impossible for me to do by myself. I need my daughter or someone else for that.

Over time, Sylvia observed her daughter performing various tasks, gradually gaining the confidence to handle more complex functions independently. However, she continues to rely on her daughter for complex digital tasks, such as managing taxes or online banking. Sylvia's intrinsic motivation to learn emerges not from imposed institutional expectations but from the immediate practical needs she encounters in her daily life. Her dynamics of self-optimization reflect an interplay between her desire for independence and the safety net provided by her daughter's support. This highlights how horizontal self-optimization enables individuals to develop digital competencies at their own pace and within their context, fostering a sense of agency that acknowledges both their capabilities and limitations.

Safety and familiarity are essential for these social dynamics to foster a positive affective experience, as shown by Ellen (59, F):

> All the time I hear that I need to go to school or develop all these digital skills. But why? I like to have my sister beside me helping me out, and as far as I know, she likes it too. We view it as a social activity, a bonding experience. We laugh about the mistakes I make, and then she helps me. I never feel stupid in front of her when I don't know how to do something. However, when I go to a school or something and make a mistake, that is when I feel stupid, and that anxiety pops up again.

Ellen's experience further exemplifies the relational nature of horizontal self-optimization. Her sister's involvement in teaching her to use a smartphone transforms digital learning into a shared activity that strengthens their bond. Ellen values this collaborative process, which alleviates her anxiety about making mistakes and helps build her confidence. For Ellen, the presence of a trusted social resource enables her to

acquire the skills she needs to use applications like WhatsApp for social interaction. Her intrinsic motivation stems from her desire to connect with friends and family, not from external pressures. This alignment with her needs and preferences fosters a sense of self-efficacy, as Ellen feels capable of navigating digital tools within the safety of her social network. Her experience underscores how horizontal self-optimization prioritizes equity by tailoring digital participation to the individual's circumstances and leveraging social resources to mitigate affective barriers.

In contrast, Andre's story highlights the challenges faced when social support is insufficient or unavailable:

> People told me that having a smartphone was beneficial and easy. My neighbor even gave me one for free and showed me how to use it. Initially, I was able to open apps, use WhatsApp, and make calls. However, soon messages about the memory being full stopped me from updating or using it properly. Eventually, I couldn't even start it. I gave it back, saying, no thanks.

While his neighbor initially helped him learn basic smartphone functions, Andre struggled to maintain access independently. Messages about memory storage overwhelmed him, and without the vocabulary or confidence to articulate the issue, he returned the phone and disengaged entirely. Unlike Sylvia and Ellen, Andre lacked sustained social resources to guide him through these challenges, leading to a diminished sense of self-efficacy. His case illustrates the fragility of self-optimization when social support is limited or temporary. The absence of ongoing assistance and the lack of intrinsic motivation to prioritize digital tools contributed to his withdrawal from digital engagement. For Andre, the potential of the horizontal dimension remains unrealized because the social resources needed to foster sustained inclusion were insufficient.

Implicit norms of self-optimization, framing digital participation as inherently valuable, also shape the attitudes of individuals from disadvantaged backgrounds. Participants, such as Sylvia, questioned the utility of digital inclusion in their lives. Sylvia doubted whether owning a smartphone would enhance her sense of belonging, echoing Andre's sentiment. For both, the perceived value of digital access did not outweigh the economic and technological constraints required to sustain it. This reliance on informal social support structures fosters vulnerability. When such resources are unavailable, individuals face increased risks of exclusion.

Sylvia, who accessed digital tools through her daughter's phone, and Andre, who learned basic smartphone skills from his neighbor, illustrate adaptive forms of self-efficacy. Their experiences highlight how disadvantaged individuals creatively navigate digital barriers, leveraging their specific capabilities and limitations. These cases collectively highlight the central role of personal needs, motivations, and social contexts in the development of digital literacy. Consequently, policies and educational programs should focus on cultivating robust social infrastructures to support disadvantaged individuals. Digital inclusion efforts must move beyond mere access and usage, acknowledging the inherently relational and personalized aspects of self-optimization.

An apparent tension exists between institutional expectations of uniform digital competence (vertical dimension) and the personalized, socially driven methods disadvantaged citizens employ (horizontal dimension). Institutional frameworks typically hold individuals accountable to standardized digital requirements without sufficiently considering their limited resources, connectivity issues, or emotional and physical barriers. When these individuals struggle to meet these standards, they are often blamed for their

perceived shortcomings. In contrast, the horizontal dimension emphasizes reliance on social networks, where support from peers, family, and neighbors helps mitigate fear and shame associated with digital tasks. Despite their practical value, these collaborative, context-sensitive learning approaches are often undervalued by institutional systems that prioritize efficiency, cost-effectiveness, and measurable outcomes.

5. Discussion

The findings demonstrate that employing the regime of self-optimization as a theoretical framework effectively reveals the experiences of disadvantaged citizens within socio-digital infrastructures. This approach highlights the tensions arising from interactions between the vertical demands imposed by institutional structures and the horizontal, socially embedded practices employed by individuals. Such frictions often undermine the intrinsic motivations, needs, and rights of disadvantaged citizens, thereby intensifying digital exclusion and inequality. By closely examining the lived experiences of low-literate individuals, the framework offers nuanced insights into how digital exclusion is intertwined with e-government systems. Additionally, the results highlight how the meritocratic principles embedded in self-optimization narratives shift responsibility to individuals, exacerbating feelings of shame and vulnerability when expectations, such as adequate broadband coverage, user-friendly interfaces, and personalized guidance, are not met. Conversely, community support networks help mitigate these adverse effects, allowing digital tasks to be perceived as collective challenges rather than solitary burdens. This finding does not romanticize informal support, as such networks are often unstable; instead, it emphasizes that digital inclusion policies focused exclusively on vertical performance metrics risk perpetuating inequalities without sustained investment in horizontal support systems, such as care, translation, and mentoring.

The dual structure of the regime of self-optimization refines traditional governmentality theory (Foucault, 2008; Henman, 2010), which predominantly emphasizes the internalization of top-down power dynamics. By incorporating the horizontal dimension, the framework reveals that outcomes are not uniform and that individual agency and social interactions significantly mediate the effects of institutional power. This perspective addresses recent critiques in social theory and digital welfare studies, advocating for a more complex understanding of neoliberal governance beyond simplistic, one-dimensional representations (Nehring & Röcke, 2023; Schou & Pors, 2019).

This perspective aligns with recent discussions on the digital welfare state, emphasizing the ambivalence inherent in digitally transforming public service provision (Kaun & Forsman, 2024; Zakharova et al., 2024). While digital technologies promise to streamline welfare services by enhancing efficiency, personalizing care, and reducing bureaucratic burdens, these technologies simultaneously introduce new surveillance and control mechanisms (Kaun & Liminga, 2023; Zakharova et al., 2024). For example, Kaun and Forsman (2024) document how public librarians undertake digital care work, revealing hidden administrative responsibilities and gendered labor divisions embedded within digitalization processes.

Our framework critically examines this context, highlighting how digital welfare policies often impose self-regulation expectations ill-suited to the realities of disadvantaged citizens. Such individuals are frequently required to adopt self-management practices that presuppose a level of digital competence they lack, exacerbating exclusion and systemic vulnerabilities.

The vertical dimension of self-optimization translates traditional self-service expectations into digital contexts, framing compliance with digital norms as a civic obligation. According to this view, being a "good citizen" equates to digital capability and autonomy (Nehring & Röcke, 2023). Noncompliance, such as failing to submit online applications on time, is treated as a personal failure, reinforcing a moral discourse that blames disadvantaged citizens for their exclusion rather than addressing structural inequalities (Henman, 2010; Kaun & Forsman, 2024). Consequently, administrative systems promoting standardized digital interactions inadvertently marginalize citizens with diverse capabilities.

In contrast, the horizontal dimension emphasizes relational approaches characterized by empathy, community support, and context-specific learning. Participants reported greater comfort and confidence when developing digital skills alongside trusted individuals, such as friends or family, rather than in impersonal, standardized training environments. These relational methods align with self-efficacy principles, enhancing empowerment through tailored, individualized support. However, the effectiveness of such approaches depends significantly on the availability of reliable social networks. Recognizing this variability highlights the importance of institutional policies that actively facilitate community-driven interventions, rather than relegating them to secondary roles. Institutions can better support individuals' self-efficacy and achieve more equitable digital inclusion outcomes by formally integrating these horizontal strategies.

Our findings challenge the assumption that digital inclusion is universally advantageous and equally relevant to all. Many low-literate individuals perceive digital inclusion initiatives as burdensome, creating additional barriers rather than mitigating existing ones. Attributing digital marginalization to personal resistance unfairly places responsibility on individuals, neglecting broader structural, political, economic, and cultural factors that perpetuate digital inequality. This argument aligns with Riddell's (2009) critique of the "moral underclass" discourse, cautioning against attributing social problems solely to individual attitudes without considering external influences. Indeed, digital exclusion cannot be reduced solely to personal choice, as technological access, educational resources, and socioeconomic conditions have a profound impact on it. This resonates with humanistic and communitarian perspectives on digital inclusion, emphasizing bottom-up, context-sensitive solutions (Alper, 2017; Helsper, 2021).

This clash intensifies when the top-down expectations of universal digital literacy implicitly treat the horizontal modes of learning and support as deficient or transitional rather than recognizing them as vital, context-specific strategies for coping with digital demands. Disadvantaged individuals often rely on trusted family members or neighbors to navigate online forms, e-government portals, or other complex digital services. While such support networks foster confidence and reduce feelings of isolation, they starkly contrast institutional logics that valorize self-reliance and measure success primarily through standardized metrics.

Additionally, research indicates that social and digital vulnerabilities are unevenly experienced, influenced by cultural, legal, policy, and technological factors (Alper, 2017; Helsper, 2021; van Dijk, 2020). Disadvantaged individuals often struggle to fully leverage the potential benefits of digital technologies, thereby limiting their effectiveness as tools for achieving equity and inclusion (Goedhart, 2021). Access and basic digital skills alone do not guarantee meaningful participation; instead, they represent an initial step toward comprehensive engagement within digital societies (Ragnedda, 2020). Thus, this study refines the notion of self-optimization by illustrating how its outcomes depend critically on the interplay between institutional structures and the availability of social support.

The regime of self-optimization framework can be extended to other marginalized groups, such as elderly populations who often face pressure to adopt digital technologies ill-suited to their needs. Friemel et al. (2021) illustrate how digital inequalities among older adults involve intersecting challenges related to age, skills, and motivation. These can be seen as conflicts between institutional expectations (e.g., using online banking) and practical realities (e.g., relying on relatives for assistance). Thus, the assumption that self-optimization inherently reduces inequalities merits scrutiny. The framework effectively identifies the balance—or imbalance—between horizontal social support and vertical institutional pressures in various contexts.

6. Conclusion

This study investigated how disadvantaged, low-literate citizens experience enforced digital inclusion within a regime of self-optimization that frames digital participation as both a moral duty and a practical necessity. To explore the tensions between these citizens and the socio-technical policies embedded in the regime of self-optimization, we identified two dimensions that structure their experiences: a vertical and a horizontal mode. The vertical mode emphasizes individual responsibility and compliance with institutional norms, thus potentially reinforcing neoliberal assumptions and overlooking structural inequalities. In contrast, the horizontal mode emphasizes community support, empathy, and culturally sensitive approaches, factors often absent in mainstream digital inclusion policies. Thus, the horizontal dimension serves both as a complement to and a challenge for the vertical dimension, emphasizing the importance of socially supported and empathetic interventions.

By clearly defining these vertical and horizontal dimensions, the study bridges macro-level critiques of neoliberal governance with micro-level lived experiences. While traditional governmentality theory predominantly emphasizes top-down control (Foucault, 2008), our findings reveal that horizontal relationships can mediate, modify, and sometimes mitigate these power dynamics, contingent upon the availability and institutional recognition of supportive social networks. For policymakers, this indicates that sustainable digital inclusion requires not only promoting individual self-reliance but also actively incorporating collective learning and supportive community practices. Practically, this involves developing accessible digital interfaces, ensuring non-digital alternatives remain available, and providing institutional support for community-based digital training.

The regime of self-optimization thus bridges critical perspectives on digital governance and more constructive aspirations for digital empowerment. Its analytical structure effectively integrates macro-level policy critique and micro-level lived realities, enhancing our understanding of digital inclusion processes. By operationalizing abstract concepts into tangible, empirically grounded dimensions, this framework offers valuable insights into the complexities of digital inclusion, highlighting the interplay between institutional power (vertical) and everyday practice (horizontal). Our findings also reveal that standardized digital demands clash with the everyday realities of low-literate individuals who rely on social networks to navigate technology, confront affective challenges, and mitigate shame or anxiety. This tension highlights how the very mechanisms intended to foster inclusion, by urging citizens to “optimize” themselves, can reinforce marginalization when structural supports are inadequate. We argue that recognizing and engaging with these experiences is crucial for designing equitable digital policies.

Although the regime of self-optimization was formulated based on research with low-literate adults in the Netherlands, its relevance extends far beyond this specific case. The framework addresses phenomena that are common across many countries and diverse populations in an era of digital transformation of public services. Many advanced welfare states are undergoing a similar shift toward digital-by-default governance, as seen in the Netherlands. The European Union's Digital Economy and Society Index (DESI) and national digital strategies consistently push for more online services and citizen e-participation. This suggests that vertical-horizontal tension is not unique to a country's bureaucratic culture, but rather a general feature of digitizing welfare systems. Researchers and policymakers in other countries can apply the regime of self-optimization as a lens to examine their digital inclusion challenges. For instance, one might investigate whether strong social safety nets and community programs (horizontal strengths) mitigate the pressures of digital-by-default (vertical demands) better than in other places.

The study's scope is necessarily bounded. It centers on a purposive sample of low-literate adults in the Netherlands; other welfare regimes, linguistic contexts, or demographic groups may display divergent constellations of vertical pressure and horizontal support. Nor does the analysis account in depth for intersecting factors such as disability, migration status, or racialization, which could compound or reshape digital vulnerability. Longitudinal data would also be required to gauge whether shifts toward participatory co-design or simplified bureaucratic procedures translate into enduring improvements in self-efficacy and inclusion.

Future work can address these limitations by applying the vertical-horizontal schema in comparative settings, examining its relevance for other marginalized constituencies, and testing interventions that combine accessible design with durable social support. Such research will refine understanding of how digital power operates and, crucially, how it might be redirected.

Addressing digital inclusion and exclusion from the perspective of disadvantaged, low-literate Dutch citizens necessitates a critical reevaluation of existing policy frameworks, educational methods, and local support systems. Central to this approach is developing inclusive policies that validate informal and community-based learning methods, rather than viewing them merely as transitional steps towards institutional compliance (Goedhart et al., 2019). Equally crucial is the provision of continuous, context-sensitive support through accessible digital tools, simplified interfaces, and supportive environments that enable individuals to develop digital skills without fear of judgment. Community-oriented and peer-learning strategies involving families, friends, or neighborhood organizations should be systematically integrated into formal educational programs. These practices help bridge the gap between institutional expectations and the realities of daily life experienced by disadvantaged groups. Additionally, it is vital to establish robust channels for dialogue and feedback between disadvantaged citizens and policymakers, ensuring that lived experiences, emotional well-being, and localized digital practices inform inclusive digital policy-making. Prioritizing these areas can effectively mitigate tensions between vertical institutional requirements and horizontal social support mechanisms, fostering a self-optimization regime that genuinely empowers individuals rather than marginalizing them.

Acknowledgments

We want to thank all the participants for sharing their experiences, perspectives, and opinions with us. We would also like to thank the libraries, community centers, and educational institutions for their assistance in recruiting participants for this study and for allowing us to gather data at their facilities.

Funding

The authors disclose receipt of the following financial support for the research, authorship, and/or publication of this article: This publication is part of the research project Informed Citizenship for All. Digital Literacy as Prerequisite for an Inclusive Society, funded by the Dutch Research Council (NWO), grant no. 410.19.008, and supported by the National Library of the Netherlands (KB), the Ministry of the Interior and Kingdom Relations (BZK), and Stichting Kinderopvang Stad Groningen (SKSG). Publication of this article in open access was made possible through the institutional membership agreement between the University of Groningen and Cogitatio Press.

Conflict of Interests

The authors declare no conflict of interests.

References

Ajzen, I. (2020). The theory of planned behavior: Frequently asked questions. *Human Behavior and Emerging Technologies*, *2*(4), 314–324. https://doi.org/10.1002/hbe2.195

Alper, M. (2017). *Giving voice: Mobile communication, disability, and inequality*. MIT Press.

Asmar, A., van Audenhove, L., & Mariën, I. (2020). Social support for digital inclusion: Towards a typology of social support patterns. *Social Inclusion*, *8*(2), 138–150. https://doi.org/10.17645/si.v8i2.2627

Ballin, E. H. (2021). *Mensenrechten als ijkpunten van artificiële intelligentie* (Working Paper No. 42). Wetenschappelijke Raad voor het Regeringsbeleid. https://www.wrr.nl/publicaties/working-papers/2021/09/29/mensenrechten-als-ijkpunten-van-artificiele-intelligentie

Beck, U. (2000). *Risk society: Towards a new modernity*. Sage.

Bovens, M., Keizer, A., & Tiemeijer, W. (2017). *Weten is nog geen doen: Een realistisch perspectief op redzaamheid* (Report No. 97). Wetenschappelijke Raad voor het Regeringsbeleid.

Broersma, M., Swart, J., Mensonides, D., Smit, A., & Rebergen, M. (2024). Digital in- and exclusion in everyday life: Practices and literacies across the lifespan. *Media and Communication*, *12*(1), 1–6. https://doi.org/10.17645/mac.v12i1.9245

Buddeberg, K. (2019). Supporters of low literate adults. *International Journal of Lifelong Education*, *38*(4), 420–432. https://doi.org/10.1080/02601370.2019.1600059

Buisman, M., Bollen, I., Jacobs, B., Huijts, T., Cornelisse, R., & van Gulik, N. (2024). *PIAAC 2023: Resultaten voor Nederland*. Research Centre for Education and the Labour Market, Maastricht University.

Carpentieri, J. D. (2015). Adding new numbers to the literacy narrative: Using PIAAC data to focus on literacy practices. In M. Hamilton, B. Maddox, & C. Addey (Eds.), *Literacy as numbers: Researching the politics and practices of international literacy assessment* (pp. 93–110). Cambridge University Press.

Ceccarini, L. (2021). *The digital citizen(ship): Politics and democracy in the networked society*. Edward Elgar Publishing.

Centraal Bureau voor de Statistiek. (2024). *Demografische gegevens populatie Nederlandse burgers: Dashboard armoede*. https://dashboards.cbs.nl/v3/appArmoede_2022

Charmaz, K. (2021). The genesis, grounds, and growth of constructivist grounded theory. In J. M. Morse, B. J. Bowers, K. Charmaz, A. E. Clarke, J. Corbin, C. J. Porr, & P. Noerager Stern (Eds.), *Developing grounded theory: The second generation revisited* (2nd ed., pp. 153–187). Routledge. https://doi.org/10.4324/9781315169170-13

Choroszewicz, M., & Mäihäniemi, B. (2020). Developing a digital welfare state: Data protection and the use of automated decision-making in the public sector across six EU countries. *Global Perspectives*, *1*(1), Article 12910. https://doi.org/10.1525/gp.2020.12910

Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology*, *12*(3), 297–298. https://doi.org/10.1080/17439760.2016.1262613

De Brabander, R. (2014). *Wie wil er nou niet zelfredzaam zijn? De mythe van zelfredzaamheid.* Garant.

Deleuze, G., & Guattari, F. (1977). *Anti-Oedipus: Capitalism and schizophrenia*. University of Minnesota Press. (Original work published 1972)

European Union. (2023). *DESI 2022: Full European analysis.* https://digital-strategy.ec.europa.eu/en/policies/desi

European Union. (2024). *DESI 2023: Full European analysis.* https://digital-strategy.ec.europa.eu/en/policies/desi

Fathali, S., & Okada, T. (2018). Technology acceptance model in technology-enhanced OCLL contexts: A self-determination theory approach. *Australasian Journal of Educational Technology*, *34*(4), 69–87. https://doi.org/10.14742/ajet.3873

Foucault, M. (1991). Governmentality. In G. Burchell, C. Gordon, & P. Miller (Eds.), *The Foucault effect: Studies in governmentality* (pp. 87–104). University of Chicago Press.

Foucault, M. (2008). *The birth of biopolitics: Lectures at the Collège de France, 1978–1979*. Palgrave Macmillan.

Frau-Meigs, D., Velez, I., & Michel, J.-F. (Eds.). (2017). *Public policies in media and information literacy in Europe: Cross-country comparisons.* Taylor & Francis.

Friemel, T., Frey, T., & Seifert, A. (2021). Multidimensional digital inequalities: Theoretical framework, empirical investigation, and policy implications of digital inequalities among older adults. *Weizenbaum Journal of the Digital Society*, *1*(1), Article w1.1.3. https://doi.org/10.34669/wi.wjds/1.1.3

Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age.* Polity Press.

Goedhart, N. S. (2021). *Social inclusion in digitizing societies: Starting from the lifeworld of people with a low socioeconomic position* [Unpublished doctoral dissertation]. Vrije Universiteit Amsterdam.

Goedhart, N. S., Broerse, J. E. W., Kattouw, R., & Dedding, C. (2019). "Just having a computer doesn't make sense": The digital divide from the perspective of mothers with a low socio-economic position. *New Media & Society*, *21*(11/12), 2347–2365. https://doi.org/10.1177/1461444819846059

Goedhart, N. S., Verdonk, P., & Dedding, C. (2022). "Never good enough": A situated understanding of the impact of digitalization on citizens living in a low socioeconomic position. *Policy & Internet*, *14*(4), 824–844. https://doi.org/10.1002/poi3.315

Grotlüschen, A., Buddeberg, K., Redmer, A., Ansen, H., & Dannath, J. (2019). Vulnerable subgroups and numeracy practices: How poverty, debt, and unemployment relate to everyday numeracy practices. *Adult Education Quarterly*, *69*(4), 251–270. https://doi.org/10.1177/0741713619841132

Helsper, E. J. (2021). *The digital disconnect: The social causes and consequences of digital inequalities.* Sage.

Helsper, E. J., & Reisdorf, B. C. (2017). The emergence of a "digital underclass" in Great Britain and Sweden: Changing reasons for digital exclusion. *New Media & Society*, *19*(8), 1253–1270. https://doi.org/10.1177/1461444816634676

Henman, P. (2010). *Governing electronically: E-government and the reconfiguration of public administration, policy and power.* Springer.

Hummel, R. P. (1994). *The bureaucratic experience: The post-modern challenge* (6th ed.). Routledge.

Illouz, E. (2007). *Cold intimacies: The making of emotional capitalism.* Polity Press.

Juhila, K., Raitakari, S., & Hall, C. (2021). *Responsibilisation at the margins of welfare services.* Routledge.

Kappeler, K. (2024). *Negotiating digital technology use in the highly digitized Swiss society: A mixed-method analysis of the digital practices of individuals* [Unpublished doctoral dissertation]. University of Zurich.

Kaun, A., & Forsman, M. (2024). Digital care work at public libraries: Making Digital First possible. *New Media & Society*, *26*(7), 3751–3766. https://doi.org/10.1177/14614448221104234

Kaun, A., & Liminga, A. (2023). Welfare service centers: Maintenance, repair, and care at the analog interfaces of the digital welfare state. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448231220362

Leurs, K. (2016). Digital divides in the era of widespread internet access: Migrant youth negotiating hierarchies in digital culture. In K. Ponnet, E. Vanderhoven, M. Walrave, J. Haers, & B. Segaert (Eds.), *Youth 2.0: Social media and adolescence—Connecting, sharing and empowering* (pp. 61–78). Routledge.

Lupton, D. (2021). Self-tracking. In R. P. Robinson & B. Jones (Eds.), *Information: Keywords* (pp. 187–198). Columbia University Press.

Maddux, J. E. (2016). Self-efficacy. In J. E. Maddux (Ed.), *Interpersonal and intrapersonal expectancies* (pp. 41–46). Routledge.

Margetts, H., & Dunleavy, P. (2013). The second wave of digital-era governance: A quasi-paradigm for government on the web. *Philosophical Transactions of the Royal Society A*, *371*(1987), Article 20120382. https://doi.org/10.1098/rsta.2012.0382

McGee, M. (2005). *Self-help, Inc.: Makeover culture in American life*. Oxford University Press.

Mensonides, D., Smit, A., Talsma, I., Swart, J., & Broersma, M. (2024). Digital literacies as socially situated pedagogical processes: Genealogically understanding media, information, and digital literacies. *Media and Communication*, *12*(1), Article 8174. https://doi.org/10.17645/mac.8174

Nehring, D. (2024). The self in self-help: A re-appraisal of therapeutic culture in a time of crisis. *Sociological Research Online*, *29*(2), Article 13607804241242345. https://doi.org/10.1177/13607804241242345

Nehring, D., & Röcke, A. (2023). Self-optimisation: Conceptual, discursive and historical perspectives. *Current Sociology*, *72*(6), 1069–1087. https://doi.org/10.1177/00113921231202809

Newman, J., & Tonkens, E. (Eds.). (2011). *Participation, responsibility and choice: Summoning the active citizen in Western European welfare states*. Amsterdam University Press.

Norris, P. (2000, April 10–13). *The worldwide digital divide* [Paper presentation]. Annual Meeting of the Political Studies Association of the UK, London School of Economics and Political Science, London, UK.

Notley, T., & Aziz, A. (2024). The unjust burden of digital inclusion for low-income migrant parents. *Policy & Internet*, *16*(2), 428–442. https://doi.org/10.1002/poi3.383

Ragnedda, M. (2020). *Enhancing digital equity: Connecting the digital underclass*. Springer Nature.

Riddell, S. (2009). Social justice, equality and inclusion in Scottish education. *Discourse: Studies in the Cultural Politics of Education*, *30*(3), 283–296. https://doi.org/10.1080/01596300903036889

Rimke, H. (2000). Governing citizens through self-help literature. *Cultural Studies*, *14*(1), 61–78. https://doi.org/10.1080/095023800334986

Ruiu, M. L., & Ragnedda, M. (2024). From poverty to digital poverty. In M. L. Ruiu & M. Ragnedda (Eds.), *Digital-environmental poverty: Digital and environmental inequalities in the post-Covid era* (pp. 15–39). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-56184-9_2

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Saikkonen, P., & Ilmakunnas, I. (2024). Reconciling welfare policy and sustainability transition: A case study of the Finnish welfare state. *Environmental Policy and Governance*, *34*(1), 53–64. https://doi.org/10.1002/eet.2055

Schou, J., & Pors, A. (2018). Digital by default? A qualitative study of exclusion in digitalized welfare. *Social Policy & Administration*, *53*(3), 464–477. https://doi.org/10.1111/spol.12470

Smit, A., Swart, J., & Broersma, M. (2023). Digital inclusion of low-literate adults: Challenging the sequential underpinnings of the digital divide. In B. Herlo & D. Irrgang (Eds.), *Practicing sovereignty—Interventions for*

open digital futures: Proceedings of the Weizenbaum Conference 2022 (pp. 72–84). Weizenbaum Institute. https://doi.org/10.34669/wi.cp/4.7

Smit, A., de Winkel, T., & Wieringa, M. (2024a). Exposing civic normativity: Applying the persona-based walkthrough method to the Dutch Happiness Meter. *Mediatization Studies*, *8*, 25–46.

Smit, A., Swart, J., & Broersma, M. (2024b). Bypassing digital literacy: Marginalized citizens' tactics for participation and inclusion in digital societies. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448231220383

Smit, A., Rebergen, M., Swart, J., & Broersma, M. (2025). Including disadvantaged citizens in smart societies: How learning in formal and informal educational settings fosters digital literacy. In B. Fung & H. Chen (Eds.), *Smart cities to smart societies* (pp. 29–50). Routledge.

State Government of the Netherlands. (2024). *Waardengedreven digitaliseren 2024: Geactualiseerde werkagenda*. Ministry of the Interior and Kingdom Relations.

Stichting Lezen en Schrijven. (2017). *Verschillen in niveau-aanduidingen voor Nederlandstaligen en anderstaligen*. www.lezenenschrijven.nl/sites/default/files/2020-08/Verschil_in_niveau-aanduidingen_Nederlandstaligen_en_anderstaligen_LS_V201701.pdf

van Assche, K., & Hornidge, A.-K. (2015). *Rural development: Knowledge & expertise in governance*. Wageningen Academic Publishers.

van Deursen, A. J. A. M., Helsper, E. J., Eynon, R., & van Dijk, J. A. G. M. (2017). The compoundness and sequentiality of digital inequality. *International Journal of Communication*, *11*, 452–473. https://ijoc.org/index.php/ijoc/article/view/5739

van Dijk, J. A. G. M. (2020). *The digital divide*. Polity Press.

Watson, R. P. (1997). Wittgenstein on language: Toward a theory (and the study) of language in organizations. *Journal of Management History*, *3*(4), 360–374. https://doi.org/10.1108/13552529710184181

Weber, S., Maurer, S., & Peters, M. A. (2006). Neoliberal governmentality: Foucault on the birth of biopolitics. In S. Weber, S. Maurer, & M. A. Peters (Eds.), *Gouvernementalität und Erziehungswissenschaft: Wissen—Macht—Transformation* (pp. 37–49). VS Verlag für Sozialwissenschaften.

Zakharova, I., Jarke, J., & Kaun, A. (2024). Tensions in digital welfare states: Three perspectives on care and control. *Journal of Sociology*, *60*(3), 540–559. https://doi.org/10.1177/14407833241210371

About the Authors

Alexander Smit is a PhD candidate at the Centre for Media and Journalism Studies at the University of Groningen. His study examines the aspects of contemporary digital literacy frameworks that disadvantaged citizens perceive as problematic or important in their civic lives, as well as their experiences of digital exclusion and inequality.

Joëlle Swart is an assistant professor at the Centre for Media and Journalism Studies at the University of Groningen. Her research focuses on changing patterns of news use and how users develop knowledge and habits around news and journalism. She is a member of the editorial board of *Digital Journalism*.

Marcel Broersma is a full professor and director of the Centre for Media and Journalism Studies and its Digital Inclusion Lab at the University of Groningen. His research focuses on the interface between the digital transformation of journalism, changing media use, and digital literacy and inclusion.

Social Inclusion
2025 • Volume 13 • Article 9950
https://doi.org/10.17645/si.9950

cogitatio

ARTICLE

Open Access Journal

Experiencing Social Exclusion and Distrust: Mental Health Rehabilitees Struggling With Digital Administrative Burdens

Hannele Palukka [1], Anne Koski [1], Jaana Parviainen [1], and Laura Eilola [2]

[1] Faculty of Social Sciences, Unit of Social Research, Tampere University, Finland
[2] Faculty of Information Technology and Communication Sciences, Tampere University, Finland

Correspondence: Hannele Palukka (hannele.palukka@tuni.fi)

Submitted: 28 January 2025 **Accepted:** 29 April 2025 **Published:** 16 July 2025

Issue: This article is part of the issue "Fostering the Socially and Ecologically Sustainable Digitalisation of Welfare States" edited by Paula Saikkonen (Finnish Institute for Health and Welfare) and Marta Choroszewicz (University of Eastern Finland), fully open access at https://doi.org/10.17645/si.i514

Abstract

The digital welfare state has been transforming into a type of state structure in which the citizen no longer faces the official in person, but interaction instead occurs with system-level bureaucracy, and decision-making takes place in an algorithmic system's frame. This article aimed to determine what kind of burden digital self-service and algorithmic decision-making poses to people with mental health problems. The article contributes to the social sustainability literature by introducing the concept of digital administrative burden and how it can create social exclusion and distrust in public administration among vulnerable citizens. Drawing on social representation theory and the concept of social identity, we examine how the digital administrative burden faced by mental health rehabilitees affects their identity and self-perception. The data gathered comprises audio-recorded group discussions with mental health rehabilitees. Based on the data, it can be demonstrated that digital citizen–state encounters create subject positions for mental health rehabilitees that portray them as dispossessed, unreliable, insignificant, and inferior. The positioning of mental health rehabilitees highlights how administrative burdens in digital citizen–state encounters contribute to social exclusion, preventing the full realisation of their citizenship. It can be stated that digital burdens imposed by automated administration on those dependent on social benefits can endanger society's social sustainability. This is particularly irritating because the welfare state's capability and commitment to social sustainability are crucial in preventing development of inequality and polarisation between different population groups.

Keywords

digital administrative burdens; digital citizen–state encounters; mental health rehabilitee; social exclusion; social sustainability

1. Introduction

The digital welfare state describes a type of state structure in which diverse digital technologies—such as automated systems, algorithms, big data analytics, and artificial intelligence (AI)—are incorporated into government agencies' management and delivery of welfare services. In some instances, these services are provided through a network of contracted providers, a concept known to social policy scholars as "the mixed economy of welfare" (van Toorn et al., 2024). Thus, the state's operations have changed through digitalisation. Bovens and Zouridis (2002) described a transition from "street-level bureaucracies," in which citizens engage directly with public officials in physical offices, to "screen-level bureaucracies," in which interactions occur in person, over the phone or through online chat with bureaucrats who are working behind screens, and ultimately system-level bureaucracy, in which decision-making on welfare benefits takes place in an algorithmic system frame. In many areas of social services, citizens are engaging more frequently with governments via digital platforms and mobile applications (Henman, 2019; Schou & Hjelholt, 2018). The goal of digital welfare states is to enhance access and efficiency in service delivery while significantly cutting costs. This is achieved by minimising reliance on manual human tasks and administrative work, and by using algorithms to automate decisions regarding eligibility, service levels, and other aspects of social resource distribution for individuals and their families (van Toorn et al., 2024).

The rise of algorithmic systems and self-transactions in digitalized public services has included the promise of providing services more efficiently, with more transparent, value-neutral, impartial, and fair decision-making (Bovens & Zouridis, 2002; Casey, 2022; Wirtz et al., 2019). While it commonly is believed that digital technology's main advantage is its ability to automate tasks and improve efficiency, several extant studies have pointed out that human involvement is crucial for its functioning (Le Ludec et al., 2023; Pinchevski, 2023). In practice, citizens, as benefit claimants, operate as human co-producers of public information systems, requiring citizens to provide all necessary information, verify calculations, and present evidence to contest algorithmic decisions. This has shifted the burden of proof, which, in cases of alleged debts, has been described as a form of extortion (Carney, 2020). In digital welfare states, it has been observed that automation leads to greater administrative burdens as agencies become more data-hungry to support complex eligibility testing, verification and compliance (Madsen et al., 2022; Peeters, 2023).

Bureaucratic processes associated with automated systems are likely to shape barriers that citizens face when interacting digitally with the state. According to Peeters (2020), these barriers are unlikely to be intentional but may arise from the complexity of service systems that cannot be adapted to simple rule-based automation (Larsson, 2021). Thus, administrative burdens can be unintended results of large-scale digitalisation and algorithmic decision-making processes in which some citizens do not fit into predefined boxes and stereotyped categories (Peeters & Widlak, 2018). For example, Griffiths (2024) demonstrated how people with irregular pay dates risk missing out on benefits for which they are eligible because automation processes do not account for irregular cases. Similarly, Widlak and Peeters (2020) have stated that some citizens face various administrative burdens in correcting errors made by these algorithmic systems. Administrative burden originally was defined as citizens' onerous experiences with policy implementation (Burden et al., 2012; Halling & Baekgaard, 2024). In the field, administrative burden has been thought to materialise as compliance, learning, and psychological costs that citizens experience in citizen–state encounters (Halling & Baekgaard, 2024). Although administrative burden increasingly is being

examined in the context of algorithmic administrative systems, only a few scholars are using the concept of digital administrative burden in their research (Peeters, 2023).

This article is motivated by the question of what kinds of costs people with mental health problems incur due to the digital administrative burdens of self-service and algorithmic decision-making. Extant research has suggested that compliance costs (meeting requirements to access government services or benefits), learning costs (the effort required to understand bureaucratic processes), and psychological costs (e.g., stress or frustration) may be prevalent in digital bureaucratic encounters, as citizens require skills to navigate the digital administrative encounters and often are blindsided as citizens confront information sources and a lack of algorithmic transparency (Larsson, 2021; Madsen et al., 2022). In our novel formulation of the digital administrative burden, research on cost is combined with an examination of social identities formed by the mental health rehabilitees' experiences with the digital administrative burden, potentially influencing how they position themselves as citizens and members of society. Drawing on social representation theory (Moscovici, 1976/2008, 2000) and the concept of social identity (Duveen, 2001; Duveen & Lloyd, 1986; Howarth, 2002), we examine how mental health rehabilitees' understandings of digital administrative burdens encompass discourses about their identity and what kinds of effects digital administrative burdens and their costs may exert on the development of social sustainability in digital citizen–state encounters. The article contributes to the social sustainability literature by introducing the concept of digital administrative burdens and how they can create social exclusion and distrust towards public administration among vulnerable citizens. To accomplish this objective, we analyse (a) mental health rehabilitees' experiences through citizen–state interactions in digital encounters and (b) how administrative burdens and their costs shape mental health rehabilitees' identity from such encounters.

2. Social Sustainability in the Digital Welfare State

Extant research has demonstrated that mental illness is associated with low household income, poverty, and long-term unemployment (e.g., Knifton & Inglis, 2020; Lund et al., 2018). Many individuals with mental illness have less access to modern technologies, such as smartphones and the internet, compared with the general public (e.g., Spanakis et al., 2021). Thus, use of online services requires additional mental and financial effort from them. Furthermore, many people with mental illness possess limited digital skills. The most commonly missing skills were handling passwords, using device settings, and solving online problems (Spanakis et al., 2021). Along with material deprivation (e.g., personal finances and living situation) and lack of digital skills, various aspects of mental health are identified as barriers to engaging with online services. Thus, compared with other population groups, people with mental illness are at greater risk of potential unfair treatment by automated systems, as they need public services more than the average population segment. Extant research conducted with individuals with mental health backgrounds in Finland has found that although high-level digital skills were observed, trust in providers of digital services within the healthcare and social welfare sector remained low, particularly among younger participants (Tetri et al., 2024).

As many vulnerable groups of citizens, such as mental health rehabilitees, are both targets of implementation of automated systems and the largest user group for public benefits, the dynamics of trust primarily concern how mutual trust is built between public authorities and these citizens. The dynamics of trust between public authorities and vulnerable citizens is crucial, particularly in the context of implementing automated systems and providing public benefits.

As citizen–state encounters increasingly take place digitally, and authorities operate through automated systems, digital platforms become important sites for enacting one's citizenship and living out social sustainability. Digital social sustainability has been discussed in the corporate world and at the community level, including in cities (Osburg & Lohrmann, 2017). When discussing sustainability in the digital welfare state, we need definitions in which social sustainability operates on a full societal level. In extant digital transformation research, digital social sustainability refers to "the ways in which digital technology is designed and used to support and promote social equity, fairness and well-being, as well as to address social challenges, such as inequality, poverty and social exclusion" in society (Nosratabadi et al., 2023, p. 11). There has been increasing critical discussion about "the dark side" of algorithmic governance in liberal democracies, casting doubt on whether algorithmic systems can follow the principles of good administration (e.g., Oswald, 2018; Veale & Brass, 2019). Equity, fairness, and well-being in the definition can be thought to refer to principles of good administration, such as equality of treatment, transparency, and responsiveness to citizens' needs (Roehl, 2023). Social sustainability intersects with institutional and social trust, and in this article, we focus in particular on experiences in which the digital administrative burden causes or affirms potentially socially exclusive identities for marginalised citizens as users of algorithmic systems. These systems' design has been known to be motivated often by policymakers' distrust towards marginalised citizens, treating them as potential fraudsters (Parviainen et al., 2025), whereas recurring experiences with digital exclusion among citizens can rupture social cohesion over time, resulting potentially in more basic distrust between citizens and the state, threatening the entire society's social sustainability. When digital citizen–state interactions are placed at the heart of the welfare state's sustainability, algorithmic governance can maintain citizens' trust only if people feel that they are being treated fairly, with dignity and respect. We argue that the welfare state's capability and commitment to these principles are crucial in preventing the emergence of inequality and polarisation between different population groups.

As noted, automated systems are viewed as increasing government efficiency and reducing compliance costs for citizens, but a growing body of literature suggests that citizens in vulnerable situations may incur costs caused by digital administrative burdens (Moynihan et al., 2015; Peeters & Widlak, 2018). Herd and Moynihan (2018) highlighted these burdens' consequential, distributive and constructed nature, arguing that failure to overcome these burdens can lead to citizens' exclusion from rights, benefits, and services to which they are formally entitled. A key insight into administrative burdens is that their consequences are not equally distributed across the population. Similarly, digital burdens are likely to affect certain target groups more than others, leading to distributive effects from government digitalisation. For example, vulnerable groups have been known to be affected disproportionally by burdensome procedures, which often arise because of decisions related to resource allocation and the design of policies, programmes, and administrative procedures (Peeters, 2020).

3. Data, Methods, and Ethical Considerations

The research data were gathered in Finland from a centre that provides training for experts by experience. The role of experts by experience, as former service users, is to support the rehabilitation of substance users and those suffering from mental health problems by drawing on their experiential knowledge. The Ethical Committee of the Tampere Region approved the present study (decision 49/2022), which followed ethical guidelines from the Finnish National Board on Research Integrity (2023).

The research data were generated from audio-recorded group discussions involving mental health rehabilitees and experts by experience. Seven two-hour group discussion sessions were conducted in autumn 2023 and spring 2024. These group discussions served as forums for mental health rehabilitees and experts by experience to come together and share their experiences with digitalised public services when claiming social benefits, searching for a job, and/or dealing with social and health services. Altogether, the group discussions comprised 35 participants, who were either on disability pension or unemployed job seekers, ranging in age from 28 to 55. All the participants had experience with digital citizen–state encounters, having claimed welfare benefits when their income and assets did not cover their essential daily expenses. Most of the participants used digital services provided by the Social Insurance Institution of Finland (Kela), which is responsible for implementing basic social and disability benefits, as well as basic unemployment allowances in Finland. Kela uses automation in routine tasks and when decisions are completely undisputable (Kela, 2025). The Finnish government has designed new legislation to make electronic notifications of public administration the primary option for citizens in administrative communication, but the objective is a gradual transition to digital services as the primary channel for accessing public services in general (Ministry of Finance Finland, 2025).

The topics of the group discussions fell under two main themes: (a) experiences that people with mental illness have in digital citizen–state encounters and (b) how algorithmic decision-making, which is used in benefit eligibility, affect people with mental illness. In the analysis of the group discussions, we used two parallel analytic strategies to examine social representations and identities: thematic content analysis (Vaismoradi et al., 2013) and discourse analysis (Hall, 2001; Potter & Wetherell, 2002). Initially, the group discussion data were classified using thematic content analysis (Vaismoradi et al., 2013). The content-analytic reading method served as a preliminary tool for examining the research material. The analysis began by reading the data set regarding mental health rehabilitees and experts by experience. The text then was condensed and grouped into subcategories based on how they described the administrative burdens that the interlocutors experienced. The subcategories were grouped further into main categories, which represented the views of experts by experience and mental health rehabilitees regarding administrative burdens in digital citizen–state encounters. Table 1 illustrates how the interlocutors' experiences with digital citizen–state encounters create various subject positions and ultimately lead to social exclusion.

Table 1 indicates that learning costs create the subject positions of an unreliable person, an insignificant person, and an inferior person for mental health rehabilitees and that this positioning engenders socially excluded citizenship among these individuals.

After using the content-analytic reading method, we analysed the group discussion material by relying on social representation theory (Moscovici, 1976/2008, 2000) and the concept of social identity (Duveen, 2001; Duveen & Lloyd, 1986; Howarth, 2002). Social representation theory enables us to study the relational and symbolic dimensions of interaction (Campbell & Jovchelovitch, 2000) and, thus, helps in understanding citizenship as an "interactional matter" (R. Barnes et al., 2004) realised in the intersubjective space between policymakers and citizens in digital public service contexts.

Social representations are systems of social knowledge collectively constructed and reconstructed through communicative interactions and social practices (Moscovici, 1976/2008). In digital citizen–state encounters,

Table 1. Analysis of the mental health rehabilitees' experiences with administrative burdens, termed "learning costs," in digital citizen–state encounters.

Learning costs			
Original expression	Condensed expression	Subcategory	Main category
We are just cases on paper, and that's why I feel mistrust towards decision-making, which is happening somewhere online by anonymous and faceless people.	*We are just cases on paper.*	The subject position of an insignificant person The term "insignificant" refers to something or someone that is not important, notable, or influential.	Socially excluded citizen
The people from the job market service who make decisions about my ability to work or other things do not trust me; they want to know about me and ask for information because they think I'm cheating.	*The people from the TE service…don't trust me. They think I'm cheating.*	The subject position of an unreliable person The term "unreliable" refers to something or someone that cannot be depended on for consistent accuracy, truthfulness, or performance.	
I always encounter such pitfalls in the system that are not even there in Kela, so no one can advise me on what should be done next; it's probably because when there is an unemployed job seeker who is unable to work and when there is no direct position for such a person in the system…	*…no one can advise what should be done because I am an unemployed job seeker who is unable to work.*	The subject position of an inferior person The term "inferior person" refers to someone who is perceived to have lesser qualities, abilities, or value compared with others.	

common symbolic resources are shared by mental health rehabilitees and the benefits system to give meaning to rehabilitees' social and material worlds, and to help them navigate in these worlds, social representations inform the group's behaviours (Moscovici, 1984). Identities become meaningful in social interactions and practices through processes of positioning the self in relation to social representations that circulate in our environment, and by appropriating, reworking, and/or contesting these representations (Duveen, 2001). Various identity positions' availability within these networks of meaning is framed and constrained by contextual norms and values (Duveen, 1993).

The relationship between how others represent the groups to which we belong and how we construct ourselves becomes clear in the case of minority and socially excluded groups (e.g., Hodgetts et al., 2007; Howarth, 2002), such as people with mental illness, who are vulnerable to specific kinds of epistemic injustice, such as testimonial injustice, stigmatisation, and discrimination. Furthermore, self-distrust felt by vulnerable people based on their experiences and knowledge as epistemic agents is a complex combination of shame, self-accusation, feelings of oppression, and lack of proper concepts to express their feelings (Auvinen et al., 2021). In the case of digital citizen–state encounters, interactions between mental health

rehabilitees and public administrators in decision-making are asymmetrical in terms of symbolic and material power (e.g., status, access to information), thereby preventing mental health rehabilitees from participating in ways that adequately reflect their own concerns and needs (see also Ansell & Gash, 2008; M. Barnes & Coelho, 2009).

4. Results

We examined digital administrative burdens' impact on citizen–state encounters among mental health rehabilitees by (a) analysing mental health rehabilitees' experiences with these interactions and (b) examining how administrative burdens shape mental health rehabilitees' identity and contribute to social exclusion. We analysed the data in terms of costs caused by digital citizen–state encounters. These costs can manifest in various ways, including compliance costs (meeting requirements to access government services or benefits), learning costs (the effort required to understand bureaucratic processes), and psychological costs (e.g., stress or frustration). However, our analysis did not focus on these costs themselves, but on how they construct mental health rehabilitees' identity as members of society.

Our analysis found that learning costs and psychological costs in particular were prevalent in digital bureaucratic encounters, as mental health rehabilitees require skills to navigate these encounters and often are blindsided by information sources and a lack of algorithmic transparency. Aside from digital proficiency, mental health rehabilitees struggle with navigating complex bureaucratic processes due to opaque information sources and a lack of algorithmic transparency. Although extant studies have found that many people with mental health problems have difficulties using digital tools (cf. Spanakis et al., 2021), the mental rehabilitees who participated in the group discussions did not mention this, except for a few older participants. Our results support Tetri et al.'s (2024) research on mental health backgrounds in Finland, which found that although high digital skills were observed, trust in providers of digital services within the healthcare and social welfare sector remained low, particularly among younger participants. In our research, participants in the group discussions noted that they find it difficult to trust benefit processors because they do not interact with them face-to-face. The participants suspected that the distrust was mutual in digital encounters regarding benefit processing.

Compliance costs, learning costs, and psychological costs produced a marginalised position for vulnerable mental health rehabilitees. Based on the data, it can be demonstrated that all three cost categories create the subject positions of a dispossessed person, unreliable person, insignificant person, and inferior person among mental health rehabilitees. Mental health rehabilitees' positioning demonstrates how digital administrative burdens in citizen–state encounters build social exclusion and produce a socially excluded citizenship. Our analysis indicates that mental health rehabilitees, as social security recipients, are punished for needing to access services. This does not mean that punitive experiences with social security and employment services are new or caused by the adoption of automated technologies in service delivery. Many writers have chronicled social services delivery experiences as being punitive, from the 19th century to the present in America and in South Africa (Headworth, 2021; Super, 2021; Trattner, 2007). Similarly, Carney (2020) pointed out that, in Australia, social security recipients historically have been treated like economic scapegoats. According to Sleep (2024), what is new is the link between this punitive manifestation and the technological promise of efficiency, particularly in welfare compliance.

4.1. Being Dispossessed

Our analysis found that the compliance costs produced by administrative burden are common among the mental health rehabilitees who participated in the group discussions. In their case, economic factors, such as poverty, create exclusion because they cannot afford to acquire and maintain the digital tools needed to apply for financial benefits in the digital welfare state. Compliance costs particularly create the subject position of a dispossessed person among mental health rehabilitees. Being "dispossessed" means being deprived of land, property, or other possessions. It often refers to situations in which individuals or groups have had their homes, possessions, or security taken away from them. In Excerpt 1, a mental health rehabilitee, MHR1, highlights the contradiction that affects vulnerable people in the digital welfare state in particular, namely that one should be able to afford digital tools and maintain an internet connection to apply for financial support, but when no money is available—not even for food, housing or health services—the internet can be out of reach:

> MHR1: "But I always ponder and wonder how it can be made possible for everyone, meaning that to do something in the digital service system, you need to have an internet connection and the equipment to access it. So, I find it unlikely that such discretionary social assistance would pay for these devices for people." (Excerpt 1)

MHR1's turn highlights the necessity of having both an internet connection and the appropriate equipment to access digital services. Digital inclusion is a fundamental issue in ensuring that everyone can benefit from digital advancements. MHR1 expressed scepticism about whether discretionary social assistance would cover these devices' costs. This reflects a concern about current social support systems' adequacy in addressing digital inclusion. Ultimately, their pondering suggests a need for more inclusive policies that ensure everyone, regardless of their economic status, can access digital services. This might involve government or community initiatives to provide necessary equipment and internet access to those in need. Overall, MHR1's statement underscores the challenges of digital equity and the importance of addressing these barriers to creating a socially sustainable digital society.

Excerpt 2 highlights how the poor financial situation often experienced by mental health rehabilitees also can contribute to vulnerability, as individuals may rely heavily on public benefits and need to trust that they will receive the necessary support. MHR2 emphasised the importance of addressing both mental health and practical barriers to improve daily functioning and access to necessary resources, as well as their determination and resourcefulness in overcoming these obstacles:

> MHR2: "When the ability to function varies greatly between days, sometimes the technology itself causes difficulties. I didn't have an internet connection, I didn't have a laptop, I didn't have money, so I had to get myself five kilometres away to the local library and then from time to time to take care of these things." (Excerpt 2)

MHR2's statement highlights several challenges that stem from both mental health issues and practical barriers in digital citizen–state encounters. First, the interlocutor's ability to function varies greatly from day to day due to mental health issues, which can be unpredictable and challenging. Second, the lack of an internet connection and a laptop creates significant obstacles to managing benefit encounters, which require digital access. Third, limited financial resources complicate the situation further, making it difficult to afford necessary

technology. Despite these challenges, the interlocutor demonstrates resilience by traveling five kilometres to the local library to access resources and manage tasks.

Excerpt 3 reflects a sense of exclusion and difficulty in adapting to digital advancements. The interlocutor acknowledged the risk of being left behind in the digital era, stemming from a lack of early exposure to technology. They expressed a feeling of not having a natural inclination towards or interest in information technology, which makes it challenging:

> MHR3: "And somehow, I recognise it. I have been at a big risk of being excluded from this digital development. Yeah. When we were young, we didn't have any computers at home, nothing…but I don't have it naturally either. I feel like I don't have the same interest in anything related to information technology, that it is in principle it is always a bit challenging." (Excerpt 3)

MHR3 feels they are at risk of being excluded from digital development, highlighting a gap between their experiences and the rapid pace of technological advancement. Growing up without computers at home has contributed to their current challenges with technology. The interlocutor does not have a natural interest in information technology, which adds to the difficulty faced in engaging with it. Regarding the digital welfare society's social sustainability, MHR3 seems to suffer from digital exclusion, passed on from one generation to the next, in which difficulties in using digital devices manifest as a transgenerational problem. Thomas et al. (2020) found that it is a key concern within the sector that increasing reliance on digital services excludes those with limited access to digital devices, data or reliable internet coverage. Our analysis also found that digital exclusion is particularly prevalent among the most vulnerable service users, such as mental health rehabilitees.

4.2. Being Unreliable

Research data from the group discussions indicated that mental health rehabilitees' experiences with digital citizen–state encounters are determined by perceptions that they are unreliable individuals. The term "unreliable" refers to something or someone that cannot be depended on for consistent accuracy, truthfulness or performance. For example, an unreliable source might provide incorrect information, or an unreliable person might fail to fulfil promises. Digital public services often are designed to contain algorithmic tools for tracing potential fraudsters (Parviainen et al., 2025), so interlocutors obviously sense the basic distrust towards citizens targeting them behind the system. The following excerpts reveal a phenomenon that we defined as digital administrative burden in the Introduction. Due to the transition to self-service and automation in decision-making, citizens increasingly are required to act as advocates for their own affairs but cannot be sure how officials will respond to their arguments or pleas about their own situations. This uncertainty causes stress and anxiety for many.

In Excerpt 4, the psychological costs produced by the digital administrative burden are well-highlighted and arise from how MHR4 must send documents about his income and expenses electronically. MHR4 described themself as a person without privacy because they are not trusted. They must prove their eligibility every month because they are viewed as an unreliable person. This incurs a psychological cost for MHR4, which they expressed as sadness. Furthermore, the excerpt portrays MHR4's frustrations with the lack of privacy, limitations, and conditions on their access to social assistance and the repetitive process they must endure

each month. It conveys a sense of disappointment and stress that accompanies the application process and the outcomes' uncertainty:

> MHR4: "I have no privacy when I apply for basic social assistance. Everything must be delivered, including bank account statements. Although it is promised there that you could get social assistance for six months, but when it is affected by certain other benefits, it saddens me that I will never get it for that long. I must make a new application every month. And always the stress of getting through it and what comes out of it…" (Excerpt 4)

The interlocutor expressed their frustration with the lack of privacy they have experienced when applying for basic social assistance. The expression stated that they are required to share personal information, including bank account statements, as part of the application process. This suggests a lack of confidentiality and boundaries between the individual and the aid agency. The interlocutor's disappointment stems from being promised social assistance for six months, but this promise is affected by certain other benefits. This implies that their access to social assistance is contingent on meeting specific criteria or having a limited eligibility period. The interlocutor expressed sadness at the realisation that they may not receive this assistance for an extended period of time due to these factors. The burden of the process is highlighted when the interlocutor mentions the need to file a new application every month. This suggests a repetitive and time-consuming task that adds stress to the interlocutor's life. The stressfulness of the system is highlighted by the fact that Kela has access, for example, to national income register. Moreover, the system could probably be much less repetitive for the benefit claimants than it is. The interlocutor's anxiety is emphasised further when they mentioned the uncertainty of the application process, indicating a lack of transparency and control over their situation and outcomes.

Excerpt 5 illustrates the learning costs that mental health rehabilitees face in digital citizen–state encounters. The excerpt ties into the phenomenon described by Bovens and Zouridis (2002) concerning the transition from "street-level bureaucracies," in which citizens engage directly with government officials in physical offices, to "screen-level bureaucracies," in which interactions occur in person, over the phone or through online chat with bureaucrats who are working behind screens. This rehabilitee, MHR5, described how they often need assistance on what information to disclose or withhold to survive within the system. To receive their entitled welfare benefits, they must learn to act correctly from the system's perspective:

> MHR5: "The fact that you have somehow had to survive this way and ask for help so much, and then always thought about what it is worth to tell someone about yourself to get that help. When you have learned the system that if you tell someone about something, it prevents someone from getting a certain social benefit, that you must be alert." (Excerpt 5)

The experiences shared by MHR5 in Excerpt 5 prove that digital welfare services are less accessible to some citizens, leading to new discriminatory effects on low-income and underserved populations. The interlocutor reflected on mental health rehabilitees' experiences regarding their reliance on assistance and the constant evaluation of the value of sharing personal information electronically to receive this help. The interlocutor stated that they have had to survive in a particular manner and frequently asked for help, implying that they have faced challenging and difficult circumstances. The interlocutor's thought process revolves around weighing the costs and benefits of disclosing personal information in the pursuit of receiving assistance.

The interlocutor was cautious about sharing personal details, having learned that disclosing certain information may prevent others from receiving social benefits. This suggests that the interlocutor is aware of a competitive element within the system, in which resources or benefits are limited and potentially could be capped through disclosure of personal information. The interlocutor's statement that "you must be alert" indicates a sense of vigilance or caution that has developed within her due to these experiences. This highlights the interlocutor's perception that navigating the system requires constant awareness of how disclosing personal information can impact both her own access to benefits and opportunities available to others.

The commitment of another rehabilitee, MHR6, in Excerpt 6 continues to express the interlocutors' mistrust in the decision-making process that occurs online, conducted by anonymous and faceless people. The excerpt highlights the interlocutor's sense of being dehumanised and doubted by the employment service. They feel that their individual circumstances and integrity are not being considered, and that their trustworthiness is being questioned without proper cause. This contributes to their overall lack of trust in the decision-making process and a perceived need to provide documentation continually to validate his legitimacy:

> MHR6: "So, maybe it's because we are just cases on paper, and because of that, I feel mistrust in this kind of decision-making, which takes place somewhere online by anonymous and faceless people. And I'm thinking that, in the same way, these people from the employment services who make decisions about my ability to work or other things don't trust me. They want to know about me and are fishing for information, thinking that I'm cheating. That's why I must deliver all kinds of files to them." (Excerpt 6)

In this excerpt, the interlocutor begins by acknowledging that people who have to apply for benefits feel like mere "cases on paper," suggesting that they are reduced to impersonal, bureaucratic documentation, rather than being viewed as individuals with unique circumstances. This depersonalisation contributes to their overall mistrust of the decision-making process conducted by anonymous and faceless individuals online. For the interlocutor, screen-level bureaucracy creates a system that is onerous and more burdensome than street-level bureaucracy. The interlocutor believes that these decision-makers from employment services do not trust them, as evidenced by their persistent "fishing for information." The interlocutor suspects that this lack of trust stems from an assumption that they are engaged in some form of dishonesty or cheating. The mention of having to deliver "all kinds of files" implies that the interlocutor is required to provide various forms of documentation or evidence to prove his compliance or eligibility for certain benefits or assistance. This adds further to the interlocutor's frustration and suggests an excessive burden of proof being placed on them.

Tetri et al. (2024) found that digital proficiency alone did not bring trust in digital welfare systems with it, particularly among younger generations of marginalised people, who did not trust service providers. The digital welfare state is not operating in a vacuum, and mental health rehabilitees have faced several reductions in welfare services and benefits in recent years (Ahonen, 2022). More basic distrust towards the state may stem from previous negative experiences in interpersonal relationships, or from a general sense of social disconnect caused by former reductions, undermining trust in public welfare systems in general.

4.3. Being Insignificant

The group discussions with mental health rehabilitees paint a picture of individuals feeling insignificant. The subject position of an insignificant person created by the digital administrative burden becomes

particularly evident when interlocutors express their frustration with digital benefit transactions. The term "insignificant" refers to something or someone that is not important, notable or influential, suggesting a lack of significance, meaning, or impact. When applied to a person, it may indicate a feeling of being unimportant or unnoticed. Excerpt 7 highlights the interlocutor's dissatisfaction with their experience applying for disability and national pensions, particularly due to the reinitiation of the matter by Kela, the Finnish pensions and social security organisation, without their involvement. The excerpt appears to examine the interplay between mental health, self-awareness, and the complexities of modern communication, particularly in sharing personal information online:

> MHR7: "I applied for a disability pension from Ilmarinen after being ill for 300 days, and at the same time, a national pension application was pending with Kela. Ilmarinen rejected my application, and so did Kela, but without asking me, Kela started this matter again." (Excerpt 7)

In this excerpt, the interlocutor discusses their experience applying for a disability pension and a national pension. They mentioned that after being ill for 300 days, they applied for a disability pension from Ilmarinen, a presumably relevant organisation. However, it is not explicitly stated *whether* this application was accepted or rejected. The interlocutor also mentioned that they have a national pension application pending with Kela, another private organisation that deals with pensions and social security in Finland. It remains unclear from the excerpt whether this application was initially accepted or rejected. The interlocutor expressed frustration because despite both Ilmarinen and Kela rejecting her application(s), Kela reopened the matter without consulting or informing them. This could imply a lack of communication or transparency in the process, as the interlocutor seems surprised by Kela's actions.

Based on Excerpt 8, it appears that citizens are experiencing difficulties within the digital benefit system. The excerpt reflects the psychological costs caused by the digital administrative burden. The interlocutor seems frustrated because they cannot find guidance or advice on what to do next. This situation possibly stems from a situation in which an unemployed job seeker is unable to work, and the system does not have a specific position to accommodate them:

> MHR8: "Yes, you see, these kinds of unreasonably burdensome situations, I would say, they happen constantly. I don't know, that might be the main reason why I always encounter such pitfalls in the system, where even at Kela [the Social Insurance Institution of Finland], no one knows what to advise next. It probably stems from being an unemployed job seeker who is also unfit for work, and since there's no direct category for that in the system….So, you have to adapt, now it's an unemployed person, but they're unfit for work. And what on earth, they've done volunteer work? Weren't they unfit for work? Actually, even now, is this the fourth or fifth time I've applied for a permanent disability pension, and yesterday was the deadline at Kela, the 11 weeks it has to be processed. I had other matters with Kela yesterday. I clicked onto their pages, and they keep you in suspense until the last moment. There's the national pension issue, and it's just, poof, disappeared, and there's nothing about it anywhere." (Excerpt 8)

The digital benefit system poses a few potential implications, as discussed in Excerpt 8. First, there is a lack of support, i.e., the system in question may not provide appropriate support or solutions for individuals in unique situations, such as unemployed job seekers who are unable to work. This situation could leave people feeling

lost and without guidance. Second, the limited options and the absence of a direct position within the system suggest potential limitations or shortcomings in available opportunities. It implies that the system does not account for people with specific circumstances or challenges. Third, frustration and confusion are evident in the interlocutor's use of terms such as "pitfalls" and their expression regarding a lack of advice, highlighting their frustration with the situation. They feel that they are encountering obstacles or difficulties within the system that they cannot navigate on their own. Their bewilderment was strengthened by finding out that their case had disappeared from their list of open cases in OmaKela service (e-service for social welfare services in Finland). The subject position of an insignificant person produced by the digital administrative burden proves, in part, that the promise of enhancing access and efficiency in service delivery of the digital welfare state does not apply to vulnerable people (van Toorn et al., 2024).

4.4. Being Inferior

The interlocutors often pointed out that they view themselves as inferior in digital interactions. The term "inferior person" typically refers to someone perceived as having lesser qualities, abilities, or value compared with others. This perception can be based on various criteria, including social status, intelligence, skills, moral values, or even personal traits. However, it should be noted that labelling an individual as inferior can be highly subjective and often reflects the biases or prejudices of the person making the judgement. Using such labels can foster negative attitudes and discrimination. It is essential to appreciate each person's individuality and unique contributions, recognising that everyone possesses inherent worth regardless of their perceived status or abilities. Compassion, empathy, and understanding are crucial in promoting a more inclusive and respectful view of others.

Excerpt 9 provides insight into the experiences of individuals receiving social assistance, highlighting a complex interplay between financial realities, societal perceptions and personal agency. The excerpt encapsulates the struggles faced by these individuals as they navigate their daily lives under the weight of societal judgement and stigma. It also emphasises the emotional and psychological burden that comes with financial oversight and the ways in which individuals reconcile their needs with others' perceptions. This dynamic invites a broader reflection on how society views and treats those reliant on others:

> MHR9: "As a social assistance client, I've done that sometimes, and I'll probably do it in the future because I have to deliver my account statements to Kela. So, I've withdrawn cash from the ATM when I go to a restaurant to eat a hamburger. Because I've had such a strong feeling like that, I don't want it, it shows on my account statement that I, as a social assistance client, go to Hesburger to eat a hamburger." (Excerpt 9)

Excerpt 9 highlights the learning costs of the digital administrative burden. MHR9 is concerned about how their spending habits, specifically dining out, may reflect poorly on their status as a social assistance client. This indicates a broader societal stigma that equates social assistance with poverty and restricts individuals' freedom to enjoy basic pleasures without judgement. The mention of the requirement to submit account statements to Kela highlights the oversight faced by social assistance clients. This monitoring system can create pressure to justify personal expenditures, leading to a feeling of being constantly scrutinised. The phrase "such a strong feeling like that" suggests an internal conflict. The interlocutor wants to enjoy experiences such as eating out but feels constrained by the social implications of his financial status. This reflects a sense of

guilt or shame associated with relying on social assistance. The act of withdrawing cash for a meal, despite possible judgement, signifies an assertion of personal agency. The interlocutor acknowledges that they might continue these behaviours, indicating a desire to assert normalcy in their life despite the label of being a social assistance client.

Excerpt 9 clearly illustrates how the design has moved towards increased self-service, requiring citizens to provide all necessary information, verify calculations, and present evidence to contest algorithmic decisions. This has shifted the burden of proof, which, in cases of alleged debts, has been described as a form of extortion (Carney, 2019). The interlocutor's digital literacy demonstrates how by switching to cash, they want to prevent the authorities from seeing and collecting information about their daily private life. According to Griffiths (2024), automated systems that impose greater compliance and administrative demands on benefit claimants can undermine the traditional democratic notion that the state is accountable to its citizens, and not vice versa.

Excerpt 10 offers poignant insight into the interlocutor's internal struggles with self-esteem and feelings of inferiority, particularly concerning her understanding of digital technology. The excerpt encapsulates the complexities of self-perception within the context of modern technology and personal struggles. It demonstrates how unfamiliarity with digital interactions can exacerbate feelings of inadequacy and hinder self-esteem. This reflection highlights the emotional and psychological barriers that many individuals encounter when navigating a rapidly evolving digital landscape:

> MHR10: "Then I always feel a little inferior because of not knowing how to do things or having to ask for help, and I feel stupid and like that, when I've had to struggle with self-esteem anyway and feelings of inferiority, weakness…so it somehow amplifies the fact that you don't know or don't understand digital interaction, and you can't familiarise yourself with it, so…" (Excerpt 10)

The interlocutor explicitly states that they "feel a little inferior," suggesting a pervasive sense of inadequacy. This reveals how their lack of knowledge about digital interactions affects their self-perception, exacerbating existing insecurities. The act of asking for help is portrayed as a source of embarrassment, as the interlocutor feels "stupid" for not knowing how to navigate digital spaces. This highlights a fear of judgement or criticism from others, which often serves as a barrier to seeking assistance. The mention of struggling with self-esteem reflects deeper psychological issues. The interplay between feeling weak and not understanding technology contributes to a cycle of negative self-assessment, in which one issue compounds the other.

The struggle to familiarise oneself with digital interactions may represent broader societal shifts towards technology in which those lacking the necessary skills can feel increasingly left behind. The interlocutor's acknowledgment of this disconnect suggests a feeling of isolation in a tech-driven world. The phrase "somehow amplifies the fact" indicates that the interlocutor's existing struggles with digital literacy, combined with an already-complicated web of emotional challenges, can magnify feelings of inadequacy. The excerpt's tone conveys a mix of frustration and resignation. There is a sense of defeat in acknowledging one's limitations while wishing for understanding and improvement. The repetition of negative self-descriptors such as "inferiority" and "weakness" intensifies the emotional weight, revealing these feelings' profound impact on the interviewee's daily life.

5. Discussion and Conclusions

Digital citizen–state encounters have become important sites for enacting citizenship and living out social sustainability, particularly for those who rely more on public services or benefits. This article has addressed an underexamined aspect of digital administrative burdens and their costs in citizen–state encounters, particularly regarding their implications for people with mental illness. The mental health rehabilitees struggled with problems of digital inclusion, problems with transparency in digital and bureaucratic processes, and lack of support. Mental health rehabilitees' digital proficiency varied. We investigated how people with mental illness experience digital administrative burdens and how these experiences produce socially exclusive identities. By analysing excerpts from discussions, we demonstrated how digital citizen–state encounters, particularly when claiming welfare benefits, create subject positions of feeling dispossessed, unreliable, insignificant, and inferior among mental health rehabilitees—qualities that shape how rehabilitees define themselves as citizens. This illustrates how digital administrative burdens in digital citizen–state encounters create or affirm identities, resulting in social exclusion.

The rise of automated systems and self-transactions in digitalised public services has included the promise of providing services more efficiently, with more transparent, value-neutral, impartial, and fair decision-making (Bovens & Zouridis, 2002; Casey, 2022; Wirtz et al., 2019). However, our research results do not support this promise from the perspective of vulnerable people. In a socially sustainable digital welfare state, algorithmic systems promote social equity, fairness, and well-being, and are designed to address social challenges in particular, such as inequality, poverty, and social exclusion. Our research results indicate that digital public services are not socially sustainable for mental health rehabilitees even though public services are associated with the social consensus that certain basic services should be available to all, regardless of income, physical ability, or mental capacity. For example, poverty creates exclusion because vulnerable people rarely can afford to acquire and maintain the digital tools needed to apply for financial benefits in the digital welfare state.

Mental health rehabilitees cited lack of face-to-face interaction as a problem with digital citizen–state encounters, as it made the digital welfare system appear onerous. Mental rehabilitees have expectations—such as confidentiality, reciprocity, and predictability—from citizen–state encounters. These expectations, which they carry over from face-to-face to digital citizen–state encounters, are often unmet by digital platform procedures. Extant research on successful face-to-face encounters between mental health rehabilitees and authorities in different Scandinavian countries indicates that positive experiences correlate with feelings of confidence and security in interactions between mental rehabilitees and professionals. These experiences are also linked to the continuity of the process and the ability to foresee future steps, the ability to build trust in the relationship, professionals' responsiveness to individual needs, and treatment of rehabilitees as unique individuals, rather than mere users of mental health services (Nouf & Ineland, 2023). The participants in the group discussions repeatedly highlighted the emotional impact of digital administrative burdens.

Mental health rehabilitees have experienced frustration and disappointment due to a lack of trust and faith in their integrity. They feel that their privacy is being invaded when they are constantly monitored or evaluated. They also feel disappointment and stress due to a lack of confidentiality and boundaries between the individual and the aid agency. They found claiming welfare benefits to be a repetitive and

time-consuming task that adds stress to their lives. They also experienced uncertainty about the outcome of the application process, indicating a lack of control over their situation and a sense of vigilance or caution. These experiences produce a sense of being dehumanised and doubted by society, contributing to an overall lack of trust in the decision-making process and society, which should care for people in vulnerable positions.

Digital welfare state information systems might be overly complicated or not user-friendly, causing frustration for individuals who already struggle with cognitive functioning or focus. They may be automated systems that do not provide adequate human interaction, leaving individuals feeling isolated, particularly if they need personalised assistance for their financial or mental needs. There is evidence that a lack of support when using digital services can trigger existing social distrust and revive memories of negative experiences that were not originally caused by digital services (Andersen et al., 2020). The mental health rehabilitees who participated in the group discussions often cited heightened concerns about data privacy, fearing that sensitive information could be misused or lead to stigmatisation. Generic or robotic responses to enquiries are distressing for them, leading to further disengagement from essential services.

The pressures of poverty exacerbate mental health conditions, making it even harder for mental health rehabilitees to use digital tools. Improving digital citizen–state encounters for people with mental illness involves addressing these challenges through better access, tailored communication, and supportive infrastructures. Most participants in the group discussions viewed themselves as somehow irregular and exceptional cases, i.e., automated systems did not provide enough human discretion and professional help to meet their needs. As demonstrated, digital services have difficulties reaching out to mental rehabilitees who are in the worst and most complicated situations (Virtanen et al., 2021). Our results suggest that digital public services and automated systems are unable to correct existing problems of exclusion in society or the underlying causes of distrust between citizens and the state. Instead, they tend to reproduce these issues in digital form or, at worst, exacerbate the problems further.

Acknowledgments

We would like to give special thanks to the mental health rehabilitees who, by participating in the group discussions, opened up a perspective for us on the position of vulnerable people as users of digital public services. We would also like to thank the journal's anonymous reviewers for their valuable and insightful comments.

Funding

This research was funded by the Kone Foundation, grant number 202102433. Publication of this article in open access was made possible through the institutional membership agreement between Tampere University and Cogitatio Press.

Conflict of Interests

The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Data Availability

The research data used in this study is not available to other researchers due to ethics and confidentiality issues.

References

Ahonen, K. (2022). Miten potilaan toimijuus rakentuu historiallisesti suomalaisessa mielenterveyspolitiikassa? In M.-L. Honkasalo, L. Jylhäkangas, & A. Leppo (Eds.), *Haavoittuva toimijuus. Sairastaminen ja hoiva hyvinvointivaltion laitamilla*. Vastapaino.

Andersen, L. M. B., Reavley, N. J., Bøggild, H., & Overgaard, C. (2020). The role of social technologies in community care—A realist evaluation of a Danish web-based citizen-to-citizen platform adopted in community care to promote belonging and mental health. *Heath and Social Care in Community*, *30*(2), 435–444. https://doi.org/10.1111/hsc.13222

Ansell, C., & Gash, A. (2008). Collaborative governance in theory and practice. *Journal of Public Administration Research and Theory*, *18*(4), 543–571. https://doi.org/10. 1093/jopart/mum032

Auvinen, P., Parviainen, J., Lahikainen, L., & Palukka, H. (2021). Discussion protocol for alleviating epistemic injustice: The case of community rehabilitation interaction and female substance abusers. *Social Sciences*, *10*(2), Article 45. https://doi.org/10.3390/socsci10020045

Barnes, M., & Coelho, V. S. (2009). Social participation in health in Brazil and England: Inclusion, representation and authority. *Health Expectations*, *12*(3), 226–236. https://doi.org/10.1111/j.1369-7625.2009.00563.x

Barnes, R., Auburn, T., & Lea, S. (2004). Citizenship in practice. *British Journal of Social Psychology*, *43*(2), 187–206. https://doi.org/10.1348/0144666041501705

Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, *6*(2), 174–184. https://www.jstor.org/stable/3109901

Burden, B. C., Canon, D. T., Mayer, K. R., & Moynihan, D. P. (2012). The effect of administrative burden on bureaucratic perception of policies: Evidence from election administration. *Public Administration Review*, *72*(5), 741–751. https://doi.org/10.1111/j.1540-6210.2012.02600.x

Campbell, C., & Jovchelovitch, S. (2000). Health, community and development: Towards a social psychology of participation. *Journal of Community & Applied Social Psychology*, *10*(4), 255–270. https://doi.org/10.1002/1099-1298(200007/08)10:4%3C255::AID-CASP582%3E3.0.CO;2-M

Carney, T. (2019). Robo-debt illegality: The seven veils of failed guarantees of the rule of law? *Alternative Law Journal*, *44*(1), 4–10. https://doi.org/10.1177/1037969X18815913

Carney, T. (2020). Automation in social security: Implications for merits review? *Australian Journal of Social Issues*, *55*(3), 260–274. https://doi.org/10.1002/ajs4.95

Casey, S. J. (2022). Towards digital dole parole: A review of digital self-service initiatives in Australian employment services. *Australian Journal of Social Issues*, *57*(2), 111–124. https://doi.org/10.1002/ajs4.156

Duveen, G. (1993). The development of social representations of gender. *Papers on Social Representations*, *2*(3), 1–7.

Duveen, G. (2001). Representations, identities, resistance. In K. Deaux & G. Philogène (Eds.), *Representations of the social: Bridging theoretical tradition* (pp. 257–270). Blackwell Publishing.

Duveen, G., & Lloyd, B. (1986). The significance of social identities. *British Journal of Social Psychology*, *25*(3), 219–230. https://doi.org/10.1111/j.2044-8309.1986.tb00728.x

Finnish National Board on Research Integrity. (2023). *The Finnish code of conduct for research integrity and procedures for handling alleged violations of research integrity in Finland 2023*. https://tenk.fi/en/research-integrity-ri

Griffiths, R. (2024). Universal credit and automated decision making: A case of the digital tail wagging the policy dog? *Social Policy and Society*, *23*(1), 1–18. https://doi.org/10.1017/S1474746421000749

Hall, S. (2001). Foucault: Power, knowledge and discourse. In M. Wetherell, S. Taylor, & S. J. Yates (Eds.), *Discourse theory and practice. A reader* (pp. 72–81). Sage.

Halling, A., & Baekgaard, M. (2024). Administrative burden in citizen–state interactions: A systematic literature review. *Journal of Public Administration Research and Theory*, *34*(2), 180–195. https://doi.org/10.1093/jopart/muad023

Headworth, S. (2021). *Policing welfare: Punitive adversarialism in public assistance*. University of Chicago Press.

Henman, P. (2019). Of algorithms, apps and advice: Digital social policy and service delivery. *Journal of Asian Public Policy*, *12*(1), 71–89. https://doi.org/10.1080/17516234.2018.1495885

Hodgetts, D., Radley, A., Chamberlain, K., & Hodgetts, A. (2007). Health inequalities and homelessness: Considering material, spatial and relational dimensions. *Journal of Health Psychology*, *12*(5), 709–725. https://doi.org/10.1177/1359105307080593

Howarth, C. (2002). Identity in whose eyes? The role of representations in identity construction. *Journal for the Theory of Social Behaviour*, *32*(2), 145–162 https://doi.org/10.1111/1468-5914.00181

Kela. (2025). *Automated decisions at Kela*. https://www.kela.fi/automated-decisions

Knifton, L., & Inglis, G. (2020). Poverty and mental health: Policy, practice and research implications. *BJPsych Bull*, *44*(5), 193–196. https://doi.org/10.1192/bjb.2020.78

Larsson, K. K. (2021). Digitization or equality: When government automation covers some, but not all citizens. *Government Information Quarterly*, *38*(1), Article 101547.

Le Ludec, C., Cornet, M., & Casilli, A. A. (2023). The problem with annotation. Human labour and outsourcing between France and Madagascar. *Big Data & Society*, *10*(2). https://doi.org/10.1177/20539517231188723

Lund, C., Brooke-Sumner, C., Baingana, F., Baron, E. C., Breuer, E., Chandra, P., Haushofer, J., Herrman, H., Jordans, M., Kieling, C., Medina-Mora, M. E., Morgan, E., Omigbodun, O., Tol, W., Patel, V., & Saxena, S. (2018). Social determinants of mental disorders and the Sustainable Development Goals: A systematic review of reviews. *Lancet Psychiatry*, *5*(4), 357–369. https://doi.org/10.1016/S2215-0366(18)30060-9

Madsen, C. Ø., Lindgren, I., & Melin, U. (2022). The accidental caseworker: How digital self-service influences citizens' administrative burden. *Government Information Quarterly*, *39*(1), Article 101653. https://doi.org/10.1016/j.giq.2021.101653

Ministry of Finance Finland. (2025). *Priority of digital communications by public authorities*. https://vm.fi/en/priority-of-digital-communications-by-public-authorities

Moscovici, S. (1984). The phenomenon of social representations. In R. Farr & S. Moscovici (Eds.), *Social representations* (pp. 3–69). Cambrigde University Press.

Moscovici, S. (2000). *Social representations: Explorations in social psychology*. Polity.

Moscovici, S. (2008). *Psychoanalysis: Its image and its public*. Polity. (Original work published in 1976)

Moynihan, D. (2018). A great schism approaching? Towards a micro and macro public administration. *Journal of Behavioral Public Administration*, *1*(1). https://doi.org/10.30636/jbpa.11.15

Moynihan, D., Herd, P., & Harvey, H. (2015). Administrative burden: Learning, psychological, and compliance costs in citizen–state interactions. *Journal of Public Administration Research and Theory*, *25*(1), 43–69. https://doi.org/10.1093/jopart/muu009

Nosratabadi, S., Atosbishi, T., & Hegedűs, S. (2023). Social sustainability of digital transformation: Empirical evidence from EU-27 countries. *Administrative Science*, *13*(5), Article 126. https://doi.org/10.3390/admsci13050126

Nouf, F., & Ineland, J. (2023). Epistemic citizenship under structural siege: A meta-analysis drawing on 544 voices of service user experience in Nordic mental health services. *Frontiers of Psychiatry*, *14*, Article 1156835. https://doi.org/10.3389/fpsyt.2023.1156835

Osburg, T., & Lohrmann, C. (2017). *Sustainability in a digital world: New opportunities through new technologies*. Springer. https://doi.org./10.1007/978-3-319-54603-2

Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: Framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A*, *376*(2128). https://doi.org/10.1098/rsta.2017.0359

Parviainen, J., Koski, A., Eilola, L., Palukka, H., Alanen, P., & Lindholm, C. (2025). Building and eroding the citizen–state relationship in the era of algorithmic decision-making: Towards a new conceptual model of institutional trust. *Social Sciences*, *14*(3), Article 178. https://doi.org/10.3390/socsci14030178

Peeters, R. (2020). The political economy of administrative burdens: A theoretical framework for analyzing the organizational origins of administrative burden. *Administration & Society*, *52*(4), 566–592. https://doi.org/10.1177/009539971985436

Peeters, R. (2023). Digital administrative burdens: An agenda for analysing the citizen experience of digital bureaucratic encounters. *Perspectives on Public Management and Governance*, *6*(1), 7–13. https://doi.org/10.1093/ppmgov/gvac024

Peeters, R., & Widlak, A. (2018). The digital cage: Administrative exclusion through information architecture—The case of the Dutch civil registry's master data management system. *Government Information Quarterly*, *35*(2), 175–183. https://doi.org/10.1016/j.giq.2018.02.003

Pinchevski, A. (2023). Social media's canaries: Content moderators between digital labor and mediated trauma. *Media, Culture & Society*, *45*(1), 212–221. https://doi.org/10.1177/01634437221122226

Potter, J., & Wetherell, M. (2002). Analyzing discourse. In A. Bryman & R. G. Burgess (Eds.), *Analyzing qualitative data* (pp. 47–66). Routledge.

Roehl, U. B. U. (2023). Automated decision-making and good administration: Views from inside the government machinery. *Government Information Quarterly*, *40*(4). https://doi.org/10.1016/j.giq.2023.101864

Schou, J., & Hjelholt, M. (2018). *Digitalization and public sector transformations*. Palgrave Macmillan.

Sleep, L. (2024). 'This is NOT human services': Counter-mapping automated decision-making in social services in Australia. *Journal of Sociology*, *60*(3), 618–642. https://doi.org/10.1177/14407833241266022

Spanakis, P., Peckham, E., Mathers, A., Shiers, D., & Gilbody, S. (2021). The digital divide: Amplifying health inequalities for people with severe mental illness in the time of COVID-19. *The British Journal of Psychiatry*, *219*(4), 529–531. https://doi.org/10.1192/bjp.2021.56

Super, G. (2021). Punitive welfare on the margins of the state: Narratives of punishment and (in) justice in Masiphumelele. *Social & Legal Studies*, *30*(3), 426–447. https://doi.org/10.1177/0964663920924764

Tetri, B., Rantanen, T., & Kouvonen, A. (2024). Digital skills and intention to use digital health care and social welfare services among socially marginalized individuals in Finland: A cross-sectional study. *Finnish Journal of EHealth and EWelfare*, *16*(2), 117–130. https://doi.org/10.23996/fjhw.143006

Thomas, J., Barraket, J., Wilson, C. K., Holcombe-James, I., Kennedy, J., Rennie, E., & MacDonald, T. (2020). *Measuring Australia's digital divide: The Australian digital inclusion index 2020*. https://doi.org/10.1093/ppmgov/gvac024

Trattner, W. I. (2007). *From poor law to welfare state: A history of social welfare in America*. Simon and Schuster.

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, *15*(3), 398–405. https://doi.org/10.1111/nhs.12048

van Toorn G., Henman P., & Soldatić, K. (2024). Introduction to the digital welfare state: Contestations, considerations and entanglements. *Journal of Sociology*, *60*, 507–522. http://doi.org/10.1177/14407833241260890

Veale, M., & Brass, I. (2019). Administration by algorithm? Public management meets public sector machine learning. In K. Yeung & M. Lodge (Eds.), *Algorithmic regulation* (pp. 1–30). Oxford University Press. https://ssrn.com/abstract=3375391

Virtanen, L., Kaihlanen, A-M., Isola, A-M., Laukka, E., & Heponiemi, T. (2021). Mielenterveyskuntoutujien kokemuksia etäpalveluiden hyödyistä COVID-19-aikataudella: Laadullinen kuvaileva tutkimus. *Sosiaalilääketieteellinen aikakauslehti*, *58*, 266–283. https://doi.org/10.23990/sa.107405

Widlak, A., & Peeters, R. (2020). Administrative errors and the burden of correction and consequence: How information technology exacerbates the consequences of bureaucratic mistakes for citizens. *International Journal of Electronic Governance*, *12*(1), 40–56. https://doi.ord/10.1504/IJEG.2020.106998

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration*, *42*(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103

About the Authors

Hannele Palukka (PhD) is a visiting fellow at Tampere University, Finland. Palukka specializes in situational interaction studies of high-reliability organizations, such as those found in aviation and healthcare. At present, she is focused on studying the interaction between the social and healthcare service system and its clients.

Anne Koski (PhD) is an interdisciplinary-oriented political scientist and visiting fellow at Tampere University. She is focusing especially on the repercussions of algorithmic governance and public-private partnerships in designing public information systems for citizen–state relations and democracy.

Jaana Parviainen (PhD) is a senior research fellow at Tampere University. She specializes in the philosophy of technology, social epistemology, and phenomenology. Her research encompasses the broad field of how technology and social sciences intersect and extend into philosophical questions about automation, robotization, automated decision-making, algorithms, disinformation, and epistemic inequalities.

Laura Eilola (PhD) is a university instructor at Tampere University. She is focusing on the use of digital public services by two vulnerable groups: mental health rehabilitees and migrants. Her main research area is multimodal interaction in both institutional and everyday life settings.

Social Inclusion
2025 • Volume 13 • Article 10289
https://doi.org/10.17645/si.10289

cogitatio

ARTICLE

Open Access Journal

Algorithmic Decision-Making and Harmonization in Multi-Level Governance Welfare Practices: Empirical Evidence From Belgium

Janne Petroons [1,2], Périne Brotcorne [3], Martin Wagener [3], Koen Hermans [1,4], and Wim Van Lancker [1]

[1] ReSPOND, KU Leuven, Belgium
[2] Centre for IT & IP Law (CiTiP), KU Leuven, Belgium
[3] Centre Interdisciplinaire de Recherche Travail, État et Société, UCLouvain, Belgium
[4] LUCAS—Centre for Care Research & Consultancy, KU Leuven, Belgium

Correspondence: Janne Petroons (janne.petroons@kuleuven.be)

Submitted: 10 March 2025 **Accepted:** 9 July 2025 **Published:** 2 September 2025

Issue: This article is part of the issue "Fostering the Socially and Ecologically Sustainable Digitalisation of Welfare States" edited by Paula Saikkonen (Finnish Institute for Health and Welfare) and Marta Choroszewicz (University of Eastern Finland), fully open access at https://doi.org/10.17645/si.i514

Abstract

Algorithmic decision-making (ADM) is increasingly used by public organizations to allocate social benefits. However, it remains unclear whether ADM leads to more harmonized decisions, especially in multi-level governance contexts like Belgium. Therefore, we ask whether, and to what extent, ADM is linked to the harmonization of welfare decisions across local social agencies. More specifically, we analyze decisions related to additional financial support in terms of (a) the likelihood of granting monthly financial support compared to other types of support, and (b) the amount of monthly support granted. In doing so, we focus on REDI, a digital rule-based algorithmic system designed to assess families' financial needs in Belgium. We draw on an online survey with respondents from 344 public centers for social welfare (PCSW), 40 interviews with respondents from 20 PCSWs, and policy guidelines from 86 PCSWs. Our findings demonstrate that the adoption of REDI can be linked to harmonization, yet only regarding the *form* and height of support, with users being more inclined to grant support monthly and provide a higher amount. Nonetheless, variance in welfare decisions persists, indicating a half-hearted harmonization. With both financial and normative considerations at the local level playing a significant role in how the ADM system is shaped and implemented, this study highlights the importance of examining the organizational and political context in which ADM systems are deployed to understand their influence on welfare decisions.

Keywords

algorithmic decision-making; harmonization; policy implementation; social assistance; social welfare; standardization; street-level bureaucracy

1. Introduction

Algorithms are nothing new. In fact, evidence for the first algorithm dates from 2500 BC (Chabert, 1999), and the first algorithm to operate a computer stems from the 19th century (Bolter, 1984). Yet, since the fourth industrial revolution, the scale and extent to which digital algorithms are used to support human decision-making have rapidly increased, and welfare practices are no exception. Digital algorithmic decision-making (ADM; for a conceptualization, see European Parliamentary Research Service, 2019) is an increasingly common feature of public organizations involved in the allocation of benefits or the provision of services (Bovens & Zouridis, 2002; Van Gerven, 2022). In Norway, for instance, applications for social security benefits (such as child benefits or unemployment benefits) are automatically processed, sometimes even "without a person having been involved" (Nav, 2024). Another example is Finland, where the decision to assess eligibility for social assistance was centralized to address unequal treatment across municipalities and was later supported by ADM (Algorithm Watch, 2020; Varjonen, 2020).

This evolution brings about opportunities and risks for our welfare states. For instance, some argue that digital ADM can decrease the non-take-up of social rights, as it enables the automatic allocation of welfare benefits and minimizes shame and stigma linked to applying for such benefits in face-to-face contact (Madsen et al., 2022; Van Lancker, 2020). In contrast, it is demonstrated that the rise of ADM and digital technologies broadens the digital divide and creates administrative burdens, especially for people who are already vulnerable (Algorithm Watch, 2020; Eubanks, 2018). A further illustration of the ambivalent effects of ADM concerns the role of discretion. Discretion can be defined as the freedom public officials have when interpreting and implementing regulations (see Evans & Hupe, 2020; Lipsky, 1980 for commonly used definitions and conceptualizations). On the one hand, ADM may decrease the number of arbitrary or discriminatory decisions caused by public officials' large discretionary powers (Arvidsson & Noll, 2023). On the other hand, empirical evidence suggests the opposite since practices of discrimination can become embedded within ADM systems, or systems running on or being trained by biased data (Larson et al., 2016). Furthermore, it is argued that the diminished capacity for case-by-case policy implementation due to ADM systems disproportionately affects those who are already vulnerable (Germundsson & Stranz, 2023).

In this article, we zoom in on what seems to be another puzzle: the harmonization of policy outcomes across local social agencies when ADM is used to assess eligibility for welfare benefits. Some authors expect ADM systems to harmonize policy implementation since they might reduce the discretion of public officials (Bovens & Zouridis, 2002), make public officials adhere to regulation, whilst focusing on routines in their enforcement style (de Boer & Raaphorst, 2023), and streamline decision-making processes (Germundsson & Stranz, 2023). Other scholars emphasize that professionals' willingness to adopt ADM, along with the organizational context and culture in which they operate, significantly influences how such technologies are accepted and implemented in practice (Venkatesh et al., 2003). Indeed, public officials are reported to deliberately oppose the digital systems they are requested to use, sometimes precisely to provide individual assistance to clients (Jørgensen & Schou, 2020). And so, the extent to which ADM leads to harmonization of welfare decisions across local social agencies, as well as the conditions under which such decisions materialize, remain open questions. Additionally, the impact of ADM on the autonomy of lower-level governments is still under-examined. This is, however, crucial, since in the majority of welfare states, social assistance policies are executed at the municipal level (Kazepov & Barberis, 2013). Understanding the

harmonizing effects of ADM is key to designing socially just and sustainable welfare states, as it affects both citizens' social rights and the multi-layered governance of social policies directly.

In Belgium, a digital system called REDI (shorthand for References Budgets for a Decent Income; Frederickx et al., 2022) was implemented at the local level on a large scale. REDI is a rule-based algorithm that allows for evaluating and calculating the need for additional financial support of welfare clients, based on a fine-grained assessment of their household and employment situation. The responsibility to determine whether, and if so, how much, additional financial support is provided to clients lies fully within the discretion of the public centers for social welfare (PCSW). In contrast, legislation governing social assistance benefits is determined at the national level. REDI is meant to evaluate clients' living standards and calculate the resources they need to sustain a decent minimum, running counter to the usually idiosyncratic rules and guidelines of local social agencies. In 2023, the Belgian federal government encouraged local governments to adopt REDI through a full, but temporary, reimbursement of the financial support given, with the only requirement being that social agencies use REDI for the specific clients to whom they decide to grant additional financial support. In doing so, they expected an increase in equal treatment of clients across local social agencies (PSS Social Integration, n.d.-a, n.d.-b). Local governments still had the autonomy to determine how they incorporated various factors into the system, such as savings, debts, or the income of children. And so, while the implementation of REDI at a large scale was meant to streamline local policies to grant additional financial support to welfare clients, at the same time, it did not change the autonomy of local authorities regarding additional financial support and allowed them to use REDI in line with local practices.

The Belgian case is a prime example of how ADM is used with the expectation of harmonizing welfare decision-making in a multi-level governance context. Using unique data to evaluate the implementation of REDI at the local level, we ask whether, and to what extent, the use of ADM in welfare decision-making is linked to the harmonization of welfare decisions across local social agencies. An online survey including five vignettes with responses from 344 local governments allows us to assess welfare decisions in specific cases with and without REDI, and before and after REDI was implemented. To contextualize these decisions, we draw on qualitative analysis of 40 semi-structured interviews with frontline workers and managers from 20 PCSWs, along with an assessment of policy guidelines from REDI-users from 86 PCSWs, allowing us to explore why harmonization might (not) be linked to REDI.

2. Literature Review

Previous research shows contradicting findings and perspectives on whether ADM harmonizes welfare practices, and various perspectives on how technology brings about societal change.

Applied to public administrations and welfare states, the impact of ADM on discretion is often framed as a shift from street-level bureaucracies, where street-level bureaucrats have degrees of freedom when implementing regulation in their direct interactions with citizens (Lipsky, 1980), to system-level bureaucracies, where discretion is completely abolished or shifted to the designers of ADM systems (Bovens & Zouridis, 2002). Zouridis et al. (2020) highlight how this shift not only occurs by directly replacing human decisions with ADM but also by the mere use of software programs as they prestructure processes. Empirical evidence is found to support the claim that ADM *can* limit discretion and streamline practices (e.g., de Boer & Raaphorst, 2023; Ranerup & Henriksen, 2022). Buffat (2015) refers to this stance, reflecting

the expectation that ADM limits human discretion, as the "curtailment thesis." Such technological determinism expects technology, by itself, to bring social or behavioral change (although various variants and nuances exist; for that, see Lauwaert & Chomanski, 2025). In legal philosophy, this argument is derived from the idea that both computational code and law regulate and affect human behavior—e.g., Lessig's (1999) "code is law" stance or Diver's (2021a, 2021b) concept of "digiprudence." Furthermore, policy outcomes are framed as inherently present in, or directly caused by ADM (Diver, 2021b; Grimmelmann, 2005). These studies articulate a clear causal relationship (Kling et al., 2000): As code regulates, the discretion of public officials decreases, with more harmonized policy outcomes as a result.

Technological determinism is challenged both in theory and by empirical evidence. Social scientists especially criticize a deterministic approach to technology, as it neglects the human and societal influence in both the design of ADM systems and their practical effects. Social construction of technology, in particular, questions digital ADM systems as external objective realities by framing them as human constructs, shaped through social actions (Berger & Luckmann, 1966). Actor-network theory also highlights the interactions between technology and humans, expecting a two-way relationship (Michael, 2017; Orlikowski, 2000). In line with this, Venkatesh et al. (2003) developed a unified theory of acceptance and use of technology, describing how individual characteristics and social contexts influence how technologies are accepted and used in practice. With humans shaping technology, and thus ADM systems, they are also expected to influence the effects ADM might bring to society. Consequently, authors call into question whether ADM in itself will ensure policy harmonization. Indeed, empirical evidence shows how public officials working with digitalized systems oppose them (Baines et al., 2010; Devlieghere & Roose, 2018, 2020) and develop parallel ways of working to still be able to individually assist clients (Jørgensen & Schou, 2020) or keep the relational aspects of social work practices (De Witte et al., 2016). In the literature, these ideas are reflected in the "enablement thesis" (Buffat, 2015) or the "continuation thesis" (Marienfeldt, 2024), showing ADM systems do not *automatically* or *just* curtail human discretion, but can bring about new ways for public officials to use discretion. Adding to that, it is contested that the creation of more rules (Evans & Harris, 2004)—for example, through code (Diver, 2021a, 2021b; Lessig, 1999)—or the standardization of practices (as shown by Nordesjö et al., 2020; even though this study did not cover standardization through ADM), actually lead to public officials having less discretion in practice.

Not only is there disagreement in the literature concerning the harmonizing effects of ADM, but the desirable extent of harmonization versus discretion also remains a subject of debate. Stances about (limiting) discretion in welfare practices are often based on normative assumptions about what public officials should be or do and what desired levels of discretion might be (Brodkin, 2016). For instance, Molander et al. (2012) frame discretion as a threat to equal treatment in welfare states, whereas Lipsky (1980, p. 161) underlines the importance of discretion in service provision. The same goes for the desired level of (de)centralization in multi-level governance welfare states. Where Frederickx et al. (2022), for instance, mention the risk of the fulfillment of social rights becoming dependent on the place of residence when local practices differ, Sellers and Lidström (2007) highlight the importance of strong local governments to counter spatial inequalities. Yet, Evans and Harris (2004) underline that discretion is neither "good" nor "bad." Whatever the desired amount of discretion in welfare decisions, when discretion becomes systematized to decide "who gets what, when [and] how," it becomes political (Brodkin, 2020, p. 64). This is particularly salient in multi-level governance contexts, where the authority to design anti-poverty policies is distributed across various layers of policymaking and political institutions, as in the case of Belgium. Moreover, given the diverse policy

challenges and political dynamics at the local level, the federal government and local social agencies may pursue different interests when designing and implementing supplementary support measures. As a result, they may aim for divergent objectives in the deployment of ADM systems. In sum, the extent to which ADM leads to harmonized decisions across different organizations, especially in a multi-level government context, invites further examination. Moreover, the expected outcomes of ADM in welfare practices are often normatively framed, driven by a tension between local autonomy and centralized legislation. In what follows, we discuss the Belgian case in which an ADM system is implemented to streamline local decisions on additional financial support in a multi-level governance context.

3. Additional Financial Support in Belgium: The Role of ADM

In Belgium's multi-level governance system, local PCSWs serve as the final safety net. Each municipality has a PCSW that provides financial support in the form of social assistance benefits on the one hand, and can provide additional cash and in-kind support on the other hand. While the first task is subject to federal legislation in terms of entitlement conditions and amounts, PCSWs are fully responsible for defining the type of additional support as well as its conditions. For example, local social agencies can autonomously decide on the type of support they provide for each client, ranging from rent allowances and medical bill contributions to food bank vouchers and cultural activity discounts. Given the full autonomy of local authorities to determine who is financially in need, various practices exist (Van Mechelen & Bogaerts, 2008). Examples derived from our interviews include using standardized Excel files to calculate financial needs, whereas other PCSWs do not adhere to such quantifications and discuss cases at team meetings.

In Belgium, some PCSWs use the digitalized rule-based algorithmic system REDI to assess the needs of vulnerable clients. REDI was developed by scientists of the Centre for Budget and Financial Well-Being (CEBUD) and is based on the reference budget method. Reference budgets are developed to determine the minimum household income necessary for social participation in a given context. Theoretically and conceptually, they are based on the theory of human need (Doyal & Gough, 1991; Goedemé et al., 2019). While reference budgets are based on "model family types" and assumptions about health and housing, REDI draws on the reference budget method to allow social professionals working in PCSWs to determine whether a *specific* client's household has sufficient financial resources to live a decent life (Frederickx et al., 2022). Reference budgets represent the minimum costs to fulfil 12 basic needs, including elements as having access to healthy food, basic hygiene, sufficient relaxation, and being able to maintain social relationships. Different parameters influence the height of the reference budget. On the one hand, they are based on (inter)national guidelines, focus groups with citizens, and budget checking in stores. Every six months, the budgets are indexed. Every two and a half years, budgets are re-evaluated. On the other hand, personal characteristics, such as age and gender of family members, and socio-financial elements, such as renting on the private or social market and the source of income (work or social security), influence the height of the reference budgets (CEBUD, n.d.-a, n.d.-b).

In encouraging PCSWs to adopt REDI, the federal government aimed to ensure "equal treatment of client situations across PCSWs and social professionals" (PSS Social Integration, n.d.-a, n.d.-b). The federal government provided a temporary subsidy—fully reimbursing REDI-based financial support—from 2023 until the end of 2024. In turn, 424 of 581 PCSWs (73%) started using it in 2023 and 45 (8%) were already using REDI before.

When using REDI, social workers start by inputting the personal information of their client, such as their national registration number, sex, and date of birth. Next, information about financial needs and household budget is added. This information is structured in four different categories of expenses: fixed costs, living expenses, savings for future expenses, and additional costs. For each of these categories, social workers can choose to use reference budgets as calculated in REDI or enter the real expenses of clients. Different types of income, stemming from both work and social security, can be registered as well. REDI then calculates a monthly amount by comparing the family income to the (adapted) reference budgets. If the result is negative, the household income is insufficient to live a dignified life (Frederickx et al., 2022). REDI provides an overview of the incomes and expenses of the household, which may prompt a response to reduce expenses that are (much) larger than the reference budgets, and reveal potential sources of income, such as allowances that are not yet claimed. Most importantly, it provides a benchmark of clients' financial needs, which can be used to determine the type and level of additional support. According to the REDI-philosophy, the entire negative balance should ideally be granted to the client every month if the budget allows, replacing other ad hoc additional types of support (CEBUD, 2024, p. 28).

Importantly, the federal government did not impose strict regulations on local governments regarding the use of REDI. The only requirement for reimbursement was that social assistance agencies use REDI for the specific clients they support with financial aid. As a result, local governments retained autonomy to determine how they incorporated various income and expense factors into the calculation, such as savings, debts, or the income of children, and hence could easily deviate from the REDI calculations. Yet, while there was no formal change in the local authority to provide additional financial support, the implementation of REDI still meant to streamline how client cases were evaluated by different PCSWs. This provides an excellent case to assess the competing expectations about the harmonization effect of ADM systems in a context of almost full discretion at the local level. Given the objective of the federal government, we will test whether the implementation of REDI in local social welfare agencies has led to harmonization in terms of the decision to grant monthly financial support and the amount of support provided.

4. Data and Methodology

To answer our research question, we used both quantitative and qualitative methods. We draw on an online survey filled in by 344 representatives of PCSWs, 40 semi-structured interviews with frontline workers and managers across 20 PCSWs, and policy guidelines from 86 PCSWs. Data were gathered in a project funded by the Belgian federal government (no. MIIS2023 06) to evaluate how PCSWs have implemented and used REDI and how their experience and decisions compared with non-users.

4.1. Quantifying Case Variation and Harmonization

All 581 PCSWs were invited to take part in the research. Each participating PCSW delegated one respondent to fill in the online survey, someone who is aware of the local policies that apply in the PCSW regarding granting additional (financial) support, and knows how these policies are applied in specific cases. Of 344 participants, 320 observations without missing values on the variable indicating the type of support for all client cases (vignettes) were retained.

To evaluate whether the use of REDI would lead to harmonization in client case assessment, we included five vignettes in the online survey. For each of these vignettes, we asked respondents if and how additional support would be granted in their PCSW. What follows is a short description of these client cases. A detailed overview of how the vignettes were presented to the respondents in their original languages (Dutch and French) can be found in the Supplementary File.

Vignette 1 (Jens): Jens, a single homeless man in his 30s, stays either at a friend's place or at his brother's. He has children, but does not have contact with them. Jens works as a bartender, which grants him a low and fluctuating income of approximately €1,220 a month. Jens contacted a PCSW because of his accumulating debts. His monthly costs include a limited compensation to cover board and lodging (€250), child support (€280), and repaying his debts. He has monthly repayment plans for a telecom provider (€50), lawyer's fees (€100), and bailiff fees (€100).

Vignette 2 (Anne): Anne is a young female student who has a history in youth care. Due to traumatic events in the past, she no longer has contact with her parents. She contacts the PCSW because she can't pay her medical bills. Anne receives a social assistance benefit. Her monthly costs include rent for a studio (€500), water and energy bills (€145), and psychiatric therapy (€300). She has unpaid hospital bills accumulating up to €1,500.

Vignette 3 (Fatima): Fatima is 40 years old and a single mother of three kids. She has no contact with the father of her children, who she believes to be in jail. She has two minor children, who are twins, and one son who is 19 years old. Her oldest son just started a handyman business, from which he receives an irregular income. Fatima became unemployed six months ago due to cut-offs at her previous job. Fatima contacts the PCSW because her savings are running out, and she wants to avoid being unable to pay her monthly costs. Her income includes an unemployment benefit and child support (€2,095). Her savings are €6,504. The fluctuating income of her son is approximately €700. Her monthly costs include a loan for her home (€844), water and energy bills (€216), her car (€410), and a fitness subscription for one of her twins (€30).

Vignette 4 (Valerie): Valerie is a woman in her 30s. She has a partner, Tuur, who is in his 30s as well. The couple has four children, all of whom are minors. Valerie is a stay-at-home mom and volunteers by taking in stray cats. Tuur is employed and has an income. Valerie is contacting the PCSW because they report having insufficient means. The couple flags that their relationship suffers from their financial situation. Their monthly income is €2,190, including Tuur's salary, child support, and Valerie's volunteer allowance. Their monthly costs include rent (€925), water and energy bills (€230), and pets (€120). They also pay a monthly rental deposit (€100). Apart from his deposit, they have no debts.

Vignette 5 (Marc): Marc is a single middle-aged man. After quitting his job due to excessive stress, he has been receiving a social assistance benefit. Marc contacted the PCSW again because he is still unable to make ends meet. Marc is reported to give the impression of being very stressed or under the influence of drugs. His monthly costs include rent (€612), water and energy bills (€161), smoking (€70), and a subscription to his favorite football team (€34).

All respondents were asked to assess these vignettes as they would do today, with REDI for the users and without for the non-users. Respondents of the user group were asked how they would have assessed these client cases before they adopted REDI. Respondents could indicate whether they would grant additional support and also which types they would grant. They were allowed to indicate multiple responses and we provided open text fields in which respondents could give more details on the specifics of the support on offer. We operationalized the responses into four categories: (a) no additional support; (b) in-kind support; (c) non-recurring financial support; and (d) monthly financial support. While respondents opting for financial support could also include material support, the category *in-kind support* excludes financial support.

In order to evaluate whether client decisions are more harmonized linked to the adoption of REDI, we define harmonization as an increase in similar outcomes of welfare decisions for similar cases across PCSWs. Note that we focus on harmonization, which we define as the reduction in variation, compared to standardization, which is related to the removal of variation altogether. We first focus on the type of additional support provided for each case, with a particular focus on monthly financial support. For those respondents, we subsequently assessed the amount of monthly financial support proposed. In doing so, we compare these outcomes for REDI-users and non-users and users before and after the implementation of REDI.

Given the literature review and the explicit expectations of the federal government, we hypothesize that harmonization will occur in two ways. The first is what is commonly referred to as sigma-convergence, a reduction in the cross-sectional dispersion of outcomes. Sigma-convergence is usually measured by a statistical measure of dispersion, such as the coefficient of variation. In our analysis, we assume harmonization if the coefficient of variation in the outcomes of the different vignettes is lower in the users compared to the non-user group, and becomes smaller after the adoption of REDI compared to before the adoption of REDI. A reduction in the coefficient of variation means that the variation in the decisions of respondents declines. Since REDI imposes an explicit standard, we also expect decisions to be more in line with the REDI norm. In this specific case, this would entail that (a) respondents are more likely to provide monthly financial support and (b) the amount of the benefits moves upward (also known as an anchoring effect; Tversky & Kahneman, 1974). We measure upward convergence through delta convergence, a concept capturing changes in the distance between the outcome and an exemplary model (Starke et al., 2008). In doing so, we ran our five vignettes through REDI to have a benchmark of how these client cases are assessed using the budget standard method (Table 1). We will analyze to what extent the amounts of additional financial support proposed by respondents are closer to the standard set by REDI.

Table 1. Overview of the client cases and their corresponding calculation in REDI.

	Case description	Monthly income	Monthly costs	Savings	Debts	REDI's output
Vignette 1: Jens	Single man 34 years old Homeless	Wage: €1,220	Drinking fee subtracted from wage: €80 Living cost: €250 Child allowance: €280	None	Payment plans for debts: • bailiff fee: €100 (€2,430 in total) • telecom: €50 (€2,00 in total) • lawyer fee: €100 (€1,720.16 in total)	€519.90
Vignette 2: Anne	Single girl Student 20 years old No contact with parents	Social assistance benefit: €1,263.17 Child benefit: €168.96	Rent: €500 Train and bus subscription: €30 Utilities: €180 Psychiatrist: €300	€407.30	Hospital debts without payment plan: €1,500	€417
Vignette 3: Fatima	Single mother of three kids, two minors 40 years old	Unemployment benefit: €1,705. 08 Child benefit: €389.57 Income of son: €700	Mortgage: €844 Fitness subscription: €30 Utilities: €216 Car: €410	€6,504.12	None	€1,571. 69
Vignette 4: Valerie	Mother of four minor kids Has a partner 32 years old	Wage of partner: €1,880 Child benefit: €1,210 Volunteering fee: €100	Rent: €925 Utilities: €230 Pets: €120	€54	Payment plan for rental deposit: €100	€1,423.94
Vignette 5: Marc	Single man 52 years old	Social assistance benefit: €1,263.17	Rent: €612 Utilities: €161 Smoking: €70 Football subscription: €34	None	None	€260

4.2. Contextualizing and Explaining Case Variation and Harmonization

Besides measuring harmonization in welfare decisions quantitatively, we draw on a qualitative analysis of policy documents and in-depth interviews to contextualize these findings and explain the remaining variation that our quantitative data show. The online survey included questions about how additional financial support is granted in PSCWs, with or without REDI. Adding to that, we analyzed policy documents of 86 PSCWs that use REDI to assess similarities and differences in rules regarding how additional support should be granted. To do so, we compared policy documents and the recommendations of the developers of REDI (CEBUD, 2024), focusing on rules stipulating eligibility criteria for additional financial support and rules indicating the amount of financial support. An overview of the specific parameters used can be found in Table A1 in the Supplementary File. Finally, we conducted semi-structured interviews with 40 respondents from 20 PSCWs in Brussels, Flanders, and Wallonia. More information about the respondents and interview conditions can be found in Table A2 in the Supplementary File. All agencies were selected to ensure geographical representation, based on a socio-economic typology of Belgian municipalities (Belfius-indicators). In each PCSW, we aimed to select one case manager responsible for assessing eligibility for additional support and one respondent at the management level of the organization. In these interviews, we discussed local policies regarding additional support with or without REDI and client vignette 3 (Fatima) in depth, as it was reported to show the most variation in our preliminary results. The interviews were analyzed thematically, according to the staged process of Braun and Clarke (2006). After familiarization with the data, the qualitative researchers jointly identified key topics. Relevant topics include ways of granting additional support and perceived advantages and disadvantages of REDI. All topics are included in Table A3 in the Supplementary File. Next, they summarized the main findings in a thematic grid. This process allows for structuring similarities and differences between PCSWs in how they allocate additional support and why, allowing the researchers to agree on findings that represent all cases and finally identify arguments explaining variation. During that process, additional insights emerged about the different normative assumptions about welfare support, which are discussed in Section 5.3.2. This research was approved by the ethical commission of the KU Leuven with references G-2023–7204-R4(MAR) and G-2023–7204-R6(AMD).

5. Results

We examine whether the use of REDI in welfare decision-making is associated with greater harmonization in case assessments among PCSWs. We structure our results by first discussing harmonization in case assessment between REDI-users and non-users, and second, by comparing case assessment before and after the implementation of REDI. We first examine differences between REDI users and non-users regarding the decision to provide support. Next, we look at harmonization by means of the coefficient of variance to assess whether users are more likely to provide monthly financial support (sigma-convergence). Finally, for those PCSWs who provide monthly financial support, we evaluate whether the level of support is closer to the benchmark amount provided by REDI (Table 1; delta-convergence).

5.1. Comparing Users and Non-Users

5.1.1. Variation and Harmonization in the Type of Additional Support

In Table 2, we compare the responses of REDI-users and non-users on the question of what type of additional support they would grant for the different client cases. Generally speaking, we find similar degrees

Table 2. Frequency of the choice between granting no support, in natura support, one-off support, and monthly recurrent support for each client situation.

	Non-users		Users	
	N	%	*N*	%
Vignette 1: Jens				
No support	16	28.1	97	36.9
In kind	23	40.1	68	25.9
Financial support, once	14	24.6	49	18.6
Financial support, monthly	4	7.0	49	18.6
X^2 (3, N = 320) = 9.1433, p = .027				
Vignette 2: Anne				
No support	4	7.0	12	4.6
In kind	1	1.8	5	1.9
Financial support, once	16	28.1	42	16.0
Financial support, monthly	36	63.2	204	77.6
X^2 (3, N = 320) = 5.6513, p = .130				
Vignette 3: Fatima				
No support	31	54.4	131	49.8
In kind	7	12.3	18	6.8
Financial support, once	7	12.3	34	12.9
Financial support, monthly	12	21.1	80	30.4
X^2 (3, N = 320) = 3.3027, p = .347				
Vignette 4: Valerie				
No support	12	21.1	48	18.3
In kind	15	26.3	48	18.3
Financial support, once	16	28.1	47	17.9
Financial support, monthly	14	24.6	120	45.6
X^2 (3, N = 320) = 9.1838, p = .027				
Vignette 5: Marc				
No support	12	21.1	45	17.1
In kind	10	17.5	38	14.5
Financial support, once	10	17.5	49	18.6
Financial support, monthly	25	43.9	131	49.8
X^2 (3, N = 320) = 1.078, p = .782				

of consensus in the type of support chosen by respondents. Across users and non-users, we see that Fatima (vignette 3) is least likely to get any support at all, while Jens (vignette 1) is least likely to get financial support. In contrast, Anne (vignette 2), Valerie (vignette 4), and Marc (vignette 5) are most likely to receive financial support. Only in the cases of Jens and Valerie, a significant difference in the proportion of the response categories is observed. In the case of Jens, a homeless man with debt, REDI users are *more* inclined to forego any support, but among those who provide support, the type of support differs. For Valerie, we see substantial shifts in the categories of in-kind support and monthly financial support. However the adoption of REDI did not lead to a substantially different pattern of decisions to provide additional support or not.

In Table 3, we focus specifically on differences between users and non-users in the probability of providing monthly financial support for each vignette, for those who opted to provide additional support. While Table 2 indicated that there are significant shifts in the type of support in only one out of five cases, here we ask a different question: if PCSWs decide to grant support, are they more likely to provide monthly financial support? The results provide evidence for sigma convergence since the coefficient of variation across all vignettes is lower for REDI users than for non-users. For all cases, a higher share of users opts for monthly financial support, with substantial and significant differences for the vignettes of Jens, Anne, and Valerie. For Jens, for instance, the share of respondents increases by 20 p.p., and the coefficient of variation declines by 50% from 3.08 to 1.55. While REDI-users are more likely to opt for no support for Jens, at the same time, they opt much more often for monthly financial support when they do choose to provide support. For Anne and Valerie, respondents were more likely to provide monthly financial support to begin with, but here too we see substantial differences between users and non-users of 15 and 25 p.p., respectively, with a strong decline in the coefficient of variation. For Fatima, the decline in the coefficient of variation is smaller, and the difference between users and non-users in the probability of providing support is not significant. For Marc, differences are negligible.

Table 3. Variation in the decision to grant monthly financial support for users and non-users.

Case	Share of respondents granting monthly financial support		*t*-value	*p*-value	Δp.p.	Coefficient of variation	
	Non-users (*n*)	Users (*n*)	*t*	*p*		Non-users	Users
Jens	10% (41)	30% (166)	−2,6269	0.0093	20	3.08	1.55
Anne	68% (53)	81% (251)	−2,1760	0.0303	13	0.69	0.48
Fatima	46% (26)	61% (132)	−1,3652	0.174	15	1.10	0.81
Valerie	31% (45)	56% (215)	−3,0576	0.0025	25	1.50	0.89
Marc	56% (45)	60% (218)	−0,5745	0.5622	4	0.90	0.82

5.1.2. Variation and Harmonization in the Amount of Support Granted

In Table 4, we show the level of monthly financial support provided by non-users and REDI-users. For vignettes 1 (Jens) and 3 (Fatima), the sample was too small to provide meaningful results since not all respondents filled in the question about the amounts. Notably, we only observe a decline in the coefficient of variation for the case of Marc, indicating sigma convergence and a more similar evaluation of client cases. In the case of Marc, however, mean amounts of support are lower amongst users than amongst non-users. This is in line with the results in Table 3, where Marc was less likely to receive monthly financial support. In this case, there is sigma convergence but no delta convergence. The differences in the amounts are not

significant, except for the case of Valerie ($p = 0.02$). Here we see a much higher mean amount in the users group, but also an increase in the coefficient of variation, meaning that there is less agreement amongst REDI users on the amount provided compared to non-users. In the cases of Anne and Valerie, we do see some evidence for delta convergence, since monthly amounts are closer to the norm set by REDI. At the same time, the levels of support remain (far)below the REDI benchmark. For instance, while the mean amount of support for Valerie is over twice as high in the users group, it still only covers 30% of the financial need identified through the REDI tool. Where REDI proposes to grant most support to Valerie (€1423.94), she receives proportionally less compared to Marc and Anne. For them, the differences between users and non-users are small, however. This indicates harmonization with the REDI norm regarding the height of support, but not regarding its assumptions about neediness.

Table 4. Variation in the amount of monthly financial support for users and non-users.

Case	Mean amount of monthly financial support		REDI-benchmark	Support as % of REDI benchmark		*t*-value	*p*-value	Δ€	Coefficient of variation	
	Non-users (*n*)	Users (*n*)		Non-users	Users	*t*	*p*		Non-users	Users
Jens			€519.90							
Anne	€238 (31)	€251 (188)	€416.97	57%	60%	–0,5300	0,55966	€13	0.47	0.49
Fatima			€1571.69							
Valerie	€197 (12)	€428 (105)	€1423.94	14%	30%	–2,3258	0,0218	€231	0.61	0.80
Marc	€173 (21)	€166 (118)	€260	67%	64%	0,2822	0,7782	€–7	0.73	0.53

Notes: The monthly amount of monthly financial support includes the amounts indicated by local social welfare agencies as monthly additional financial support, plus specific recurring amounts for rent and utilities; for Jens and Fatima, the number of respondents in the non-users group is too small to make any meaningful comparison.

5.2. Variation and Harmonization Before and After Adopting REDI

To make a comparison on the levels of variation and harmonization before and after the implementation of REDI, we compare case assessments of REDI users before and after they started using REDI. Here, we include only respondents who provided information about how they assess clients' needs before the adoption of REDI (with a smaller number of observations across cases as a result). To gauge the situation before adopting REDI, we asked respondents to assess these cases a second time, taking into account the practices and guidelines in place before the implementation of REDI. Table 5 shows the share of respondents granting monthly financial support compared to all decisions (including no support), while Table 6 shows the mean level of monthly financial support for those who opted for monthly financial support.

For the subsample of PCSWs who use REDI and provided full information on their decisions before adopting REDI, we see a reduction in the coefficient of variation across all cases, suggesting sigma convergence. For all cases, we also notice an increase in the probability of support, although the differences are small and insignificant in the case of Jens and Fatima. For Anne, we see a substantial increase in the probability of granting monthly financial support (22 p.p.) and a large reduction of the coefficient of variation (–40%). Across cases, we observe that the patterns of agreement remain similar before and after adopting REDI, with most agreement on the type of support in the case of Anne, and least agreement in the cases of Jens and Fatima.

Table 5. Variation in the decision to grant monthly financial support, before and after REDI.

Case	Share of agencies granting monthly financial support			*t*-value	*p*-value	Δp.p.	Coefficient of variation	
	Before	After	*n*	*t*	*p*		Before	After
Jens	13%	17%	159	−1,6140	0,1085	4	2,64	2,22
Anne	56%	78%	154	−5,0798	0,0000	22	0,88	0,53
Fatima	28%	32%	133	−1,0919	0,2769	4	1,62	1,48
Valerie	37%	49%	150	−2,7827	0,0061	12	1,30	1,03
Marc	43%	52%	153	−3,0656	0,0026	9	1,15	0,96

Note: Only includes local social welfare agencies that adopted REDI and provided full information on how they evaluated requests for additional financial support before, including agencies that provided no support.

Table 6. Variation in the amount of monthly financial support, before and after REDI.

Case	Mean amount of monthly financial support			*t*-value	*p*-value	Δ€	Coefficient of variation	
	Before	After	*n*	*t*	*p*		Before	After
Jens								
Anne	€211	€258	77	−1.2988	0.0025	€47	0.70	0.54
Fatima	€243	€318	28	−1.2659	0.2164	€75	1.11	1.01
Valerie	€272	€369	43	−2.4778	0.0173	€97	0.84	0.84
Marc	€161	€176	59	−0.8961	0.3739	€15	0.78	0.61

Notes: Only includes local social welfare agencies that adopted REDI and provided full information on how they evaluated requests for additional financial support before; the amount of monthly financial support includes the amounts indicated by local social welfare agencies as monthly additional financial support plus specific recurring amounts for rent and utilities; for Jens, the number of respondents is too small to make any meaningful comparison.

Finally, regarding the level of monthly support, our data shows significant and substantial upward changes for Anne (with on average €47 more support granted) and Valerie (with on average €97 more support granted) after the adoption of REDI, moving closer to the benchmark set by REDI (Table 4). Across all cases, we observe that mean amounts tend to be higher after adopting REDI (delta convergence), while in most cases the coefficient of variation decreases (sigma convergence). However, in the case of Mark, the difference in mean amounts is small, and the disagreement in the amount given to Valerie is similar before and after the adoption of REDI.

5.3. Qualitative Contextualization Explaining Half-Hearted Harmonization

5.3.1. (Variation in) Making Rules About Using ADM

While our quantitative analyses do suggest that the use of ADM harmonizes outcomes in welfare decisions to a certain extent, at the same time, it is clear that there is still substantial variation within and across cases. How can we understand such half-hearted harmonization? In this section, we provide more evidence on this, drawing on policy documents of REDI users, and interviews with managers and frontline workers.

Once PCSWs decided to use REDI, most of them (79%, $n = 288$) made agreements on how to do so. Written agreements often stipulate detailed rules regarding (a) the target audience of the policy measure, (b) how the

REDI-benchmark should be calculated, i.e., what kind of expenses should be included, and (c) the amount of support to grant when REDI indicates clients have insufficient means. Whereas differences in the second type of rules result in differences in how the REDI-benchmark is computed, differences regarding the third type of rules result in what support is provided when the REDI-benchmark indicates a client does not have sufficient means at their disposal.

In the agreements we analyzed ($n = 86$), many PCSWs deviate from the recommendations provided by REDI's developer. For example, the recommendation of granting the full negative result generated by REDI was only found in 15% of the analyzed policy documents. In fact, none of the recommendations were adopted as rules in the majority of our sample, and, more importantly, a wide range of different practices were found regarding all three types of rules. Some policy documents even explicitly mention the (continued) use of alternative tools or calculations to decide whether a specific client has sufficient means.

5.3.2. Reasons to Use Local Autonomy When Deploying ADM in Welfare Practices

The policy documents revealed substantial variation among PCSWs concerning the formal rules governing the use of REDI. During the interviews, respondents in management positions elaborated on the reasons why they chose to diverge from the developers' guidelines and use their local autonomy, either by introducing locally adapted practices or by maintaining pre-existing procedures. Two primary lines of argumentation can be identified.

First and foremost, respondents report that financial uncertainty and the financial position of local PCSWs play a crucial role in deciding whether or not to use REDI in the first place, and second, how to use the tool. In fact, the subsidy was the reason most reported (80%, $n = 290$) by REDI-users to implement the system. Because of the link with government subsidies, respondents reported feeling "forced" to use the tool, even though they did not always want to integrate this new way of working into their local practice. This was especially mentioned in municipalities facing debts or having financial problems. Adding to that, interviewees mentioned it was unclear whether—and if so, how long—the federal subsidy would continue, which is why they were hesitant to increase the amount of support with REDI, as it would be "unjust" to later cut back financial support. This argument was also used to restrict the target group for REDI support in the early stage, with the possibility of enlarging it when more clarity was provided about the continuation of the project. Adding to that, respondents mentioned not only uncertainty about the financial aspects of the policy initiative, but also about the policy aims of the central government. More specifically, respondents from multiple PCSWs stated it was not clear whether the federal government wanted to strengthen control over the local practices and reduce local autonomy.

The second argument is linked to normative assumptions of local PCSWs on how welfare support should be organized. For example, some respondents mentioned they consider the amount proposed by REDI too high to be "fair." They refer to the fact that they believe a high amount of support would disincentivize people from participating in the labor market and that there should be a (larger) difference between the minimum wage and the amount of financial welfare support. Respondents also stated that they do not agree with "granting a blank check," without additional conditions for clients or ways to control how clients spend their money. This way, respondents refer directly to the "philosophy" underpinning REDI: instead of granting different additional (financial) benefits that are linked to specific costs, such as access to food stamps or paying pharmaceutical

bills, REDI advises granting the entire negative sum that stems from its calculation of a client's family neediness. Yet, as one respondent states, "because of the financial situation of the PCSWs, this philosophy cannot always be realized."

6. Discussion

In line with the technological deterministic approach, it was expected that an ADM system as REDI would increase harmonization in welfare decisions across PCSWs. However, in our case, characterized by strong local autonomy, REDI appears to foster a partial, half-hearted harmonization. REDI users are not more likely than non-users to provide support, but when they do provide support, they are more likely to provide monthly financial support (sigma conversion). In other words, harmonization is primarily observed regarding the *form* of support. When PCSWs provide monthly financial support, the amounts tend to be higher for those who adopted REDI, and there is some evidence for delta-convergence towards the REDI benchmark. At the same time, dispersion within the group of users sometimes increases, and the amounts are still far off from the benchmark.

The qualitative part of the study reveals why harmonization appears to be half-hearted. Local welfare agencies use their autonomy to shape welfare decisions at the organizational level, also when ADM is used. In our case, this resulted in PCSWs both adhering to previous practices and deriving from the system's philosophy. Two arguments were mentioned to do so: (financial) uncertainty about the federal policy initiative and normative considerations about welfare support. While the federal subsidy incentivized many PCSWs to adopt the tool, its unclear continuity led to reluctance, particularly in financially constrained municipalities. Aside from REDI not being in line with the financial reality of the local organizations, the system was also opposed by local governments due to normative considerations, including the adequacy of the proposed support amounts and the lack of conditionality. In line with other studies (e.g., Baines et al., 2010; Devlieghere et al., 2020), we found that REDI was implemented corresponding to the goals and normative beliefs of the organization. While other studies identify the relational aspects of social work to be a driving factor to circumvent standardization (e.g., De Witte et al., 2016; Jørgensen & Schou, 2020), our qualitative findings suggest that normative beliefs of who should get what at the organizational level are equally important to understand why ADM does not lead by default to harmonization. Our study shows that ADM systems used to allocate welfare benefits, as REDI, are not just tools used to implement regulations; they contain normative assumptions on how additional financial support should be provided. Thus, the deployment of ADM cannot be separated from normative discussions about welfare and the political power dynamics involved, showing ADM systems are indeed political (Brodkin, 2020). Consequently, our findings problematize notions of technological determinism, demonstrating that ADM does not impose uniformity by design but is instead mediated by organizational dynamics, financial constraints, and normative interpretations of welfare provision. Our findings thus support the continuation and enablement thesis and are in contrast with the curtailment thesis (Buffat, 2015; Marienfeldt, 2024). In addition, our study illustrates how ADM does not necessarily lead to system-level bureaucracies (Bovens & Zouridis, 2002), highlighting the importance of contextualizing both the modalities of the digital systems and their policy implementation to fully understand their impact.

Despite showing some evidence for partial harmonization, we only studied REDI after its implementation, which brings two limitations. First, this might prevent us from measuring policy changes, as these take time.

Moreover, respondents were hesitant to implement policy changes due to uncertainty regarding the program's continuation. Second, this inhibits causal interpretation. Harmonization might stem not directly from REDI but from the introduction of the subsidy, the policy guidelines, or the exposure to information about REDI. We relied on respondents' memory to compare welfare decisions before and after the adoption of REDI, introducing risks of recall and expectation bias. Adding to that, since local social welfare agencies were free to opt in on the policy initiative, REDI users might already favor generous, aligned interpretations of financial need.

7. Conclusion

We present evidence of partial, half-hearted harmonization in case assessment between local agencies linked to the adoption of an ADM system. Yet, substantial variation in local practices remains firmly in place as local governments still use their autonomy to shape welfare decisions. They do so by either adhering to their local practices as before the introduction of the system or by adapting the system to the financial and normative reality of their organization. In sum, in a multi-level governance context in which a federal government aims to streamline the autonomy of local organizations, our results suggest that the implementation of an ADM system to standardize practices will not, by itself, lead to a harmonization of welfare decisions. This way, we show the added value of approaching ADM systems from a socio-constructivist angle to understand the effects of ADM on welfare decisions in a multi-level government context.

Acknowledgments

The authors thank everyone who contributed to the finalization of the research project requested by the Belgian federal government (no. MIIS2023 06), including the researchers from CEBUD, our respondents, the Federal Public Planning Service for Social Integration, and Dr. Marjolijn De Wilde. We are grateful for the insightful comments we received from the anonymous peer reviewers, professor Sonja Bekker for her feedback during the Early Career Researchers' Day of ESPAnet Low Countries, and our colleagues at ReSPOND and CiTiP for their comments, especially Abrar Bawati, Joanna Geerts Danau, and Victoria Hendrickx.

Funding

Data were gathered in a project funded by the Belgian federal government (no. MIIS2023 06). Publication of this article in open access was made possible through the institutional membership agreement between KU Leuven and Cogitatio Press.

Conflict of Interests

The author declares no conflict of interest.

Data Availability

The authors can provide a replication package for the quantitative analysis upon request.

LLMs Disclosure

LLM tools (such as Grammarly, ChatGPT, and DeepL) were only used as a way of editing original text and writing code.

Supplementary Material

Supplementary material for this article is available online in the format provided by the authors (unedited).

References

Algorithm Watch. (2020). *Automating society report 2020*. https://automatingsociety.algorithmwatch.org

Arvidsson, M., & Noll, G. (2023). Decision making in asylum law and machine learning. Autoethnographic lessons learned on data wrangling and human discretion. *Nordic Journal of International Law*, *94*(1), 56–92.

Baines, S., Wilson, R., & Walsh, S. (2010). Seeing the full picture? Technologically enabled multi-agency working in health and social care. *New Technology, Work and Employment*, *25*(1), 19–33.

Berger, P. L., & Luckmann, T. (1966). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Anchor Books.

Bolter, J. D. (1984). *Turing's man: Western culture in the computer age*. University of North Carolina Press.

Bovens, M., & Zouridis, Z. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, *62*(2), 174–184.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Brodkin, E. Z. (2016). Street-level organizations, inequality, and the future of human services. *Management, Leadership & Governance*, *40*(2), 444–450.

Brodkin, E. Z. (2020). Discretion in the welfare state. In T. Evans & P. Hupe (Eds.), *Discretion and the quest for controlled freedom* (pp. 63–78). Springer.

Buffat, A. (2015). Street-level bureaucracy and e-government. *Public Management Review*, *17*(1), 149–161.

Centre for Budget and Financial Well-Being. (n.d.-a). *Budgets de reference pour une vie digne*. https://www.budgetsdereference.be

Centre for Budget and Financial Well-Being. (n.d.-b). *Referentiebudget voor een menswaardig leven*. https://www.referentiebudget.be

Centre for Budget and Financial Well-Being. (2024). *Afsprakenkader—sjabloon. Expertisecentrum Budget & Financieel Welzijn*. https://www.referentiebudget.be/_files/ugd/4ab716_fbce9732f1004d3ca7e70e8eae91677c.pdf

Chabert, J.-L. (1999). *A history of algorithms. From the pebble to the microchip*. Springer.

de Boer, N., & Raaphorst, N. (2023). Automation and discretion: Explaining the effect of automation on how street-level bureaucrats enforce. *Public Management Review*, *25*(1), 42–62.

De Witte, J., Declerq, A., & Hermans, K. (2016). Street-level strategies of child welfare social workers in Flanders: The use of electronic client records in practice. *British Journal of Social Work*, *46*(5), 1249–1265.

Devlieghere, J., & Roose, R. (2018). Electronic information systems: In search of responsive social work. *Journal of Social Work*, *18* (6), 650–665.

Devlieghere, J., Roose, R., & Evans, T. (2020). Managing the electronic turn. *European Journal of Social Work*, *23*(5), 767–778.

Diver, E. L. (2021a). *Digiprudence: Code as law rebooted*. Edinburgh University Press.

Diver, E. L. (2021b). Digiprudence: The design of legitimate code. *Law, Innovation and Technology*, *13*(2), 325–354.

Doyal, L., & Gough, I. (1991). *A theory of human need*. Palgrave Macmillan.

Eubanks, V. (2018). Automating inequality. How high-tech tools profile, police and punish the poor. St. Martin's Press.

European Parliamentary Research Service. (2019). *Understanding algorithmic decision-making: Opportunities and Challenges* (PE 624.261). https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf

Evans, T., & Harris, J. (2004). Street-level bureaucracy, social work and the (exaggerated) death of discretion. *British Journal of Social Work*, *34*(6), 871–895.

Evans, T., & Hupe, P. (2020). *Discretion and the quest for controlled freedom*. Springer.

Frederickx, M., Storms, B., & Cornelis, I. (2022). Implementing the right to a decent living standard in Belgium: The use of the REDI tool. *Journal of Social Intervention: Theory and Practice*, *31*(3), 21–36.

Germundsson, N., & Stranz, H. (2023). Automating social assistance: Exploring the use of robotic process automation in the Swedish personal social services. *International Journal of Social Welfare*, *33*(3), 647–658.

Goedemé, T., Penne, T., Hufkens, T., Karakitsios, A., Bernát, A., Simonovits, B., Carillo Alvarez, E., Kanavitsa, E., Cussó Parcerisas, I., Riera Romaní, J., Mäkinen, L., Matsaganis, M., Arlotti, M., Kopasz, M., Szivós, P., Ritakallio, V.-M., Kazepov, Y., Van den Bosch, K, & Storms, B. (2019). *Decent incomes for all: Improving policies in Europe*. Oxford University Press.

Grimmelmann, J. (2005). Regulation by software. *The Yale Law Journal*, *144* (7), 1719–1758.

Jørgensen, B., & Schou, J. (2020). Helping or intervening? Modes of ordering in public sector digitalization. *Journal of Organizational Ethnography*, *9*(3), 265–279.

Kazepov, Y., & Barberis, E. (2013). Social assistance governance in Europe: Towards a multilevel perspective. In I. Marx & K. Nelson (Eds.), *Minimum income protection in flux* (pp. 217–248). Palgrave Macmillan.

Kling, R., Crawford, H., Rosenbaum, H., Sawyer, S., & Weisband, S. (2000). *Learning from social informatics: Information and communication technologies in human contexts*. Center for Social Informatics, Indiana University.

Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016). *How we analyzed the COMPAS recidivism algorithm*. ProPublica. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Lauwaert, L., & Chomanski, B. (2025). *We, robots. Questioning the neutrality of technology, ethical AI and technological determinism*. Springer.

Lessig, L. (1999). *Code and other laws of cyberspace*. Basic Books.

Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. Russell Sage Foundation.

Madsen, C. Ø., Lindgren, I., & Melin, U. (2022). The accidental caseworker—How digital selfservice influences citizens' administrative burden. *Government Information Quarterly*, *39*(1), Article 101653.

Marienfeldt, J. (2024). Does digital government hollow out the essence of street-level bureaucracy? A systematic literature review of how digital tools' foster curtailment, enablement and continuation of street-level decision-making. *Social Policy & Administration*, *58*(5), 831–855.

Michael, M. (2017). *Actor network theory: Trials, trails and translations*. Sage.

Molander, A., Grimen, H., & Eriksen, E. O. (2012).Professional discretion and accountability in the welfare state. *Journal of Applied Philosophy*, *29*(3), 214–230.

Nav. (2024). *Automated evaluations of applications in Nav*. https://www.nav.no/automatisk-behandling/en

Nordesjö, K., Ulmestig, R., & Denvall, V. (2020). Coping tensions between standardization and individualization in social assistance. *Nordic Social Work Research*, *12*(4), 435–449.

Orlikowski, W. J. (2000). Using technology and constituting structures. A practical lens for studying technology in organizations. *Organization Science*, *11*(4), 404–428.

PSS Social Integration. (n.d.-a). *REDI*. https://www.mi-is.be/fr/themes/aide-sociale/redi

PPS Social Integration. (n.d.-b). *REMI*. https://www.mi-is.be/nl/themas/maatschappelijke-hulp/remi

Ranerup, A., & Henriksen, H. Z. (2022). Digital discretion: Unpacking human and technological agency in automated decision making in Sweden's social services. *Social Science Computer Review*, *40*(2), 445–461.

Sellers, J. M., & Lidström, A. (2007). Decentralization, local government, and the welfare state. *Governance*, *20*(4), 609–632.

Starke, P., Obinger, H., & Casttles, F. G. (2008). Convergence towards where: In what ways, if any, are welfare states becoming more similar? *Journal of European Public Policy*, *15*(7), 975–1000.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *785*(4157), 1124–1131.

Van Gerven, M. (2022). Studying social policy in the digital age. In K. Nelsen, M. Yerkes & R. Nieuwenhuis (Eds.), *Social policy in changing European societies: Research agendas for the 21st century* (pp. 251–264). Edward Elgar Publishing.

Van Lancker, W. (2020). Automating the welfare state: Consequences and challenges for the organisation of solidarity. In I. Van Hoyweghen, V. Pulignano & G. Meyers (Eds.), *Shifting solidarities. Trends and developments in European societies* (pp. 153–173). Springer.

Van Mechelen, N., & Bogaerts, K. (2008). *Aanvullende financiële steun in Vlaamse OCMW's. CSB-berichten 6/2008*. Antwerp University.

Varjonen, S. (2020). Institutional evolution and abrupt change: Reforming the administration of social assistance in Finland. *International Journal of Social Welfare*, *29*(1), 62–70.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478.

Zouridis, S., van Eck, M., & Bovens, M. (2020). Automated discretion. In T. Evans & P. Hupe (Eds.), *Discretion and the quest for controlled freedom* (pp. 313–329). Springer.

About the Authors

Janne Petroons is a doctoral researcher at KU Leuven. She is affiliated with ReSPOND at the Faculty of Social Sciences and with CiTiP at the Faculty of Law and Criminology. In her work, she focuses on digital welfare states from a socio-legal perspective.

Périne Brotcorne is a postdoctoral researcher at the Centre Interdisciplinaire de Recherche Travail, État, Société at the Faculté Ouverte de Politique Economique et Sociale at UCLouvain.

Martin Wagener is an associate professor in sociology at the Centre Interdisciplinaire de Recherche Travail, État, Société at the Faculté Ouverte de Politique Economique et Sociale at UCLouvain.

Koen Hermans is an associate professor in social work and social policy at ReSPOND, as well as the head of LUCAS—Centre for Care Research & Consultancy at KU Leuven.

Wim Van Lancker is an associate professor in social work and social policy at ReSPOND at KU Leuven.

Social Inclusion
2025 • Volume 13 • Article 10001
https://doi.org/10.17645/si.10001

cogitatio

ARTICLE

Open Access Journal

Trust-Affording Action: Citizens' Everyday Relations With Algorithmized Public Services

Antti Rannisto [1] **and Fanny Vainionpää** [2]

[1] Department of Computer Science, Aalto University, Finland
[2] Department of Information Technology and Electrical Engineering, University of Oulu, Finland

Correspondence: Antti Rannisto (antti.rannisto@aalto.fi)

Submitted: 30 January 2025 **Accepted:** 23 June 2025 **Published:** 2 September 2025

Issue: This article is part of the issue "Fostering the Socially and Ecologically Sustainable Digitalisation of Welfare States," edited by Paula Saikkonen (Finnish Institute for Health and Welfare) and Marta Choroszewicz (University of Eastern Finland), fully open access at https://doi.org/10.17645/si.i514

Abstract

Finland is considered a society of high trust. Finnish citizens' trust in public administration and institutions is clearly above the OECD average, and there is broad consensus on the virtues of maintaining high societal trust. As high-trust public institutions are now turning to new efficiency-promising AI technologies, it is important to ask: "How are these [technologies] then capable of upholding that trust?" This question is a direct quote from our fieldwork following Finnish citizens' everyday trust-building with new AI-infused services. Based on a trust-focused reading of our qualitative data, we propose an approach to trust that affords greater empirical nuance than alternative conceptions, which we see as limited for following the evolving dynamics of citizens' trust in new technologies. The approach we are developing situates trust within a processual conception of action and highlights the need to also grasp the "quieter"—embodied, habitualized, and intuitive—forms of trust as part of its living dynamic. We then apply this approach to examine the adoption of the Finnish Covid-19 tracing application and citizens' perceptions of algorithmically infused services provided by Finland's social insurance institution. We highlight creative tactics that citizens use to establish trust: the agent heuristic, the case heuristic, the social heuristic, and the interaction heuristic. Our research contributes to a nuanced understanding of trust and its situated dynamics from the citizen perspective, a focus we consider crucial at a time of unprecedented excitement around the transformation of high-trust institutions through algorithmic technologies.

Keywords

artificial intelligence; citizens; public services; trust; trust in AI; trust in technology

1. Introduction

In March 2024, Finland's newly inaugurated president, Alexander Stubb, took to X to celebrate another year of Finland topping *the World Happiness Report* (Stubb, 2024; see also Helliwell et al., 2024):

> Finland has been ranked the Happiest Country in the World, now for seven years in a row. Many reasons for it, here are my three: 1. Nature. 2. Trust. 3. Education. There are naturally many other reasons for collective happiness. What would be your top three?

Finland's top ranking was followed by the other Nordic countries—Denmark, Iceland, and Sweden, in that order—with Norway placing seventh. By highlighting trust as a key ingredient in the Nordic mix of happiness, President Stubb echoed a long-standing discourse linking trust not only to happiness but also to broader social goods such as cohesion, stability, and institutional legitimacy. Finnish citizens' trust in public institutions is well above the OECD average (OECD, 2021, 2024), and there is broad consensus on the "importance of maintaining the Finnish trust capital as a core value of the Finnish administrative culture and a cornerstone of institutional legitimacy" (OECD, 2021, p. 11). According to the OECD country report *Drivers of Trust in Public Institutions in Finland*, "The government of Finland sees trust as a fundamental and guiding value that underpins the functioning of the public administration and the development of people-oriented public services" (OECD, 2021, p. 16).

The development of such people-oriented public services has recently enacted a new imaginary of AI-assisted services tailored to individual citizens' needs (Räisänen, 2024), coupled with technology-solutionist promises of state efficiency. The current Finnish government identifies AI as a key element of its data economy strategy and promises to "enable the use of artificial intelligence to automate decisions made by public authorities" (Government of Finland, 2023, p. 125) as part of its larger efforts to support the "full use of the opportunities provided by digitalisation and artificial intelligence" (p. 6). With high-trust Nordic institutions now turning to new efficiency-promising but unstable algorithmic technologies (Kaun, 2021, p. 2047) like generative AI, it is important to ask: "How are these [technologies] then capable of upholding that trust?" This question is a direct quote from our fieldwork following Finnish citizens' everyday sensemaking and trust-building with the new AI-infused services that have entered their lives in recent years.

Our study focuses on the everyday dynamics of trust-building and breakdown in citizens' encounters with AI-infused public services. We ask the following research questions:

> RQ1: How does trust in algorithmic technologies emerge in everyday encounters with these technologies?
>
> RQ2: How is trust questioned and constructed in the process?

Drawing from four rounds of qualitative data collection, we develop an empirically grounded framework of trust-affording action. We see trust as a nuanced navigational device in everyday settings, shaped by the constant flux of new (AI) technologies that bring both promise and excitement but also lead to frequent experiences of uncertainty. Our qualitative reading of this uncertainty is corroborated by a 2023 survey conducted by Finland's Digital and Population Data Services Agency (DVV), which found that "only 15% of

Finns trust AI services" (DVV, 2023; the figure rose to 19% in the 2024 survey; see also Gerlich, 2024; Rousku, 2024, pp. 59–62). This kind of low trust is precisely the problem that the EU's AI Act seeks to address. According to Laux (2024), "its overarching aim…[is] to create European Union ('EU') citizens' trust in AI" (p. 1), which serves "a strategic purpose: [to] induce people to place trust in AI so that they will use it more and, hence, unlock the technology's economic and social potential" (Laux et al., 2023, p. 3).

We begin by looking at research on trust in social and technology studies and propose an alternative conceptual approach that is directly motivated by our qualitative engagements with citizens, which force us to look beyond dichotomies such as cognitive versus emotive concepts of trust. Our data unveils trust not only in its cognitive and active form—reflective and deliberative, commonly seen as "rational"—that is customary especially in computer science and ICT approaches but also in its intuitive, implicit, and latent forms, which are sometimes proposed in social scientific approaches countering the assumption of rational choice theory, something we call quiet trust. We approach trust as affecting the flow, direction, and continuation of action processes and place its elusive varieties into a unified processual frame that enables us to incorporate both active and quiet, or cognitive and intuitive, trust into one action-based framework. We then apply this approach to unpack specific cases of trust formation in citizen encounters with new algorithmic technologies. Finally, we circle back to the question of trust as fundamental to the social sustainability of welfare states and highlight its brittle nature at a time of overwhelming societal excitement around AI technologies. Our study contributes to the multidisciplinary research on citizens' trust in AI by highlighting the living dynamic and variety of trust. We propose a synthetic interpretation that integrates different modes of trust into a unified framework of processual action, showing how citizens' trust shifts between quiet and active forms as they move through a dynamic of established and exploratory phases within human-machine interactions.

2. Trust in Technology and AI

Approaching our empirical observations of people's relationships with AI technologies, we reviewed existing research and approaches to trust in both technological and social studies and found much of it difficult to work with. Indeed, our participants' relationship with new technologies called for a far more nuanced, formative, and fluid approach to trust than the dichotomous, mechanistic models found in much of the literature. Our search for more congruent approaches turned the research process into an open, critical conversation between available conceptualizations and our empirical reflections, and ultimately into an attempt to formulate a more generous approach to trust—one that allows for greater variety in how trust is lived and enacted as part of people's everyday experience and conduct. In our search, we draw on multidisciplinary studies of trust. First, we conduct an overview of empirical studies mainly in information systems, human–computer interaction, and software engineering to explore how trust in technology and AI is studied in ICT research. Much of this research borrows approaches from the more firmly established behavioral and social scientific studies on trust, to which we then turn in our search for alternative conceptions from the sociology and anthropology of trust.

2.1. Antecedents of Trust

Researchers have theorized trust in technology (Gille et al., 2020) and the formation of trust (Gillath et al., 2021; Jacovi et al., 2021), along with searching for ways to analyze trust specifically in human–AI

interactions (Ferrario et al., 2020). Some researchers have suggested that trust in AI should be considered separately from trust in more traditional technologies due to fundamental differences in these technologies (Bedué & Fritzsche, 2022). A literature review on the antecedents of trust in AI identified five trust challenges "unique to or exacerbated by AI": transparency and explainability, accuracy and reliability, automation, anthropomorphism, and mass data extraction (Lockey et al., 2021, pp. 5463–5467). Much of the technical research on trust also focuses on factors predicting trust. According to Li and Hahn (2022), research has explored the effects of human characteristics (e.g., age, gender, region, and socioeconomic factors), technological features (performance, transparency, reliability, etc.), and interactional aspects (building process, calibration, interaction quality, etc.; see also Bao et al., 2021; Kaplan et al., 2021). Other studies (e.g., Ferrario et al., 2020) categorize trust-affecting factors into specific characteristics at the human (e.g., personality), environmental (e.g., culture and structural assurances such as contracts), and technological levels (performance, process, and purpose; Siau, 2018). Trust in AI is often studied in relation to a specific use case such as chatbots in customer service (Zierau et al., 2020), AI in financial risk management (Fritz-Morgenthal et al., 2022), and AI in healthcare (Gille et al., 2020). Importantly, contextual factors affecting trust include which person or organization is deploying the technology and for what purpose. Citizen perspectives on AI have been found to be affected by the context and application of the technology (Crockett et al., 2022; Habib et al., 2020). In the context of AI in healthcare, Baldauf et al. (2020) found perceptions of AI trustworthiness to be significantly affected by institutional factors such as the associated healthcare institution or medical certification. The trust-affecting sociocultural context is also related to national institutions and differences in their cultural position (Chen et al., 2021; Julsrud & Krogstad, 2020; B. Zhang & Dafoe, 2020). An earlier study found that citizens' perceptions of shared values with people who are affiliated with an institution's services are instrumental (Warkentin et al., 2012). The relatively recent European AI Act regulation is intended to foster trustworthy AI, and a literature review of the critiques and challenges of the regulation paints a picture where citizens' trust in AI is actually trust that organizations and institutions using AI are trustworthy (Vainionpää et al., 2023). Robinson (2020) has examined how Nordic nations have positioned themselves in national policy strategies on AI and how they emphasize democracy, ethics, privacy, and autonomy to support trust in AI.

It should be noted that the institutional underpinnings of trust more generally have long been recognized in sociology. In the sociology of organizations, Zucker (1986) distinguishes between three modes of trust production: In addition to process-based trust built on past or expected exchange, Zucker focuses on characteristic-based trust, which is tied to the individuals involved, and institutional-based trust, which is connected to societal structures, institutions, and their intermediary extensions such as certification and regulation. More recently, Callon (2021) highlighted similar process-based, personal, and institutional mechanisms of trust in the context of markets: "Trust can come from repeated interactions the agents want to prolong"; it can be "guaranteed by belonging to a group whose members are known [to] each other, are interdependent, and share the same norms"; and it can be "encouraged by the existence of devices and ad hoc procedures (for example, warranties, labels, or an innovation such as blockchain, which logs transactions)" (p. 251). Studying people's perceptions of data-driven systems in the context of public service media, Steedman et al. (2020, pp. 823–824) find similar institutional underpinnings of trust when citizens consider trustworthiness in terms of the institutional and organizational sources of these data-driven services.

Technical researchers have often approached trust as a problem to solve and aligned with interests to support trust in and acceptance of AI. Trust in AI is viewed as a hurdle to successfully integrating AI into society (Li &

Hahn, 2022), and there is extensive research on trust requirements in AI adoption and how to promote trust in AI (Bangui & Buhnova, 2024; Bedué & Fritzsche, 2022; Ezer et al., 2019; Hasija & Esper, 2022; Knowles & Richards, 2021; Naiseh et al., 2021; Okamura & Yamada, 2020; Rebensky et al., 2021; Weitz et al., 2019; Zierau et al., 2020). Researchers have provided design principles for interaction design (Naiseh et al., 2021; Zierau et al., 2020) and on crafting design dimensions for the evaluation of explainable AI approaches (Sperrle et al., 2020). Transparency in and about AI decision making can affect how the public views its trustworthiness; according to these studies, decision makers should provide reasons for the goals and priorities chosen for AI, make them available to the public, and be clear about who is accountable for a decision and how to intervene (e.g., de Fine Licht & de Fine Licht, 2020). Additionally, certification, self-imposed standards, and guidelines are often highlighted as ways to increase trust in AI (Baldauf et al., 2020; Bedué & Fritzsche, 2022). In sum, this research on trust in AI paints a mechanistic and instrumentalized picture of trust as a function of specific features that can be increased or decreased. This literature can help define specific elements of people's trust constructs—to which we turn later in our empirical findings—but it is not especially useful for approaching the everyday phenomenological underpinnings of lived trust as something enacted and experienced in interactions with AI technologies. In addition, in its unreflexive approach, this research does not reveal a great deal about what trust is in the first place: how we should critically understand it, how it is constituted, how it is lived, and how we should look for and approach it.

2.2. Conceptions of Trust

The Oxford English Dictionary defines trust as a "firm belief in the reliability, truth, or ability of someone or something; confidence or faith in a person or thing, or in an attribute of a person or thing" (Trust, n.d.). This general description does not take any position on the constitution of this firm belief, but much of the literature on trust in the social sciences defines it as a cognitive construct (Cook & Santana, 2020, p. 191). To us, these definitions do not seem far removed in their activist and utilitarian orientation from how, for example, transactional interests and calculative evaluations typically appear in rational choice theory and in neoclassical economists' *Homo economicus* models of agency (Urbina & Ruiz-Villaverde, 2019). For example, Hardin (2002, pp. 3–7) describes trust as "encapsulated interests," as articulated by Cook and Santana (2020, p. 191): "A *trusts* B with respect to x when A believes that her interests are included in B's utility function, so that B values what A desires *because* B wants to maintain good relations with A." However, as Ramírez-i-Ollé (2019), drawing on Misztal (1996; see also Zucker, 1986, pp. 6–7), highlights, conceptualizations of trust vary greatly between traditions in social theory and at least three different approaches to trust can be distinguished: trust as a predisposition enabling a "collaborative order" between calculative individuals; trust as normatively based, enabling a "cohesive order" among functional systems in society; and trust as an implicit background assumption enabling a "stable order" achieved in and for interaction (Ramírez-i-Ollé, 2019, p. 2). It is broadly within these two latter traditions that we find more nuanced approaches to trust, as in Luhmann's (1988) elaborations on the relations of familiarity, confidence, and trust (see also Seligman, 2021), according to which "a rational explanation of other people's interests can hardly serve as an adequate substitute for trust" (Gambetta, 1988, p. xii), and especially in the practice-oriented approaches of the interactionists. This line of thought on trust draws on the symbolic interactionism of Erving Goffman and the ethnomethodology of Harold Garfinkel, rendering it better suited for following trust on the ground, in its lived natural variety as part of interactions within sociotechnical ecologies. The interactionist approach shares roots with our approach, introduced below, in which we draw inspiration from the early 20th-century American pragmatists.

Another highly valuable critique of rationalistic models of trust—one that arguably comes close to the interactionist approach—is provided in the (design) anthropology of Sarah Pink (2021, 2022a, 2022b; Pink et al., 2018; Quilty & Pink, 2024). In Pink's (2021, p. 193) work, trust is approached as a nonrepresentational, affective, sensory, and anticipatory phenomenon built on the ground, in and for everyday action, contributing to our sense of "what might be going to happen next." Pink proposes that "trust, like hope, can be thought of as a *feeling*, or category of feeling, which describes anticipatory sensations," and that "trust is experienced when things 'feel right'" (2021, pp. 194, 196). By contrast, "trust has been commonly conceptualized in organization studies, and in technology design disciplines such as human-computer-interaction research, as a rational, transactional, and interactional relationship between two entities" (Pink, 2021, p. 194). This narrow relational approach has also been challenged in the context of data-driven systems by Steedman et al. (2020), who highlight the "complex ecologies of trust" and the need for a way of "conceptualising trust in data-driven systems that accounts for complexity" (p. 828), because simplistic patterns and the human-technology interaction of trust-building and trust-breaking are difficult to find when approached qualitatively.

At the same time, as Pink willingly admits, "trust is notoriously difficult to define, and whenever it is defined, a competing definition is not far away" (2022a, p. 29). We agree and propose a step away from competing definitions. As important as we find Pink's suggestion that "to trust involves 'a sensory experience of feeling or disposition towards something' rather than an explicit cognitive decision made in relation to a specific technology" (2022b, p. 47)—and share her insistence on sensory experiences and dispositions—we propose looking at things slightly differently: through a framework that moves away from the rather-than and toward a pluralistic, synthetic approach combining various aspects and constituents of trust under a single pragmatic framework that highlights the processual nature of trust-affording action. This shift has another advantage over the aforementioned approaches in diversifying our conceptual vocabulary of trust; because these approaches each highlight a specific character of trust, sometimes as a corrective to earlier models, they do not show how these different elements—cognitive, affective, sensory, or otherwise—might be connected and related. In this regard, and based on our empirical findings, we propose an extension of these approaches as well. We now elaborate on our approach and its foundations in pragmatist action theory.

2.3. Trust-Affording Action: A Processual Interpretation of Trust and Its Varieties

We approach trust-affording action in the context of emerging algorithmic technologies in two ways: By examining trust as it emerges from everyday actions with these technologies, and by analyzing specific technologically entangled everyday actions as they emerge from trust. In short, we study action as affording trust, and trust as affording action. This is our action-based proposal for a synthetic approach that integrates the cognitive, affective, and habitual components of trust within a unified framework of processual action. We deliberately connect to Gibson's (1979) work by using the concept of affordances to highlight the functions of trust for action, treating trust as a form of action possibility. For Gibson (1979, p. 127), famously, "the affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill." Later research has reworked the concept to include mental affordances (McClelland, 2020), under which aspects like trust could be placed—though we do not confine trust merely to the "mental," but also acknowledge its embeddedness in the behavioral dispositions of the body.

Trust as an affordance for action becomes more active and deliberate during moments of uncertainty. In Callon's (2021) formulation, "the importance of trust lies in the fact that it enables actors to make decisions in a complex and uncertain universe" and thus continue to act:

> It is not only a mental and affective disposition that influences behavior. It is also a form of knowledge and a way of grasping reality. In the face of unknown risks, the rational behavior might be to abstain from acting. If, however, I trust an agent or an institution offering to collaborate or transmit information, then I can decide to commit to a course of action that I might otherwise have hesitated to undertake. (p. 252)

This grounding of trust in conditions of uncertainty is also highlighted by Steedman et al. (2020, p. 817): "It has been argued that trust is crucial for dealing with uncertain, uncontrollable or risky situations," a "description that could be applied to the data practices (that is, organisations collecting, analysing and sharing data and the outcomes of these processes) that characterise contemporary digital life." This certainly applies to the fundamentally uncertain AI practices entrenched by the current sweeping enthusiasm to apply generative AI in services across society.

Dealing with uncertainty in action was a key concern for the early 20th-century American pragmatist philosophers William James, Charles S. Peirce, John Dewey, and George Herbert Mead. Peirce's (1877) canonical *The Fixation of Belief* introduced the doubt–belief model of inquiry, highlighting the dynamic of belief formation to replace the discomfort of doubt as generating action. For the classical pragmatists, uncertainty was the condition of living processes in general, and the "quest for certainty" (Dewey, 1929) was what motivated action, especially creative action. Certainty, or belief, in its pragmatist conception, is a habit, and according to Peirce, "the world is characterized by 'the tendency of all things to take habits'" (as cited in Kohn, 2013, p. 59). In what follows, we propose that this is how we should approach trust: as certainty aspiring to take the form of habit. Dewey (1920, 1929), for his part, insisted that philosophy and social thought desperately needed a re-grounding in concrete and practical processes of life from which, and for which, all kinds of complex and abstract human phenomena (like reflective and deliberative thought) functionally emerge (Dewey, 1922). Grounding thought in action situates abstract ideas like generalized ends and values—and similarly, we suggest, trust—as dispositional "instrumentalities in judgment of particular cases as the latter arise; they are, in effect, tools that direct and facilitate examination of things in the concrete while they are also developed and tested by the results of their application in these cases" (Dewey, 1939, p. 44).

This general pragmatist approach to action and thought was later reintroduced in the context of social theory by Hans Joas (1996) in his foundational book *The Creativity of Action*, which focuses on uncertain processes of action where crises provoked by changes in the environment spark situated creative action to establish new habits better suited to the ever-evolving social and material conditions of action (see also Joas & Kilpinen, 2006, p. 326; Kilpinen, 2000, pp. 24, 37–39, 58–60). This kind of creativity is not free-floating but always situated and closely tied to concrete processes of action. More recently, in the context of technology studies, similar approaches with explicit reference to classical pragmatism can be found in Steven Jackson's (2014) influential work on "broken world thinking," which asserts that "breakdown, dissolution, and change, rather than innovation, development, or design as conventionally practiced and thought about are the key themes and problems facing new media and technology scholarship today" (p. 222) and that with the ever-present possibility for breakdown "comes a second and more hopeful approach":

> [N]amely, a deep wonder and appreciation for the ongoing activities by which stability (such as it is) is maintained, the subtle acts of repair by which rich and robust lives are sustained against the weight of centrifugal odds, and how sociotechnical forms and infrastructures, large and small, get not only broken but *restored*. (p. 222)

Similarly, Ananny (2022, 2024), with reference to Dewey, has proposed studying the social entanglement of technologies through errors and uncertainties, focusing on algorithmic mistakes as avenues for unveiling and deliberating on public problems related to the entry of technologies into social spheres.

This approach of emanating inquiry from the dynamics of action provides a fruitful lens through which to view trust. Below, we approach the fractures and uncertainties of the social life of technologies through citizens' perspectives on, and practices of, making sense of and coming to terms with various kinds of AI technologies entering their everyday lives. With this action-based framing, trust can be seen in its "normal" state as quietly (but very actively) embedded and embodied in the agent's generative habits of action. By connecting trust with habit, we join a long history of sociological and economic theorizing utilizing the idea of socially embedded dispositional regularities (Camic, 1986; Gronow, 2008; Hodgson, 2004), made perhaps most famous by Pierre Bourdieu's theory of the *habitus*. For Bourdieu (2005), habitus is "a system of dispositions, that is, of permanent manners of being, seeing, acting, and thinking, or a system of long-lasting (rather than permanent) schemes or schemata or structures of perception, conception, and action" (p. 43):

> [S]ystems of durable, transposable dispositions, structured structures predisposed to function as structuring structures, that is, as principles which generate and organize practices and representations that can be objectively adapted to their outcomes without presupposing a conscious aiming at ends or an express mastery of the operations necessary in order to attain them. (Bourdieu, 1990, p. 53)

We see trust as settling into these kinds of systems of organized dispositional practices that do not require conscious reflection or articulation. Trust remains in this dispositional, quiet form embedded in practice unless errors or uncertainties of action call for reflection on it. Reflexive, critical, and creative capacities for deliberating on trust and trustworthiness are activated not by chance but by specific troubling and uncertain moments in action, when things go wrong or that possibility is recognized by the agent, when action and trust as its disposition runs into a crisis, or when something new (e.g., a new algorithmic technology or its application) emerges in people's lives that needs to be made sense of because there is no guidance from our previous habits and the dispositional trust they embed. It is in these moments that habit and its encapsulated trust can be reflexively recognized, inspected, possibly questioned, and sometimes altered. Once made sense of, creatively reworked, and established, trust once again settles back into its quiet, dispositional state. It is important to note that this kind of habitual trust is not blind faith but is always open to reflexive interrogation and creative change under uncertainty; habits are reflexive and living entities (Kilpinen, 2009). While avoiding rigid dichotomies, another point to highlight is that reflexive and deliberative trust-building necessarily relies on habitual modes of inference or heuristics of deliberation—something the ecological psychologist Gigerenzer (2008, p. 22) aptly calls "rationality for mortals," defined as "a strategy, conscious or unconscious, that searches for minimal information and consists of building blocks that exploit evolved capacities and environmental structures"—as will become clear when we turn below to specific tactics our interview participants use in their reflexive trust-building.

3. Empirical Approach: Data and Methods

Our empirical analysis is based on a trust-focused reading of qualitative data collected between 2020 and 2023, consisting primarily of participant diaries, semi-structured interviews, and three rounds of follow-up interviews. All interviews were conducted remotely using video conferencing software. In August 2020, the first author oversaw the recruitment of ten Finnish citizens to follow their perceptions of and practices with the Finnish Covid-19 contact tracing application, Koronavilkku. We commissioned a data collection agency to handle participant recruitment using a screening tool designed to ensure diversity in age (21–66), gender, geography (urban and rural), education, occupation, and household structure. We also recruited participants to reflect differences in attitudes toward digital technologies, perceived technical skills, and views on health and Covid-19. Participants first kept a pre-interview diary capturing daily observations, impressions, and reactions related to the new app. This was followed by a one-on-one interview (60–90 minutes) in which they reviewed diary entries and discussed their sensemaking of the new algorithmic technology entering their lives. These materials were supplemented by media coverage during Koronavilkku's launch, particularly around trust-building and public justifications by authorities and experts. After six months, in spring 2021, the same participants completed a second diary and interview round, now addressing broader encounters with AI-infused services. These two rounds and the initial thematic analysis were led by the first author with two other researchers (not listed as co-authors; see Acknowledgments Section). In the final two rounds, held a year apart, most participants continued, and five new ones were added based on case-relevant profiles. In spring 2022, the interviews focused on a customer-facing, AI-based banking service seen by participants as straddling public and private sectors due to its regulatory context. This round included eight returning participants and five new ones recruited with the help of a Finnish bank, which provided contacts from its customer base using our screener (excluding health attitudes). The final round (spring 2023) centered on a fictional AI use case in the provision of social assistance by Kela, Finland's Social Insurance Institution, and involved five returning participants with prior experience using Kela's services. By this point, data saturation was becoming obvious: participants' trust-related framings, practices, and heuristics were recurring across cases to such an extent that we decided to discontinue further interviews. The final two rounds and their thematic analysis were conducted by the two authors with the help of a research assistant who transcribed the interviews and assisted in early analysis. All participants gave informed consent, following the ethical guidelines of participating organizations and the research project.

Thematic analysis across rounds was guided by two RQs:

> RQ1: How does trust in algorithmic technologies emerge in citizens' everyday encounters with these technologies?
>
> RQ2: How is trust questioned and constructed in the process?

Each round of analysis was conducted immediately after data collection and later revisited by the first author, then developed jointly with the second author in 2023–2024. Our analysis began with identifying patterns in participants' trust-evaluating accounts and gradually organizing them into heuristics through which trust was actively constructed. We then developed a thematic distinction between different modes of trust to better capture its more latent forms that emerged across our data. Finally, we analyzed these modes in relation to existing conceptualizations of trust in the literature, situating them within a synthetic framework that allowed

for their integration under a unified conception of trust-affording action. The final round of analysis was a theory-informed thematic analysis, drawing on the conceptual framework presented above in Section 2.3, which was itself motivated by and developed in response to the limitations we encountered when applying existing trust concepts to our data.

4. Empirical Findings: Trust in Action

We present our findings by connecting the empirical variety of trust in citizens' action processes with algorithmic technologies. We begin by discussing the challenges our participants face in reflecting on their relationships with, and trust in, algorithmically infused tools and public services that are embedded in their everyday lives. Trust, once established, often seemed to escape the discursive. We then examine instances of reflective trust and heuristic tactics of trust-building in moments of uncertainty.

4.1. Ambiguity of Trust

Experiences and manifestations of trust were often difficult to situate, verbalize, or reflect upon, even as a sense of trust appeared to guide participants' choices and actions—quietly operating as habitual dispositions that afforded certain paths and disrupted others. More generally, we repeatedly found our participants struggling to reflect on and verbalize the role of AI and related risks in their everyday lives, especially during the early interview rounds conducted between autumn 2020 and spring 2022 (before the release of ChatGPT in November 2022). It seemed easier for participants to avoid these questions by steering the discussion toward high-level reflections on, e.g., the importance of technological innovation for a thriving Finnish economy, or more generally, on "efficiency and removing the risk of human errors." Closer to mundane, everyday encounters with AI technologies, participants often expressed technological goodwill, highlighting personal curiosity about new AI tools, being "very positive about technology," and willing "to learn new things" (as exemplified here by a financial administration professional in her late forties). Most participants initially found it difficult to spontaneously give examples of AI in their everyday lives; the technology seemed to be experienced as everywhere and nowhere at once, diffused in digital environments and difficult to detect. Upon prompting, AI was often described as highly advanced software, automation, algorithms, and robots. As concrete examples of AI applications, participants frequently mentioned social media platforms and their algorithms, Google search, chatbots, mobile applications, and smart devices such as smart TVs.

Our questions regarding trust in AI were often initially met with varying degrees of confusion. While understood as legitimate within the interview context, participants often found it difficult to reflect on mundane practices involving technologies that were inconspicuously scattered throughout their material environments. Participants grappled with expressing, or even forming, opinions about their trust or distrust in AI, as illustrated by a retired teacher in her late sixties, who explained her hesitation:

> You don't notice it in your everyday….When things go well, you don't need to think about these things….Pretty much every process is digitalized; it is very mundane. Information technology is assisting in everyday life, and when it is well built, you don't even notice it.

Here, engineered interactional fluency, combined with the habitual and dispositional nature of trust, made it difficult to reflect or verbalize. A theology student in his mid-twenties, by contrast, explained that certain institutions in society are so fundamentally based on trust that even their algorithmically mediated services are, by association, granted that same trust—unless something triggers concern and disrupts this quiet confidence (cf. Luhmann, 1988). An electrician, in his early forties, noted the seeming inconsistency with which people extend this inherent trust across different algorithmic services and devices: "With my workmate, we were discussing how some people worry about being surveilled by this [Finnish Covid-19 contact tracing application], but these same people have a smartphone in their pocket."

However, in this context, the Covid-19 contact tracing application was something entirely new—introduced abruptly and with global alarm—and as such, it called for critical reflection and sense-making. In contrast, people's relationships with their smartphones had formed over years and decades of gradually intensifying data extraction and carefully engineered behavioral monitoring (Zuboff, 2019). One participant, reflecting on her own confidence in AI technologies, concluded: "Unless one feels unsafe, there is no reason to worry," evoking Pink's (2021, pp. 194–196) view of trust as anticipatory sensation—or, perhaps more relevantly here, as the absence of anticipatory unease.

Participants also recalled moments of unease or friction in which their quiet confidence fractured and the question of trust surfaced. These moments of uncertainty were often tied to suspicions about the data practices behind AI-infused tools and services, echoing findings by Steedman et al. (2020, p. 825), whose participants "felt insecure, but the precise nature of their insecurities was hard to express" (see also Pink et al., 2018). Sometimes, these fractures prompted deeper reflection, as in the case of a 33-year-old single mother whose child played the augmented reality game Pokémon Go. A sudden change in the app's behavior led to her critical questioning:

> It [suddenly] was focusing many times. Like when you open the camera and it looks like what you would see through the camera…and then it appears as if the Pokémon would be there, and when you take the picture, it looks like the Pokémon is in your home, for example. So, several times, especially on [the child's] phone, like, what it did was that it focused more on the environment than on the Pokémon….Well, I had the feeling of…like what the fuck is this now! Wasn't this [game] the thing that I thought it was….Yeah, and then when I don't even know where the pictures are going, what are they going to take? Do they go to Niantic [the company] and what are they…like…taking? I don't know.
>
> It just seems weird, because yes, the picture…it goes somewhere, when you take it with the game's software…you take it in the game—so does the picture go somewhere? I don't know. So, then [the question] is—what's there in the background, so to speak?

These experiences fractured her prior dispositions and directly impacted how the now low-trust software was used. She described adopting countermeasures to resist data surveillance:

> Then I just said that I'll have it [under observation] in the future, and if you have some kind of tasks that you have to do [in the game], then you just put it against a white wall, or you let me do it.

Participants' critical reflections on trustworthiness were also provoked by entirely new algorithmic technologies entering their lives—technologies for which they lacked established understandings and

habitual orientations. One such case was the launch of Koronavilkku, the Covid-19 contact tracing application, introduced at the very start of our data collection in autumn 2020. Others included the widespread emergence of generative AI tools like ChatGPT during 2022 and 2023, leaving some participants wondering whether society was placing too much trust in new, impressive, but still uncertain technologies. These concerns were often triggered by a sense that AI was beginning to encroach upon domains of personal autonomy and agency (Savolainen & Ruckenstein, 2022). As a law student in his early twenties put it: "I want to use it. I don't want it to use me."

4.2. Action Under Uncertainty: Situated Tactics of Trust Building

Next, we examine the reflective and creative tactics our participants used in establishing or withdrawing trust during uncertain phases in human-AI interactions. We present four recurring and intertwined trust heuristics that participants employed in moments when trustworthiness was in question: (a) the agent heuristic, (b) the case heuristic, (c) the social heuristic, and (d) the interaction heuristic. These include foundational heuristics that contribute to the construction of trust by inferring qualities of the technology itself, as well as supporting heuristics that define conditions, parameters, and case-specific standards of trustworthiness within particular contexts. We explore these heuristics in action by analyzing the adoption of the Finnish Covid-19 contact tracing application Koronavilkku, drawing from participant diaries, interviews, and media coverage collected during the application's release and dissemination. The development of the application, published in August 2020, was led by the Finnish Institute for Health and Welfare in collaboration with the Ministry of Social Affairs and Health, and its design and technology development were done in collaboration with private technology consultancies. The case provides a unique lens for examining trust, as Koronavilkku became one of the world's most widely adopted Covid-19 apps relative to population size. We further complement this analysis with later interview data, including citizens' reflections on fictional cases of AI-infused services at Finland's social insurance institution, Kela. The heuristics are not presented as exhaustive or definitive but rather as illustrative examples of the kind of cognitive deliberation and creative trust-building that emerge through action under uncertainty. Figure 1 presents these heuristics as situated within the broader process of trust-affording action.

The agent heuristic (a) was clearly the most prominent. Here, our participants' trust-evaluating focus was not on the algorithmic application itself, but the agent responsible for deploying it. This was trust by association: Institutions' or individuals' perceived trustworthiness was inferred to extend to associated services and their technologies (Cook & Santana, 2020, pp. 197–199; Steedman et al., 2020, pp. 823–824; Zucker, 1986). When the Koronavilkku application was released at the end of August 2020, the Finnish Broadcasting Company (YLE) aired the event, during which leading public authorities from the Finnish Institute for Health and Welfare and the Ministry of Social Affairs and Health took to the stage to introduce the app. This was quickly followed by widespread activity in mainstream and social media, where public authorities and officials were joined by politicians, journalists, technologists, cybersecurity experts, privacy activists, researchers, and lay citizens to encourage the downloading of the application. This diverse constellation of agents—representing a spectrum of institutional roles and interests—collectively vouched for the app's adoption (this broad and diverse public deliberation process is also closely tied to the social heuristic). This discussion was not missed by our research participants, who in interviews referred to testimonials from various agents and highlighted the application's close association with public authorities.

A similar agent heuristic was repeatedly observed during our subsequent interview rounds, for example, when participants deliberated on the trustworthiness of AI used by Kela, the Finnish Social Insurance Institution. When presented with various possible uses of AI in its citizen-facing services, a 34-year-old animal care worker said: "Well, Kela is owned by the government, so maybe if it [AI] is used there, it should be trustworthy in principle." Another participant, a graphic designer in her early forties, said how she "would trust Kela using AI more than, say, a newly established business, like a private company." When we met her a year later, she added:

> I do think that public services like these, and agents like these—their [AI applications] are tested more thoroughly and thought through more thoroughly, like, they must be; they maybe have a different responsibility compared to, say, the world of business.

The electrician in his early forties, who often expressed skepticism about AI and criticized the hype surrounding it, explained: "I trust the state more, in the sense that I believe they are more precise in how they process information, and it's more strictly defined where they are used." Stronger expressions of the agent heuristic were not limited to comparisons between public and private spheres (or to recognition of diverse social roles and identities, as seen in the testimonials above); participants expanded this relational thinking to a broader sociopolitical and geographical scale, associating varying interests and governance orientations with the perceived trustworthiness of AI systems in Europe ("public interest," "regulation"), the United States ("uncontrolled capitalism"), and China ("state surveillance," "control").

The case heuristic (b) and social heuristic (c) were strongly present in participants' deliberations around the launch of Koronavilkku. Here, the focus was on situational factors—such as the acuteness and severity of the pandemic—the pressing need for safety and hope, and framings of the pandemic as a shared social condition. One participant, writing about his elderly parents in his research diary, explained that "they said [Koronavilkku] is a major relief for them," and added that he felt "Koronavilkku, on its part, is about caring" and brings "a special sense of relief" for him. It was a device of hope, "something that unites us all, a safety net of sorts, something with which to work our way through these difficult times." Such hopeful framings were common immediately after the application's release, although disappointment and critical voices began surfacing more prominently over time. These heuristics, operating through the specific case and its social context, illustrate how participants defined conditions and case-specific standards of trustworthiness. They also reflect a kind of normative pressure ("If only everyone would use Koronavilkku, this could have been avoided") that fueled swift adoption rather than critical scrutiny. An important element of the social heuristic was also the broader social deliberation and testimonials among diverse and trusted agents, roles, and social positions (as noted earlier, these heuristics often appeared intertwined), not only through the media but also among neighbors and within close communities. This is exemplified by a law student in his early twenties:

> Outside, talking with my neighbor, while tamping some rugs, we were chatting this and that, then the Coronavirus and Koronavilkku came up, because we both had heard about it in the news….I was a bit skeptical about its functionality, but this discussion [with the neighbor] got me to finally register for the app.

> I was listening to a program on YLE Puhe [national radio]…the experts were talking about the Coronavirus and how we might pull through and how Koronavilkku could help us get over the disease.

Finally, the interaction heuristic (d) refers to a trust-establishing tactic that operates by negation. It is characterized by human-machine interaction without critical awareness, and a lack of trust-reflection by design: Sleek and frictionless usability and smooth functionality of the algorithmic service or device that supports the habitualization of trust. This heuristic was previously illustrated by the retired teacher, who explained: "When things go well, you don't need to think about these things," and "when it is well built, you don't even notice it." In the case of Koronavilkku, participants referred to its smooth usability, the ease of setup, and how this frictionless experience let it "sink" into the background. Others described specific glitches and errors that forced the application back into reflective awareness, sometimes prompting them to reconsider its trustworthiness more broadly.

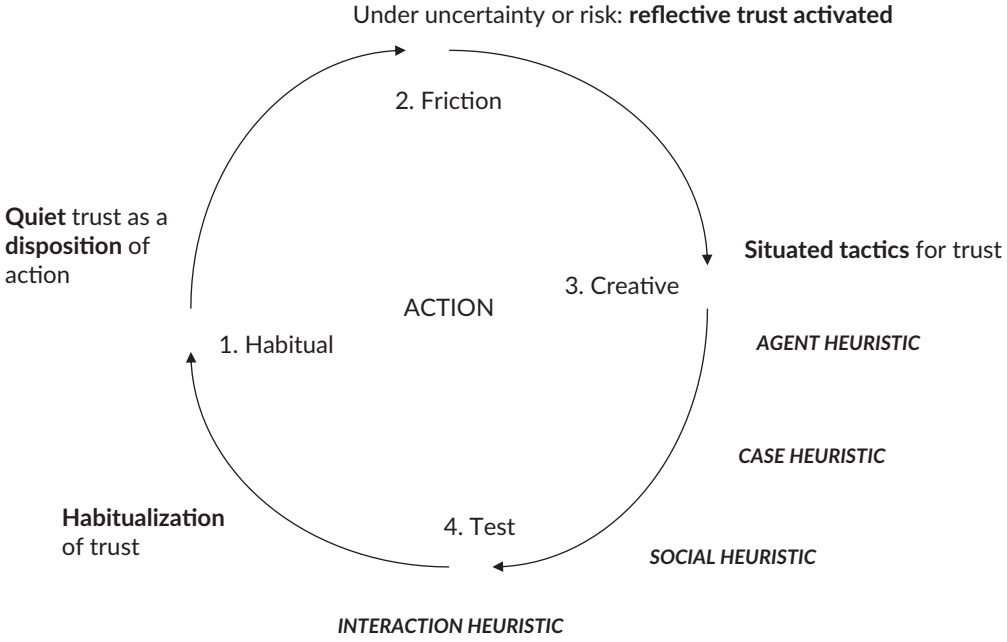

Figure 1. Process view: Dispositional and deliberative trust as phases in action and heuristic tactics of trust formation.

5. Discussion and Conclusion

As high-trust Nordic institutions increasingly embrace intriguing yet uncertain algorithmic technologies, such as generative AI, for more impactful and efficient public services, it is becoming increasingly important to frame questions of sustainable social order as questions of sustainable sociotechnical order, where the scope of trust, as a key mechanism of that order, extends beyond humans to include institutions' material presence through these technologies. Many traditional approaches to trust fail us here, as they approach trust by focusing too narrowly on only some of its aspects, making trust slip out of focus once it exceeds the boundaries of their chosen categories. In computer science and ICT research, trust is often approached in a non-reflexive and uncritical manner by mechanistically measuring and instrumentally tweaking it as a problem to be solved for the purpose of technology acceptance. The various available sociological approaches to trust and its role in social order—such as those based on the concepts of "calculative individuals," "functional systems," or "interaction"—all highlight important aspects of trust, but simultaneously downplay others, leaving the scope of trust-tracing empirical work rather limited. Pink's (e.g., Pink, 2021; Pink et al., 2018) anthropological reconceptualization of trust shows trust as an anticipatory emotion and

points to an action-oriented and dialectical approach. However, to our understanding, it does not extend its dialectics to those between emotive and cognitive trust processes.

The theory of trust-affording action we outlined in this article aims to overcome persistent dichotomies in trust research by showing the connections between cognitive trust-building and the bodily inscription of trust in and for action. We presented a complementary approach to trust that moves from embodied dispositions to cognitive decisions and back, highlighting how trust emerges through situated, everyday encounters with algorithmic technologies. This addresses RQ1 regarding how trust emerges in citizens' everyday encounters with algorithmic technologies. Trust manifests in multiple modes, and these modes are interconnected. Our theory of trust-affording action captures this dynamic variety by placing trust within a pragmatist framework of habit formation and creative adaptation. We understand trust as a constitutive element of both habit-making and habit-breaking: embedded in established patterns of action yet activated and reshaped in moments of disruption or doubt. This processual understanding also enables us to examine how trust is called into question and reconstructed in practice. Methodologically, our approach foregrounds the importance of studying trust as it is enacted in the flow of action—often quietly embedded in routine practices, but sometimes consciously reflected upon.

Our analysis highlighted the habitual underpinnings of trust in both action and reflection, the latter illustrated through four recurring heuristics that participants used to evaluate trustworthiness or that influenced its evaluation: the agent heuristic, the case heuristic, the social heuristic, and the interaction heuristic. This speaks directly to RQ2, concerning how trust is constructed in the process of everyday encounters. Our framework and analysis help illuminate both the emergence of trust in citizens' mundane engagements with algorithmic systems and the moments when trust becomes a site of reflection, judgment, and reconfiguration. Our findings here resonate with prior research that emphasizes the complex social conditions and institutional underpinnings of trust, particularly when citizens evaluate trustworthiness based on institutional and organizational sources (e.g., Steedman et al., 2020; Zucker, 1986). Yet, we emphasize that such active, deliberative, and reflective forms of trust represent only part of the picture. Our analysis underscores that trust is not merely a matter of explicit evaluation but is grounded in embodied, habitual practices that often remain unspoken yet are vital to how citizens navigate algorithmic systems in everyday life.

Building on this analysis, we outline policy considerations for fostering trust in algorithmic public services. Rather than approaching trust instrumentally as something to be engineered for technology acceptance, we argue for the importance of creating conditions under which experienced and enacted trust can take root, or be critically withdrawn. In this sense, designing for trustworthy systems means not engaging in trust engineering, but making space for trust to form, evolve, or dissolve. Among the heuristics we identified, the social heuristic stands out as particularly significant. It underscores the need for infrastructures and practices that support the social deliberation of algorithmic systems—that is, spaces in which trust can be surfaced and reflexively examined from various social positions. The agent heuristic plays a central role in this dynamic. Trust is often inferred from the perceived trustworthiness of institutions, groups, and individuals. Enabling opportunities for diverse agents to engage in public reflection on new technologies and their application—across roles, domains, and experiences—is therefore central. The case heuristic further reminds us of the importance of attending to the very situated and shifting demands of trust, as shaped by the varying contexts in which these technologies are applied. Trust is not simply embedded in technological systems once and for all, but is continually shaped by the social and institutional contexts in which it emerges.

A recent extensive survey on AI adoption in the European public sector highlights that "AI adoption is no longer a promise; it is a reality, in particular in service delivery and internal operations" (Grimmelikhuijsen & Tangi, 2024, p. 26). According to the same report, a majority of public organizations in Europe are either planning one or more AI projects (63.1%) or already have projects fully adopted and in use (51.8%). In Finland, the Social Insurance Institution Kela has set a "future target level of automation" at "70–80 percent" in its strategic vision (Kela, 2024). Yet as AI technologies have repeatedly failed to meet the exaggerated expectations placed on them (Mitchell, 2021), it is worth recalling the OECD's (2021, p. 21) reminder on the nature of trust: "Trust is…a fragile societal asset; while it takes time to establish, it can be lost quickly."

Our research contributes to a more nuanced understanding of trust and its situated dynamics in human-technology interaction, approached from the lived perspective of citizens. We consider this perspective crucial at a time of unprecedented excitement around the transformation of welfare societies through algorithmic technologies and a discourse that operates overwhelmingly in technology-first terms. Despite the growing role of AI in public services, algorithmic governance has not yet sparked widespread public reflection in Finland. The situation is markedly different in some other northern European countries. In the Netherlands, the exposure of biased algorithmic systems—most notably in the welfare fraud detection system SyRI (*Systeem Risico Indicatie*) case (Rachovitsa & Johann, 2022) and the childcare benefits scandal (Peeters & Widlak, 2023)—has turned automated decision-making into a high-stakes political and legal issue, illustrating what Ananny (2022, p. 346) describes as "algorithmic breakdowns as public problems" rather than "idiosyncratic quirk[s] requiring private troubleshooting." Public reckonings are now starting to surface in Sweden and Denmark, where recent investigations by Lighthouse Reports ("Sweden's suspicion machine," 2024) and Amnesty International ("Coded injustice," 2024) have raised concerns about discriminatory AI systems in welfare provision (see also Kaun, 2021). These developments underscore that algorithmic systems do not travel uniformly across social contexts. Their deployment, contestation, and the trust they command are always locally mediated. In our empirical material, the trust heuristics were mobilized in ways that reflect the particularities of the Finnish context and the specific encounters reported by our participants. Further research is needed to examine how trust is managed in other settings, especially where algorithmic systems have already triggered public controversy or resistance.

Acknowledgments

We gratefully acknowledge Anni Ojajärvi and Topias Tuomisto, colleagues of the first author at Solita, for their valuable contributions to data collection and analysis in the first two research rounds (autumn 2020 and spring 2021). We further acknowledge research assistant Teemu Saravirta, who transcribed the later interview rounds and contributed to the early analysis.

Funding

This work was supported by the Research Council of Finland through funding for the projects Civic Agency in AI? [357349] and GenZ [318930], and by Business Finland through funding for the project Artificial Intelligence Governance and Auditing.

Conflict of Interests

Antti Rannisto is a part-time employee at Solita, a technology consultancy that was closely involved in the design and development of the Covid-19 contact tracing application Koronavilkku. Although he was employed by Solita at the time, he had no involvement in the Koronavilkku project and no direct connection

to the development teams. His role at Solita was strictly tied to research activities conducted under the Artificial Intelligence Governance and Auditing (AIGA) project, a multi-year initiative (2020–2022) in which Solita participated as a consortium partner. The project was funded by Business Finland and the participating consortium organisations. His employment at Solita provided access to fieldwork opportunities relevant to the research, which influenced the decision to include this case within the scope of the AIGA project. Fanny Vainionpää declares no conflicts of interest.

References

Ananny, M. (2022). Seeing like an algorithmic error: What are algorithmic mistakes, why do they matter, how might they be public problems? *Yale Journal of Law and Technology*, *24*(1), 342–364. https://yjolt.org/sites/default/files/0_-_ananny_-_seeing_like_an_algorithmic_error.pdf

Ananny, M. (2024). Making generative artificial intelligence a public problem: Seeing publics and sociotechnical problem-making in three scenes of AI failure. *Javnost—The Public*, *31*(1), 89–105. https://doi.org/10.1080/13183222.2024.2319000

Baldauf, M., Fröehlich, P., & Endl, R. (2020). Trust me, I'm a doctor—User perceptions of AI-driven apps for mobile health diagnosis. In J. Cauchard & M. Löchtefeld (Eds.), *MUM '20: Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia* (pp. 167–178). Association for Computing Machinery. https://doi.org/10.1145/3428361.3428362

Bangui, H., & Buhnova, B. (2024). Towards anthropomorphic trust management for digital society. In *2024 IEEE 21st International Conference on Software Architecture Companion (ICSA-C)* (pp. 87–91). IEEE. https://doi.org/10.1109/ICSA-C63560.2024.00022

Bao, Y., Cheng, X., de Vreede, T., & de Vreede, G.-J. (2021). Investigating the relationship between AI and trust in human-AI collaboration. In T. X. Bui (Ed.), *Proceedings of the 54th Hawaii International Conference on System Sciences* (pp. 607–616). Association for Information Systems. https://hdl.handle.net/10125/70685

Bedué, P., & Fritzsche, A. (2022). Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption. *Journal of Enterprise Information Management*, *35*(2), 530–549. https://doi.org/10.1108/JEIM-06-2020-0233

Bourdieu, P. (1990). *The logic of practice*. Polity Press.

Bourdieu, P. (2005). Habitus. In E. Rooksby & J. Hiller (Eds), *Habitus: A sense of place* (2nd ed., pp. 43–52). Routledge. https://doi.org/10.4324/9781315253701

Callon, M. (2021). *Markets in the making: Rethinking competition, goods, and innovation*. Zone Books.

Camic, C. (1986). The matter of habit. *American Journal of Sociology*, *91*(5), 1039–1087. https://doi.org/10.1086/228386

Chen, T., Guo, W., Gao, X., & Liang, Z. (2021). AI-based self-service technology in public service delivery: User experience and influencing factors. *Government Information Quarterly*, *38*(4), Article 101520. https://doi.org/10.1016/j.giq.2020.101520

Coded injustice: Surveillance and discrimination in Denmark's automated welfare state. (2024, November 12). *Amnesty International*. https://www.amnesty.org/en/latest/news/2024/11/denmark-ai-powered-welfare-system-fuels-mass-surveillance-and-risks-discriminating-against-marginalized-groups-report

Cook, K., & Santana, J. (2020). Trust: Perspectives in sociology. In J. Simon (Ed.), *The Routledge handbook of trust and philosophy* (1st ed., pp. 189–204). Routledge. https://doi.org/10.4324/9781315542294

Crockett, K., Colyer, E., & Latham, A. (2022). The ethical landscape of data and artificial intelligence: Citizen perspectives. In K. Crockett & S. Mostaghim (Eds.), *2021 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1–9). IEEE. https://doi.org/10.1109/ssci50451.2021.9660153

de Fine Licht, K., & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. *AI and Society*, *35*(4), 917–926. https://doi.org/10.1007/s00146-020-00960-w

Dewey, J. (1920). *Reconstruction in philosophy*. H. Holt and Company.

Dewey, J. (1922). *Human nature and conduct: An introduction to social psychology*. H. Holt and Company.

Dewey, J. (1929). *The quest for certainty: A study of the relation of knowledge and action*. Putnam.

Dewey, J. (1939). *Theory of valuation*. In O. Neurath (Ed.), *International encyclopedia of unified science* (Vol. 2, No. 4, pp. 1–66). University of Chicago Press.

Digital and Population Data Services Agency. (2023, September 21). *Digiturvabarometri: Suomalaisista vain 15 % luottaa tekoälypalveluihin: "Digiturvakoulutus kiinnostaa yhä useampia"* [Press release]. https://dvv.fi/-/digiturvabarometri-suomalaisista-vain-15-prosenttia-luottaa-tekoalypalveluihin-digiturvakoulutus-kiinnostaa-yha-useampia

Ezer, N., Bruni, S., Cai, Y., Hepenstal, S. J., Miller, C. A., & Schmorrow, D. D. (2019). Trust engineering for human-AI teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 322–326. https://doi.org/10.1177/1071181319631264

Ferrario, A., Loi, M., & Viganò, E. (2020). In AI we trust incrementally: A multi-layer model of trust to analyze human-artificial intelligence interactions. *Philosophy and Technology*, *33*(3), 523–539. https://doi.org/10.1007/s13347-019-00378-3

Fritz-Morgenthal, S., Hein, B., & Papenbrock, J. (2022). Financial risk management and explainable, trustworthy, responsible AI. *Frontiers in Artificial Intelligence*, *5*, Article 779799. https://doi.org/10.3389/frai.2022.779799

Gambetta, D. (1988). Foreword. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. ix–xii). Basil Blackwell. https://www.nuffield.ox.ac.uk/users/gambetta/Trust_making%20and%20breaking%20cooperative%20relations.pdf

Gerlich, M. (2024). Public anxieties about AI: Implications for corporate strategy and societal Impact. *Administrative Sciences*, *14*(11), Article 288. https://doi.org/10.3390/admsci14110288

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton, Mifflin and Company.

Gigerenzer, G. (2008). *Rationality for mortals: How people cope with uncertainty*. Oxford University Press.

Gillath, O., Ai, T., Branicky, M., Keshmiri, S., Davison, R., & Spaulding, R. (2021). Attachment and trust in artificial intelligence. *Computers in Human Behavior*, *115*, Article 106607. https://doi.org/10.1016/j.chb.2020.106607

Gille, F., Jobin, A., & Ienca, M. (2020). What we talk about when we talk about trust: Theory of trust for AI in healthcare. *Intelligence-Based Medicine*, *1/2*, Article 100001. https://doi.org/10.1016/j.ibmed.2020.100001

Government of Finland. (2023). *A strong and committed Finland: Programme of Prime Minister Petteri Orpo's government*. Publications of The Finnish Government. https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/165044/Programme-of-Prime-Minister-Petteri-Orpos-Government-20062023.pdf?sequence=4

Grimmelikhuijsen, S., & Tangi, L. (2024). *What factors influence perceived artificial intelligence adoption by public managers?* Publications Office of the European Union. https://doi.org/10.2760/0179285

Gronow, A. (2008). Not by rules or choice alone: A pragmatist critique of institution theories in economics and sociology. *Journal of Institutional Economics*, *4*(3), 351–373. http://dx.doi.org/10.1017/S1744137408001124

Habib, A., Alsmadi, D., & Prybutok, V. R. (2020). Factors that determine residents' acceptance of smart city

technologies. *Behaviour & Information Technology*, *39*(6), 610–623. https://doi.org/10.1080/0144929X.2019.1693629

Hardin, R. (2002). *Trust and trustworthiness*. Russell Sage Foundation.

Hasija, A., & Esper, T. L. (2022). In artificial intelligence (AI) we trust: A qualitative investigation of AI technology acceptance. *Journal of Business Logistics*, *48*(3), 388–412. https://doi.org/10.1111/jbl.12301

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2024). *World happiness report 2024*. University of Oxford; Wellbeing Research Centre.

Hodgson, G. M. (2004). *The evolution of institutional economics*. Routledge. https://doi.org/10.4324/9780203300350

Jackson, S. J. (2014). Rethinking repair. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 221–239). MIT Press.

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 624–635). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445923

Joas, H. (1996). *The creativity of action*. University of Chicago Press.

Joas, H., & Kilpinen, E. (2006). Creativity and society. In J. R. Shook & J. Margolis (Eds.), *A companion to pragmatism* (pp. 323–335). Wiley.

Julsrud, D. T. E., & Krogstad, D. J. R. (2020). Is there enough trust for the smart city? Exploring acceptance for use of mobile phone data in Oslo and Tallinn. *Technological Forecasting and Social Change*, *161*, Article 120314. https://doi.org/10.1016/j.techfore.2020.120314

Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2021). Trust in artificial intelligence: Meta-analytic findings. *Human Factors*, *65*(2), 337–359. https://doi.org/10.1177/00187208211013988

Kaun, A. (2021). Suing the algorithm: The mundanization of automated decision-making in public services through litigation. *Information, Communication & Society*, *25*(14), 2046–2062. https://doi.org/10.1080/1369118X.2021.1924827

Kela. (2024, June 28). *Kela's strategic vision for the future influences management and development of social security services* [Press Release]. https://www.kela.fi/news/kela-s-strategic-vision-for-the-future-influences-management-and-development-of-social-security-services

Kilpinen, E. (2000). *The enormous fly-wheel of society: Pragmatism's habitual conception of action and social theory* (Research report No. 235). University of Helsinki.

Kilpinen, E. (2009). The habitual conception of action and social theory. *Semiotica*, *173*, 99–128. https://doi.org/10.1515/SEMI.2009.004

Knowles, B., & Richards, J. T. (2021). The sanction of authority: Promoting public trust in AI. In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 262–271). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445890

Kohn, E. (2013). *How forests think: Toward an anthropology beyond the human*. University of California Press.

Laux, J. (2024). Institutionalised distrust and human oversight of artificial intelligence: Towards a democratic design of AI governance under the European Union AI Act. *AI & Society*, *39*, 2853–2866. https://doi.org/10.1007/s00146-023-01777-z

Laux, J., Wachter, S., & Mittelstadt, B. (2023). Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance*, *18*(1), 3–32. https://doi.org/10.1111/rego.12512

Li, Y., & Hahn, J. (2022). Review of research on human trust in artificial intelligence. *ICIS 2022 Proceedings*, *8*, Article 1822. https://aisel.aisnet.org/icis2022/ai_business/ai_business/8

Lockey, S., Gillespie, N., Holm, D., & Someh, I. A. (2021). A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions. In T. X. Bui (Ed.), *Proceedings of the 54th Hawaii International Conference on System Sciences* (pp. 5463–5472). HICSS. https://aisel.aisnet.org/hicss-54/os/trust/2

Luhmann, N. (1988). Familiarity, confidence, trust: Problems and alternatives. In C. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 94–107). Basil Blackwell. https://www.nuffield.ox.ac.uk/users/gambetta/Trust_making%20and%20breaking%20cooperative%20relations.pdf

McClelland, T. (2020). The mental affordance hypothesis. *Mind*, *129*(514), 401–427. https://doi.org/10.1093/mind/fzz036

Misztal, B. A. (1996). *Trust in modern societies. The search for the bases of social order*. Polity Press.

Mitchell, M. (2021). *Why AI is harder than we think*. arXiv. https://doi.org/10.48550/arXiv.2104.12871

Naiseh, M., Al-Thani, D., Jiang, N., & Ali, R. (2021). Explainable recommendation: When design meets trust calibration. *World Wide Web*, *24*(5), 1857–1884. https://doi.org/10.1007/s11280-021-00916-0

OECD. (2021). *Drivers of trust in public institutions in Finland*. OECD Publishing. https://doi.org/10.1787/52600c9e-en

OECD. (2024). *OECD survey on drivers of trust in public institutions—2024 results: Building trust in a complex policy environment*. OECD Publishing. https://doi.org/10.1787/9a20554b-en

Okamura, K., & Yamada, S. (2020). Adaptive trust calibration for human-AI collaboration. *PLoS ONE*, *15*(2), Article e0229132. https://doi.org/10.1371/journal.pone.0229132

Peeters, R., & Widlak, A. C. (2023). Administrative exclusion in the infrastructure-level bureaucracy: The case of the Dutch daycare benefit scandal. *Public Administration Review*, *83*(4), 863–877. https://doi.org/10.1111/puar.13615

Peirce, C. S. (1877). The fixation of belief. *Popular Science Monthly*, *12*(1), 1–15. https://philarchive.org/rec/PEITFO

Pink, S. (2021). Sensuous futures: Re-thinking the concept of trust in design anthropology. *The Senses and Society*, *16*(2), 193–202. https://doi.org/10.1080/17458927.2020.1858655

Pink, S. (2022a). *Emerging technologies/Life at the edge of the future*. Routledge.

Pink, S. (2022b). Trust, ethics and automation: Anticipatory imaginaries in everyday life. In S. Pink, M. Berg, D. Lupton, & M. Ruckenstein (Eds.), *Everyday automation: Experiencing and anticipating emerging technologies* (pp. 44–58). Routledge. https://www.taylorfrancis.com/chapters/oa-edit/10.4324/9781003170884-4

Pink, S., Lanzeni, D., & Horst, H. (2018). Data anxieties: Finding trust in everyday digital mess. *Big Data & Society*, *5*(1). https://doi.org/10.1177/2053951718756685

Quilty, E., & Pink, S. (2024). Trust as a sensory mode of engaging culturally diverse communities in net zero futures. *The Senses and Society*, *20*(1), 1–15. https://doi.org/10.1080/17458927.2024.2350810

Rachovitsa, A., & Johann, N. (2022). The human rights implications of the use of AI in the digital welfare state: Lessons learned from the Dutch SyRI case. *Human Rights Law Review*, *22*(2), Article ngac010. https://doi.org/10.1093/hrlr/ngac010

Räisänen, S. (2024). Machinic, inadequate, entrepreneurial: Uncovering the citizen subject of the human-centric welfare state. *European Journal of Cultural Studies*, *27*(6), 1211–1232. https://doi.org/10.1177/13675494231213943

Ramírez-i-Ollé, M. (2019). Trust, scepticism, and social order: A contribution from the sociology of scientific knowledge. *Sociology Compass*, *13*(2), Article e12653. https://doi.org/10.1111/soc4.12653

Rebensky, S., Carmody, K., Ficke, C., Nguyen, D., Carroll, M., Wildman, J., & Thayer, A. (2021). Whoops! Something went wrong: Errors, trust, and trust repair strategies in human agent teaming. In H. Degen & S. Ntoa (Eds.), *Artificial intelligence in HCI: HCII 2021: Lecture notes in computer science* (Vol. 2, pp. 95–106). Springer Nature. https://doi.org/10.1007/978-3-030-77772-2_7

Robinson, S. C. (2020). Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). *Technology in Society*, *63*, Article 101421. https://doi.org/10.1016/j.techsoc.2020.101421

Rousku, K. (2024). *Digiturvabarometri-raportti: Millaisena kansalaiset kokevat digitaalisen maailman elokuussa 2024?* Digi- ja väestötietovirasto. https://cdn.verkkopalvelu.suomi.fi/files/Digiturvabarometri_raportti_2024-27ff385c4db3df2ab328c67258a3667d.pdf

Savolainen, L., & Ruckenstein, M. (2022). Dimensions of autonomy in human-algorithm relations. *New Media & Society*, *26*(6), 3472–3490. https://doi.org/10.1177/14614448221100802

Seligman, A. (2021). Trust, experience and embodied knowledge or lessons from John Dewey on the dangers of abstraction. *Journal of Trust Research*, *11*(1), 5–21. https://doi.org/10.1080/21515581.2021.1946821

Siau, K. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal*, *31*(2), 47–53. https://www.cutter.com/article/building-trust-artificial-intelligence-machine-learning-and-robotics-498981

Sperrle, F., El-Assady, M., Guo, G., Chau, P., Endert, A., & Keim, D. (2020). *Should we trust (X)AI? Design dimensions for structured experimental evaluations*. arXiv. https://doi.org/10.48550/arXiv.2009.06433

Steedman, R., Kennedy, H., & Jones, R. (2020). Complex ecologies of trust in data practices and data-driven systems. *Information, Communication & Society*, *23*(6), 817–832. https://doi.org/10.1080/1369118X.2020.1748090

Stubb, A. [@alexstubb]. (2024, March 20). *Finland has been ranked the Happiest Country in the World, now for seven years in a row…* [Post]. X. https://x.com/alexstubb/status/1770335834543124770?

Sweden's suspicion machine. (2024, November 27). *Lighthouse Reports*. https://www.lighthousereports.com/investigation/swedens-suspicion-machine

Trust. (n.d.). In *Oxford English Dictionary*. https://doi.org/10.1093/OED/5777528687

Urbina, D., & Ruiz-Villaverde, A. (2019). A critical review of *Homo economicus* from five approaches. *American Journal of Economics and Sociology*, *78*(1), 63–93. https://doi.org/10.1111/ajes.12258

Vainionpää, F., Väyrynen, K., Lanamäki, A., & Bhandari, A. (2023). A Review of challenges and critiques of the European Artificial Intelligence Act (AIA). In R. De (Ed.), *ICIS 2023 Proceedings*. ICIS.

Warkentin, M., Sharma, S., Gefen, D., Pavlou, P., & Rose, G. (2012). Government of the people, by the people: A look at trust in eGovernment. In K. D. Joshi & Y. Yoo (Eds.), *AMCIS 2012 Proceedings*. Association for Information Systems. https://aisel.aisnet.org/amcis2012/proceedings/EGovernment/20

Weitz, K., Schiller, D., Schlagowski, R., Huber, T., & André, E. (2019). "Do you trust me?" Increasing user-trust by integrating virtual agents in explainable AI interaction design. In C. Pelachaud & J.-C. Martin (Eds.), *IVA '19: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents* (pp. 7–9). Association for Computing Machinery. https://doi.org/10.1145/3308532.3329441

Zhang, B., & Dafoe, A. (2020). U.S. public opinion on the governance of artificial intelligence. In A. Markham, J. Powles, T. Walsh, & A. L. Washington (Eds.), *AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 187–193). Association for Computing Machinery. https://doi.org/10.1145/3375627.3375827

Zierau, N., Hausch, M., Bruhin, O., & Söllner, M. (2020). Towards developing trust-supporting design features for AI-based chatbots in customer service. *ICIS 2020 Proceedings*, *2*, Article 1325. https://aisel.aisnet.org/icis2020/digital_commerce/digital_commerce/2

Zuboff, S. (2019). *The age of surveillance capitalism*. Profile Books.

Zucker, L. G. (1986). Production of trust: Institutional sources of economic structure, 1840–1920. *Research in Organizational Behavior*, *8*, 53–111.

About the Authors

Antti Rannisto is a sociologist, ethnographer, and PhD researcher at Aalto University. His doctoral research examines AI innovation processes within the Finnish public sector. He also works as insight lead in the Design & Strategy unit at Solita, a Nordic technology consultancy.

Fanny Vainionpää is a postdoctoral researcher in the INTERACT Research Unit at the University of Oulu, Finland. Her research started with a focus on ICT career choice and education and has moved towards sustainable technology, focusing on systemic, social, and cultural perspectives. Vainionpää currently works on artificial intelligence regulation.

Social Inclusion
2025 • Volume 13 • Article 10087
https://doi.org/10.17645/si.10087

cogitatio

ARTICLE

Open Access Journal

Exploring the Futures of Datafied Welfare State Education: Thematic Analysis of Sociotechnical Imaginaries

Lauri Palsa, Janne Fagerlund, and Pekka Mertala

Department of Teacher Education, University of Jyväskylä, Finland

Correspondence: Lauri Palsa (lauri.p.palsa@jyu.fi)

Submitted: 9 February 2025 **Accepted:** 26 March 2025 **Published:** 21 April 2025

Issue: This article is part of the issue "Fostering the Socially and Ecologically Sustainable Digitalisation of Welfare States" edited by Paula Saikkonen (Finnish Institute for Health and Welfare) and Marta Choroszewicz (University of Eastern Finland), fully open access at https://doi.org/10.17645/si.i514

Abstract

Technological developments have inspired many scholars and other professionals to envision the possibilities that digital data technologies bring to the future of education. However, some aspects of this so-called datafication may conflict with the local characteristics of education systems. In this study, we investigate the future imaginaries of datafied education in the context of a welfare state education system. By interviewing Finnish experts ($N = 25$) from various sectors, we looked beyond official policies to explore the multi-perspective views of national-level stakeholders. Through a thematic analysis, we constructed four 1st order sociotechnical imaginaries that illustrate the anticipated impacts of datafication on education, specifically that it makes education (a) easier and (b) more coherent, (c) maximises learning, and (d) enhances visibility. However, some future visions involve recognising and assessing the consequences of the 1st order imaginaries. These 2nd order imaginaries broaden the spectrum of perspectives and highlight the pluralism of educational futures. Thus, the results did not give rise to a single coherent or holistic "imaginary" of datafied education but, instead, highlighted different aspects of datafied education. Through these imaginaries, we were able to identify the key characteristics of the Finnish welfare state education system, such as trust and the strong role of the public sector, that may be subject to negotiation in the datafication process. Through the development of sociotechnical imaginaries, our goal is to create a space for an inclusive debate on the future of education and thereby contribute to the promotion of sustainable development in education.

Keywords

datafication; education; experts; imaginary; sociotechnical

1. Introduction

In 1910, French artist Villemard envisioned what education might look like in the year 2000, resulting in one of the most famous historical images of the future of education. In this "paleofuturistic" picture, a teacher feeds textbooks into a machine, from which information is transmitted directly via wires to the pupils' headsets. While this imaginary might appear somewhat clumsy today, there remain similar aspirations in current "imaginaries" about the future of education, such as the desire to make education easier, automated, and more efficient. Accordingly, current imaginaries often revolve around the possibilities of digital data, rather than the prospect of feeding physical books to a machine. As the OECD (2025) conveyed, "data and digital technologies are among the most powerful drivers of innovation in education."

Datafication has inspired stakeholders, such as policymakers and researchers, to imagine its desirable and undesirable future potentials (Marcetic & Nolin, 2023). Utopian imaginaries of datafication often concern the ways in which data technologies, such as learning analytics and digital platforms, can improve education by personalising learning or automating administrative processes. For instance, combining learning analytics with facial recognition technology can enable the "continuous analysis of instructional dynamics and evaluating student effort," in turn making "instantaneous" feedback and "reporting [of] student progress and warnings of misbehaviour" possible (OECD, 2020, p. 43). However, continuous analysis requires large-scale data generation, which can threaten children's privacy and personal data protection (Silvennoinen et al., 2024). Similarly, for some, the imagined benefits of data carry the risk of "negative impacts on learning relationships, exploitative commercial uses of collected student data, discriminatory practices" (Ross & Wilson, 2023, p. 19).

Future imaginaries are not limited to the use of (data) technologies and their technical properties; rather, they depict changes in social processes. According to the OECD (2023, p. 3), digital transformation implies a "fundamental change in some educational processes…as a way to reshape teaching methodologies, learning processes, and the educational ecosystem at large, to make it more effective." The dynamic relationship between the technological (datafication) and the social (education) illustrates the sociotechnical nature of future imaginaries (Jasanoff, 2015).

Imagining the future is thus not only about the properties of technologies themselves but also about the impact they are imagined to have on different social practices, such as the way education is organised. As Burbules et al. (2020, p. 96) put it, the datafication of education "can help remedy certain disabilities; or it can create *new* disabilities. It can be liberating or coercive, and it raises new questions about surveillance and privacy." In other words, datafication can appear either as a treasure chest or as Pandora's box, depending on the imagineer.

As various stakeholders (e.g., policymakers, companies, researchers, civil society actors) can have different viewpoints regarding education, their visions about the desired and undesired futures of datafication would also differ. Several imaginaries, including utopian hopes and dystopian risks, can be present simultaneously. Moreover, what may seem utopian from a technical perspective (e.g., data-based feedback improving the efficiency of education) may seem dystopian from a social perspective (e.g., increased risk of data privacy), suggesting that differing visions can overlap and connect in intricate, tense ways.

Imagining the future is about the present—how the future is imagined can influence the decisions and actions taken today. Future visions create "a strong foundation to calls for action in the name of the future" (Mertanen & Brunila, 2024, p. 950). For example, the European Commission (2020, p. 10) has envisioned "a high-performing digital education ecosystem" that "must be a common endeavour across society" as a call to action to implement what they see as essential changes in national educational systems. International policy documents, such as the European Commission's action plan referred to above, are commonly used research materials on (future) imaginaries (e.g., Williamson, 2015). While we acknowledge the significance of future-visioning policies as instruments for guiding educational practice, our approach is more nuanced. Rather than focusing on the explicit (and often rigid) "formal" imaginaries presented in official education policy documents, we are interested in making visible the implicit nonformal future visions present in society. Accordingly, our study involved interviewing Finnish experts in different sectors (further detailed in Section 3). Experts often have the networks, capacity, and resources to make their own and their organisation's voices heard; they understand the factors and elements that are relevant for impacting the future, and they are often in positions where their decisions impact other stakeholders. Thus, because of their status within social structures, they have the "authority to construct reality" (Meuser & Nagel, 2009, p. 19).

Through our study, we addressed the following research question: How do experts at the national level imagine the future of datafied education in Finland? Analytically, we used the concept of "sociotechnical imaginary" (Jasanoff, 2015) as a heuristic device to consider the technological and social nature of datafication and education as well as the interconnections between them. This concept enabled us to analyse collectively held visions of futures that are "animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology" (Jasanoff, 2015, p. 3). Constructing such imaginaries also allowed us to identify the features that characterise the education system in the Finnish welfare state, which datafication might impose on the negotiation, thus influencing the sustainability of the education system.

Finland provides an excellent context for exploring the research topic. First, there is a strong belief in the potential of datafied education among Finnish policymakers, and their approaches resonate with international policies. In fact, the aforementioned European Commission action plan is one of the guiding documents behind the *Desired Outcomes for Digitalisation of Early Childhood Education and Care, Pre-Primary, Primary and Lower Secondary Education* proposed by the Finnish Ministry of Education and Culture (2023). Second, education policies in Finland are commonly prepared in a participatory manner, with cooperation among various relevant stakeholders through surveys, workshops, and commenting (Halinen, 2018), which means that there are opportunities for experts, such as civil society organisations, researchers, and companies, to have their voices heard.

2. Sustainable Education in the Digital Age

The Finnish education system has been argued to "focu[s] on sustainable development" (Messina, 2023, p. 1). However, future aspirations of datafication may place this central feature under negotiation. What makes Finland an interesting context for research in terms of sustainability is the paradoxical situation between future aspirations and existing characteristics. The use of digital data is highly prioritised in educational policy, even though features of datafication, such as accountability (Grek et al., 2020) and commercialisation (Cone &

Lai, 2024), are at odds with the sustainable development of the education system in a welfare society based on trust (Välimaa, 2021) and the strong role of the public sector (Dovemark et al., 2018).

Sustainability relates both to the aims of education as part of society and to the way education is organised. The public sector plays an important role in both cases. First, educational equality and equity, in line with the SDGs presented by the UN (2015), are said to be "at the heart of educational policy in Finland" (Välijärvi & Sulkunen, 2016, p. 17). In structural terms, equity is rooted in the educational reforms of the 1970s, following which the previously parallel school system was transformed into a comprehensive system (Kortekangas et al., 2019). In practice, this means that schools are for all pupils, regardless of their backgrounds (Toom & Husu, 2016). Allowing pupils to participate in attaining the same education regardless of their socioeconomic situation, among other things, is thought to enable social mobility in society (Välimaa, 2021). Second, trust has been identified as playing an important role in the Finnish school system; it is evident in the way education is governed as well as in the opportunities for teachers to use their own agency (Välimaa, 2021). Trust helps in building a culture of cooperation and common good, which in turn supports the achievement of the SDGs (UN, 2015). Concerning education policymaking, trust is reflected in the democratic ideals of openness and participation. In addition to public actors, many other stakeholders are involved in the development of education policy. At the national level, the national core curriculum (EDUFI, 2014), which provides a basis for participatory planning of local curricula, is prepared in a participatory manner based on cooperation among key stakeholders (Halinen, 2018). In addition to policymakers, educational authorities, and professionals, these include teacher education departments of universities, civil society organisations, trade unions, and private companies. Lavonen (2017, p. 7) explained that "participation of several parties in education…made the process and product [of the curriculum] more equal." It is expected that involving a wide range of actors in the preparation of education policies would lead to sustainable outcomes without radical changes, which may contribute to increasing trust. As explained by Välimaa (2021, p. 173), "it can be assumed that the stability of educational policies helps to strengthen trust in education because the policies are predictable. Citizens, schools, and teachers know what to expect today and in the future."

Rather than dictating how education should be organised (top-down), decentralised education governance (Lavonen, 2017) ensures that "schools are given almost full autonomy in developing their daily delivery of education services" (Toom & Husu, 2016, p. 49). As educational authorities and policymakers trust teachers, there is no need to pre-evaluate learning materials, inspect or supervise schools, conduct national-level testing, or systematically evaluate teachers (Lavonen, 2017). As explained by Toom and Husu (2016, p. 44), the pedagogical situations that teachers come across are so varied and pervasive that "all the criteria for teachers' pedagogical action cannot be stated explicitly." Teachers can choose their teaching methods, materials, and assessments according to what they consider most appropriate (Lavonen, 2017; see also Toom & Husu, 2016). Trust offers conditions that support teachers' pedagogical autonomy and, in turn, their professional agency. Despite their autonomy, teachers may be expected to use digital technologies in their work (see Mertala, 2020).

Over the past decades, Finnish education policymakers have made a significant push towards digitalisation through several policy documents, developmental programmes, and funding allocations. Kaarakainen and Kaarakainen (2018) have identified three phases of this endeavour: Between 1998 and 2004, education policies focused on equipping schools with information technology. Between 2005 and 2010, the policies

were aimed at evaluating the benefits of technological tools. Finally, between 2010 and 2018, the policies were focused on diversifying the use of digital tools. Despite the broad visions presented in educational policies, digitalisation has not reached educational practice to an equivalent extent. The results of national evaluations of digitalisation in education, which include classroom observations conducted in 2022 (Oinas & Hotulainen, 2024) as well as recent large-scale assessment studies such as the *International Computer and Information Literacy Study* (Fagerlund et al., 2024), have shown that learning in schools is still primarily teacher-led and that digitalisation has not brought about a significant change in pedagogy. Thus, despite decades of effort, digitalisation does not play as significant a role in everyday schooling as educational policies and visions might suggest.

3. Conducting the Study

3.1. Data Collection

Data were collected through semi-structured interviews with 25 experts working on issues related to datafied education on a national scale. Although the Finnish education system is decentralised, actors at the national level play a key role in the overall system, such as in creating structures that enable trust, participation, and decentralised organisation. The experts were selected through purposive sampling (Gill, 2020) based on their relevant but differing approaches to the datafication of education. They represented education policy, educational technology, and so-called data justice, including researchers and professionals working in civil society organisations (Dencik et al., 2019). The interviews were conducted remotely via Zoom (n = 21) or face-to-face (n = 4) between November 2023 and February 2024. The length of these audio-recorded interviews varied between 47 minutes and 122 minutes, and the total length of the data was 31 hours and 18 minutes (227,427 words). The transcriptions were pseudonymised by removing or changing identifying information, including the names of the individuals, services, and organisations.

The interview protocol (see the Supplementary File) covered a wide range of topics related to the role of data in education, with a specific focus on the future of datafied education. The interviewees were asked about their views on (a) what they would like the future of datafied education to be (utopia), (b) what their most undesirable future would be (dystopia), and (c) what they saw as the most likely future (realistic imaginary). To facilitate potentially realistic imaginaries, we instructed the interviewees to imagine the future within a concrete timeframe of five years. The micro utopias (Cooper, 2013) that result from such a process are arguably more concrete and contextualised than those that address a distant future (Selwyn et al., 2020).

3.2. Thematic Analysis

We conducted a thematic analysis (Braun & Clarke, 2006) to construct sociotechnical imaginaries of datafied education from the interviews. The analysis process is detailed in Table 1.

Table 1. Thematic analysis of the process employed in the present study.

Phase	Performed analysis	Outcome
1. Familiarisation with the data	The transcribed interviews were read thoroughly. The sections concerning utopian, dystopian, and realistic futures were identified and marked using Atlas.ti software. The sections on how the future can be influenced (i.e., the relevant actors and effective means) were also identified and marked.	The data sections identified for further analysis comprised 18,743 transcribed words.
2. Initial coding	We identified specific future-addressing citations and open-coded them to form "data simplifications," which we then categorised as focusing primarily on either education (socio) or datafication (technical).	The simplifications helped us "translate" the data into a more consistent form, supporting further analysis. Among the data simplifications, 142 concerned education, and 59 concerned datafication.
3. Constructing themes	We identified coherent and meaningful patterns (Clarke & Braun, 2013) by interpreting and comparing the data simplifications to find connecting similarities and distinctive differences. We reviewed the potential themes by reflecting their meanings on the original transcribed data to ensure that the interpreted meanings corresponded with the data.	Twelve themes were constructed.
4. Defining and naming themes	The themes were compared to identify the relationships between them. This step revealed that some of the themes directly concerned imagined educational change, while others were related to the estimated consequences of these imagined changes. Based on this notion, we differentiated between 1st and 2nd order themes. Themes referring to the imagined future of datafied education were defined as sociotechnical imaginaries.	Four main 1st order and eight related 2nd order sociotechnical imaginaries of datafied education were identified.

The following data extract exemplifies the analysis process. Here, interviewee Policy 7 imagines a future from the perspective of national-level education development:

> In an ideal world, education providers produce a certain kind of data that we can use without, for example, bothering education providers with these continuous surveys on different topics, which sometimes really overload the field….In the future, I see…that we will have a certain type of database. In their normal activities, education providers produce data for the national data bank, from which we can look at certain things and act accordingly. Without us burdening schools to death by asking once a week or once a month for all sorts of odd things.

Since the expert explicitly describes an "ideal world," this section was identified as describing a utopia (Phase 1), which relates to both education and datafication (Phase 2). In terms of datafication (technical), the expert describes the automated generation of data that occurs in normal, everyday school life as well as the data bank that integrates these data at the national level. New data infrastructure would allow education administrators to look at the different school issues for the whole country. The areas to look at come across as being comprehensive, as data production would cover "all sorts of odd things"—thematised as visibility

(Phase 3). In terms of education (socio), this would make the work of school professionals easier, as they would not have to be "burdened to death" by surveys. The mention of burdening indicates that the interviewee recognises the workload involved in data generation; this was thematised as data work.

Although an imagined future may seem straightforward in the context of a single interview, imaginaries become more diverse and relational when multiple interviews are considered as a whole. Therefore, in Phase 3, we defined four 1st order sociotechnical imaginaries that describe how datafication will make education easy, coherent, maximised, and visible, as (partly) exemplified by the previous excerpt.

Some interviewees expressed doubts about the direct changes that datafication was imagined to have in education. The following was considered by Data Justice 13:

> As I have now followed different kinds of data things, I have formed this vision that the terribly broad datafication dreams…might not really produce terribly good results.

Here, the term "datafication dreams" refers to the 1st order imaginaries that are desired, while the consideration that they "might not really produce terribly good results" illustrates the interviewee's doubt regarding these imaginaries. During the analysis, we noted that each of the 1st order imaginaries was considered to have various societal consequences. For instance, making education more visible was imagined as leading to increased surveillance and control. Therefore, in Phase 4, we constructed 2nd order imaginaries derived from the consequences of 1st order imaginaries. All the constructed imaginaries are summarised in Figure 1, illustrating how the expert interviews did not result in a single coherent or holistic

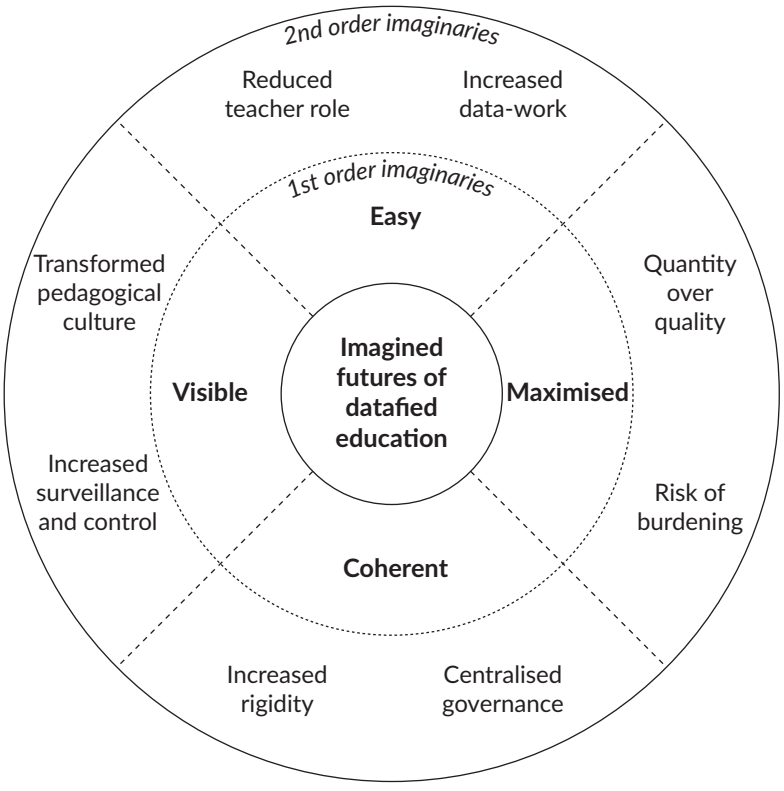

Figure 1. Sociotechnical imaginaries of datafied education.

imaginary of datafied education but, instead, highlighted different aspects of it. For example, the imaginary of "education made visible" had imagined consequences that the 1st order imaginaries did not include (e.g., increased surveillance of transformed pedagogical culture). This set of imaginaries was then diversified by determining 2nd order imaginaries, which emphasised the pluralism of educational futures. In the following section, the 1st order imaginaries are introduced, followed by a discussion of the related 2nd order imaginaries.

4. Findings

4.1. Imaginary 1: Education Made Easy (With Data)

The experts imagined different ways in which data technologies could make teachers' work easier (1st order imaginary), such as a personalised learning platform doing the teaching for them. By allowing pupils to study the content on their own using a platform, teachers can save time for other activities, such as interacting with pupils. However, this imaginary involves tension because when a teacher shares the teaching responsibility with the data technology, their *role* (2nd order imaginary) in the classroom reduces from an agentic and autonomous subject to an assistant for data technology (see Selwyn, 2019). Moreover, data technologies may require additional *work* (2nd order imaginary), in which case the teacher's time is not saved.

The following excerpt exemplifies the theme of ease. Here, the expert imagines a situation where, instead of learning from a teacher, pupils learn a subject independently using data technology that identifies their competence levels and adjusts their tasks accordingly:

> The lesson begins. Each pupil takes their own learning [platform], welcoming [them] with: "Hello, welcome [pupil's name]. Last time, we studied photosynthesis. Do you have any questions about it? Do you want me to help you with that? Do you feel you understand what photosynthesis means?" And I say: "I don't really understand it yet. Can you tell me a little bit more about it and what I need photosynthesis for?" Whereas another pupil in the class might ask: "This is quite simple for me; tell me more." And a third will ask: "What does photosynthesis mean in Finnish? I don't understand the word flower. What is a flower in Finnish?" (Technology 5)

The time "free[d] up" (Policy 14) by automated teaching was imagined as providing teachers with more opportunities to hold discussions with their pupils and get to know them better. As Technology 5 stated: "Then you [as a teacher] get to walk up to those pupils, you get to talk to them: 'Hey, how are you doing?' "

If data technologies are given a major role in education, even these dialogical interactions would serve the purpose of "fine-tuning" the algorithm in the end, as elaborated by Technology 25:

> [We] learn to read those children, to recognise, and that more teachers would have time to give it—let's call it empirical information. That the teacher would be present, they would have time to be more present to those children, to get to know them, to help the machine rather to develop so that we learn faster to understand that child, so that the programmes can support that child faster and easier, in a more straightforward manner.

The interviewee described how sharing teaching responsibilities with a "machine" would (appear to) increase the teacher's agency by enabling them to "be there" for the pupils. The latter part of the extract reveals how the teacher would not use the new information they obtain to inform their own teaching but, instead, would have to help develop the machine so that the software can support children more effectively. Thus, the teacher becomes an assistant to the machine, providing it with empirical information. In contrast to technologies "empowering and supporting teachers and educators to have confidence in what they are doing" (Policy 9), this transformation would limit teachers' pedagogical agency, highlighting their deprofessionalisation (Selwyn, 2016). This reduction in the role of a teacher illustrates the first 2nd order imaginary related to Imaginary 1: Education made easy (with data).

The next 2nd order imaginary is related to the amount of work that data technologies require. While it was imagined that technology would "free up people's time for more essential things" (Policy 8), many experts recognised that technology could also lead to more "data work" (see Hoeyer, 2023) for teachers. Policy 16 stated that they value educational data technologies based on the amount of work they require from teachers at a concrete level. When they had commonly asked technology providers about the sources of data, "quite often the answer is that, well, the teacher makes records" (Policy 16), and they continued to illustrate the problematic situation with a practical example:

> There is a system used in class, so you can observe that when [a pupil] is there, you click here, click there, and then make a record here. Thus, this accumulates certain types of data that can be used. For example, a certain kind of picture is formed about this pupil. (Policy 16)

Here lies the risk of increased data work: According to the technology provider, entering a single pupil's data would just be "a small thing. It's five minutes when you click on this page and put it like this" (Policy 16). However, the amount of work will quickly accumulate. As explained by Policy 16, "if a teacher has 25 pupils, and you have this one task that you do, so it took five times 25…it's a huge 125 minutes, more than two hours. That's a huge amount of work." If a teacher spends more than two hours manually entering data, they presumably would not have any more freed-up time left to get to know their pupils better, which is what data technologies are often proposed to offer. The problem is not the data but the way in which the data are generated. The expert conveyed their preference for the data to:

> By default…be of the kind that is generated automatically. I understand that if you are working in a digital learning environment, the data generated in it, if it can use it automatically, without you having to enter the data into it, then it is already a thing that you can get something from. (Policy 16)

It should also be noted that the use of digital environments enabled by automated data collection may require data work. Data work is not only related to whether a teacher has to manually enter data but is also a matter of the meta-level work required, such as developing the necessary skills and competencies. The integration of data technologies raises new demands not only for data competencies among learners but also for those teaching the competencies, such as for educational institutions. Accordingly, Data Justice 11 imagined a desirable future in which "teacher education should be really prepared for this and ready to teach these things to future teachers."

4.2. Imaginary 2: Data-Maximised Learning

Data technologies were imagined to increase the amount of learning among pupils by challenging them and ensuring that all curricular contents are addressed. These are aspects of data-maximised learning—an imaginary that concerns the objectives of education. A closer exploration revealed two 2nd order imaginaries that created tension with data-maximised learning. With the desire to increase learning comes the risk of *burdening* the pupils. In addition, maximisation can emphasise the *quantitative* aspects of learning at the expense of the qualitative aspects.

The experts conveyed that data technologies, such as personalised learning platforms, have the potential to maximise the amount of learning. Technology 4 envisioned the following:

> If [data-based artificial intelligence] can be harnessed, for example, in learning environments in such a way that it learns from what the user is doing and offers the right kind of questions, asks the pupils certain types of questions, and in a way challenges them to learn more and more, then I think that's desirable.

A similar "expansionist mind-set" (Selwyn et al., 2020, p. 101) was also evident when data technologies were imagined as ensuring that curricular content would be addressed in its entirety. In this regard, Data Justice 2 imagined that "the pupil would see a heatmap, that 'hey, now in history we have covered things like this. There are still these unaddressed issues.' So that this issue can be brought to the fore." In this scenario, with the help of a "heatmap," pupils would be able to follow the progress of their curriculum-based learning and check whether any aspects are missing.

Some experts highlighted the importance of emphasising the quality of learning, rather than simply maximising learning, which led to tension in the imaginary of data-maximised learning. The 2nd order imaginary in question is evident in the following scenario, where Technology 21 envisioned a national policy that would aim to improve the overall quality of the learning experience rather than purely quantity:

> [The policy] doesn't have to define what data will be used, but it would be a kind of national data programme to improve the overall quality of learning. That's probably the right term. It doesn't mean that the numbers should go up, but that we should have [the] well-being [of] pupils and graduates who leave school believing that [they] can make it in this world.

This interviewee clearly opposed the "the logic of output maximisation" (Zomer, 2024, p. 7), according to which numbers are expected to grow. The interviewee recognised the difference between the actual phenomena and the data that represent them. As data technologies tend to operate with quantitative measures, which are often technically easier to process, they are unable to account for the complexity of multifaceted issues such as learning. If the focus is on measurable data, some vital aspects may be overshadowed (Zomer, 2024). The expert noted that simply increasing numbers is not enough; the actual well-being of pupils is more than just a quantitative measure. The viewpoint was also reinforced by Data Justice 12, who argued that by limiting (instead of maximising) the curricula, education providers would be "able to focus on the right things":

> Because somehow it seems that we always say that pupils should be taught this, that and the other, [and] we say that it's important to teach them data protection, but in the end it can be quite a heavy burden what they should learn. So that [learning] would remain in moderation, however.

The interviewee seemed to note a common mindset: That various societal challenges, such as data security, can be addressed through education. As new challenges, concerns, and issues appear, schools end up covering a wide range of topics: "this, that, and the other." The challenge is that individual phenomena can be vast, making even a single topic "quite a heavy burden." This notion represents another 2nd order imaginary: the risk of burdening. To address the imbalance between the numerous teachable topics and the limited resources available without compromising pupils' well-being, the expert suggested limiting the learning content rather than maximising it.

The question about burdening goes beyond the current era. Since the end of the 19th century, the amount of content taught in schools has increased, "piling up more work for the teacher and the pupil to do" (Lamb, 1903, p. 67). A glance at the three latest Finnish national core curricula for basic education indicate that this trend has stayed the same in the 20th and 21st centuries: The length of the core curriculum was 111 pages in 1994 and 320 pages in 2004, and the efficient 2014 core curriculum is 472 pages long. Additionally, recent large-scale studies have evidenced a steady growth in the number of students experiencing school burnout (Read et al., 2022; THL, 2023), indicating the intensification of educational content. However, it should be noted that as digitalisation increases access to different types of content, the number of topics covered may also increase.

4.3. Imaginary 3: Coherent Education

Many of the interviewed experts stated that data technologies have the potential to increase coherence in education. Datafication was imagined as facilitating alignment between the amount and quality of education received by pupils across the country. However, the requirement of coherence can lead to *centralised* educational governance and create perceived *rigidity* and associated social consequences, leading to a tensioned imaginary.

Policy 6 imagined educational coherence based on equality, one of the core values of the Finnish educational system:

> Well, at least my thinking is very strongly guided by this kind of equality perspective. I think it would be equal, for example, that pupils receive the same amount of math instruction. So then, if it seems that there are big differences in the number of hours of teaching…we would be able to observe, a little bit to think about what is going on.

In other words, data technologies can enable national-level monitoring of the amount of teaching that pupils receive, which would then allow for identifying and addressing possible disparities. Data Justice 2 provided an example of what this could mean in practice. They referred to the vision paper of the DigiOne platform (under development in Finland at the time of writing), which is aimed at connecting several educational elements, from lesson planning to assessment, on the same digital platform. DigiOne (2025) is a "national digital platform for education" built by Finnish municipalities (2019–2025), around which an "education ecosystem" will be

created. The development project is funded by the municipalities involved and by Business Finland, a public organisation (funds for innovation). With the help of this system:

> [A] teacher [would be] able to plan the lesson so that when I start to build an exercise and the learning session [based on the curricular criteria integrated in the system] and the pupils go to the system, [they] do the tasks [which are then] stored there. So it all connects everything all the way from the curriculum to what happened there. So it, in that way, of course, makes it also more visible in the information system that now these curriculum things had been dealt with today in this way and what kind of results came from there. (Data Justice 2)

According to the expert, the system "makes the teacher's work a little more goal-oriented" and also enables "the principal to see whether our teachers have gone through the whole curriculum and what has been left" (Data Justice 2). Therefore, the datafication of education could help align the amount of teaching and teaching methods. The interviewees also imagined that dissimilar integrations of data technologies across the country would lead to differing educational realities for pupils. Thus, in the name of educational coherence, a push for technological coherence was identified. To achieve the expected impact of data technology on educational practice similarly at the national level, technologies should be integrated to the same extent across all contexts. For Data Justice 11, in an undesirable future, "some teachers and some municipalities prioritise these things [i.e., datafication] really high or do a lot of things and pupils get a completely different education than what pupils in other schools or classrooms get. It's just that inequality…so that it doesn't become such a deepening…data gap."

The quest for coherence creates tension in a decentralised education system where local autonomy is expected to lead to relevant but possibly different educational practices (Lavonen, 2017). As explained by Technology 17, the national education administration could support schools by guiding them towards common solutions with "a recommendation on how to create interoperable learning analytics." In relation to the 2nd order imaginary of centralised governance, local education providers and schools in Finland have the autonomy to choose the services they use. Thus, the ministry "does not want to choose for [municipalities] what is the way [to use learning analytics]" or "does not want to dictate from above, because that is not how our education system works" (Technology 17).

Some experts highlighted the importance of pedagogical freedom instead of a centrally administered and coherent education system, even though it would lead to variance in educational practices and decrease coherence:

> I'm somewhat confident that I believe that we have this pedagogical freedom and that teachers have had the space to experiment and take action and work, so I do believe that good things are happening in Finnish schools, certainly a lot of things that I haven't heard of, but that enable us to find new paths and do new things. (Policy 14)

In addition to the tension between centralised and decentralised educational administration, the experts feared that the coherence brought about by technology (e.g., in teaching, curriculum, and assessment) would increase rigidity and thereby have undesirable consequences. Regarding educational practice, Data Justice 19 stated that "the concern is that now that there is not yet such strict legislation, Google and the

like will get the marketplace. And then it's really slow and difficult to move away from that to other tools." In this imaginary, schools would have difficulties changing the platforms they use if the role of a certain technology provider became highly essential. This would increase rigidity in the way education is organised, representing "infrastructural dependency in the datafied welfare state" (Cone & Lai, 2024, p. 1). On a generic level, a "horror scenario" for Data Justice 23 would be the "really strong social control and classification of people into different tracks based on what their results look like, which would then of course very much maintain the social class." This could happen by using data "in such a way that different gates are closed, and people are somehow classified and categorised in such a way that it becomes more difficult to make a social class jump or something."

4.4. Imaginary 4: Data-Visible School

The fourth 1st order imaginary involves making education visible through data. This refers to the multiple ways in which data technologies may help different stakeholders monitor different educational aspects. Two tensions were uncovered in relation to this imaginary. First, increased visibility was seen as a threat to current pedagogical *culture* based on autonomy and agency. Second, increased visibility was seen as increasing the risk of *surveillance and control*.

Data technologies were imagined as making aspects of education more visible, as they could enable data collection, aggregation, and discoverability. Data thus serve as visible representations of schools and the education system at large. The potential of data technologies to increase visibility was not only related to things already visible but also to other aspects that could be monitored. Imagining new potentials for visibility creates the logic of an ever-expanding "data gaze" (Beer, 2019). This was evident when Technology 21 envisioned the possibility of broadening monitoring from the perspective of educational governance:

> If we forget about the limitations and think about [the desirable future], then [the Ministry of Education] could bring in a clear, data-based map of more than just skills. Competence data is easier to deal with, but then what is the well-being or social exclusion, or where are the most positive pupils-type system, so it could be seen there.

Further, Technology 21 imagined that data technologies would have the potential to open a window into the future. With data, "you could sort of see if we should put a little bit more money into potential drop-outs, because then it will come back many times over if you get them on the right side of the fence, as it were." The data were imagined as ways to predict the future trajectories of pupils, thus facilitating decisions that could influence these speculative futures.

Increased visibility creates tension when the possible consequences for the current pedagogical culture are considered. According to the experts, trust strengthens agency and individual responsibility on many levels. This idea is often used to explain how trust encourages the development of pedagogical practices (e.g., Toom & Husu, 2016; Välimaa, 2021), whereas monitoring has the opposite effect:

> I have learned that Finnish primary school pedagogy produces good results for some strange reason when you trust children and adults and don't measure them too closely. It is the psychological phenomenon that when you start to measure how many lessons you have missed and whether the

> homework is done and so on, so in a way many times people have a bit of a childish reaction, a kind of rebellion that "this kind of thing, so what is this, now it is forced," that in a way it loses its own agency in the process and its own sense of responsibility and do now this way when it is forced. (Data Justice 2)

Shifting to data-based measurements could take the education system to a new stage. Data Justice 2 saw a risk associated with this: "If we transit to this type of measurement culture in education and the autonomy and agency and experimentation disappear…many things can fall apart."

The previous data extracts illustrate that it is important not only to consider data-based visibility in general but also to determine which issues are made visible, for whom, and for what purposes. Increased visibility was perceived as a form of control that diminishes individual agency, leading to things "falling apart" (Data Justice 2). Similarly, Technology 25 imagined a dystopia in which data would be used to control pupils by monitoring them.

> What I certainly don't want to see is that it's a control tool—the data. I mean, really. One time at a [large corporate event], a representative came to explain to me, in a fit of enthusiasm, what a great system they have, that they can monitor all the time that the pupils are on the computer, don't cheat and do this and that.

The expert's mention of "all the time" directs attention to the amount of visibility. Some experts noted that an increased amount of data can lead to the risk of surveillance. For Technology 4, in a dystopian situation, data would be collected "as a precautionary measure just in case something is discovered." Data Justice 12 stated that in a dystopia, "you collect a lot of data, and then there is a lot of surveillance. That's what the General Data Protection Regulation is trying to prevent, the formation of this kind of surveillance society." This expert noted and contradicted the logic of surveillance capitalism (Zuboff, 2015)—that is, of using accumulated data to predict and control human behaviour—by highlighting the role and purpose of the EU regulation to strengthen data protection rights.

5. Concluding Remarks: Negotiating the Future of the Welfare State Education System

The next data extract succinctly summarises the main findings of our study:

> Data collection [has] insane potential that can be used for very bad, or it can be used for very good. (Data Justice 3)

It reinforces the notion that data will play a strong role in the future of education while simultaneously addressing the potentials and pitfalls of datafication. While the interviewed experts imagined datafication as making education easier, maximising learning, and making education coherent and visible, they also imagined several consequences. The tensions between 1st and 2nd order imaginaries direct attention to the relationship between datafication and the key features of sustainability of the education system in the Finnish welfare state, such as the culture of trust and the strong role of the public sector.

5.1. From Trust to Accountability?

In light of our findings, trust—one of the key features of the Finnish education system (Välimaa, 2021)—becomes subject to negotiation in various ways. On one hand, trust relates to both the imagined visibility of datafication and to making education easier. On the other hand, education is trusted to enable social mobility. In these contexts, datafication may also have unexpected consequences.

Data technologies were imagined as ensuring that pupils across the country receive the same amount of instruction. This highlights the imagined visibility and coherence of education. In Finland, the minimum number of lessons for different subjects has already been defined by law. Pupils starting their nine-year basic education in autumn 2025 will undergo 1254 hours of mathematics lessons (Finnish Government, 2012). Thus, it appears that the imaginary is prompted by the perceived affordances of data technologies that enable more nuanced monitoring. Lesson time monitoring is also in contrast to the fact that there are no extensive accountability measures, such as the systematic school supervision system in Finland—a system that is datafied in other contexts (e.g., Dabisch, 2023). Gulson et al. (2022, p. 16) pointed out that data "transformations are closely tied to the reconfiguring of relations between trust, discretionary judgment, and personal and systemic accountability in education," and some experts have recognised that increased monitoring, which is experienced as mistrust, may reduce teachers' agency. Data-based monitoring can also influence power relations, creating asymmetry within a decentralised educational system. Although teachers may still have autonomy in local education decisions, increased data generation would lead to intensified monitoring, which may have the unintended consequence of reducing teachers' actual educational leeway. Thus, some aspects of the Finnish pedagogical culture might even fall apart, necessitating negotiations in the future to find balance.

Imaginaries that involve the use of technologies to solve problems on behalf of teachers (i.e., making things easier) give rise to questions about trust in teachers' agency in their own work. For instance, the imagined possibility of technology making education easier by reducing teachers' teaching workloads conflicted with the 2nd order imaginary of data work (see Hoeyer, 2023), which concerns the need for data technologies for maintenance and manual data generation, along with meta-level work for learning how to use these technologies. Instead of increasing the time available to teachers, data technologies may change where time is spent. Furthermore, some "technosolutions" (Bass, 2018, p. 37) can limit teachers' agency and pedagogical potentialities. Selwyn (2022), for instance, has discussed the consequences of monitorining pupils' attendance. The use of facial recognition system was stated to add 2.5 hours of weekly learning time for each pupil and free teachers from the burden of roll calls, thus making their job easier (1st order imaginary). Regarding the consequences, teachers identified that the system actually takes away many pedagogically meaningful opportunities in addition to routine manual work (2nd order imaginaries). Teachers described how the traditional roll-call routine enables them to interact with pupils and provides them with valuable information about pupils' readiness for a lesson. As one teacher in the study stated: "It's a pretty perfunctory process, but as a teacher you can use it in a whole range of ways that are actually about ensuring that the pupils are in the best place to learn" (Selwyn, 2022, p. 81). This practical example illustrates that if technology and its purpose are viewed too narrowly, it is easy to miss the complexity of the pedagogical realm. Moreover, to do the things that teachers have been doing during the traditional roll call, they need to create other new activities.

The question of trust also relates to making the future visible and controllable, which can increase rigidity in social mobility. Data technologies were imagined as enabling educational professionals to predict pupils' potential futures and make decisions to prevent school dropouts. Smithers (2023) addressed the mismatch of the imagined potentials of data-based identification and the consequences by arguing that "the promise and the function of predictive analytics is to freeze students' futures through a freezing of the past that takes as its object the permanent optimization of the present" (p. 110). This freezing was reflected in the 2nd order imaginaries, where datafication was feared to prevent people's mobility among social classes, regardless of their socioeconomic background, which went against the equity principle of welfare states. Predictions made exclusively based on existing data (indicating possibly unfavourable results) carry the risk of excluding futures that might otherwise be possible. Thus, to avoid unnecessarily closing off potential futures to people in advance, trust in the possibility of futures other than those predicted by data is needed.

5.2. From Public Goods to Private Benefits?

In Finland, formal education is provided by the public sector, which is in contrast to the more privatised systems even in other Nordic welfare states (Dovemark et al., 2018). The future imaginaries of datafication leave the current role of the public sector in education open to negotiation. This could mean integrating private services into the education system, which would expose pupils to the commercial digital ecosystem, or redefining the tasks, responsibilities, and role of the public sector.

On one hand, the tension between the public and private sectors is related to the division of labour between sectors. Although Finnish education is largely the responsibility of public actors, schools and education providers have the autonomy to choose the materials and services they use. As many data technologies used in schools are provided by private companies, their use represents the growing role of private influence within the public education system. Some interviewees suspected that these services may reflect a corporate perspective in addition to a pedagogical perspective. Others were concerned that the platformisation and concentration of services would make it difficult to break away from individual service providers, leading to "infrastructural dependency" (Cone & Lai, 2024). Some experts were hopeful that the public sector would get more involved in the development of technology, which represents a new avenue for the division of labour. In practice, this hopefulness is evident in the DigiOne development project mentioned earlier, which involves the municipalities, as public education providers, preparing a digital ecosystem for schools to use.

A renegotiated relationship between private and public interests is also related to pupils. Traditionally, welfare states have sought to protect pupils from commercial influences. The Finnish core curriculum states that "school and education must not be used as a channel for commercial influence" (EDUFI, 2014, p. 15). However, the intensifying integration of (privately produced) data technologies into everyday school practices may make it difficult to limit possible commercial influences. The digital tools and services used by teachers can collect a wide variety of data about their users, which companies can then use for purposes such as advertising, profiling, and optimising their own services (Lai et al., 2023). Given the diversity of actors and the changing roles, as mentioned in the previous paragraph, it is important that the backgrounds of the producers of data technologies are made transparent so that teachers and educational authorities have sufficient opportunities to assess the implications of the technologies they use and for the commercialisation of education. Technological development can be so rapid that all its consequences are not

sufficiently considered. As explained by Lai et al. (2023, p. 14), through "datafication, students' (school) life is commodified to an extent that has not yet been fully uncovered or debated within the context of the Danish welfare state." This indicates that more research and discussion on the future(s) of datafied education in welfare states is needed.

5.3. Sustainable Imagination Needs Inclusive Language

Future imagination is needed because educational systems face new and complex challenges. We agree with Nexø (2023, p. 1458) that, currently, "the dominant mood is not this [state of things] can never change, but, rather, this cannot go on." Concerning datafied education within a welfare state, this emphasises the need to imagine the direction and ways in which change occurs. Overall, the results illustrate that datafied education does not have a single imagined future but multiple. This diversity does not mean that the imaginaries presented here cover the whole range of possible futures. One limitation of our study was the scope of the interviews. The imaginaries we constructed from the data may have been different if we'd interviewed experts from additional relevant fields. After all, the future of datafied education is not only created by those who are involved particularly in datafication but by a broader constellation of stakeholders in education. In terms of sustainability, it would be beneficial to broaden our understanding of how local educational leaders, teachers, other professionals, and pupils themselves view the role and nature of datafication in the future. Would future imaginaries be shaped by themes similar to those in this study or by different ones?

One way to lower the threshold for participation in a discussion is to pay attention to the language and terminology used. So far, education policy as well as technology companies have maintained strong footholds in influencing how the future of education is discussed. This is reflected in strong and normative visions in education policy and technical features and affordances in educational technology. It is important to note that focusing on particular perspectives may emphasise them at the expense of others. For example, the Finnish education authority's prediction that data will be produced "everywhere" (EDUFI, 2019, p. 15) illustrates that central role that datafication is expected to play in the future. This imaginary has such powerful gravitation that it might be difficult to imagine education without various data technologies. On the flip side, if attention is focused on micro-level issues, such as the technological features and affordances of a single digital service or tool, complex structural issues of education may be neglected, or the language used may be so specialised that it excludes those who are actually affected. The question surrounding the future of datafied education is not just about technology but also about the potential and consequences of technology in the organisation and practice of education. Through our construction of sociotechnical imaginaries related to education (ease, coherence, maximised, and visible), we aim to create a space for an inclusive debate on the future of education, which could, in turn, contribute to strengthening sustainable development in education.

The value of imagining the future cannot necessarily be assessed based on accuracy. The complexity of the social world deems imaginaries to always be incomplete and contradictory to a certain extent. According to Jameson's (2005, p. xvi) idea of the "anti-anti-utopia," the strength of future imaginings may lie in their inconsistencies and limitations. Discussions about the future may be more relevant than trying to imagine a perfect picture. Imperfections can inspire new and unexpected ways of thinking, which can lead to sustainable decisions and actions that ultimately influence the way the future shapes up.

Funding

This work was supported by the Research Council of Finland, project no. 355523, "Movement for Data Literacy" (MODALITY). Publication of this article in open access was made possible through the institutional membership agreement between the University of Jyväskylä and Cogitatio Press.

Conflict of Interests

The authors declare no conflict of interests.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Supplementary Material

Supplementary material for this article is available online in the format provided by the author (unedited).

References

Bass, R. (2018). The impact of technology on the future of human learning. *Change: The Magazine of Higher Learning*, 50(3/4), 34–39. https://doi.org/10.1080/00091383.2018.1507380

Beer, D. (2019). *The data gaze: Capitalism, power and perception*. Sage Publications. https://doi.org/10.4135/9781526463210

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Burbules, N. C., Fan, G., & Repp, P. (2020). Five trends of education and technology in a sustainable future. *Geography and Sustainability*, *1*(2), 93–97. https://doi.org/10.1016/j.geosus.2020.05.001

Clarke, V., & Braun, V. (2013). Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *Psychologist*, *26*(2), 120–123.

Cone, L., & Lai, S. S. (2024). Infrastructural dependency in the datafied welfare state: The case of Google Chromebooks. *Learning, Media and Technology*, *50*(1), 44–60. https://doi.org/10.1080/17439884.2024.2437680

Cooper, D. (2013). *Everyday utopias: The conceptual life of promising spaces*. Duke University Press. https://doi.org/10.1215/9780822377153

Dabisch, V. (2023). The practices of data-based governance: German school supervision, professionalism and datafied structurations. *Tertium Comparationis*, *29*(1), 48–72. https://doi.org/10.31244/tc.2023.01.03

Dencik, L., Hintz, A., Redden, J., & Treré, E. (2019). Exploring data justice: Conceptions, applications and directions. *Information, Communication & Society*, *22*(7), 873–881. https://doi.org/10.1080/1369118X.2019.1606268

DigiOne. (2025). *Description of the DigiOne project*. https://www.digione.fi/digione-hanke

Dovemark, M., Kosunen, S., Kauko, J., Magnúsdóttir, B., Hansen, P., & Rasmussen, P. (2018). Deregulation, privatisation and marketisation of Nordic comprehensive education: Social changes reflected in schooling. *Education Inquiry*, *9*(1), 122–141. https://doi.org/10.1080/20004508.2018.1429768

EDUFI. (2014). *Perusopetuksen opetussuunnitelman perusteet 2014*. National Board of Education.

EDUFI. (2019). *Osaaminen 2035: Osaamisen ennakointifoorumin ensimmäisiä ennakointituloksia*. National Agency for Education.

European Commission. (2020). *Digital education action plan 2021-2027: Resetting education and training for the digital age*.

Fagerlund, J., Leino, K., Niilo-Rämä, M., Puhakka, E., & Markkanen, I. (2024). *Kohti digiosaamisen strategista kehittämistä: Kansainvälinen monilukutaidon ja ohjelmoinnillisen ajattelun tutkimus* (ICILS 2023 Report). University of Jyväskylä, Institute for Educational Research. https://doi.org/10.17011/ktl-t/40

Finnish Government. (2012). *Government decree on the national objectives of education under the basic education act and on the distribution of basic education lessons, 422/2012*.

Finnish Ministry of Education and Culture. (2023). *Desired outcomes for digitalisation of early childhood education and care, pre-primary, primary and lower secondary education*.

Gill, S. L. (2020). Qualitative sampling methods. *Journal of Human Lactation*, *36*(4), 579–581. https://doi.org/10.1177/0890334420949218

Grek, S., Maroy, C., & Verger, A. (Eds.). (2020). *World yearbook of education 2021: Accountability and datafication in the governance of education* (1st ed.). Routledge. https://doi.org/10.4324/9781003014164

Gulson, K., Sellar, S., & Webb, T. (2022). *Algorithms of education: How datafication and artificial intelligence shape policy*. University of Minnesota Press.

Halinen, I. (2018). The new educational curriculum in Finland. In M. Matthes, L. Pulkkinen, C. Clouder, & B. Heyes (Eds.), *Improving the quality of childhood in europe* (pp. 75–89). Alliance for Childhood European Network Foundation.

Hoeyer, K. (2023). *Data paradoxes: The politics of intensified data sourcing in contemporary healthcare*. The MIT Press.

Jameson, F. (2005). *Archaeologies of the future: The desire called utopia and other science fictions*. Verso.

Jasanoff, S. (2015). Future imperfect: Science, technology, and the imaginations of modernity. In S. Jasanoff & S.-H. Kim (Eds.), *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226276663.001.0001

Kaarakainen, S.-S., & Kaarakainen, M.-T. (2018). Tulevaisuuden kansalaisia rakentamassa. *AFinLAn Vuosikirja*, *76*(1), 22–40. https://doi.org/10.30661/afinlavk.69269

Kortekangas, O., Paksuniemi, M., & Ervast, H. (2019). Milestones of basic education in Finland: Pedagogy, structure and language. In M. Paksuniemi & P. Keskitalo (Eds.), *Introduction to the Finnish educational system* (pp. 14–27). Brill.

Lai, S. S., Andelsman, V., & Flensburg, S. (2023). Datafied school life: The hidden commodification of digital learning. *Learning, Media and Technology*, *49*(3), 1–17. https://doi.org/10.1080/17439884.2023.2219063

Lamb, G. (1903). Is the elementary course of study overcrowses? *The Journal of Education*, *57*(5), 67–68.

Lavonen, J. (2017). Governance decentralisation in education: Finnish innovation in education. *Revista de educación a distancia (RED)*, *53*. https://doi.org/10.6018/red/53/1

Marcetic, H., & Nolin, J. (2023). Utopian and dystopian sociotechnical imaginaries of Big Data. *Journal of Digital Social Research*, *5*(4), 93–125. https://doi.org/10.33621/jdsr.v5i4.180

Mertala, P. (2020). Paradoxes of participation in the digitalization of education: A narrative account. *Learning, Media and Technology*, *45*(2), 179–192. https://doi.org/10.1080/17439884.2020.1696362

Mertanen, K., & Brunila, K. (2024). Fragile utopias and dystopias? Governing the future(s) in the OECD youth education policies. *Globalisation, Societies and Education*, *22*(5), 942–953. https://doi.org/10.1080/14767724.2022.2121687

Messina, N. (2023). Sustainable development: A comparison between the Finnish and the Italian education systems. *Sustainability*, *15*(10), Article 8077. https://doi.org/10.3390/su15108077

Meuser, M., & Nagel, U. (2009). The expert interview and changes in knowledge production. In A. Bogner, B. Littig, & W. Menz (Eds.), *Interviewing experts* (pp. 17–42). Palgrave Macmillan. https://doi.org/10.1057/9780230244276_2

Nexø, T. A. (2023). Utopia against the welfare state: Rethinking utopia in an age of reproductive crises. *Textual Practice*, *37*(9), 1456–1474. https://doi.org/10.1080/0950236X.2023.2248796

OECD. (2020). *Back to the future(s) of education: The OECD schooling scenarios revisited*. https://doi.org/10.1787/178ef527-en

OECD. (2023). *OECD digital education outlook 2023: Towards an effective digital education ecosystem*. https://doi.org/10.1787/c74f03de-en

OECD. (2025). *Smart data and digital technology in education: Artificial intelligence, learning analytics and beyond*. https://www.oecd.org/en/about/projects/smart-data-and-digital-technology-in-education--artifical-intelligence,-learning-analytics-and-beyond.html#publications

Oinas, S., & Hotulainen, R. (2024). Havaintoja ja yläkoululaisten ajatuksia digitalisaatiosta oppitunneilla. In S. Oinas & M.-P. Vainikainen (Eds.), *Digitalisaatio oppimisen ja oppimistulosten selittäjänä* (pp. 427–452). Suomen kasvatustieteellinen seura.

Read, S., Hietajärvi, L., & Salmela-Aro, K. (2022). School burnout trends and sociodemographic factors in Finland 2006–2019. *Social Psychiatry and Psychiatric Epidemiology*, *57*(8), 1659–1669. https://doi.org/10.1007/s00127-022-02268-0

Ross, J., & Wilson, A. (2023). Reconfiguring surveillance futures for higher education using speculative data stories. In N. B. Dohn, J. Jaldemark, L.-M. Öberg, M. Håkansson Lindqvist, T. Ryberg, & M. De Laat (Eds.), *Sustainable networked learning* (pp. 19–33). Springer Nature. https://doi.org/10.1007/978-3-031-42718-3_2

Selwyn, N. (2016). The dystopian futures. In N. Rushby & D. W. Surry (Eds.), *The Wiley handbook of learning technology* (1st ed., pp. 542–556). Wiley. https://doi.org/10.1002/9781118736494.ch28

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Selwyn, N. (2022). Less work for teacher? The ironies of automated decision-making in schools. In S. Pink, M. Berg, D. Lupton, & M. Ruckenstein (Eds.), *Everyday automation: Experiencing and anticipating emerging technologies* (1st ed., pp. 73–86). Routledge. https://doi.org/10.4324/9781003170884

Selwyn, N., Pangrazio, L., Nemorin, S., & Perrotta, C. (2020). What might the school of 2030 be like? An exercise in social science fiction. *Learning, Media and Technology*, *45*(1), 90–106. https://doi.org/10.1080/17439884.2020.1694944

Silvennoinen, E., Tedre, M., & Valtonen, T. (2024). Datafikoituva peruskoulu—tasapainoilua lapsen henkilötietojen suojan ja opetuksen tavoitteiden välillä. *Lakimies*, *5*, 655–678.

Smithers, L. (2023). Predictive analytics and the creation of the permanent present. *Learning, Media and Technology*, *48*(1), 109–121. https://doi.org/10.1080/17439884.2022.2036757

THL. (2023). *Well-being of children and young people—School health promotion study*. Finnish Institute for Health and Wellbeing. https://thl.fi/tutkimus-ja-kehittaminen/tutkimukset-ja-hankkeet/kouluterveyskysely/kouluterveyskyselyn-tulokset#alueittain

Toom, A., & Husu, J. (2016). Finnish teachers as 'makers of the many.' In H. Niemi, A. Toom, & A. Kallioniemi (Eds.), *Miracle of education* (pp. 41–55). Springer Nature. https://doi.org/10.1007/978-94-6300-776-4_3

UN. (2015). *Transforming our world: The 2030 agenda for sustainable development*.

Välijärvi, J., & Sulkunen, S. (2016). Finnish school in international comparison. In H. Niemi, A. Toom, & A. Kallioniemi (Eds.), *Miracle of education* (pp. 3–21). Springer Nature. https://doi.org/10.1007/978-94-6300-776-4_1

Välimaa, J. (2021). Trust in Finnish education: A historical perspective. *European Education*, 53(3/4), 168–180. https://doi.org/10.1080/10564934.2022.2080563

Williamson, B. (2015, December 10). *Smarter learning software: Education and the big data imaginary* [Paper presentation]. Big Data—Social Data, University of Warwick, UK.

Zomer, C. (2024). How 'much' engaged are you? A case-study of the datafication of student engagement. *Learning, Media and Technology*. Advance online publication. https://doi.org/10.1080/17439884.2024.2437547

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

About the Authors

Lauri Palsa is a postdoctoral researcher at the Department of Teacher Education, University of Jyväskylä, Finland. His research interests focus on datafication and emerging and contextualised literacies, such as data literacy, media literacy, and multiliteracies, both within and outside formal education.

Janne Fagerlund is a postdoctoral researcher at the Department of Teacher Education, University of Jyväskylä, Finland. His research interests are in education related to digital topics such as programming, computational thinking, artificial intelligence, and datafication. He is also actively involved in developing pedagogical approaches through teacher training and school activities.

Pekka Mertala is an associate professor in the Department of Teacher Education at the University of Jyväskylä, Finland. His main research interests focus on critical studies of the role of digital technologies in both formal (i.e., school) and non-formal (i.e., hobby-related) learning contexts.

SOCIAL INCLUSION
ISSN: 2183-2803

Social Inclusion is a peer-reviewed open access journal which provides academics and policymakers with a forum to discuss and promote a more socially inclusive society.

The journal encourages researchers to publish their results on topics concerning social and cultural cohesiveness, marginalized social groups, social stratification, minority-majority interaction, cultural diversity, national identity, and core-periphery relations, while making significant contributions to the understanding and enhancement of social inclusion worldwide.

cogitatio

www.cogitatiopress.com/socialinclusion