Politics and Governance
2026 • Volume 14 • Article 11234
https://doi.org/10.17645/pag.11234

cogitatio

ARTICLE

Open Access Journal

# In Pursuit of Informed Voters: Three Experimental Studies on Enhanced Voting Advice Applications

Naomi Kamoen [1], Christine Liebrecht [1], and Rieke van Lieshout [2]

[1] Department of Communication and Cognition, Tilburg University, The Netherlands
[2] Faculty of Arts, Radboud University, The Netherlands

**Correspondence:** Naomi Kamoen (n.kamoen@tilburguniversity.edu)

**Submitted:** 29 August 2025 **Accepted:** 27 November 2025 **Published:** 11 February 2026

**Issue:** This article is part of the issue "Voting Advice Applications: Methodological Innovations, Behavioural Effects, and Research Perspectives" edited by Diego Garzia (University of Lausanne / Bologna University), Stefan Marschall (Heinrich Heine University Düsseldorf), Mathias Wessel Tromborg (Aarhus University), and Andreas Albertsen (Aarhus University), fully open access at https://doi.org/10.17645/pag.i485

## Abstract

Voters frequently struggle to understand political attitude statements in voting advice applications (VAAs) and often invest limited effort in resolving these difficulties. Conversational agent VAAs (CAVAAs) aim to reduce the cognitive effort involved in searching for relevant information by integrating chatbots that can provide contextual support. This article presents findings from three studies comparing CAVAAs to standard VAAs without additional information (Studies 1 and 2) and to a VAA with static clickable information (VAA+, Study 3). Study 1 ($N$ = 93) was a laboratory experiment conducted during the 2023 Dutch parliamentary elections. University students were assigned to a standard VAA or a CAVAA. The chatbot in the CAVAA was used in approximately 45% of cases, with users showing a preference for opinion-based and status quo information. Compared to the VAA, CAVAAs reduced non-directional responses to the VAA statements and increased evaluations of the tool's usability. Moreover, users' perceived knowledge was higher, whereas no differences were found for factual knowledge and turnout intention. Study 2 ($N$ = 144) largely replicated these results outside the lab with a more diverse sample and showed that the effects hold across different levels of political sophistication. Study 3 ($N$ = 159), conducted during the 2024 European elections, compared a CAVAA to a VAA+. While VAA+ users requested information more frequently, both tools received similar evaluations, and this finding was again consistent across groups of different political sophistication. In the manuscript, we discuss the implications of these findings for theory and practice.

## Keywords

conversational agents; political knowledge; political sophistication; voting advice applications

## 1. Introduction

The 2022 Dutch municipal elections saw a historically low turnout, with only 50.99% of eligible voters voting. In a letter to the House of Representatives, the Minister of Interior and Kingdom Relations identified a lack of political knowledge as one of the primary reasons for this outcome (Bruins Slot, 2022). This conclusion aligns with Delli Carpini and Keeter's (1996) seminal work, which argues that a lack of political knowledge is among the most important reasons to abstain from voting.

Low levels of political knowledge and voter turnout are not exclusive to Dutch municipal elections; these issues are also prevalent in other countries and types of elections (e.g., Fivaz & Nadig, 2010). In today's increasingly complex political landscapes—particularly in multiparty systems like the Netherlands—voters often struggle to make informed decisions. One reason for this is the increasing number of political parties competing in elections. To illustrate this, the 2021 Dutch parliamentary elections featured 37 parties, 20 of which were new (Sipma et al., 2021). Other contributing factors include a growing sense of distance between citizens and politics, declining motivation to seek out information (see the literature review in Kamoen et al., 2015), and the rising influence of misinformation complicating voters' ability to assess the reliability of political content (Mason et al., 2018).

Parallel to these developments runs the steep rise of voting advice applications (VAAs) as a source of political information (Garzia & Marschall, 2012). In a VAA, voters express their opinions on approximately 30 political statements. The tool then compares the users' stances to those of the political parties or candidates, depending on what is relevant for the elections at stake, and calculates the strength of the match with each. The goal of a VAA, however, is not to calculate matches and make suggestions about whom voters could vote for. Instead, a VAA is an information resource aiming to provide reliable information about the political parties and their stances (De Graaf, 2010).

Empirical studies suggest that VAAs indeed increase users' (perceived) political knowledge and likelihood to vote (Munzert & Ramirez-Ruiz, 2021), especially among theoretically educated users (Heinsohn et al., 2019). At the same time, however, users have been found to encounter comprehension difficulties while interacting with these tools (Kamoen & Holleman, 2017). Although additional information to resolve these comprehension problems is available online, users rarely search for it and instead rely on assumptions to fill in the gaps of information.

To help VAA users obtain reliable political information at low cognitive costs, researchers have introduced a new type of VAA that includes a conversational agent (CA) in the form of a chatbot that is trained to respond to user queries related to political attitude statements in the VAA (e.g., Dieing, 2025; Hankel et al., 2024; Kamoen & Liebrecht, 2022; Van Veggel et al., 2025; Van Zanten & Boumans, 2023). Although CAVAAs, to the best of our knowledge, have not (yet) been launched in real-life elections, a first experimental comparison between a CAVAA and a VAA showed that the former yielded better user evaluations and higher political knowledge scores (Kamoen & Liebrecht, 2022). While these results are promising, the study had certain limitations, particularly in terms of ecological and construct validity, highlighting the need for further research.

Therefore, the present series of three studies further explores the usage of chatbots in the context of VAAs and builds on this earlier work by addressing its limitations and extending it in several ways. More specifically,

Study 1 compares a CAVAA to a VAA among theoretically-educated voters, using construct-valid experimental materials in the context of the 2023 Dutch parliamentary elections. Study 2 makes the same comparison, but among a diverse user audience. Finally, while Studies 1 and 2 compare a CAVAA to a VAA without supplementary information, Study 3 contrasts it with an enhanced VAA (VAA+) to understand if adding additional information via buttons in the standard web environment works as well as adding information via a chatbot. Across all studies, both tool evaluation measures (ease of use, usefulness, and playfulness) and political engagement measures (perceived and factual political knowledge, internal political efficacy, and turnout intention) are assessed as outcome measures. Additionally, the response behavior to VAA statements is included as an unobtrusive indicator of comprehension difficulties. As such, the aim is to advance theories of information processing in contexts where citizens invest minimal effort, while also providing insights for VAA practice regarding whether and how chatbots can be effectively integrated into this context.

## 2. Theoretical Framework

### *2.1. Comprehension Problems in VAAs*

In a large-scale think-aloud study, Kamoen and Holleman (2017) found that VAA users encounter comprehension difficulties with roughly one in five statements. These comprehension problems ranged from difficulties with word meaning ("What is the VAT rate?"), to a lack of knowledge about the status quo ("How high is the VAT rate?"), to missing contextual arguments ("Why increasing the VAT rate?"). While VAAs aim to inform users (De Graaf, 2010), they do not always provide structured background information to resolve such problems. As a result, users must consult search engines or other external resources to fill these gaps. VAA users, however, have been found to display so-called satisficing behavior, opting to answer with limited effort, and thus limited information, rather than seeking clarification (Kamoen & Holleman, 2017).

The concept of satisficing stems from Krosnick's satisficing-optimizing model (Krosnick, 1991; Roberts et al., 2019), which explains variation in how respondents answer questions in surveys. While some aim to match their true opinion (optimizers), others opt for the quickest plausible answer (satisficers). Satisficing behavior is more likely when tasks are difficult, motivation is low, and when respondents have limited cognitive ability (Krosnick, 1991). While VAA users are motivated enough to enter a VAA survey voluntarily, these users, too, have been found to satisfice. In a VAA context, the satisficing behavior manifests in at least two ways. First, when facing comprehension problems, users tend to make assumptions rather than look up information (Kamoen & Holleman, 2017). Second, it amounts to the interchangeable usage of the "neutral" and "no opinion" response option to the VAA statements (Baka et al., 2012; Van Outersterp et al., 2016). This latter behavior is particularly problematic for those VAAs that assign different value to these responses in result calculations; for example, in the Dutch Stemwijzer, neutral answers are matched to parties, while no opinions are excluded from the calculation of the advice (De Graaf, 2010). As the voting advice has been found to affect the vote choice made (Wall et al., 2014), the interchangeable usage of these response options may lead to distorted matches and, ultimately, biased voting decisions.

### *2.2. Conversational Agent Voting Advice Applications*

The comprehension challenges that VAA users face, combined with their limited motivation to consult external sources, provide two key reasons for integrating additional information directly into the VAA

interface. Researchers have therefore recently proposed integrating a CA in the form of a chatbot into a VAA, resulting in what is known as a Conversational Agent Voting Advice Application (CAVAA; e.g., Kamoen & Liebrecht, 2022).

Chatbots can take different forms, but most previous CAVAA studies (Hankel et al., 2024; Kamoen & Liebrecht, 2022; Van Veggel et al., 2025) have included buttons within the chatbot interface to access frequently asked questions. For example, for a statement like "The VAT rate on fruit should be increased," users could click a button to learn what "VAT rate" means, another one to access information about the current rate, and a third one to learn about the arguments used in the political debate. In addition, users could type questions in an open chat. AI technology was then used to recognize these questions and return predetermined researcher-controlled responses (e.g., "Party X supports VAT reduction"). In case of unrecognized input, the chatbot would return an error message (e.g., "I do not understand your question").

The theoretical advantage of such a chatbot, relative to a VAA without additional information, is that it mitigates users' tendency to minimize effort. Since the information is embedded within the tool, users do not need to switch screens or search elsewhere. Moreover, the information is tailored to user needs: for example, those interested in the stances of Party A are not shown irrelevant information about Party B. Because of these presumed advantages, Kamoen and Liebrecht (2022) hypothesized that a CAVAA would be evaluated better than a VAA without information on tool evaluation measures (higher perceived usefulness, ease of use, and playfulness), as well as on political engagement measures (increased perceived and factual political knowledge, and turnout intention). In their Study 1, they indeed observed higher tool evaluation scores as well as higher perceived and factual political knowledge scores for the CAVAA. Counter to expectations, however, the CAVAA did not increase turnout intention.

While promising, a threat to the construct validity of that study was that the VAA and CAVAA had different layouts, which may have contributed to the observed differences. Furthermore, the study was conducted outside of an election context, whereas VAAs are typically used in the run-up to elections (Van de Pol et al., 2014). Therefore, the current Study 1 re-examines the hypotheses formulated by Kamoen and Liebrecht (2022), now using ecologically valid materials within the context of the 2023 Dutch parliamentary elections.

> H1: Compared to users of a traditional VAA without information, CAVAA users will provide more positive tool evaluations.

> H2: Compared to users of a traditional VAA without information, CAVAA users will be more politically engaged.

In addition to comparing VAAs to CAVAAs on these measures, the current research introduces the users' response behavior to political attitude statements as an outcome measure. Based on previous studies (Baka et al., 2012; Van Outersterp et al., 2016), we expect that CAVAA users will provide neutral and no-opinion answers (jointly referred to as non-directional answers) less frequently to the political attitude statements, as these answers are both a sign of the comprehension difficulties and of the satisficing behavior CAVAAs aim to combat.

> H3: Compared to users of a traditional VAA without information, CAVAA users will provide non-directional answers to VAA statements less frequently.

### 2.3. Political Sophistication

While Study 1 will test H1–H3 for theoretically educated voters, the goal of Study 2 is to compare the effect of CAVAAs versus VAAs for two types of users: those with higher and lower levels of political sophistication. Political sophistication is defined (e.g., Luskin, 1990) as a combination of cognitive ability (often proxied by education level), existing political knowledge (operationalized by a knowledge test), and motivation to acquire political information (typically operationalized as political interest). Distinguishing between users with higher and lower levels of sophistication is relevant because these groups are likely to differ in several respects, such as in the number and type of comprehension problems they encounter, their ability to recognize and articulate such problems (Elling et al., 2012), and their susceptibility to satisficing behavior (Krosnick, 1991). Moreover, real-world VAAs attract a broad user base (Van de Pol et al., 2014), making this distinction practically meaningful too.

In their study, Kamoen and Liebrecht (2022) also explored differences between users with higher and lower political sophistication. They did not formulate a hypothesis regarding this interaction, as theoretical arguments could be made in both directions: on the one hand, lower-sophisticated users might benefit more from CAVAAs due to experiencing more comprehension difficulties; on the other hand, these users might struggle to recognize their information needs or to formulate appropriate questions for the chatbot, which could limit their ability to benefit from the tool. Results of their study found no interaction effect between user group and tool type on both tool evaluation measures and political engagement measures, suggesting that CAVAAs may be equally effective for different types of users. However, the authors noted that their sample was skewed toward theoretically-educated participants (80%) and recommended replication with a more diverse population.

Therefore, Study 2 compares a CAVAA and a VAA among a heterogeneous population with more variation in educational level. Following earlier research, we do not formulate an hypothesis about the interaction of the effect of tool type and the level of political sophistication, but instead explore the relationship. Moreover, to better understand any possible effects of political sophistication, we will also explore any possible differences in the amount of information requested in the CAVAA between users with higher and lower levels of political sophistication.

### 2.4. Providing Information Within an Enhanced VAA+

While some VAA brands, such as the Dutch Kieskompas and the German Wahl-O-Mat, typically present only political attitude statements, other VAA brands offer additional information along with the political attitude statements in their VAAs. The Dutch VAA StemWijzer, for example, offers clickable explanations of key terms, a “Do you want to know more?” button that provides background information for each statement and a “party stances” section where users can view parties’ positions along with their justifications. Similar features are found in other European tools, such as Smartwielen in Luxembourg and Smartvote in Switzerland. This variation raises a question whether the CAVAA continues to provide added value when it is compared with a so-called VAA+, which offers political background information in a more conventional format through clickable buttons in the web environment.

This question is relevant for VAA design practice, but also from a theoretical perspective. Research on CAVAAs can be linked to Krosnick's (1991) satisficing-optimizing model from survey research, as well as to the task-technology fit (TTF) theory from the field of information systems (Goodhue & Thompson, 1995). According to TTF, technology is most effective when its functionalities align with both the demands of the task and the characteristics of the user. In the context of VAAs, the task of forming an opinion on complex political statements can be cognitively demanding, especially for users with lower levels of political sophistication. While both CAVAAs and VAA+s aim to support understanding by offering information, they differ fundamentally in structure and interactivity, so on the level of the "technology" in the TTF. VAA+s present predefined, static content through a familiar interface with clickable explanations. In contrast, CAVAAs enable dynamic information retrieval, which may be more flexible but also more cognitively demanding. Study 3 explores how the format used to provide background information, chatbot (CAVAA) versus clickable web interface (VAA+), may affect user evaluation measures and political engagement measures, and whether this effect depends on users' political sophistication.

## 3. Method Study 1

### *3.1. Design*

Study 1 was an experimental study conducted in the run-up to the 2023 Dutch parliamentary elections. Participants were randomly assigned to a traditional VAA without supplementary information or a CAVAA. In the original study set-up, there were two CAVAA presentation forms that we offered: either the chatbot was opened by default (proactive) or it was available through a mouse click (reactive). For consistency between the three studies, we only report on the comparison between the proactive CAVAA and the VAA; a comparison between all three conditions is reported in Kamoen et al. (2024). After using a VAA or a CAVAA, participants completed a survey that included both tool evaluation measures and political engagement measures. In addition, for each VAA statement, we recorded whether participants provided a directional or non-directional response to the VAA statements, and we logged which information was requested via the CA. The study was approved by the Research Ethics and Data Management Committee of the Tilburg School of Humanities and Digital Sciences at Tilburg University (reference number: REDC 2023.13a).

### *3.2. Materials*

#### 3.2.1. The VAA

The VAA was a copy of the Dutch StemWijzer. This copy contained 30 statements that VAA developer ProDemos created to represent the relevant issues in the Dutch political environment. In the tool, users are to indicate to what degree they agree with 30 political statements (e.g., "People aged 65 and older should be able to travel for free on trains") using a three-point scale (agree/neither agree nor disagree/disagree), or they may indicate that they want to skip the question. After having answered all statements, a "voting advice" is provided, depicted as a percentage of agreement with the parties running in the elections. The three most agreed-upon parties are highlighted at the top of the page.

The VAA differed from the original StemWijzer only at two points. First, the additional information about each VAA statement that was provided in the original was deleted from the VAA, as the goal was to make a comparison with an empty VAA. Second, the name, logo, and house style of the original StemWijzer were altered so that it was clear to participants that they were filling out an experimental VAA. ProDemos granted permission to use the copy in this altered form for the current study (Figure 1).

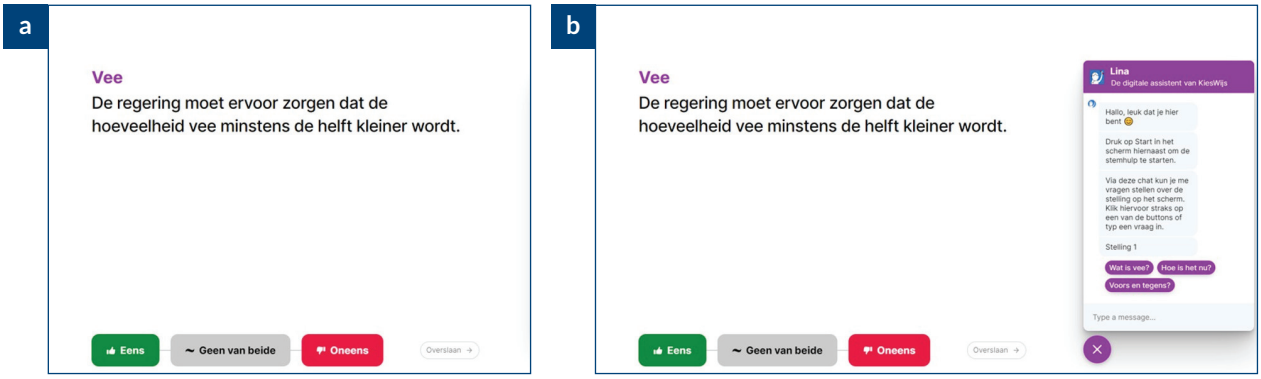


**Figure 1.** Screenshot of the VAA (a) and CAVAA (b). Notes: The header reads "Cattle" and the statement is "The government must ensure that livestock numbers are reduced by half." In the CAVAA, the chatbot introduced itself and then provided three buttons and the open chat.

### 3.2.2. The CAVAA

In the CAVAA, a CA in the form of a chatbot was added to the VAA. The chatbot introduced itself on the first page of the tool and invited users to ask questions about the political attitude statements. For each statement, users could click on one of three suggestion buttons to access additional information (Figure 1b). The first button read "What is X?," where "X" was replaced by a key term (such as "Cattle"). The second button read "What is the situation now?" and clicking would lead to information about the current state of the issue. The third button read "Pros and cons?" and clicking would open a balanced overview of no more than three arguments per side. These pro-con arguments were based on the arguments the political parties mentioned most frequently for and against each statement. We chose to always present an equal number of arguments pro and con, as users may otherwise be led by the number of arguments.

Besides using the suggestion buttons, users could request information by typing a question in the open chat window. An AI model was trained using open-source software Rasa to answer such user questions. In deciding which output to provide, the chatbot could recognize various intents: Semantic questions ("What is/are…?"), pragmatic questions ("How much…?"), reasons ("Why?"), party stances ("What does Y think?"), lead candidates ("Who's the leader of Y?"), and all combinations of pros and cons ("What are the pros/cons/pros and cons"). Moreover, the chatbot could recognize VAA-related questions ("What does neither agree nor disagree mean?"), and it was able to respond to conversation-related expressions (reactions to "Hello"). Finally, the chatbot was also equipped with repair strategies in case it did not understand the question asked or in case the user said something rude.

While we used AI to recognize the user intent, no AI was used to generate the chatbot responses. Instead, the responses were written by humans. Information in the CA was based on reliable information resources (e.g., dictionaries and government websites) and was cross-checked by all members of the research team to

ascertain that the information was as correct and objective as possible (for a demonstration, see https://surfdrive.surf.nl/files/index.php/s/UiuTsI3WDtdlCOq).

### *3.3. Measures*

#### 3.3.1. Tool Evaluation Measures

Participants' evaluation of the tool was measured in a survey using a set of 7-point Likert-scales with questions measuring three aspects of the technology acceptance model (see Davis, 1989; Marangunić & Granić, 2015), namely ease of use (e.g., "I found the voting advice application easy to use"), usefulness (e.g., "This voting advice application was helpful"), and playfulness (e.g., "Using the voting advice application was fun"). These items were based on the items used in Kamoen and Liebrecht (2022; Study 2), who tailored the original survey questions of Ahn et al. (2007) to fit the context. The reliability of all scales was acceptable (ease of use: Cronbach's $\alpha = .79$; usefulness: $\alpha = .66$; playfulness: $\alpha = .84$).

#### 3.3.2. Political Engagement Measures

We measured several different aspects of political engagement. First, participants' perceived political knowledge after tool usage was measured on a 7-point Likert-scale with three statements (e.g., "Using this voting advice application helps me better understand political parties' positions"). The same items were also used in the study of Kamoen and Liebrecht (2022, Study 1), who based these items on Ladner (2012). The reliability of this scale was acceptable only if the item that measured "the motivation to learn more about politics" was excluded ($\alpha = .74$).

Second, factual political knowledge was measured by letting participants judge seven statements about the political issues in the (CA)VAA, such as "In the Netherlands, currently no $CO_2$ is stored under the surface. True–false–don't know." For the analysis, an index of the number of correct answers was analyzed; "don't know" answers were scored as incorrect.

Third, internal political efficacy was measured using a single item: "I feel sufficiently informed to vote in the upcoming parliamentary elections." Finally, voting intention in terms of turnout was assessed with the item: "I am planning to vote in the elections of the Parliament on November 22, 2023." These questions were both based on Kamoen and Liebrecht (2022), who based these measures on Glynn et al. (2009).

#### 3.3.3. Response Behavior

In both conditions, participants responded to the political attitude statements on a three-point scale, with an additional fourth "skip this question" option. Due to privacy regulations, we could not access participants' exact responses but were permitted to record whether each answer was directional (agree/disagree) or non-directional (neutral/skip). To compare the proportion of non-directional responses, we estimated a cross-classified multilevel logistic model with random effects for both items and participants (Quené & Van den Bergh, 2004, 2008). A contrast test ($\chi^2$) was used to compare means in the fixed part.

### 3.4. Participants

Data were gathered between 31 October and 7 November 2023, about three weeks prior to the Dutch parliamentary elections on 22 November. We asked students from the participant pool of Tilburg University to participate in exchange for participant credits.

In total, 93 participants took part. Of them, 33 identified as male (32.4%) and 60 as female (66.9%). The average age was 20.89 years ($SD = 2.40$), ranging from 18 to 28. Almost all participants were students at the university level (97.8%), one participant preferred not to say (1.1%), and one was a student at the level of a higher vocational education (1.1%). All participants were allowed to vote in the upcoming elections. We compared the experimental groups for age and gender and found no differences (Gender: $\chi^2(1) = 0.31$, $p = .39$; Age: $t(91) = 0.77$, $p = .22$), which indicates that there is no reason to assume the groups differed *a priori*.

### 3.5. Procedure

Participants were seated in soundproof research cabins of the Dante laboratory at Tilburg University and were instructed on how to navigate between two tabs: one in which the tool (VAA or CAVAA) was open, and another tab that featured a survey. Participants started in the survey tab, where they read the instructions for how to fill out the (CA)VAA and, depending on the assigned condition, how to use the chatbot. After providing informed consent, they were instructed to go to the next tab in the browser. Here, they used the (CA)VAA, and after having received voting advice, participants were instructed to return to the previous tab to fill out the survey.

The survey measured tool evaluation and political engagement. Moreover, participants were also asked several other questions in the survey, including demographic variables and their general political interest. After completion, we informally asked the participants a couple of questions orally, mainly about their experience with the (CA)VAA. The purpose of this was to gather qualitative feedback on the workings of the tool. After having answered these interview questions, participants were thanked for their participation.

## 4. Results Study 1

Before discussing the results related to tool evaluation measures, political engagement measures, and the users' response behavior, we will first provide descriptive statistics related to the usage of the chatbot in the CAVAA.

### 4.1. Descriptive Statistics: Chatbot Usage

In the CAVAA condition, we monitored how participants ($N = 46$) used the chatbot. To quantify usage, we calculated the proportion of respondent-item combinations in which at least one suggestion button was clicked. Across all participants and items, the chatbot was used in 44.9% of the cases. The most frequently used button was "Pros and cons" (33.0%), followed by "What is the situation now?" (24.0%). The "What is X?" button, explaining terminology, was the least used (5.2%).

Participants submitted a total of 54 open-ended questions via the chatbot. In eight instances, multiple questions were asked by the same participant for the same VAA item, resulting in an overall open-chat usage rate of 3.3%. User questions were about various topics, and most of them related to party positions ($N = 16$). The chatbot provided correct responses to 48.1% of all queries.

### 4.2. Tool Evaluation and Political Engagement Measures

Table 1 shows the scores for the tool evaluation and political engagement measures for the VAA and the CAVAA.

**Table 1.** Means and standard deviations on a scale from 1 (*low score*) to 7 (*high score*) in Study 1.

| Tool | Ease of use | Usefulness* | Playfulness | Perceived knowledge* | Factual knowledge | Internal political efficacy * | Turnout intention |
|---|---|---|---|---|---|---|---|
| CAVAA | 6.49 (0.76) | 5.93 (0.83) | 5.51 (1.06) | 4.68 (1.26) | 3.11 (1.45) | 4.78 (1.78) | 6.63 (0.88) |
| VAA | 6.18 (0.96) | 5.48 (0.94) | 5.52 (0.77) | 4.12 (1.27) | 3.02 (1.21) | 3.72 (1.66) | 6.38 (1.11) |

Note: * = CAVAA > VAA with $p < .05$

As for tool evaluations, results show that the CAVAA is more useful than the VAA ($t(91) = 2.49$; $p = .014$, Cohen's $d = .51$). The size of this effect can be classified as medium-sized relative to the standard deviation. The differences for ease of use are not significant, although there is a tendency for an effect in the hypothesized direction with a small effect size ($t(91) = 1.68$; $p = .097$, Cohen's $d = .36$). No differences are observed for playfulness ($t(91) = 0.02$; $p = .99$).

Regarding the political engagement measures, results indicate that CAVAA users report higher levels of perceived political knowledge, and their internal political efficacy is also higher ($t(91) = 2.17$, $p = .03$, Cohen's $d = .44$; $t(91) = 2.97$, $p = .004$, Cohen's $d = .62$). The effects are small and medium-sized relative to the standard deviation. However, no significant differences are found for factual political knowledge or the extent to which participants were planning to vote ($t(91) = 0.32$, $p = .75$; $t(91) = 1.19$, $p = .24$).

### 4.3. Response Behavior

In the CAVAA condition, users provided non-directional answers in 13.6% of the respondent-item combinations, compared to 18.8% in the VAA condition. A contrast test indicates that this difference is statistically significant and large in terms of effect size: respondents gave fewer non-directional answers when using the CAVAA compared to the VAA ($\chi^2(1) = 4.33$, $p = .04$, Cohen's $d_{items} = 1.05$).

### 4.4. Summary

Results suggest that the CAVAA outperforms the VAA on several dimensions. In terms of tool evaluation, the CAVAA was rated higher on ease of use and usefulness, though not on playfulness. Regarding political engagement measures, CAVAA users reported higher perceived knowledge, and their internal political efficacy

was higher, but no significant differences were found in factual knowledge or turnout intention. Additionally, CAVAA users provided non-directional answers less frequently than VAA users.

## 5. Method Study 2

### 5.1. Design

Study 2 had a 2 (tool type: VAA or CAVAA) × 2 (user type: higher or lower levels of political sophistication) between-subjects design and was again conducted in the run-up to the 2023 Dutch parliamentary elections. This study aimed to replicate Study 1, but now with a more diverse sample of voters, and to explore whether the effects of VAAs versus CAVAAs are also dependent on the user's level of political sophistication. Respondents were randomly assigned to either the VAA or the CAVAA condition, and their level of political sophistication was measured as a quasi-experimental factor. The Research Ethics and Data Management Committee of approved the study (also under reference number REDC 2023.13a).

### 5.2. Materials and Measures

The same materials and measures as in Study 1 were used. Reliabilities were acceptable for all DVs that were measured with more than one item (ease of use: $\alpha = .82$; usefulness: $\alpha = .79$; playfulness: $\alpha = .88$; perceived political knowledge, 2 items: $\alpha = .82$).

To distinguish between participants with relatively higher and lower levels of political sophistication, we combined educational level and political interest into an aggregate index (Kamoen & Liebrecht, 2022). Political interest was measured using three items ($\alpha = .89$) based on Lachat (2008). The mean interest score was rescaled to a 4-point scale to match the educational scale, ensuring equal weight. We then averaged the two measures and performed a median split (cut-off = 2.50) to classify participants into higher and lower political sophistication groups.

### 5.3. Participants

A total of 149 participants took part in Study 2. For one of these participants, we were unable to trace whether the participant filled out the tool, and these data were therefore not considered for the analyses. Four other participants did not fill out any demographic questions in the survey. Because two such demographic questions were used to construct an indicator for political sophistication, it was not possible to determine political sophistication for these respondents, and their data were consequently excluded from the analyses. Of the remaining participants ($N = 144$), 74 identified as male (51.4%), 64 as female (44.4%), and six participants did not identify as either male or female (4.2%). The average age was 33.12 years ($SD = 15.85$).

We checked the gender and age distribution across the tool types and the two groups based on the level of political sophistication. No differences were observed in mean age (in all cases: $F(1, 139) < 1.28$; $p > .26$), and the male/female ratio was comparable in these groups (in both cases: $\chi^2(1) < .23$; $p > .63$).

### 5.4. Procedure

This study was conducted outside the lab between 9 November and 22 November 2023. Participants were approached in person by one of the researchers in several locations, including hairdressers, tattoo shops, and municipal buildings. In all these locations, we had consent to approach people, and if participants agreed to take part, they were seated in a quiet location within the building. From that point onwards, the procedure was the same as the procedure followed in Study 1.

## 6. Results Study 2

### 6.1. Tool Evaluations and Political Engagement Measures

Table 2 shows the means and standard deviations per condition. For both usefulness and playfulness, a medium-sized main effect of tool type is observed (usefulness: $F(1, 140) = 13.79$, $p < .001$, Cohen's $d = .63$; playfulness: $F(1, 140) = 8.96$, $p = .003$, Cohen's $d = .53$). Participants perceive the CAVAA as more useful and playful than the regular VAA. These effects are consistent across user groups, as both the main effect of sophistication and the interaction between tool type and sophistication were not significant (usefulness: $F(1, 140) = 0.21$, $p = .62$ and $F(1, 140) = 0.03$, $p = .86$; playfulness: $F(1, 140) = 0.33$, $p = .57$ and $F(1, 140) = 0.05$, $p = .83$). In line with findings in Study 1, the effect for ease of use is only marginally significant ($F(1, 140) = 3.12$, $p = .08$), which is consistent again across user groups since no significant main or interaction effect was found ($F(1, 140) = 0.20$, $p = .90$; $F(1, 140) = 0.77$, $p = .38$).

**Table 2.** Means and standard deviations on a scale from 1 (*low score*) to 7 (*high score*) in Study 2.

| Political sophistication | Tool | Ease of use | Useful-ness* | Playful-ness* | Perceived knowledge* | Factual knowledge+ | Internal political efficacy+ | Turnout intention+ |
|---|---|---|---|---|---|---|---|---|
| Low | CAVAA | 6.38 (0.76) | 5.67 (0.94) | 5.49 (1.16) | 4.79 (1.35) | 2.69 (1.75) | 4.75 (1.78) | 6.06 (1.64) |
| | VAA | 5.90 (1.22) | 5.02 (1.23) | 4.97 (1.28) | 4.36 (1.32) | 2.37 (1.80) | 4.49 (1.76) | 5.97 (1.52) |
| High | CAVAA | 6.25 (1.41) | 5.79 (1.20) | 5.64 (0.96) | 4.94 (1.60) | 4.08 (1.36) | 5.42 (1.65) | 6.87 (0.53) |
| | VAA | 6.09 (0.75) | 5.08 (0.99) | 5.04 (1.09) | 4.06 (1.33) | 3.94 (1.41) | 5.14 (1.77) | 6.69 (0.76) |

Notes: + = Main effect of political sophistication: higher scores for higher politically sophisticated users, * = main effect of tool type: higher scores in the CAVAA condition compared to the VAA condition.

As for the political engagement measures, results show a small-sized main effect of tool type for perceived knowledge ($F(1, 140) = 7.92$, $p = .006$, Cohen's $d = .47$), with scores being higher for CAVAA users. This effect is consistent across participants' levels of political sophistication, as neither the main effect of sophistication ($F(1, 140) = 0.00$, $p = .995$) nor the interaction reached statistical significance ($F(1, 140) = 0.71$, $p = .40$).

Factual knowledge scores do not differ between the CAVAA and VAA ($F(1, 138) = 0.72$, $p = .40$). However, results show a large significant effect of political sophistication: participants with lower levels of political sophistication score substantially lower on factual knowledge compared to their more sophisticated peers

($F(1, 138) = 30.81$, $p < .001$, Cohen's $d = 1.00$). This difference does not appear to be driven by tool usage, as the interaction between tool type and sophistication was not significant ($F(1, 138) = 0.72$, $p = .40$).

Finally, results for internal political efficacy and turnout intention show a consistent pattern: no significant differences appear between tool types ($F(1, 140) = 0.88$, $p = .35$; $F(1, 140) = 0.72$, $p = .40$). In contrast, political sophistication has a significant effect on these scores, as more politically sophisticated participants report higher internal efficacy scores and also report being more likely to vote ($F(1, 140) = 5.25$, $p = .02$, Cohen's $d = .38$; $F(1, 140) = 14.48$, $p < .001$, Cohen's $d = .69$). Again, these effects are not moderated by tool type, as no significant interactions are found (self-efficacy: $F(1, 140) = 0.001$, $p = .98$; turnout intention: $F(1, 140) = 0.11$, $p = .74$).

### 6.2. Response Behavior

Regarding the proportion of non-directional responses (see Table 3), results show tendencies for main effects: both lower politically sophisticated users ($\chi^2(1) = 3.67$, $p = .056$) and VAA users are more likely to provide non-directional responses ($\chi^2(1) = 2.88$, $p = .09$). The interaction between tool type and respondent type is not statistically significant ($\chi^2(1) = 0.11$, $p = .75$). The differences between conditions and levels of political sophistication are large relative to the between-item standard deviation (Cohen's $d > 1.23$) and also large in terms of percentages (roughly a 4% difference between tool types and user groups).

**Table 3.** Percentage non-directional answers for CAVAA and VAA users with higher and lower political sophistication in Study 2.

| Political sophistication | Tool | % non-directional answers (Logit) | $S^2$ items | $S^2$ participants |
|---|---|---|---|---|
| Low | CAVAA | 15.2% (−1.72) | 0.199 | 0.67 |
| | VAA | 20.0% (−1.39) | | 0.40 |
| High | CAVAA | 12.0% (−1.98) | | 0.97 |
| | VAA | 14.7% (−1.76) | | 0.85 |

### 6.3. Chatbot Usage

Table 4 displays chatbot usage by user type. Results indicate that less politically sophisticated participants more frequently requested semantic information ($\chi^2(1) = 9.44$, $p = .003$, Cohen's $d_{items} = 1.09$), while higher politically sophisticated users more often consulted pro and con arguments ($\chi^2(1) = 12.95$, $p < .001$, Cohen's $d_{items} = 1.04$). No significant difference was found for the use of pragmatic information ($\chi^2(1) = 1.13$, $p = .32$). This suggests that the type of information users seek varies with their level of political sophistication, even though the overall proportion of users requesting information via buttons at least once did not differ between these groups ($\chi^2(1) = 0.53$, $p = .47$).

In addition to button usage, participants submitted a total of 67 questions via open chat. After accounting for multiple open-ended questions related to the same respondent and item, this corresponds to 2.3% of respondent-item combinations. Including these open-ended questions in the overall measure of information requests, information was accessed in approximately 24% of cases, with no significant differences between more and less politically sophisticated users.

**Table 4.** Proportion of respondent-item combinations in which a specific button is clicked at least once in Study 2 (Logit between brackets).

| Political sophistication | Semantic information button* | Pragmatic information button | Pro and con argument button* | Clicking at least 1 button | Requesting information (incl. chat) |
|---|---|---|---|---|---|
| Low | 4.3% (−3.09) | 10.5% (−2.14) | 13.9% (−1.82) | 24.8% (−1.19) | 24.8% (−1.19) |
| High | 1.8% (−3.88) | 9.1% (−2.29) | 19.8% (−1.40) | 23.4% (−1.11) | 23.4% (−1.12) |

Note: * = $p < .05$

### *6.4. Summary*

The overall positive effects of the CAVAA compared to the VAA, as found in Study 1, are also observed in a more diverse sample. Moreover, Study 2 shows that tool type does not interact with user type across tool evaluation measures, political engagement measures, or response behavior to VAA statements. Regarding information-seeking behavior within the CAVAA, differences between lower and higher politically sophisticated users emerged: less sophisticated users more frequently request explanations of terms, while more sophisticated users more often review pro and con arguments. Given these different behavioral patterns, a relevant question that will be addressed in Study 3 is how both users with high and low political sophistication evaluate and interact with a CAVAA compared to an enhanced VAA (VAA+) that includes supplementary information in the web environment of the tool.

## 7. Method Study 3

### *7.1. Design*

Study 3 employed a 2 (tool type: VAA+ or CAVAA) × 2 (user type: higher or lower levels of political sophistication) between-subjects design, comparable to the design in Study 2. Participants were randomly assigned to one of the two tool types, while political sophistication was measured quasi-experimentally. The study was conducted in the lead-up to the European parliament election in the Netherlands on 6 June 2024. Ethical approval was obtained from the Research Ethics and Data Management Committee of the Tilburg School of Humanities and Digital Sciences at Tilburg University (reference number: REDC 2023.13b).

### *7.2. Materials*

As in Studies 1 and 2, we developed two versions of the tool based on a copy of StemWijzer (Figure 2). The tools featured 30 political attitude statements relevant to the 2024 European parliament elections. In the VAA+ version, users could access four types of additional information within the web interface. Semantic explanations of difficult terms were available by clicking on highlighted words in the statements. Information about the status quo, pro and con arguments, and party stances were accessible through a button.

In the CAVAA condition, a chatbot was integrated into an otherwise empty version of the VAA, without highlighted terms or clickable buttons. To maintain functional equivalence with the VAA+, the chatbot included four preset buttons corresponding to the same information types. Additionally, it featured an open

text field, allowing users to type their own questions and access a wider range of information than the static VAA+ interface. The chatbot was built using the same approach as in previous studies.

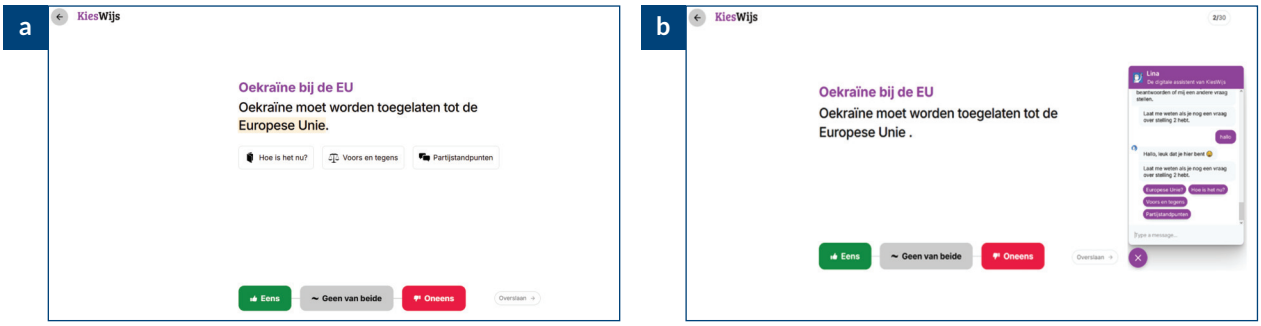


**Figure 2.** Screenshot of the VAA+ (a) and CAVAA (b). Notes: The header reads "Ukraine's accession to the EU" and the statement is "Ukrain should join the EU." In the VAA+, the highlighted term ("European Union") is clickable, and three additional clickable buttons are displayed below the statement; in the CAVAA, four buttons are displayed, each leading to the same information as in the VAA+ (i.e., via the three buttons and the highlighted term).

### *7.3. Measures*

Political sophistication was measured as in Study 2 by combining political interest ($\alpha = .82$) and education. Tool evaluation and political engagement measures were identical to previous studies, adapted to the European election context. Just like in Study 1 and 2, the scales showed acceptable consistency (ease of use: $\alpha = .93$; usefulness: $\alpha = .76$; playfulness: $\alpha = .80$; perceived knowledge: 2 items; $\alpha = .87$). We again measured users' factual political knowledge, logged directional versus non-directional responses, and tracked button clicks and open chat questions in the CAVAA condition. Differences between conditions were analyzed using multilevel models, following the approach from Studies 1 and 2.

### *7.4. Participants and Procedure*

Study 3 was conducted outside the lab between 22 May and 6 June 2024, following a comparable procedure to that used in Study 2. A total of 159 participants took part: 92 participants identified as male (57.9%), 66 as female (41.5%), and one as other (0.6%). The average age was 45.06 years ($SD = 19.63$). We compared gender and age distribution across conditions. No differences were observed in mean age (in all cases: $F(1, 139) < 1.28$; $p > .26$) and gender ($\chi^2(1) < .63$; $p > .73$).

## 8. Results Study 3

### *8.1. Tool Evaluation and Political Engagement Measures*

The scores for ease of use, usefulness, and playfulness do not differ between the two tool types, nor between participants with higher or lower levels of political sophistication (in all cases: $F(1, 155) < 2.62$, $p > .11$; see Table 5).

For the political engagement measures, only the main effects of political sophistication are observed. Consistent with findings from Study 2, more politically sophisticated users' internal self-efficacy is higher, they report a higher turnout intention and score higher on factual knowledge questions (in all cases: $F(1, 155) > 5.28$, $p < .02$). No other main effects or interactions between tool type and political sophistication are observed (in all cases: $F(1, 155) < 1.36$, $p > .25$).

**Table 5.** Means and standard deviations on a scale from 1 (*low score*) to 7 (*high score*) in Study 3.

| | Ease of use | Usefulness | Playfulness | Perceived knowledge | Factual knowledge | Internal political efficacy | Turnout intention |
|---|---|---|---|---|---|---|---|
| VAA+, low sophistication | 6.27 (0.55) | 5.63 (0.83) | 5.41 (1.01) | 4.96 (1.10) | 2.44 (1.36) | 5.17 (1.48) | 5.72 (1.70) |
| VAA+, high sophistication | 6.45 (1.03) | 5.73 (1.02) | 5.68 (0.91) | 5.10 (1.21) | 3.58 (1.31) | 5.60 (1.10) | 6.56 (1.04) |
| CAVAA, low sophistication | 6.36 (1.06) | 5.89 (0.97) | 5.60 (1.10) | 4.72 (1.56) | 2.74 (1.85) | 5.38 (1.35) | 5.56 (1.79) |
| CAVAA, high sophistication | 5.99 (1.43) | 5.68 (1.06) | 5.66 (1.07) | 4.85 (1.40) | 3.31 (1.67) | 5.87 (1.11) | 6.44 (1.19) |

### *8.2. Response Behavior*

The proportion of non-directional answers (Table 6) did not differ between participants with relatively high versus low levels of political sophistication, between tool types, or in their interaction (in all cases: $\chi^2(1) < 0.82$, $p > .37$).

**Table 6.** Percentage non-directional answers in Study 3 (Logit used in the analyses between brackets).

| | Non-directional answers | Semantic information button | Pragmatic information button | Pro and con arguments button | Party stances button |
|---|---|---|---|---|---|
| VAA+, low sophistication | 11.1% (−2.08) | 1.3% (−4.32) | 13.3% (−1.88) | 21.6% (−1.29) | 6.6% (−2.64) |
| VAA+, high sophistication | 10% (−2.20) | 2.0% (−3.89) | 12.7% (−1.92) | 17.6% (−1.54) | 15.4% (−1.70) |
| CAVAA, low sophistication | 11.9% (−2.00) | 5.6% (−2.82) | 7.6% (−2.49) | 11.9% (−2.00) | 3.8% (−3.22) |
| CAVAA, high sophistication | 10.1% (−2.19) | 2.5% (−3.67) | 4.3% (−3.11) | 6.3% (−2.70) | 2.9% (−3.51) |

### *8.3. Information Requests Made*

There was a tendency for semantic information to be requested more often in the CAVAA than in the VAA+ ($\chi^2(1) = 3.15$, $p = .08$), but neither political sophistication ($\chi^2(1) = 0.19$, $p = .66$) nor its interaction with tool type ($\chi^2(1) = 1.77$, $p = .32$) was significant.

For other information types, a clear pattern emerged: pragmatic information, pro and con arguments, and party stances were all requested significantly more often in the VAA+ than in the CAVAA (in all cases $\chi^2 > 7.49$,

$p < .01$), with no significant effects for political sophistication or its interaction (all $\chi^2$s < 2.74, $p$'s > .16). The size of the observed differences between tool types is consistently large relative to the between-item variance, and also substantial in practical terms, as some information types are requested twice or even three times as often in the VAA+ condition (Table 6).

Among the 78 CAVAA users, 40 open-ended questions were submitted (Response accuracy: 27.5%), corresponding to 27 unique respondent-item combinations (1.15%). Including these in the total measure of information-seeking behavior confirms the same overall pattern: participants requested information more frequently in the VAA+ than in the CAVAA.

### *8.4. Summary*

No differences can be observed between the VAA+ and CAVAA in tool evaluation, political engagement measures, or the proportion of non-directional answers—regardless of political sophistication. Differences do emerge in the information requests made. There is tendency for semantic information to be requested more often in the CAVAA, while status quo information, pro and con arguments, and party stances are more frequently accessed in the VAA+.

## 9. General Discussion

In three experimental studies, we examined the effects of CAVAAs compared to traditional VAAs without additional information (Studies 1 and 2), and to an enhanced VAA with clickable information buttons (VAA+; Study 3). The studies differed in sampling strategy and context: Study 1 was conducted in a controlled lab setting with a sample of highly educated participants, whereas Studies 2 and 3 targeted more diverse populations, allowing for a comparison between groups in their level of political sophistication. In terms of timing, Studies 1 and 2 took place in the run-up to the 2023 Dutch parliamentary elections (first-order elections), while Study 3 was conducted during the 2024 European elections (second-order elections). In the next sub-sections, we reflect on the most important findings across these studies and relate the findings to theory and practice.

### *9.1. VAAs versus CAVAAs*

Studies 1 and 2 show that CAVAA users actively engaged with the chatbot, using buttons in 25% (Study 2) to 43% (Study 1) of respondent-item combinations, while typing questions in 2–3% of the cases. This supports Krosnick's (1991) satisficing-optimizing model: the less effort required to access information, the more likely voters are to engage. In other words, easily accessible information encourages voters to read it, which aligns well with the core purpose of VAAs—to inform voters.

CAVAAs were also evaluated more favorably than VAAs in terms of usefulness and perceived knowledge (Studies 1 and 2), internal political efficacy (Study 1), and perceived playfulness (Study 2). Despite this variation in the exact effects observed, the overall pattern is clear and in line with H1 and H2, CAVAAs outperform traditional VAAs without information in terms of tool evaluation and political engagement measures. This conclusion aligns with Study 1 in Kamoen and Liebrecht (2022). Notably, our studies used more realistic and construct-valid materials.

As in Kamoen and Liebrecht (2022), current studies did not indicate an effect of tool type on the extent to which users plan to cast a vote. Kamoen and Liebrecht (2022) attributed this to the lack of ecological validity in their study, which is not an explanation here. In current studies, the scores for voting intention were consistently high (6.5 on a 7-point scale), which might point to a ceiling effect and perhaps to providing socially desirable answers. A suggestion for future research is therefore to construct a survey item playing into this potential bias.

Current studies also included the response behavior to political attitude statements as an unobtrusive indicator of comprehension difficulties. CAVAA users provided fewer non-directional responses compared to users of traditional VAAs, although the effect in Study 2 was only marginally significant. Hence, in line with H3, providing additional information in a chatbot is likely to mitigate comprehension challenges and satisficing behavior when filling out a VAA.

### *9.2. VAA+ Versus CAVAA*

A question is whether the benefits of the CAVAA stem from the chatbot interface itself or from the availability of additional information. Results of Study 3 showed no significant differences between CAVAA and VAA+ in terms of tool evaluations, political engagement measures, or response behavior. This suggests that the presence of information, regardless of how it is accessed, is the main driver of the effects observed in Studies 1 and 2.

Interestingly, Study 3 did reveal differences in the type and frequency of information accessed through the CAVAA versus the VAA+, as there was a tendency for semantic information to be requested more frequently in the CAVAA, whereas status quo information, pro and con arguments, and party stances were more frequently accessed in the VAA+. These findings raise several points for discussion. First, why would one type of information be requested more frequently in the CAVAA, whereas all other types are more often consulted in the VAA+? A likely explanation lies in the way information was presented. In the VAA+, semantic information was available via highlighted terms within the statements, while other types of information were accessible via buttons. Post-study interviews indicated that some users interpreted the highlighted terms as indicators of importance rather than as interactive elements, potentially explaining the lower access rate. It would be worthwhile to test this explanation in a future study.

Another point concerns how to interpret the finding that information was accessed more frequently in the VAA+ than in the CAVAA. One explanation is that the VAA+ aligns better with the low-effort information processing strategies (Krosnick, 1991) that many users apply. Alternatively, VAA+s might facilitate information access more effectively because users are more familiar with this interface. If so, we may expect differences in usage patterns between CAVAAs and VAA+s to diminish over time as users are expected to become more accustomed to conversational interfaces. A third interpretation is that VAA+ users may have clicked multiple buttons in search of specific information, while CAVAA users were able to formulate more targeted queries through open chat. In addition, in the CAVAA it was possible to re-read texts that were requested, as it was always possible to scroll through the chat, whereas an additional button-click was required to re-read button-information from the VAA+. If this explanation is valid, the higher number of clicks in the VAA+ condition might underscore the potential value of developing CAVAAs. Future research should explore these mechanisms in greater depth to better understand the differing engagement patterns across formats.

### *9.3. Political Sophistication*

In Studies 2 and 3, we distinguished users based on their level of political sophistication. A key finding is that tool type effects were not moderated by political sophistication, suggesting that the benefits of contextual information are consistent across user groups. The studies did reveal several main effects of sophistication: more politically sophisticated users scored higher on factual knowledge and reported a higher likelihood of voting. These outcomes align with expectations of what one might expect for more and less politically sophisticated users, and therefore merely give confidence in the construct validity of the measurement.

In Study 2, we also observed differences in information-seeking behavior: less politically sophisticated users more often accessed semantic information, whereas more sophisticated users more frequently consulted pro and con arguments. Interestingly, these patterns did not emerge in Study 3. We attribute this discrepancy to the nature of the election, and more specifically, to how this may have affected the sample we were able to recruit. While Studies 1 and 2 were conducted during first-order national elections, Study 3 focused on the second-order European parliament elections, which typically receive less public attention, media coverage, and voters also tend to know less about such elections (Reif & Schmitt, 1980). We suspect that differences in public interest between first- and second-order elections affected not only who participated in the study, but also the nature of the participant groups. Although recruitment took place in similar locations, recruiting participants during the European elections was noticeably more difficult. The samples from Studies 2 and 3, therefore, differed, not just in composition; for instance, participants in Study 3 were generally older, but may have also differed in the underlying characteristics of the higher and lower-sophisticated groups.

## 10. Conclusion

In conclusion, our findings suggest that embedding contextual assistance in a VAA, whether through a chatbot or clickable buttons in the web environment, improves tool evaluation and political engagement measures, and also helps users provide directional answers to the statements. These benefits occur across levels of political sophistication, making both VAA+s and CAVAAs promising tools to encourage users to inform themselves about the political issues at stake. At the same time, more research is needed to understand why voters request more information in a VAA+ than in a CAVAA. Based on current findings, it is not possible to determine whether this is an argument in favor of CAVAAs or VAA+s. We therefore encourage further research on the integration of (controlled) chatbots in the context of VAAs, but in a way that balances both the information needed and the ethical considerations of voters, governments, and supra-national agreements such as the AI Act.

## Acknowledgments

We thank ProDemos for granting us the right to use a copy of their tool. We would like to thank Elke van Veggel, Lotte Tiebosch, and Doris Verbunt for their assistance with collecting the data. We are also very grateful to Diana Schmalzried for proofreading our manuscript.

## Funding

We received funding from our department for conducting Studies 1 and 2. We want to thank our head of department for this. Additionally, for Study 3, we received funding from ProDemos for developing the CA and for hiring student assistants for data collection. Publication of this article in open access was made possible through the institutional membership agreement between Tilburg University and Cogitatio Press.

## Conflict of Interests

The authors declare no conflict of interests.

## Data Availability

The surveys and data files of all studies can be found via https://doi.org/10.34894/R2YWEL

## LLMs Disclosure

The authors used GPT-4o solely to rephrase sentences for conciseness and grammatical correctness.

## References

Ahn, T., Ryu, S., & Han, I. (2007). The impact of web quality and playfulness on user acceptance of online retailing. *Information & Management*, *44*(3), 263–275. https://doi.org/10.1016/j.im.2006.12.008

Baka, A. S., Figgou, L., & Triga, V. (2012). 'Neither agree nor disagree.' A critical analysis of the middle response option in voting advice applications. *International Journal of Electronic Governance*, *5*(3/4), 244–263.

Bruins Slot, H. (2022, October 10). *Uitkomsten onderzoek naar opkomst bij gemeentereaadsverkiezingen 2022*. Rijksoverheid. https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2022D40761&did=2022D40761

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*, 319–340. https://doi.org/10.2307/249008

De Graaf, J. (2010). The irresistible rise of Stemwijzer. In L. Cedroni & D. Garzia (Eds.), *Voting advice applications in Europe. The state of the art* (pp. 35–60). ScriptaWeb.

Delli Carpini, M. X., & Keeter, S. (1996). *What Americans know about politics and why it matters*. Yale University Press.

Dieing, T. I. (2025). AI-driven dialogue: Leveraging generative AI in conversational agent voting advice applications (CAVAAs). In A. Følstad, S. Papadopoulos, T. Araujo, E. L.-C. Law, E. Luger, S. Hobert, & P. B. Brandtzaeg (Eds.), *Chatbots and human-centered AI* (pp. 161–180). Springer. https://doi.org/10.1007/978-3-031-88045-2_11

Elling, S., Lentz, L., & De Jong, M. (2012). Users' abilities to review web site pages. *Journal of Business Technical Communication*, *26*(2), 171–201. https://doi.org/10.1177/1050651911429920

Fivaz, J., & Nadig, G. (2010). The impact of voting advice applications (VAAs) on voter turnout and their potential use for civic education. *Policy & Internet*, *2*(4), 167–200. https://doi.org/10.2202/1944-2866.1025

Garzia, D., & Marschall, S. (2012). Voting advice applications under review: The state of research. *International Journal of Electronic Governance*, *5*, 203–222.

Glynn, C. J., Huge, M. E., & Lunney, C. A. (2009). The influence of perceived social norms on college students' intention to vote. *Political Communication*, *26*(1), 48–64. https://doi.org/10.1080/10584600802622860

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, 213–236.

Hankel, S., Liebrecht, C., & Kamoen, N. (2024). 'Hi chatbot, let's talk about politics!' Examining the impact of verbal anthropomorphism in Conversational Agent Voting Advice Applications (CAVAAs) on higher and lower politically sophisticated users. *Interacting With Computers*, *37*(4), 221–233. https://doi.org/10.1093/iwc/iwae031

Heinsohn, T., Fatke, M., Israel, J., Marschall, S., & Schultze, M. (2019). Effects of voting advice applications during election campaigns. Evidence from a panel study at the 2014 European elections. *Journal of Information Technology & Politics*, *16*(3), 250–264. https://doi.org/10.1080/19331681.2019.1644265

Kamoen, N., & Holleman, B. (2017). I don't get it. Response difficulties in answering political attitude statements in voting advice applications. *Survey Research Methods*, *11*(2), 125–140. https://doi.org/10.18148/srm/2017.v11i2.6728

Kamoen, N., Holleman, B., Krouwel, A., van de Pol, J., & de Vreese, C. (2015). The effect of voting advice applications on political knowledge and vote choice. *Irish Political Studies*, *30*(4), 595–618.

Kamoen, N., & Liebrecht, C. (2022). I need a CAVAA: How conversational agent voting advice applications (CAVAAs) affect users' political knowledge and tool experience. *Frontiers in Artificial Intelligence*, *5*, Article 835505. https://doi.org/10.3389/frai.2022.835505

Kamoen, N., Liebrecht, C., Van Lieshout, R., & Van Veggel, E. (2024, August 12–14). "Hi voter, I can assist you in filling out this voting advice application!" How CAVAAs may assist voters in informing themselves about elections [Paper presentation]. ECPR General Conference, Dublin, Ireland.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Lachat, R. (2008). The impact of party polarization on ideological voting. *Electoral Studies*, *27*(4), 687–698. https://doi.org/10.1016/j.electstud.2008.06.002

Ladner, A. (2012, April 3–5). *Voting advice applications—Impact on voting decisions in the 2011 Swiss national elections* [Paper presentation]. The 62nd Political Studies Association Annual International Conference, Belfast, Northern Ireland. https://www.researchgate.net/publication/261495685_Voting_advice_applications--impact_on_voting_decisions_in_the_2011_Swiss_national_elections

Luskin, R. C. (1990). Explaining political sophistication. *Politicial Behavavior*, *12*(4), 331–361. https://doi.org/10.1007/BF00992793

Marangunić, N., & Granić, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society*, *14*, 81–95. https://doi.org/10.1007/s10209-014-0348-1

Mason, L. E., Krutka, D. G., & Stoddard, J. (2018). Media literacy, democracy, and the challenge of fake news. *Journal of Media Literacy Education*, *20*(2), 1–10.

Munzert, S., & Ramirez-Ruiz, S. (2021). Meta-analysis of the effects of voting advice applications. *Political Communication*, *38*(6), 691–706. https://doi.org/10.1080/10584609.2020.1843572

Quené, H., & Van den Bergh, H. (2004). On multilevel modeling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*, 103–121.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413–442.

Reif, K., & Schmitt, H. (1980). Nine second-order national elections: A conceptual framework for the analysis of European election results. *European Journal of Political Research*, *8*(1), 3–44.

Roberts, C., Gilbert, E., Allum, N., & Eisner, L. (2019). Research synthesis. Satisficing in surveys: A systematic review of the literature. *Public Opinion Quarterly*, *83*(3), 598–626. https://doi.org/10.1093/poq/nfz035

Sipma, T., Lubbers, M., Van der Meer, T. W., Spierings, D., & Jacobs, K. (Eds.). (2021). *Versplinterde vertegenwoordiging: Nationaal Kiezersonderzoek 2021*. Stichting KiezersOnderzoek Nederland. https://dspace.library.uu.nl/bitstream/handle/1874/421298/NKO_2021_Versplinterde_vertegenwoordiging.pdf?sequence=1

Van de Pol, J., Holleman, B. C., Kamoen, N., Krouwel, A. P. M., & De Vreese, C. H. (2014). Beyond young, higher educated males: A typology of VAA users. *Journal of Information, Technology and Politics*, *11*, 397–411. https://doi.org/10.1080/19331681.2014.958794

Van Outersterp, R., Kamoen, N., & Holleman, B. (2016). Inwisselbaar of niet? De middelste antwoordoptie en de non-responsoptie in stemhulpen. *Tijdschrift voor Taalbeheersing*, *38*(1), 27–55.

Van Veggel, E., Liebrecht, C., & Kamoen, N. (2025). How life-like digital humans in voting advice applications can stimulate young voters to inform themselves about politics: The role of familiarity and expertise. *International Journal of Human-Computer Interaction*, *41*(20), 12816–12830. https://doi.org/10.1080/10447318.2025.2465859

Van Zanten, N., & Boumans, R. (2023). Voting assistant chatbot for increasing voter turnout at local elections: An exploratory study. In A. Folstad, T. Araujo & S. Papadopoulos (Eds.), *International Workshop on Chatbot Research and Design* (pp. 3–22). Springer. https://doi.org/10.1007/978-3-031-54975-5_1

Wall, M., Krouwel, A. P. M., & Vitiello, T. (2014). Do voters follow the recommendations of voter advice application websites? A study of the effect of kieskompas.nl on its users' vote choices in the 2010 Dutch Legislative Elections. *Party Politics*, *20*(3), 416–428. https://doi.org/10.1177/1354068811436053

## About the Authors

Naomi Kamoen is an assistant professor at Tilburg University. Her current research focuses on the uses and effects of VAAs. She collaborates with societal partners in real-life election settings, using a wide range of research methods such as experiments, content analyses, and big data analyses.

Christine Liebrecht is an assistant professor in the Department of Communication and Cognition at Tilburg University. Her research examines the transformation of human communication through technology across different user types and contexts. Her approach incorporates both quantitative and qualitative research methodologies, frequently linking insights to practical applications.

Rieke van Lieshout is a research master's student in linguistics at Radboud University Nijmegen. Her current research focuses on simple language and on how large language models can improve (linguistic) accessibility. Combining computational and theoretical expertise, she builds bridges across disciplines and in society.