

APPENDIX A

Table A.1.: Introduction for human coders to the manual codebook

Codebook for manual content analysis

1) Introduction into the Codebook

This codebook is designed to support a structured and consistent approach to annotate Twitter comments, mostly related to discussions about climate change related topics. It includes detailed category definitions and information on the coding context with illustrative examples and coding instructions. In addition to the definitions and coding instructions, the codebook also provides clarifications of overarching phenomena and concepts, such as irony and sarcasm, that are mentioned within specific categories. For these more complex constructs, the codebook includes not only definitions but also step-by-step guidance to support their identification. However, the following overarching instructions should be considered before using the codebook. These instructions/principles are particularly important for coders using the codebook for the first time, as they provide guidance in handling recurring challenges.

1. Use NA when you are uncertain

Due to the complexity of certain categories in the codebook, particularly when applied to short, sometimes ambiguous, Twitter comments, the following should be considered when coding: Coders should not feel obliged to assign a label to every dimension of every comment, particularly when the information is unclear or ambiguous. Rather than making uncertain decisions (e.g., selecting 'absent' with little confidence), it is preferable to select **NA (not applicable)**. This helps preserve the validity of the data, particularly in a binary coding system where there are limited options when faced with interpretive uncertainty. For instance, if a comment lacks context, is difficult to interpret or contains insufficient information to make a confident coding decision, it should be labelled NA. Additional and more specific guidance on the use of NA is provided for some categories.

2. Avoid interpretation (in most cases)

Coders should rely on the **manifest content**. They should focus on what is explicitly stated in the comment rather than attempting to interpret implicit meanings. If a category cannot be applied with confidence based on the manifest content alone, use the **NA** (*see instruction above*).

However, some phenomena that are part of a category (e.g. irony or sarcasm) require limited interpretation, such cases are supported by detailed instructions. Unless otherwise specified, do not infer intent or meaning beyond what is visible in the text.

3. Consider the Context

All comments in the dataset (excluding the initial posts by public speakers) are replies to a **previous post**, often forming reply trees, that serve as **context for the annotation** of most categories. In many cases, this is an initial post by a public speaker, but it may also be a comment by another user as part of an ongoing reply tree. In the provided dataset, each comment includes metadata specifying the ID of the comment it is replying to (see the column "referenced_tweets_replied_to_id" for the replied comment's "id"). Sometimes, a reply is contextually linked not only to the specific comment it directly responds to, but also to the entire reply tree of prior comments, including the public speaker post. If such a situation occurs, you can trace the trees by looking at the columns with the name "reply_tree_id n", which are tracing back the sequence of comments that respond to one another (in the ICR dataset, these reply

trees are indicated by different colours and numbers to make this clear). During annotation, always consider this broader context, if it is indicated in the coding instructions (“Context”), to ensure you assign the appropriate label and achieve accurate data annotation.

Sometimes, users reply to themselves, to post a follow-up comment to their own previous one, often due to space limitations. In such cases, the **self-reply** should be coded as **NA** for every category but considered as **part of the previous comment**. This particularly applies for public speaker initial tweets, as they mostly include more than one comment.

As many categories (e.g., Substantial Dialogic Exchange, Topic Reference) are interaction-based, it is essential that coders identify the relevant context before making any coding decisions. This especially includes understanding the topic of discussion and referenced content (e.g., previous comments, external links).

However, even when context is available, it can be difficult to assess certain categories (*see instruction above*). **For example**, a user replying 'Thank you' to a post in which a public speaker mentions their engagement in climate protection may implicitly refer to the topic of the public speaker's post. However, since this is not explicitly clear, the category 'Topic reference' should be labelled **NA** in this case.

4. Multimodal Content (Images, Videos, and Links)

Many comments include **multimodal elements**, such as images or videos, which are embedded in the comment but can also be accessed via links (in the dataset, they look like this: <https://t.co/...>). If such elements are present and accessible, they should not be considered part of the comment and therefore excluded from the annotation process (with exception of links to external sources as described in next section “External Sources and Credibility Assessment”).

If access to the link is **not possible**, base your coding decision solely on the available text. Always check which categories can still be assessed with this limited information; if a category cannot be evaluated, it should be labelled **NA**. If a comment **consists solely of a link** with no accompanying text and the link is inaccessible or leads to an image or video, all dimensions of the comment should be labelled **NA**.

Considering images and videos raises further questions for **LLM validation**, as it is currently unclear to what extent such models can access and interpret linked media. Until this issue is resolved, coders should follow the previously provided instructions, even though this decision may reduce annotation accuracy. **For example**, when a comment presents factual information supported by a relevant graphic from a credible source embedded within the tweet.

5. External Sources and Credibility Assessment

As explained above, comments may contain external links to sources such as news articles or other Twitter posts. In such cases, coders should open and assess the linked content, if possible. This is particularly relevant for categories such as External Justification. **For example**, here, the emphasis is not on fact-checking the content, but on ensuring that the information provided can be traced by other users and that the comment cites a credible source. If external links are present and accessible, they are treated as part of the **contextual information** for all categories that rely on **inter-comment reference**. Coders should therefore consider different types of **metadata** of the linked source as an addition to the comment itself. However, due to potential limitations in the automated scraping of Twitter data, our manual annotation is restricted to the consideration of the **URL**, the **headline** and the **source (e.g. FAZ)** of the linked page. If the link leads to a **Twitter comment or a post from another social media platform** consider the whole comment. But with exception of visual information, as sometimes a link just leads to the same comment, if it contains visual information (e.g., pictures), which could not be displayed in the dataset. We treat **both forms of external information as equivalent** in coding[PF1] ; the only difference is the amount of accessible information (see above). Whether and to what extent external sources or social media posts are included in the classification of a category is described in a **separate section for each category** (see “Links to external sources or social media” in the codebook).

External information is only included in the following categories:

External Justification | Complete Justification | Substantial dialogic exchange | Stance on Prior Argument (Viewpoint alignment) | Civility | Constructiveness | Common good orientation

Example: If a public speaker shares a link to a FAZ article titled "*Klimaschutz ist teuer, aber dennoch notwendig*", and another user responds by acknowledging this cost-benefit reasoning, the linked article becomes essential for understanding different categories (e.g. external justification, etc.).

Whether future implementations will allow for the extraction of richer contextual data (e.g., the full content of the linked website or article) remains to be clarified. Thus, we have supplemented the dataset by flagging all rows **containing links and adding the full URL, the source, as well as the headline of the linked website or article**. This will allow us to decide later whether to include this information in the LLM validation.

If access to the link is **not possible**, base your coding decision solely on the available text. Always check which categories can still be assessed with this limited information; if a category cannot be evaluated, it should be labelled **NA**. If a comment **consists solely of a link** with no accompanying text and the link is inaccessible or leads to an image or video, all dimensions of the comment should be labelled **NA**.

6. Documentation of Difficult Coding Decisions

If you are unsure about how to code a category but eventually decide to assign a label, consider **documenting your reasoning**, especially if it may help improve the codebook or the coding process. After annotating a set of comments, take your time to review your labels. Reflect on any difficult or ambiguous cases, and assess whether your coding decisions are consistent with the codebook instructions and accurately applied.

2) Definition of relevant concepts mentioned in coding instructions

Def.: Argumentum ad hominem:

Argumentum ad hominem is a rhetorical strategy that attacks a person's character instead of (solely) debating the issue or the substance of an argument. Attacks may include derogatory statements about personal traits, moral standing, or expertise, as well as speculation about motives and special interests.

Def.: Whataboutisms:

Whataboutisms can occur in various forms. These are explained as follows:

- **Inconsistency (second-personal form):** Comment accuses another user or entity of inconsistency

For example: Sie predigen Klimaschutz, fliegen aber zur jeder Klimakonferenz mit dem Flugzeug.

- **Contrastivism (third-personal form):** Comment does not accuse another user or entity of inconsistency, but rather raises a contrasting consideration whose existence alters the evidential backing for the original claim. This means that the user does not necessarily argue that the argument of another user is wrong/good or not, but does insist on the importance of another topic or aspect in contrast.

For example: User 1: Wir müssen dringend unsere Klimaschutzanstrengungen erhöhen, um Mensch und Umwelt zu schützen! **User 2:** Sie reden hier die ganze Zeit über Klimaschutz und dessen Gefahren, aber haben Sie nur eine Sekunde ihrer politischen Arbeit dem Problem der Massenmigration und dessen Folgen für unser Land gewidmet?

Def.: Irony and sarcasm:

Only consider this definition, if irony or sarcasm may be present. Because Twitter comments are often context-poor, identifying sarcasm or irony can be difficult, especially since both are typically implicit. However, there are different markers that can help detect irony and sarcasm. We treat them as two variants of the same phenomenon.

Hashtags: If a comment contains Hashtags which signal the use of irony or sarcasm, such as #irony or #ironie, #ironyoff or #ironieoff, #saracsm or #sarkasmus, the comment should be considered as ironic or sarcastic.

Content: A comment may be ironic or sarcastic if it includes an evaluative judgment, for example, of a person, statement or situation and not just a simple description. It targets the object often in a critical way, and often contains an unexpected twist or draws attention to something unusual. To finally assess whether a comment shows irony or sarcasm, follow these steps:

1. Description or evaluation: If parts of the comment contain an evaluative statement or a statement with an evaluative connotation, it may contain irony or sarcasm. Also consider emojis as potential indicators. If a comment is only of purely descriptive nature, the comment is not labelled as ironic/sarcastic.

2. Does the literal evaluation fit the (co-)text: Determine what the literal (written) evaluation is. If the literal evaluation fits the co-text, which is constituted by the prior comment the user is referring to, the comment is not labelled as ironic/sarcastic. If it does not fit the co-text, **e.g.**, by saying something unexpected considering the prior comment, the comment may be ironic/sarcastic.

3. Comparison of literal and intended evaluation: Determine whether the literal evaluative part of the comment shows a positive or negative tone and what the target of reference is. If a plausible reading can be found in which the intended (ironic) evaluation does not match the literal evaluation in the comment, **e.g.**, it sounds positive but the user could actually mean something negative, the comment is considered ironic/sarcastic. Also consider emojis as potential indicators, **for example**, the clown emoji 🤡. If you are not able to assess this difference, the comment is labelled as **NA**.

For example:

Context: Trump nennt Merz "großartigen Kanzler" und lobt Verteidigungsausgaben, dem kann ich nur zustimmen.

Comment: Merz ein großartiger Kanzler? Und das kann Trump schon jetzt beurteilen, wo Merz quasi gerade erst vereidigt wurde? Donnerwetter, das ist Sachverstand!

Def.: Rhetorical question:

A rhetorical question presents the author's opinion and implies an obvious answer that is evident to both the addressor and the addressee. It is intended to convince the audience rather than evoking a verbal response from the addressee. Rhetorical questions exist in various forms, **e.g.**, they can include ridiculous expressions, or irony.

Hashtags: As rhetorical questions are rarely marked by hashtags like #rhetorischefrage or #rhetoricalquestion, identification relies on combining indicators with the provided definition. But it is difficult to identify rhetorical questions, as their syntactic form is indistinguishable from standard questions. And only a small number of rhetorical questions show specific linguistic markers, while most of them only can be recognised through context.

Indicators and context: Therefore, the identification of rhetorical questions relies on their contextual relationship with previous comments, as they try to challenge previously expressed opinions. This requires a thorough understanding of prior comments (**e.g.**, from the public speaker) to ensure accurate identification. If you are not able to assess or identify the context or unsure about your assessment of the following linguistic indicators, label the comment **NA**.

Linguistic indicators:

1. **The use of polarity items in questions:** (**e.g.**, Does that bring you one inch closer to what is driving that woman?)
2. **Introducing questions with a lexical item incompatible with asking for information:** Questions introduced with words atypical for seeking information (**e.g.**, Because, if you give up your pride, what are you left with, really?)
3. **Questions accompanied by ridiculous answers:** Questions followed by sarcastic or absurd responses (**e.g.**, Who do you think I am, a dog?)
4. **Questions realized in why + lexical verb form:** (**e.g.**, Why talk about her?)
5. **Questions incorporated into declarative or imperative sentences:** Questions within statements or commands (**e.g.**, What's to lose, go ahead)
6. **Auto-responsive questions:** Questions immediately followed by an answer (**e.g.**, What are a friend's books for if not to be borrowed?)

Def.: Offensive language & hate speech

o **Def. offensive language:** Offensive language is a language that uses rude, profane or demeaning comments against individuals or groups. It may include personal attacks on individuals such as cyberbullying, harassment or other. It may also include hate speech that targets people based on their protected characteristics (see definition below).

o **Def. hate speech:** Hate speech includes all forms of expression that incite violence or hatred, attack or diminish people based on their protected characteristics such as ethnicity, national origin, race, religious affiliation, caste, sexual orientation, gender, gender identity, age, disability, or disease.

Further instructions on hate speech/ offensive language:

- Abusive language and insults can also be directed against people generally considered bad guys; it does not even matter whether they are dead or alive (**e.g.** “Das Arschloch Helmut Kohl” – INSULT).
- It does not matter whether an utterance considered an insult or abuse is actually noticed by the person or group of persons targeted.
- Offensive language/hate speech may come in sentence types other than plain assertions, **for example**, rhetorical questions: “Wenn ich diese Ausländer als Asylschmarotzer bezeichne, bin ich dann ein Nazi?”. Annotators should ask themselves what they think the speaker of the utterance is actually thinking. In the above utterance, it is fairly obvious that the speaker opposes asylum seekers, if it is unclear, label the comment **NA**.

Table A.2.: Complete manual codebook with definitions and coder instructions for type I and type II deliberative quality dimensions

External Justification		
1	Variable shortcut	Jus_ext
2	Definition	External justification of a claim (see definition “Internal justification”) with relevant, traceable evidence (from a high-quality source) . For example, with figures, facts, statistics, or links to studies.
3	Explanation of concepts used in the definition	Relevance: The evidence provided must be directly relevant to the claim made in the comment. This means that the information should clearly support or explain the specific point being argued.
		Traceability: The evidence must be traceable to a specific source. It is not necessary for the comment to include a direct link, but it must at least mention the source or provide enough context information for it to be identified. For example , a comment may refer to a YouTube video without providing a link, but mentions the channel name and the topic of the video. Comments that do not provide any identifiable information about the source, should not be considered as showing traceability, even if they appear to contain relevant information.
		High-quality source: The evidence must come from high-quality sources, such as scientific studies, established newspapers (e.g., FAZ), public-service and legacy media (e.g., ARD), governmental institutions, or recognized experts and organizations. For example , a comment might refer to a study conducted by the IPCC, a specific law paper/directive by the government or state that a claim was reported by ARD. Statements that come from other (e.g., Social Media Posts or YouTube Videos) or even unreliable sources, or that do not provide any identifiable information about the source, should not be considered as using high-quality sources, even if they appear to contain factual information.
4	Special cases/ Additional information	Specific institutions as source: In some cases, a user is only referring to a specific institution (e.g., the Federal Constitutional Court), to justify a claim. For external justification to be considered present, the following three aspects <u>must be included</u> in the analysed comment: - Specification of the <u>institution or entity</u> (e.g., BVerfG; YouTube Channel of Rezo), - Clarification of the <u>subject</u> (e.g., the law being referenced; the political topic being discussed in the video), - Provision of a <u>relevant fact</u> that supports the claim For example: “Die CDU ignoriert die Klimakrise seit Jahrzehnten. Zur Erinnerung das YouTube Video von Rezo zum katastrophalen Klimaschutz in Deutschland.“ For example: “Es ist @KlausMindrup. Das in Teilen verfassungswidrige (Claim) Klimaschutzgesetz (Subject), dass das BVerfG (Institution) bemängelt (Fact) hatte, verteidigte er auch.“
		Relevance of context for new relevant and traceable information: If a source has been clearly named or linked in a previous contribution and it is clear that the comment refers to this source, then this context makes the information count as “traceable” even if the source is not mentioned in the coded comment.
5	Context	If a comment’s external justification refers to one or more prior comments from another user or a public speaker, the referenced prior comment(s) must be considered as part of the assessment of internal justification. But this is <u>limited to</u> expressed claims by other users . This is crucial because some arguments may only demonstrate the necessary characteristics for external justification when examined within their conversational context. For

		example , if the claim is already provided in a prior comment, the analysed comment may only present a relevant and credible evidence. If the prior comment is not accessible, consider the comment in isolation.	
6	Scale level	2	High-quality (Relevant, traceable and high-quality source): The user comment fulfils <u>all of the three conditions</u> of the definition.
		1	Low quality (Relevant and traceable): The user comment fulfils <u>the condition of relevance and traceability</u> of the definition.
		0	Absent: The user comment fulfils only <u>one or less</u> of the three conditions of the definition. And/or the comment does <u>not</u> show <u>traceability</u> .
7	Distinction from or relation with other variables	When external justification is present, it does not exclude the possibility that internal or complete justification may also be present. Therefore, all three categories should be examined, even when external justification is present.	
8	Coding instructions	-	
9	Links to external sources or social media	Links to external sources: If links to external sources (e.g., newspapers) are present and accessible, they are treated as part of the contextual information to assess external justification. In such cases, coders should therefore consider different types of metadata of the linked source to assess whether the comment shows external justification. However, due to potential limitations in the automated scraping of Twitter data, our manual annotation is in that case restricted to the consideration of the URL, the headline and the source (e.g. FAZ) of the linked page. Therefore, you sometimes have to assume, for example, if the title of an article, indicates that the article could contain information, that supports a claim of the user. If unsure, label the category with NA. Links to social media posts (e.g., X posts): If social media posts from X, or other platforms are linked/shared in a comment and accessible, they are treated as part of the contextual information to assess external justification. In such cases, coders should therefore consider the whole text of the linked post as well as the author of the post, but <u>not</u> present (audio-)visual content. <u>Variant 1: Shared post provides relevant information and source</u> If a linked/shared social media contains relevant information for the discussion topic from a traceable high-quality or low-quality source, external justification is labelled as <u>present</u> (Scale level: 1 or 2; see “Definition” and “Scale level”). Example (Discussion about climate change): Relevant information from a traceable high-quality source User 1: “Solche dramatischen Meldungen werden leider kaum von Medien aufgegriffen... @DLF” Linked X post by user 1: „Von der #THG-Senke zur Quelle: Der Sektor LULUCF emittiert in im zweiten Jahr in Folge mehr Treibhausgas als er aufnimmt. THG-Emissionen in der #Landwirtschaft moderat gesunken. Mehr aktuelle Zahlen & Fakten in unserem Emissions-Dossier: https://thuenen.de/de/themenfelder/klima-und-luft/emissionsinventare-buchhaltung-fuer-den-klimaschutz/treibhausgas-emissionen-aus-der-landwirtschaft“	

		<u>Variant 2: Shared post provides relevant information, but no source</u> If a linked/shared social media post contains relevant information for the discussion topic, but does not provide a specific source for the information, the annotation depends on the identifiability and type of the author, as indicated by the username. a) If the username of the author cannot be classified (e.g., fictitious username), external justification is labelled as <u>absent</u> (Scale level: 0 – Absent). b) If the linked/shared post is posted by a user, for example a private person (with an apparently real name) or an account that does not indicate any specific affiliation, external justification is labelled as <u>present</u> (Scale level: 1 – “Low-quality”). c) If the linked/shared post is posted by a clearly identifiable “high-quality” account, such as the account of a governmental organisation, or established newspaper, external justification is labelled as <u>present</u> (Scale level: 2 – “High-quality”). d) If you are unsure about the username or cannot assess it, label the category with NA. Example (Discussion about climate change mitigation): Relevant information from a high-quality account User 1: “Sie haben Kernkraft weggelassen. Wie soll Sie jemand Ernst nehmen? Wie soll man euch wählen, wenn ihr gegen die Wissenschaft arbeitet? Bei allem, was auf dem Spiel steht: Nicht zu entschuldigen.” Linked X post by user 1: Username: U.S. Department of Energy; Post: “This says it all! The Biden administration is making BIG investments that will help combat the climate crisis and provide cheap, reliable electricity by:  driving next-gen nuclear reactor deployment [...]”
10	Examples	Absent (States a relevant “fact”, but does not provide a source for traceability): User 1 (Context): Was viele bei der Klimakrise falsch verstehen und was uns die Lage in Frankreich gerade so brutal vor Augen führt: Wenn der Staat uns heute nicht vorschreibt, wie wir heizen dürfen, wird er uns morgen vorschreiben müssen, wann wir den Wasserhahn aufdrehen dürfen und wann nicht. User 2: Es gibt keine Klimakrise. Die findet nur in Ihrem Kopf statt. Das Problem mit dem Wasser, ist schon Jahrzehnte bekannt, die Ursache hat nix mit Klima zu tun, sondern mit Privatisierung von Süßwasservorkommen. Partial (Provides relevant evidence from a directly traceable, but not a high-quality source via a link): User 1 (Context): Der wirksamste Klimaschutz sind Innovationen; Technologien, die CO2 einsparen, und gleichzeitig nachhaltig Jobs schaffen. Das geht nur gemeinsam! Welchen Beitrag Digitalisierung zur Nachhaltigkeit leisten kann: Darüber sprechen @RuedigerKruse; @KatharinaGehra. #CDigitallyUnited: https://t.co/8sFxp2wrZz. User 2: @CDU @RuedigerKruse @KatharinaGehra Handeln! Nicht nur daherschwätzen. Geschwätz ist seit 16 Jahren genug. Passiert ist nämlich nix. Die CDU ist schlichtweg unfähig. Zur Erinnerung: https://t.co/U4cMjL8yyg. Present (Provides relevant and traceable evidence from a high-quality source): User 1 (Context): In #Hamburg kaufen wir seit 2020 nur noch emissionsfreie Busse. Aktuell sind wir laut #pwc bei 159 Bussen. Hunderte weitere sind bestellt. Möglich auch dank Förderung durch das #bmdv. #Busse sparen etwa 80-85 % der CO2 Emissionen ggü. #Dieselbussen über den #Lebenszyklus. User 2: Laut unserer Rechnung basierend auf Daten von Hochbahn und VHH war #Hamburg zu Jahresbeginn sogar schon bei 226 und hat damit Berlin überholt: https://www.nahverkehrhamburg.de/elektrobusse-hamburg-ueberholt-berlin-deutlich-197187/

11	Sources operationalization	Behrendt et al., 2024; Esau et al., 2023; Reveilhac, 2023; Esau et al., 2021; Oswald. 2022; Jaidka, 2022; Schroll & Huber, 2022; Taddickenier-Tombs und MacKenzie, 2021; Friess & Eilders, 2015	
Follow-Up-Questions			
1	Variable shortcut	Jus_que	
2	Definition	Asking genuine (<i>see coding instructions</i>) follow-up questions about prior discussion topics or arguments (e.g., requesting additional information, asking for reasons behind proposed claims, etc.).	
3	Explanation of concepts used in the definition	Genuineness: A question or response is considered genuine if it <u>neither contains uncivil and impolite language</u> (see "Absence of incivility" for definitions) <u>nor signs of insincerity</u> (e.g., provoking others via trolling, sarcasm, irony, rhetorical questions). If this is <u>not fulfilled</u> , e.g., if a rhetorical question was used, this category should be labelled as absent .	
4	Special cases/ Additional information	-	
5	Context	Consider the immediate context of the comment when assessing Follow-Up Questions. This includes the comment to which the evaluated comment is replying. If additional context is needed, earlier replies from the same reply tree (including the public speaker post) may be consulted as sometimes, more context is needed than the mere comment, the user is replying to.	
6	Scale level	1	Present
		0	Absent
7	Distinction from or relation with other variables	-	
8	Coding instructions	If unsure, label comment with NA: If you <u>do not have sufficient certainty</u> about whether a comment contains, for example, a rhetorical question, irony, or sarcasm, the comment should be labelled with NA . Therefore, <u>only questions that clearly show all the defining characteristics of</u> , for example, a rhetorical question, allowing us to be reasonably certain of the author's intention, should be labelled as absent .	
9	Links to external sources or social media	Links to external sources or social media are <u>not</u> part of the assessment of this category.	
10	Examples	Present (Genuine follow-up-question to challenge prior argument): User 1 (Context): Wie viel darf #Klimaschutz kosten, fragt der @ADAC @andreasscheuer & mich Dienstag live im #ADACTalk (16:00 Uhr). Ohne schon alles zu verraten, sage ich: Ein kaputter Planet ist teurer. Das sind Schulden, die Kinder & Enkelkinder nie zurückzahlen können. User 2: Aber können wir es bei den momentanen Ausgaben?	

		Absent (Rhetorical question): User 1 (Context): Wärmepumpen, E-Autos, Solar- und Windenergie machen uns immer unabhängiger. Warum werden diese Zukunftstechnologien permanent schlechtgeredet? Das verunsichert die Leute, schadet der Wirtschaft und vernichtet Jobs. https://t.co/Reta27gH1g User 2: Eher abhängigen von vergügbarem Strom. Da ist flatterstrom eine geniale Sache oder?	
11	Sources operationalization	Behrendt et al., 2024	
Internal justification			
1	Variable shortcut	Jus_int	
2	Definition	Internal justification of a statement through own reasoning, which includes, besides other optional components, two core elements in accordance with Toulmin’s model of argumentation: claim, reason . Keep in mind, that we <u>do not rate the quality of the reasoning</u> , only whether the mentioned elements are present or not.	
3	Explanation of concepts used in the definition	Claim: The point of view or normative assertion proposed in the comment, which can occur as a claim at the beginning or as a conclusion at a later stage of an argument. For example: "Als Unternehmer würde ich mich freuen, wenn die Schuldenbremse endlich fallen würde. [...]"	
		Reason (also “data” or “grounds”): Explanation of the author, why a claim should be accepted. For example: „Sparen um jeden Preis, sei es für Infrastruktur, Klimawandelbekämpfung, Bildung, schadet dem Wirtschaftsstandort Deutschland.“	
4	Special cases/ Additional information	A claim can be supported by one or more reasons. A comment only needs to contain one “claim-reason pair” for it to be labelled with internal justification. Claim and Reason are linked by a warrant (<i>see “Complete Justification”</i>), which is a missing piece of information that is taken for granted by the author and is assumed common knowledge, so it is left out in many cases and thus Implicit. Therefore, in comments where a reason supports the claim, a trivial warrant of the form ‘ if reason then claim ’ is <u>sufficient</u> .	
5	Context	If a comment’s argument refers to one or more prior comments from another user or a public speaker, the referenced prior comment(’s) must be considered as part of the assessment of internal justification. But this is <u>limited to</u> expressed claims by other users . This is crucial because some arguments may only demonstrate the necessary characteristics for internal justification when examined within their conversational context. For example , if the claim is already provided in a prior comment, the analysed comment may only present a reason. If the prior comment is not accessible, consider the comment in isolation.	
6	Scale level	1	Present
		0	Absent
7	Distinction from or relation with other variables	-	

8	Coding instructions	Comments that contain (strong) sarcasm/ irony, which makes it difficult to identify elements of Toulmin's model: Label comments with NA if (strong) sarcasm or irony prevents clear identification of the claim or reasoning according to Toulmin's model (e.g., "Genau, du hast es erkannt, nicht die große fossile Industrie, sondern die Demobanner und Fahrten für bzw. zu Aktion dagegen heizen die Klimakrise an!")
		Comments with more than one argument: If a single comment contains more than one argument, the comment should be labelled as showing internal justification as long as at least one of its arguments fulfils the criteria described in the definition. The presence of additional arguments that do not fulfil these requirements does not change this label.
		Comments that contain ad hominem arguments or whataboutisms: If a comment contains an argument that fulfils the criteria described in the definition, it should be labelled as showing internal justification, even if it also includes ad hominem arguments or whataboutisms. Therefore, the presence of these techniques, as defined in the next column, does not a not change this label.
9	Links to external sources or social media	Links to external sources or social media are <u>not</u> part of the assessment of this category.
10	Examples	Present (One full Reason-Conclusion-(Implicit)Warrant-pair): User 1 (Context): Berlin, NRW, Schleswig-Holstein. A100, A45, A20, A23. Überall, wo Grüne an der Macht sind, wird der Ausbau notwendiger Straßen behindert. Das sorgt für CO2-Emissionen durch Staus. Jetzt werfen Grüne Volker Wissing vor, im Verkehrssektor die Klimaziele zu verfehlen. Zynisch! User 2: Das ist ziemlich unterkomplex gedacht. Mehr Straßen - mehr Verkehr. Außerdem sind die Bau-Emissionen nicht zu unterschätzen (<i>Reason</i>). Wer #Klimaschutz ernst nimmt, sollte Straßenneu- und Ausbau stoppen und nur noch den Bestand sanieren (<i>Conclusion</i>).
		Absent (Claim but no Reason): User 1 (Context): Berlin, NRW, Schleswig-Holstein. A100, A45, A20, A23. Überall, wo Grüne an der Macht sind, wird der Ausbau notwendiger Straßen behindert. Das sorgt für CO2-Emissionen durch Staus. Jetzt werfen Grüne Volker Wissing vor, im Verkehrssektor die Klimaziele zu verfehlen. Zynisch! User 2: Wie wäre es mal mit sanieren! Und Ausbau von Bahnstrecken (<i>Claim</i>).
11	Sources operationalization	Duncan et al., 2024; Behrendt et al., 2024; Esau et al., 2023; Reveilhac, 2023; Oswald. 2022: Jaidka, 2022; Schroll & Huber, 2022 Fournier-Tombs und MacKenzie, 2021; Friess & Eilders, 2015
Formal dialogic exchange		
1	Variable shortcut	Int_form
2	Definition	Direct or indirect reference to other users' comments or other users (e.g., addressing practices such as the use of personal pronouns like "du", "Sie", "wir" etc., or by mentioning the party of the prior user if it is a politician, e.g., "die FDP"). Only the use of the reply function (@) at the beginning of the comment does not count as a "formal dialogic exchange", as these mentions are automatically included in the dataset. In contrast, @ mentions

		occurring within the body of the comment should be coded as present, as they indicate an intentional reference to another user and constitute a dialogic element.	
3	Explanation of concepts used in the definition	-	
4	Special cases/ Additional information	Use of personal pronouns: It does not matter how the address may be intended, it is sufficient that, for example, a personal pronoun is present.	
		Indirect reference to other users: An exception applies when a reference to another user occurs in a more indirect form. Even without the use of explicit personal pronouns, the comment can be coded as present if it implies interpersonal reference. For example , through imperatives ("Sag mir, was ist nochmal besser geworden?"). If it is unclear, whether an indirect reference is present or not, the comment should be labelled with NA .	
5	Context	The comment currently being coded should be evaluated independently , <u>without</u> taking into account any prior comments in the reply tree or external links when assessing formal dialogic exchange. However, if the comment lacks information needed to assess the category, <u>review previous comments</u> in the reply tree to clarify the context to potentially gain the necessary information (e.g. , to identify if a public speaker is part of a party which was directly addressed in a comment). This includes the comment to which the evaluated comment is replying, and, if necessary, earlier comments in the same reply tree (including the public speaker post).	
6	Scale level	1	Present
		0	Absent
7	Distinction from or relation with other variables	-	
8	Coding instructions	Use of hashtags: When analysing this criterion, hashtags used in the post are also taken into account, such as #CDU if a politician from that party was the author of a prior post.	
9	Links to external sources or social media	Links to external sources or social media are <u>not</u> part of the assessment of this category.	
10	Examples	Present (Reference to another user via personal pronouns): User 1 (Context): Schöne Studie; Erklärung von @foes_news zu den negativen (!) ökologischen, ökonomischen und sozialen Wirkungen vom #Dienstwagenprivileg. Reformvorschlag: u.a. pauschale Besteuerung von 1% auf 2% anheben User 2: Sie wollen die pauschale Besteuerung der privaten Nutzung von 1% auf 2% erhöhen und als weiteren Faktor die tatsächliche Fahrleistung berücksichtigen? · Kann es sein, dass Sie im Expertenrat die Regelung nicht ganz verstanden haben?	
		Present (Reference to another user by mentioning username): User 1 (Context): Schöne Studie; Erklärung von @foes_news zu den negativen (!) ökologischen, ökonomischen und sozialen Wirkungen vom #Dienstwagenprivileg. Reformvorschlag: u.a. pauschale Besteuerung von 1% auf 2% anheben	

		User 2: Fr. Knopf, allein das Wort "Dienstwagenprivileg" ist eine dreiste Lüge. Ohne die Pauschalbesteuerung auch privat nutzbarer Firmen-Kfz gäbe es noch mehr zugelassene Kfz!	
11	Sources operationalization	Duncan et al., 2024; Esau et al., 2023; Reveilhac, 2023; Oswald. 2022; Friess & Eilders, 2015; Bächtiger et al., 2010	
Substantial dialogic exchange			
1	Variable shortcut	Int_sub	
2	Definition	Reference to prior users' arguments, claims, or questions or only parts of their arguments, claims, or questions. This includes, for example : responding to existing arguments, presenting counterarguments, asking questions.	
3	Explanation of concepts used in the definition	-	
4	Special cases/ Additional information	This also includes meta-discussions about, for example , the argumentation style or behaviour of other users, provided these are an <u>explicit</u> part of the ongoing exchange. For Example: A comment is addressing another user's argumentation technique (e.g., “ User 1: You seem to be replacing expertise with ad hominem attacks”; User 2: “People who accuse others of unfair attacks without sticking to the topic are all too common here”).	
5	Context	Consider the immediate context of the comment when assessing Substantial Dialogic Exchange. This includes the comment to which the evaluated comment is replying. If additional context is needed, earlier replies from the same reply tree (including the public speaker post) may be consulted (e.g., to identify claims referenced in the context comment). This approach follows the same logic as stated in the “Internal Justification” category.	
6	Scale level	1	Present
		0	Absent
7	Distinction from or relation with other variables	-	
8	Coding instructions	Tone of the comment is irrelevant: The tone or assumed intention of the comment is irrelevant for category assignment. The category solely examines whether there is an <u>interactive exchange of opinions, views or ideas</u> between users, without evaluating whether users respect other viewpoints. For example , the presence of irony, sarcasm or whataboutism in a comment does not affect its labelling, provided the category definition is fulfilled.	
		Comment only refers to parts of a prior argument:	

		Comments that refer only to parts of a prior argument, claim or question, rather than the entire argument, should also be labelled as present, provided the category definition is fulfilled. Therefore, is not necessary for a comment to reference the full argument, as this is not to be expected in an online environment.
		Reference to arguments whose context is not explicit: If a comment's reference to the argument of a prior user or public speaker is only clear when the coder is already familiar with highly specific (e.g., political) context, the comment should be coded as NA . Such context-dependent references may be interpreted differently by different coders, especially if the debate or event took place some time ago.
9	Links to external sources or social media	Links to external sources: If links to external sources (e.g., newspapers) are present and accessible, they are treated as part of the contextual information to assess "Substantial dialogic exchange" (see "Definition"). In such cases, coders should therefore consider different types of metadata of the linked source to assess whether the comment shows substantial dialogic exchange (e.g., if the external source leads to a newspaper article about the discussion topic, which addresses a prior argument from another user). However, due to potential limitations in the automated scraping of Twitter data, our manual annotation is in that case restricted to the consideration of the URL , the headline and the source (e.g. FAZ) of the linked page. Therefore, you sometimes have to assume, for example, if the title of an article, indicates that the article addresses the argument of a prior user in the X conversation. If unsure, label the category with NA .
		Links to social media posts (e.g., X posts): If social media posts from X, or other platforms are linked/shared in a comment and accessible, they are treated as part of the contextual information to assess "Substantial dialogic exchange" (see "Definition"). In such cases, coders should therefore consider the whole text (e.g., to assess whether the linked/shared comment includes information about the discussion topic, which addresses a prior argument from another user) of the linked post, but <u>not</u> present (audio-)visual content.
10	Examples	Present (Reference to parts of another user's argument, similar view): User 1 (Context): #Lützerath ist kein Symbol. Unter Lützerath liegt Kohle, die nie verbrannt werden darf, denn dabei wird mehr CO2 freigesetzt, als wir uns leisten können. Es geht hier darum unser Versprechen einzulösen, das schlimmste der Klimakatastrophe zu verhindern! User 2: Ja genau, wenn Ihr das nicht gedreht kriegt, seid Ihr politisch moralisch tot - in meinen Augen fing es im Bund damit an, das Verkehrsministerium dem irrelevanten Aussenamt zu opfern (und damit schöne Reden ohne Umsetzungsrelevanz zu bekommen)
		Absent (No reference to previous argument): User 1 (Context): Herzlichen Glückwunsch natürlich auch in den „Norden“ (von uns @GrueneBW aus gesehen 😊) an Anne #Spiegel & @gruene_rlp. Es zahlt sich aus, wenn Grüne in der Regierung gestalten & #Klimaschutz praktisch umsetzen! #ltwbw21 #ltwrlp21 User 2: @cem_oezdemir @GrueneBW Bitte Ampel!!!
11	Sources operationalization	Behrendt et al., 2024; Esau et al., 2023; Jaidka, 2022

Civility			
1	Variable shortcut	Civ	
2	Definition	No signs of impolite language and no signs of uncivil language .	
3	Explanation of concepts used in the definition	Impolite language , which includes language that maximizes interpersonal conflict: · mocking (e.g., by strong cynicism) or devaluing others (e.g., by derogatory language such as “Systemling”, “Fake News”) · provoking others (e.g., “It’s hilarious how you think your opinion matters”), · or disregarding the rules of etiquette (e.g., by using vulgar language, or screaming, which is indicated by exclamation marks and by large parts of the comment or entire sentences written in CAPITAL LETTERS; one or two capitalized words for emphasis do not qualify as screaming)	
		Uncivil language , includes for example: · hate speech and other attacks on individuals or groups, for example, based on their gender, social or ethnic background (e.g., The refugees are descending on us like locusts”), · as well as threats to democracy as a whole (e.g., “We need a dictator to solve this whole mess”), · offensive language (e.g., insults: “You bastard!”), · stereotypes (e.g., “All refugees are criminal”) · threatening others (e.g., “If I ever meet you in real life, I’ll kill you”)	
4	Special cases/ Additional information	When evaluating the aspect of impolite language, consider the conversational context of the conversation and the tone of the comment. Avoid to be too strict in your classification. However, the use of uncivil language , even in the context of a critique, will result in the classification of the post as <u>absent</u> (for the only exception, see below).	
		Uncivil tongue or responsible incivility: Uncivil utterances (besides hate speech and threats to democracy, see “Explanation of concepts used in the definition”) “[...] against social injustice can sometimes be considered legitimate, [but] only when its positive democratic consequences outweigh the negative ones.” (Bormann und Ziegele, 2023, p. 209). Example: “[M]arginalized groups can use incivility to draw attention to their problems and fight for their right” (Bormann und Ziegele, 2023, p. 209).	
5	Context	The comment currently being coded should be evaluated independently , when assessing civility. However, if the comment lacks information needed to assess the category, <u>review previous comments</u> in the reply tree to clarify the context to potentially gain the necessary information (e.g. , to identify if a comment contains a strong opinion or not). This includes the comment to which the evaluated comment is replying, and, if necessary, earlier comments in the same reply tree (including the public speaker post).	
6	Scale level	1	Present
		0	Absent

7	Distinction from or relation with other variables	-
8	Coding instructions	Distinction between a strong opinion and incivility: Therefore, a post can still be classified as civil, if it critiques or dismisses others' viewpoints or arguments <u>without targeting the other user and without using uncivil language</u>. For example, by employing mild sarcasm or by expressing strong opinions. Annotators should always ask themselves whether the utterance is really meant as an offense. Example: The adjective "dumm" may be offensive in the first utterance below (notice that the speaker generally despises the party), but it should only be considered as some criticism in the second utterance (here the speaker just criticizes some particular action of the party): 1. "Die dumme SPD, immer von gestern, hat noch nie etwas Sinnvolles zu Stande gebracht." 2. "Die SPD ist dumm, wenn sie glaubt einfach mit dem Schlagwort ‚soziale Gerechtigkeit‘ die Wahl zu gewinnen."
9	Links to external sources or social media	Links to external sources: If links to external sources (e.g., blogs) are present and accessible, they are treated as part of the contextual information to assess "Civility" (see "Definition"). In such cases, coders should therefore consider different types of metadata of the linked source to assess whether the comment shows civility or not (e.g., if the external source leads to a blog article, where immigrants are described as "animals" in the title). However, due to potential limitations in the automated scraping of Twitter data, our manual annotation is in that case restricted to the consideration of the URL, the headline and the source (e.g. FAZ) of the linked page. Therefore, you can only decide by looking at the title of an article, whether it is uncivil or not. If unsure, label the category with NA. Links to social media posts (e.g., X posts): If social media posts from X, or other platforms are linked/shared in a comment and accessible, they are treated as part of the contextual information to assess "Civility" (see "Definition"). In such cases, coders should therefore consider the whole text (e.g., if the linked/shared comment contains insults directed at a minority) of the linked post, but <u>not</u> present (audio-)visual content.
10	Examples	Present (Strong opinion with impolite language, but no incivility): User 1 (Context): Berlin, NRW, Schleswig-Holstein. A100, A45, A20, A23. Überall, wo Grüne an der Macht sind, wird der Ausbau notwendiger Straßen behindert. Das sorgt für CO2-Emissionen durch Staus. Jetzt werfen Grüne Volker Wissing vor, im Verkehrssektor die Klimaziele zu verfehlen. Zynisch! User 2: Leider Blödsinn 🙄 Absent (Impolite language, but targeted at individuals) User 1 (Context): Wärmepumpen, E-Autos, Solar- und Windenergie machen uns immer unabhängiger. Warum werden diese Zukunftstechnologien permanent schlechtgeredet? Das verunsichert die Leute, schadet der Wirtschaft und vernichtet Jobs. User 2: Was seid ihr doch für grüne MINT Versager...
11	Sources operationalization	Behrendt et al., 2024; Duncan et al., 2024; Bormann & Ziegele, 2023; Esau et al., 2023; Esau et al., 2021; Jaidka, 2022; Oswald, 2022; Esau et al., 2023; Friess et al., 2020; Friess & Eilders, 2015

Constructiveness			
1	Variable shortcut	Con	
2	Definition	Comment contains one or more constructive elements; the quality and depth of the contribution is not relevant. This includes: · proposing a <u>serious solution</u> or <u>demand</u> related to a prior discussion topic with no signs of sarcasm and irony, · identifying <u>common ground</u> between users with opposing views (e.g., formulating a partial agreement), · trying to clarify open questions <u>related to a prior topic or argument</u> (it is not relevant, if these questions were serious or not), · introducing new relevant and traceable information (see "External justification" for definition) on the discussion topic, · or a <u>serious call for objectivity</u> (e.g., a request for a fact-based discussion when users become very emotional about a topic)	
3	Explanation of concepts used in the definition	Minimal requirements of a serious solutions: A serious solution is present if a <u>non-ironic</u> or <u>non-sarcastic</u> suggestion or appeal is made that contains a <u>demand</u> or <u>suggestion</u>, even if it is <u>not clear</u> who should implement it or how it should be carried out. However, if the provided solution specifically <u>targets an individual in a negative way</u>, constructiveness should be labelled as absent.	
		Indicators of finding common ground: If a comment shows <u>any willingness or attempt to find a common ground</u>, constructiveness should be labelled as present. Indicators of common ground seeking can include, for example, formulating a <u>partial agreement</u> (e.g., "Ich stimme zu, dass erneuerbare Energien wichtig sind, aber Windräder sind nicht die Lösung unseres Energieproblems"); <u>suggesting a compromise</u> (e.g., "Wir könnten sowohl Windräder als auch die von Ihnen vorgeschlagenen E-Fuels zur Umsetzung der Energiewende nutzen").	
4	Special cases/ Additional information	NA: If it is <u>unclear</u> whether the contribution is serious (e.g., the reply is ambiguous, is sarcastic or could be interpreted in multiple ways), it should be labelled with NA.	
		Relevance of context for new relevant and traceable information: If a source has been clearly named or linked in a previous contribution and it is clear that the comment refers to this source, then this context makes the information count as "traceable" even if the source is not mentioned in the coded comment.	
5	Context	Consider the immediate context of the comment when assessing constructiveness. This includes the comment to which the evaluated comment is replying (e.g., to assess whether a comment refers to a traceable source and provides new relevant information from it without explicitly mentioning the source again) and, if necessary, earlier comments in the same reply tree (including the public speaker post). Because, if the direct parent comment lacks information needed to assess constructiveness, review previous comments in the reply tree to clarify the context.	
6	Scale level	1	Present
		0	Absent
7	Distinction from or relation with other variables	In contrast to the "common good orientation" category, this category does <u>not</u> require the post to demonstrate an attempt to take common interests into account while proposing solutions or engaging in other forms of constructiveness, which are mentioned in the definition.	

8	Coding instructions	Use of hashtags while providing a serious solution: Hashtags may be included in the analysis only if they are <u>directly linked</u> to a demand or suggestion. This is typically indicated by specific signal words in the text (e.g. , “fordere”, “will”, “soll”, “muss”, “müsste”, “braucht”, “wünsche”, “würde gerne”, “möchte”, “sollte”, “plädiere für”, “appelliere”, „befürworte“, etc.). Hashtags <u>without</u> such signal words should not be considered as indicators.
9	Links to external sources or social media	Links to external sources: If links to external sources (e.g., newspaper articles) are present and accessible, they are treated as part of the contextual information to assess “Constructiveness” (see “Definition”). In such cases, coders should therefore consider different types of metadata of the linked source to assess whether the comment shows constructiveness (e.g., if the external source leads to an article, where new relevant information on the discussion topic is provided, indicated by the title of the article). However, due to potential limitations in the automated scraping of Twitter data, our manual annotation is in that case restricted to the consideration of the URL , the headline and the source (e.g. FAZ) of the linked page. Therefore, you sometimes have to assume, for example, if the title of an article, indicates whether it could contain new relevant information about the discussed topic in a X conversation. If unsure, label the category with NA .
		Links to social media posts (e.g., X posts): If social media posts from X, or other platforms are linked/shared in a comment and accessible, they are treated as part of the contextual information to assess “Constructiveness” (see “Definition”). In such cases, coders should therefore consider the whole text (e.g., if the linked/shared comment contains insults directed at a minority) of the linked post, but <u>not</u> present (audio-)visual content.
10	Examples	Present (new relevant and credible information): User 1 (Context): Fossiles Flüssiggas importieren? Are you fracking kidding me? 🤪❌❌❌ Das Unverständnis der Klimabewegung teile ich. Und #EndeGelände geht jetzt los und tut was dagegen - ich bin heute wieder dabei als Parlamentarischer Beobachter. https://t.co/RY6iMfCc5x User 2: Es gibt keine rechtliche Grundlage für den Begriff ""parlamentarischer Beobachter"". Siehe z.B. hier: https://t.co/ysXjHKuswh
		Present (clarifying open questions related to a prior topic or argument): User 1 (Context): Wärmepumpen, E-Autos, Solar- und Windenergie machen uns immer unabhängiger. Warum werden diese Zukunftstechnologien permanent schlechtgeredet? Das verunsichert die Leute, schadet der Wirtschaft und vernichtet Jobs. https://t.co/Reta27gH1g User 2: Vielleicht weil sie teilweise abhängiger machen anstelle unabhängiger?
		Present (serious solution related to a prior discussion topic): User 1 (Context): Die Vernunft hat sich durchgesetzt! Moderne Kernkraft, Forschung in Kernfusion muss unser Weg sein! Klimaschutz durch Forschung ist der Weg! #AKWs #Kernenergie #Blackout #Energie #Emsland User 2: [...] In Sachen #Kernfusion muss mehr geforscht werden bis zur Marktreife – das ist doch ganz klar!
		Present (identifying common ground between users): User 1 (Context): Bayern und Baden-Württemberg fallen beim Ausbau der Wind Energie stark zurück. Immer mehr entwickelt sich ein Nord-Süd-Gefälle für die Zukunft der Standorte mit guter Energieversorgung. User 2: Ich sehe bei uns oft Windräder wegen Flaute stehen. Vielleicht sollte man besser Windräder und Photovoltaik zusammen betrachten

11	Sources operationalization	Esau et al., 2023; Reveilhac, 2023; Esau, 2022; Jaidka, 2022; Schroll & Huber, 2022; Esau et al., 2021; Fournier-Tombs und MacKenzie, 2021; Fournier-Tombs & Di Marzo Serugendo, 2020; Friess et al., 2020; Friess & Eilders, 2015; Bächtiger et al., 2010; Steenbergen et al, 2003	
Common good orientation			
1	Variable shortcut	Comgo	
2	Definition	The user links claims, policies, recommendations, or solutions to the (perceived) interests or well-being of the society, or particularities for regional issues, or specific groups (e.g., minorities or marginalized groups). This also includes reference to broader entities such as the environment, climate, humanity, or creation. However, it is <u>not necessary</u> to mention why the specific topic is of interest to this group, i.e., whether it has negative or positive consequences. Focusing only on one's own interests or an unspecified interest group <u>does not</u> fulfil the criteria of common good orientation. Furthermore, the comment must <u>not</u> actively dismiss the interests of other groups.	
3	Explanation of concepts used in the definition	Explicitly and other markers of common good orientation, if no specific group is mentioned: If the comment shows the following markers, common good orientation is still labelled as present even if <u>no</u> specific group is mentioned: - o ... the use of plural possessive pronouns (e.g., “wir”, “unsere”) <u>in combination with</u> calls for justice, fairness, equality, etc. - o ... the use of high numerals (e.g., “tausende”), <u>in combination with</u> mentioning something that is commonly considered positive or negative (e.g., “tausende Jobs gingen verloren”) For example: „@CDU Wie lange soll es denn noch dauern, bis diese Innovationen und Technologien Marktreife erlangt haben? Verdammt - wir haben keine Zeit mehr für so ein Brimborium. Solar- & Windenergie wären da gewesen - aber der Ausbau wurde abgewürgt, Tausende Arbeitsplätze gingen verloren.“	
4	Special cases/ Additional information	-	
5	Context	The comment currently being coded should generally be evaluated independently, without considering prior comments in the reply tree when assessing common good orientation. However, if the comment lacks information needed to assess the category, <u>review previous comments</u> in the reply tree to clarify the context to potentially gain the necessary information (e.g., to identify a specific claim or group the comment is referring to). This includes the comment to which the evaluated comment is replying, and, if necessary, earlier comments in the same reply tree (including the public speaker post).	
6	Scale level	1	Present
		0	Absent
7	Distinction from or relation with other variables	-	

8	Coding instructions	Use of hashtags: If present, hashtags should be considered in the analysis only if they clearly reinforce the argument's focus on the interests or well-being of a larger group, society, or collective entity. This is typically indicated by the provided markers. Hashtags without such signal words or other explicit markers should not be considered as evidence for showing common good orientation.
9	Links to external sources or social media	Links to external sources: If links to external sources (e.g., newspaper articles) are present and accessible, they are treated as part of the contextual information to assess "Common Good Orientation" (see "Definition"). In such cases, coders should therefore consider different types of metadata of the linked source to assess whether the comment shows common good orientation (e.g., if the external source leads to an article, where interests of a minority group related to the discussed topic are mentioned, indicated by the title of the article). However, due to potential limitations in the automated scraping of Twitter data, our manual annotation is in that case restricted to the consideration of the URL, the headline and the source (e.g. FAZ) of the linked page. Therefore, you sometimes have to assume, for example, if the title of an article, indicates whether the interests of a specific group are addresses by the article. If unsure, label the category with NA. Links to social media posts (e.g., X posts): If social media posts from X, or other platforms are linked/shared in a comment and accessible, they are treated as part of the contextual information to assess "Common Good Orientation" (see "Definition"). In such cases, coders should therefore consider the whole text (e.g., if the linked/shared comment addresses interests of a minority group related to the discussed topic are mentioned) of the linked post, but <u>not</u> present (audio-)visual content.
10	Examples	Present (Solution with respect to social justice): User 1 (Context): Es sind die Ärmeren, die zuerst unter der #Klimakrise leiden, in #Kühlräume müssen, nicht ausweichen können, unter teurerem Essen leiden... wir wissen es lange, wir reden schon lange drüber. Zeit für eine Regierung, die endlich anpackt statt zu negieren. #bereitweillihresseid User 2: Das Energiegeld kann doch gerade für den sozialen Ausgleich sorgen. Und darüber hinaus kann jeder durch Co2 Einsparungen weiter Kosten minimieren. Dafür braucht es nicht gleich ein neues Haus, oder neues Auto. Present (Claim with connection to interests of society as a whole): User 1 (Context): Schlimmste Nachricht heute: Schwere Dürre in #Niedersachsen. Die #Klimakatastrophe ist da und wir müssen alles tun, um fossile Energien in der Erde zu lassen und Sonne und Wind auszubauen. #Klimaschutz Jetzt! User 2: Alle profitieren von Solar- und Windkraft durch günstigere Preise und sichere Versorgung.
11	Sources operationalization	Fournier-Tombs und MacKenzie, 2021; Friess & Eilders, 2015
Emotions (negative/positive)		
1	Variable shortcut	Emo_negpos
2	Definition	Only explicitly stated negative or positive emotions should be coded. Emojis serve as <u>main indicators</u> of emotions. See coding instructions for further descriptions and examples.

3	Explanation of concepts used in the definition	Def. Emotions: Emotions can be categorized as positive or negative, if identifiable. Fundamental positive emotions include anticipation, trust, joy, and surprise, while fundamental negative emotions include anger, fear, sadness, and disgust. Emotions also vary in their intensity, for example , anger is less intense than rage, but both belong to the same spectrum (aggressiveness).	
4	Special cases/ Additional information	-	
5	Context	The comment currently being coded should generally be evaluated independently , without considering prior comments in the reply tree when assessing common good orientation. However, if the comment lacks information needed to assess the category, <u>review previous comments</u> in the reply tree to clarify the context to potentially gain the necessary information (e.g. , to identify the context for the use of an emoji, if the comment contains no other content). This includes the comment to which the evaluated comment is replying.	
6	Scale level	3	Unclassifiable emotion: The user comment expresses an emotion, but it is not clear, whether the emotion is positive or negative.
		2	Positive Emotion: The user comment expresses <u>an explicit positive</u> emotion.
		1	Negative Emotion: The user comment expresses <u>an explicit negative</u> emotion.
		0	No emotion: The user comment expresses <u>no</u> explicit negative or positive emotion.
7	Distinction from or relation with other variables	-	
8	Coding instructions	Linguistic features of explicitly stated emotions: By considering the following linguistic features and emojis, it is possible to assess whether a comment contains an explicit emotion. If you do not have sufficient certainty about whether a comment contains an emotion or not, the comment should be labelled with NA. Signal words such as "wütend" (angry), "traurig" (sad), "haha" (joy), as well as synonyms (e.g., "furious" or "angry") indicate emotions. But only, if they are used in combination with personal pronouns (e.g., "ich", "wir", "mich") and verbs such as "fühle" or "macht" (e.g., "macht mich wütend"; "ich fühle mich schlecht"), suggesting that the emotion is experienced or shared by the speaker.	
		Emojis: Furthermore, emotions can be conveyed through the use of emojis, such as ":(" or "😞" (neutral or gesture-based emojis or symbols like "thumbs-up":👍 should be deprioritized if they are too ambiguous in the context of the comment), when considered in the context of the tone/sentiment of	

		the comment. If this is not sufficiently clear from the comment, previous comments can be included as context. If this is still not enough to identify the type of emotion, it should be coded as “Unclassifiable emotion” (3). For example: "Mit vielen Windrädern und Wärmepumpen wird es wohl keine Stille mehr geben. Das macht uns dann sicher auch krank...es ist zum 🤢."
9	Links to external sources or social media	Links to external sources or social media are <u>not</u> part of the assessment of this category.
10	Examples	Present (Negative emotion indicated by an emoji): „Sie wollen die pauschale Besteuerung der privaten Nutzung von 1% auf 2% erhöhen und als weiteren Faktor die tatsächliche Fahrleistung berücksichtigen? .“
11	Sources operationalization	Esau et al., 2023; Oswald, 2022; Esau, 2022; Esau et al., 2021; Friess et al., 2020
Empathy (Communicative empathy)		
Important: It is sufficient for one of the two indicators (“Communicative empathy” or “Cognitive/emotional perspective taking”) to be fulfilled to label “ Empathy ” as <u>present</u> .		
1	Variable shortcut	Emp
2	Definition	Comment shows signs of empathy such as <u>expression of sympathy, emotional support and/or solidarity (e.g., “Danke, dass du da [auf der Demo] bist 🍌”); help offers</u> or other <u>empathetic linguistic features (e.g., "Das tut mir leid")</u> .
3	Explanation of concepts used in the definition	-
4	Special cases/ Additional information	This can be directed to a prior user of the <u>same conversation</u> , but also to entities or groups, which are <u>not part of the conversation</u> (e.g., “Mir tuen die nachfolgenden Generationen leid, die unser Versagen im Klimaschutz ausbaden müssen).
5	Context	Consider the immediate context of the comment when assessing communicative empathy. This includes the comment to which the evaluated comment is replying, and, if necessary, earlier comments in the same reply tree (including the public speaker post). Because, if the direct parent comment lacks information needed to understand the empathetic response, review previous comments in the reply tree to clarify the context.
6	Scale level	1 Present
		0 Absent
7	Distinction from or relation with other variables	-

8	Coding instructions	Emojis: Also consider emojis when analysing this criterion.	
9	Links to external sources or social media	Links to external sources or social media are <u>not</u> part of the assessment of this category.	
10	Examples	Present (Expression of emotional support/solidarity): User: Das tut mir leid, Gösta! Bleib aufrecht und mach weiter! Als Mitglied von B'90/Die Grünen und Klima-Bewegter bedauere ich sehr, dass gerade du es nicht geschafft hast.	
11	Sources operationalization	Esau et al., 2023; Esau, 2022; Oswald, 2022	
Empathy (Cognitive/emotional perspective taking)			
1	Variable shortcut	Emp	
2	Definition	Cognitive/emotional perspective taking: Comment shows signs of perspective-taking , which represents a highly interactive phenomenon , so it only can be present when a user directly addresses a comment of a <u>prior user</u> .	
3	Explanation of concepts used in the definition	Cognitive perspective taking: For example, attempts to <u>understand another user's stance</u> or reasoning (<i>Cognitive perspective taking</i>): "Warum denkst du das?".	
		Emotional perspective taking: For example, user shows <u>understanding for the emotions of another user</u> or tries to understand them (<i>Emotional perspective taking</i>): "Ich verstehe warum du so wütend bist".	
4	Special cases/ Additional information	-	
5	Context	Consider the immediate context of the comment when assessing communicative empathy. This includes the comment to which the evaluated comment is replying, and, if necessary, earlier comments in the same reply tree (including the public speaker post). Because, if the direct parent comment lacks information needed to understand the empathetic response, review previous comments in the reply tree to clarify the context.	
6	Scale level	1	Present
		0	Absent
7	Distinction from or relation with other variables	-	
8	Coding	Emojis: Also consider emojis when analysing this criterion.	

	instructions	
9	Links to external sources or social media	Links to external sources or social media are <u>not</u> part of the assessment of this category.
9	Examples	Present (Emotional perspective taking): User 1 (Context): Aus Respekt? Nein, weil ihr euch wegduckt. Weil ihr keine Antworten habt. Weil all die tollen künftigen Erfindungen, auf die ihr setzt, gegen diese Fluten nicht helfen. Weil der Markt es nicht richtet. Weil euer Freiheitsbegriff in der Krise nackt dasteht und für nichts taugt. User 2: „Ich verstehe diese wütenden Tweets einfach nicht. Also ... ich verstehe die Emotionen, aber nicht, wieso man überhaupt nicht nachprüft, was man sagt. [...]
10	Sources operationalization	Esau et al., 2023; Esau, 2022; Oswald, 202
Storytelling		
1	Variable shortcut	Story
2	Definition	The user tells a story or about an experience of oneself or others , this also includes short or brief depictions. These techniques can have various purposes, for example , they can be used to convey messages, present justifications, support arguments, draw conclusions, make recommendations, or evoke emotions. If a user tells a story, even if it does not contain a real or believable experience or event , this fictitious story it is also labelled as storytelling.
3	Explanation of concepts used in the definition	The <u>sole</u> use of conventional phrases (commonly used fixed expressions; e.g. , “Die Kirche im Dorf lassen”) or idioms (a phrase or a group of words whose meaning is different from the literal meaning of the individual words; e.g. , “Ins Gras beißen”) is coded as Absent .
4	Special cases/ Additional information	-
5	Context	The comment currently being coded should be evaluated independently , <u>without</u> taking into account any prior comments in the reply tree when assessing storytelling.
6	Scale level	1 Present
		0 Absent
7	Distinction from or relation with other variables	-
8	Coding	Credibility of narratives and toxicity:

	instructions	Described narratives, stories or experiences do not have to be true or believable to count as storytelling. However, if the storytelling technique is being used to spread discriminatory, malicious, or toxic content (e.g. , promoting hate towards individuals or groups), the category is labelled as absent.	
9	Links to external sources or social media	Links to external sources or social media are <u>not</u> part of the assessment of this category.	
10	Examples	Present (Narrative with vivid picture): User 1 (Context): Erklärung von @foes_news zu den negativen (!) ökologischen, ökonomischen und sozialen Wirkungen vom #Dienstwagenprivileg. Reformvorschlag: u.a. pauschale Besteuerung von 1% auf 2% anheben User 2: Ich sag meinem Handwerker Kollegen mit seinem Opel Astra das er sein Privileg 15% auf sein Gehalt extra versteuern muss, aufgeben soll. Ganz böser Typ. Und der 3 Zylinder Diesel erst, man ist der Reich.	
		Present (Story/Experience of oneself): User 1 (Context): Bayern und Baden-Württemberg fallen beim Ausbau der Wind Energie stark zurück. Immer mehr entwickelt sich ein Nord-Süd-Gefälle für die Zukunft der Standorte mit guter Energieversorgung. User 2: Windkraft? Nein, danke. Heute zig Windräder an der Autobahn gesehen, max. ein Prozent davon drehte sich im Wind! Da soll mir mal einer erklären, wo in Zukunft die Energie herkommen soll!	
11	Sources operationalization	Esau et al., 2023; Esau, 2022; Schroll & Huber, 2022; Esau et al., 2021; Fournier-Tombs und MacKenzie, 2021; Bächtiger et al., 2010	
Reflexivity			
1	Variable shortcut	Ref	
2	Definition	The comment contains signs of reflexivity , indicating that the user has not only responded to other claims etc., but has also <u>modified his/her own claims, arguments or opinions</u> with respect to these (counter) arguments.	
3	Explanation of concepts used in the definition	This includes, for example , the demonstration of the willingness to consider alternative perspectives or opposing viewpoints, the acknowledgment of potential biases or limitations in own arguments, or the usage of language that indicates a change in stance or perspective.	
4	Special cases/ Additional information	-	
5	Context	Consider the whole context of the comment when assessing communicative reflexivity. This includes prior comments of the reply tree (including the public speaker post). As reflexivity can most of the time only be properly assessed with knowledge of the entire reply tree.	
6	Scale level	1	Present
		0	Absent

7	Distinction from or relation with other variables	-
8	Coding instructions	-
9	Links to external sources or social media	Links to external sources or social media are <u>not</u> part of the assessment of this category.
10	Examples	Present (modified opinion): User 1 (Context): Es gibt keine rechtliche Grundlage für den Begriff ""parlamentarischer Beobachter"". Siehe z.B. hier: https://t.co/ysXjHKuswh User 2: Vielen Dank für den aufschlussreichen Artikel aus dem Weser-Kurier. Wir haben uns mit sog. ""parlamentarischen Beobachtern"" nie tiefergehend beschäftigt. Dass es wohl für diesen besonderen Status gar keine rechtliche Grundlage gibt, hatten wir nicht gedacht. Wieder was gelernt:
11	Sources operationalization	Mendonça et al., 2022; Esau, 2022; Esau et al., 2023
Humour		
1	Variable shortcut	Hum
2	Definition	The comment contains humour, which includes a “witty, playful or clearly not seriously meant statement that is supposed to make others laugh” (Esau & Frieß, 2022, p. 8).
3	Explanation of concepts used in the definition	-
4	Special cases/ Additional information	-
5	Context	The comment currently being coded should generally be evaluated independently , without considering prior comments in the reply tree when assessing humour. However, if the comment lacks information needed to assess the category, review previous comments in the reply tree to clarify the context to potentially gain the necessary information (e.g. , to understand whether a comment contains fun content). This includes the comment to which the evaluated comment is replying, and, if necessary, earlier comments in the same reply tree (including the public speaker post).
6		1 Present
		0 Absent

7	Distinction from or relation with other variables	-
8	Coding instructions	<u>Only</u> the following indicators are used for coding humour (Ziegele & Frieß, 2018): • Jokes: short stories that provoke laughter with an unexpected twist, a surprising effect or a punchline at the end • Fool around: deliberately talking nonsense or saying silly things • Wordplay: Playing with words, humour is created by the linguistic ambiguity of the word used or by the identical or similar sound of two related words with different meanings • Fun content: information that are intended to be or come across as funny, amusing or entertaining in the context of the discussion • Downplaying: Playful or humorous portrayal of another user's post or linked articles as insignificant • Sarcasm/Irony: Use of sarcasm or irony in a playful or humorous way (see Definition of relevant concepts "<i>Irony and Sarcasm</i>")
9	Links to external sources or social media	Links to external sources or social media are <u>not</u> part of the assessment of this category.
9	Examples	Jokes (Present): "Jetzt aber wirklich! All die anderen Male war nur Joke, jetzt Naht der Untergang! Tut Buße, kauft Ablassbriefe!" Fool around (Present): "Ich finde man könnte der FDP entgegenkommen. Auf allen beschleunigt genehmigten Autobahnen gilt eine Mindestgeschwindigkeit von 200 km/h für PKW und LKW." Wordplay (Present): ":-innen oder auch :außen?" Fun content (Present): ??? Downplaying (Present): "Nötig :Kompetenz Unnötig : Julia Verlinden" Sarcasm/Irony: "Der Braunkohleabbau ist zum Beispiel bekannt dafür, innovativ CO2 einzusparen und gleichzeitig nachhaltig Jobs zu schaffen."
10	Sources operationalization	Ziegele & Frieß, 2018; Esau & Frieß, 2022

APPENDIX B

Table B.1: Results Inter-Coder-Reliability for Random Sample (n= 320): Validation Manual Content Analysis Dataset Statistics

Dimension	Operationalisation	Type	Scale	N (valid cases)	Agreement	Krippendorff's α	Brennan–Prediger	Gwet's AC1
Rationality	External Justification	I	Ordinal	280	97,64 %	.855	.965	.981
	Follow-up-question	I	Dichotomous	285	95,12 %	.755	.902	.959
	Internal Justification	I	Dichotomous	286	87,30 %	.742	.746	.791
Reciprocity	Formal dialogic exchange	I	Dichotomous	289	89,68 %	.776	.794	.835
	Substantial dialogic exchange	I	Dichotomous	272	90,06 %	.698	.801	.914
Civility	Absence of Incivility	I	Dichotomous	294	96,60 %	.676	.932	.960
Constructiveness	Constructive contribution	I	Dichotomous	279	88,00 %	.634	.760	.846
Common Good Orientation	Orientation on common interests	I	Dichotomous	287	91,89 %	.598	.838	.913
Emotions	Explicit expression of emotions	II	Nominal	283	95,14 %	.641	.935	.956
Empathy	Communicative empathy or perspective taking	II	Dichotomous	293	98,30 %	.877	.966	.984

Storytelling	Sourcing or narration	II	Dichotomous	290	93,97 %	.498	.879	.941
Reflexivity	Reflexive contribution	II	Dichotomous	292	98,87 %	.688	.977	.995
Humour	Humorous statement	II	Dichotomous	289	91,13 %	.602	.823	.898

Note: Intercoder reliability metrics for manual content analysis of validation subset (n=280 to 294 valid cases per category) coded by n=3 human coders, including percentage inter-coder-agreement, Krippendorff's α , Brennan–Prediger κ , and Gwet's AC1. Metrics are stratified by codebook dimension, subdimension, type (I/II), and scale.

Table B.2: Human Validation data distribution of labels for all categories for random sample (n= 320)

Category	Class 0	Class 1	Class 2	Class 3	Majority%
External Justification	258	7	15	-	92,1 % (0)
Follow-up-questions	258	27	-	-	90,5 % (0)
Internal Justification	180	106	-	-	62,5 % (0)
Formal Dialogic Exchange	195	94	-	-	67,5 % (0)
Substantial Dialogic Exchange	47	225	-	-	82,7 % (1)
Civility	18	276	-	-	93,9 % (1)
Constructiveness	225	54	-	-	80,6 % (0)
Common Good Orientation	259	28	-	-	90,2 % (0)
Emotions	267	2	11	3	94,3 % (0)
Empathy	273	20	-	-	93,2 % (0)
Storytelling	275	15	-	-	94,8 % (0)
Reflexivity	288	4	-	-	98,6 % (0)
Humour	255	34	-	-	88, 2 % (0)

Note: Distribution of labels for all categories in human validation dataset (before merging of labels of External Justification and Emotions for LLM validation)

Table B.3: Results Inter-Coder-Reliability for Stratified Sample (n= 266): Validation Manual Content Analysis Dataset Statistics

Dimension	Operationalisation	Type	Scale	N (valid cases)	Agreement	Krippendorff's α
Rationality	External Justification	I	Ordinal	234	97,0 %	.877
	Follow-up-question	I	Dichotomous	247	96,8 %	.878
	Internal Justification	I	Dichotomous	254	98,4%	.964
Reciprocity	Formal dialogic exchange	I	Dichotomous	262	94,3 %	.883
	Substantial dialogic exchange	I	Dichotomous	249	94,4 %	.867
Civility	Absence of Incivility	I	Dichotomous	261	96,9 %	.749
Constructiveness	Constructive contribution	I	Dichotomous	257	95,3 %	.883
Common Good Orientation	Orientation on common interests	I	Dichotomous	260	96,9 %	.944
Emotions	Explicit expression of emotions	II	Nominal	251	97,2 %	.936
Empathy	Communicative empathy or perspective taking	II	Dichotomous	261	97,7 %	.942

Storytelling	Sourcing or narration	II	Dichotomous	261	91,6 %	.760
Reflexivity	Reflexive contribution	II	Dichotomous	261	99,2 %	.969
Humour	Humorous statement	II	Dichotomous	253	98,8%	.965

Note: Inter-coder reliability metrics for manual content analysis of stratified validation subset (n=234 to 262 valid cases per category) coded by n=2 human coders, including percentage inter-coder-agreement, and Krippendorff's α . Metrics are stratified by codebook dimension, subdimension, type (I/II), and scale.

Table B.4: Human Validation data distribution of labels for all categories for stratified sample (n= 266)

Category	Class 0	Class 1	Class 2	Class 3	Majority%
External Justification	199	7	21	0	87,7 % (0)
Follow up questions	204	35	0	0	85,4 % (0)
Internal Justification	169	81	0	0	67,6 % (0)
Formal Dialogic Exchange	145	102	0	0	58,7 % (0)
Substantial Dialogic Exchange	68	167	0	0	71,1 % (1)
Civility	13	240	0	0	94,9 % (1)
Constructiveness	180	65	0	0	73,5 % (0)
Common Good Orientation	200	55	0	0	78,4 % (0)
Emotions	181	15	41	7	74,2 % (0)
Empathy	188	67	0	0	73,7 % (0)
Storytelling	191	48	0	0	79,9 % (0)
Reflexivity	223	37	0	0	85,8 % (0)
Humour	198	52	0	0	79,2 % (0)

Note: Distribution of labels for all categories in human validation dataset (before merging of labels of External Justification and Emotions for LLM validation)

APPENDIX C

Table C.1: LLM Validation data distribution of labels for all categories (n= 586)

Category	Class 0	Class 1	Class 2	Majority%
External Justification	457	49	0	90,3 % (0)
Follow up questions	414	62	0	87,0 % (0)
Internal Justification	349	187	0	65,1 % (0)
Formal Dialogic Exchange	340	196	0	63,4 % (0)
Substantial Dialogic Exchange	69	392	0	85,0 % (1)
Civility	31	516	0	94,3 % (1)
Constructiveness	405	118	0	77,3 % (0)
Common Good Orientation	459	83	0	84,7 % (0)
Emotions	458	17	52	86,9 % (0)
Empathy	461	87	0	84,1 % (0)
Storytelling	466	63	0	88,1 % (0)
Reflexivity	511	41	0	92,6 % (0)
Humour	453	86	0	84,0 % (0)

Note: Distribution of labels for all categories in the LLM-validation dataset (after merging of labels of External Justification and Emotions for LLM validation)

Figure C.1: Label distribution per category (After merging – LLM validation)



Table C.2: Example LLM-prompted Codebook:

CODEBOOK = {

"**External_Justification**": ""

You are an expert annotator. Classify the TARGET comment for: External Justification (Jus_ext).

INPUT

- TARGET: {text}

- REPLY_TO (optional context): {reply_to_text}

DEFINITION

External justification = a claim supported with evidence that is:

- (1) RELEVANT to the claim,
- (2) TRACEABLE to a specific, identifiable source,
- (3) from a HIGH-QUALITY source (scientific studies, established newspapers, government institutions, recognized experts).

CODING INSTRUCTIONS

- Over 90% of posts have NO external justification → default is 0.
- "Some study says" or "experts say" without naming the source → not traceable → 0.
- YouTube or personal website links → traceable but low-quality → 1.
- Named established newspaper, institution, or expert with a specific fact → 2.

CLASSES

- 2: High-quality (relevant + traceable + high-quality source)
- 1: Low-quality (relevant + traceable, but low-quality source)
- 0: Absent ← default for most posts

EXAMPLES

Label 0:

Context: Was viele bei der Klimakrise falsch verstehen: Wenn der Staat uns heute nicht vorschreibt, wie wir heizen dürfen, wird er uns morgen vorschreiben müssen, wann wir den Wasserhahn aufdrehen dürfen.

User: Es gibt keine Klimakrise. Die findet nur in Ihrem Kopf statt. Das Problem mit dem Wasser ist schon Jahrzehnte bekannt, die Ursache hat nix mit Klima zu tun, sondern mit Privatisierung von Süßwasservorkommen.

→ 0

Label 1:

Context: NA

User: Wie retten wir das Klima? Programmdebatte 10.03.2021 <https://t.co/fgYWy5lxsO> via @YouTube

→ 1

Label 2:

Context: Jetzt: Debatte zum #Klimaschutz im #Bundestag.

User: Ihr habt echt keine Ahnung, oder? Und trotzdem labert Ihr die ganze Zeit irgendwas! Warum informiert ihr euch nicht? Ihr tragt Verantwortung!

(Link leads to Website spectrum.de with Headline: „Wie viel CO2 kann Deutschland noch ausstoßen?“)

→ 2

Reply ONLY with 2, 1, or 0.

""", [...] }

Table C.3: LLM Validation performance for all models category-wise

Note: For postprocessing, raw model outputs are normalized by extracting the first valid label token. Unresolvable outputs were coded as the majority label for each dimension to guarantee that the process always outputs a label decision. For thinking models, internal chain-of-thought blocks are stripped before label extraction.

External Justification				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.94	.82	.85	.83
GPT-OSS:20B	.93	.82	.80	.80
Llama3.3:70b	.94	.83	.89	.86
Llama3.1-8b-instruct	.85	.63	.66	.64
Qwen3:30b-thinking	.96	.77	.78	.77
Qwen3:30b-instruct	.69	.57	.68	.55
Gemma3-27b	.76	.63	.84	.64
Gemma-4-26B	.94	.81	.87	.84

Follow-up questions				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.92	.82	.85	.84
GPT-OSS:20B	.91	.80	.88	.83
Llama3.3:70b	.90	.79	.72	.74
Llama3.1-8b-instruct	.71	.60	.56	.44
Qwen3:30b-thinking	.92	.88	.73	.79
Qwen3:30b-instruct	.86	.70	.63	.65
Gemma3-27b	.90	.80	.79	.79
Gemma-4-26B	.92	.92	.70	.76

Internal Justification				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.77	.77	.79	.78
GPT-OSS:20B	.80	.79	.75	.76
Llama3.3:70b	.80	.78	.80	.79
Llama3.1-8b-instruct	.46	.60	.56	.44
Qwen3:30b-thinking	.85	.87	.76	.79
Qwen3:30b-instruct	.59	.66	.66	.59
Gemma3-27b	.57	.67	.66	.57
Gemma-4-26B	.82	.80	.83	.81

Formal Dialogic Exchange				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.75	.75	.77	.75
GPT-OSS:20B	.66	.67	.68	.66
Llama3.3:70b	.45	.59	.55	.42
Llama3.1-8b-instruct	.41	.56	.52	.37
Qwen3:30b-thinking	.67	.71	.75	.67
Qwen3:30b-instruct	.41	.67	.53	.34
Gemma3-27b	.41	.59	.535	.36
Gemma-4-26B	.74	.73	.74	.72

Substantial Dialogic Exchange				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.80	.63	.65	.64
GPT-OSS:20B	.77	.61	.64	.62
Llama3.3:70b	.81	.47	.49	.47
Llama3.1-8b-instruct	.85	.42	.50	.45
Qwen3:30b-thinking	.85	.70	.70	.70
Qwen3:30b-instruct	.84	.42	.49	.45
Gemma3-27b	.82	.49	.50	.48
Gemma-4-26B	.85	.59	.52	.51

Civility				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.90	.68	.87	.73
GPT-OSS:20B	.94	.75	.75	.75
Llama3.3:70b	.93	.70	.72	.71
Llama3.1-8b-instruct	.95	.81	.59	.64
Qwen3:30b-thinking	.97	.88	.73	.79
Qwen3:30b-instruct	.94	.70	.65	.67
Gemma3-27b	.88	.63	.87	.81
Gemma-4-26B	.92	.70	.87	.75

Constructiveness				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.75	.67	.70	.68
GPT-OSS:20B	.71	.62	.64	.62
Llama3.3:70b	.79	.73	.81	.74
Llama3.1-8b-instruct	.44	.63	.63	.44
Qwen3:30b-thinking	.83	.75	.72	.73
Qwen3:30b-instruct	.66	.66	.74	.64
Gemma3-27b	.75	.71	.78	.71
Gemma-4-26B	.74	.71	.79	.71

Common Good Orientation				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.91	.84	.89	.86
GPT-OSS:20B	.91	.81	.82	.82
Llama3.3:70b	.83	.73	.87	.76
Llama3.1-8b-instruct	.61	.62	.74	.56
Qwen3:30b-thinking	.92	.69	.83	.72
Qwen3:30b-instruct	.79	.69	.84	.72
Gemma3-27b	.73	.67	.83	.67
Gemma-4-26B	.83	.74	.89	.77

Emotions				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.81	.54	.68	.59
GPT-OSS:20B	.77	.52	.71	.56
Llama3.3:70b	.56	.47	.72	.45
Llama3.1-8b-instruct	.20	.55	.56	.30
Qwen3:30b-thinking	.87	.59	.68	.63
Qwen3:30b-instruct	.45	.49	.61	.40
Gemma3-27b	.67	.52	.82	.53
Gemma-4-26B	.88	.65	.63	.63

Empathy				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.92	.88	.82	.85
GPT-OSS:20B	.93	.88	.84	.86
Llama3.3:70b	.92	.87	.82	.84
Llama3.1-8b-instruct	.80	.69	.79	.72
Qwen3:30b-thinking	.93	.89	.81	.85
Qwen3:30b-instruct	.89	.79	.80	.79
Gemma3-27b	.89	.81	.89	.84
Gemma-4-26B	.92	.84	.82	.83

Storytelling				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.91	.79	.74	.76
GPT-OSS:20B	.90	.77	.76	.76
Llama3.3:70b	.93	.83	.82	.83
Llama3.1-8b-instruct	.70	.61	.77	.60
Qwen3:30b-thinking	.93	.88	.74	.79
Qwen3:30b-instruct	.87	.72	.79	.75
Gemma3-27b	.90	.76	.86	.80
Gemma-4-26B	.89	.74	.82	.77

Reflexivity				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.98	.92	.92	.92
GPT-OSS:20B	.97	.94	.85	.89
Llama3.3:70b	.96	.94	.77	.83
Llama3.1-8b-instruct	.92	.79	.54	.56
Qwen3:30b-thinking	.97	.85	.61	.67
Qwen3:30b-instruct	.96	.95	.75	.82
Gemma3-27b	.97	.95	.87	.91
Gemma-4-26B	.97	.93	.86	.89

Humor				
Model	Accuracy	Precision	Recall	F1
GPT-OSS:120B	.88	.78	.77	.77
GPT-OSS:20B	.87	.77	.73	.74
Llama3.3:70b	.83	.72	.84	.75
Llama3.1-8b-instruct	.72	.64	.73	.64
Qwen3:30b-thinking	.87	.72	.82	.74
Qwen3:30b-instruct	.83	.70	.74	.72
Gemma3-27b	.79	.69	.81	.72
Gemma-4-26B	.87	.75	.81	.78

Table C.4: LLM Validation performance for best performing model (Gemma-4-26B) for Label “present” (1) category-wise (for Emotions all three label values are reported)

Label 1					
Model	Deliberation Type	Category	Precision	Recall	F1-score
Gemma-4-26B	Type I	External Justification	.64	.79	.71
		Follow-up Questions	.92	.40	.56
		Internal Justification	.69	.85	.76
		Formal Dialogic Exchange	.59	.80	.68
		Substantial Dialogic Exchange	.85	.98	.91
		Civility	.98	.93	.95
		Constructiveness	.46	.85	.60
		Common Good Orientation	.49	.95	.65
	Type II	Emotions (absent)	.93	.92	.92
		Emotions (negative)	.54	.59	.59
		Emotions (positive)	.46	.35	.40
		Empathy	.74	.69	.71
		Storytelling	.51	.72	.60
		Reflexivity	.88	.73	.80
		Humour	.56	.72	.63

Table C.5: LLM Validation performance for best performing model (Gemma-4-26B) category-wise (weighted average)

Weighted Average					
Model	Deliberation Type	Category	Precision	Recall	F1-score
Gemma-4-26B	Type I	External Justification	.95	.94	.94
		Follow-up Questions	.92	.92	.91
		Internal Justification	.84	.82	.82
		Formal Dialogic Exchange	.76	.73	.73
		Substantial Dialogic Exchange	.77	.84	.79
		Civility	.95	.92	.93
		Constructiveness	.83	.74	.77
		Common Good Orientation	.91	.84	.86
	Type II	Emotions	.87	.87	.87
		Empathy	.91	.91	.91
		Storytelling	.91	.88	.89
		Reflexivity	.97	.97	.97
		Humour	.88	.86	.87

Table C.6: Quantization information for all the models

Model	Quantization	LLM server
GPT-OSS:120B (OpenAI et al., 2025)	MXFP4	Ollama
GPT-OSS:20B (OpenAI et al., 2025)	MXFP4	Ollama
Llama3.3:70b (Grattafiori et al., 2024)	Q4_K_M	Ollama
Llama3.1-8b-instruct (Grattafiori et al., 2024)	Q8_0	Ollama
Qwen3:30b-thinking (Yang et al., 2025)	Q4_K_M	Ollama
Qwen3:30b-instruct (Yang et al., 2025)	Q4_K_M	Ollama
Gemma3-27b (Gemma Team et al., 2025)	Q4_K_M	Ollama
Gemma-4-26B (Gemma Team & Google DeepMind, 2026)	AWQ-4bit	vllm

Table C.7 Software and Hardware specifications:

Software and Hardware specifications

The annotation pipeline was implemented in Python 3.10.18. Smaller models were served locally using Ollama and VLLM on a server equipped with a single NVIDIA GeForce RTX 3090 (24GB VRAM), running CUDA 13.0. Models exceeding 30B parameters were evaluated on a bigger computing infrastructure utilizing NVIDIA A100 and H100 GPUs. The Ollama version is 0.30.7, and the VLLM version is 0.22.0. The prevalence visualizations were produced in R(v4.5.3) using the tidyverse package.

APPENDIX D

Table D.1: Regular Expression for thematic selection

Original Version

```
r"(?i)Klima|Erderwärmung|Global.*Erwärmung|Treibhauseffekt|Treibhausgas|CO2-Emission|Kohlendioxid-Emission|Methanemission|Fossile.*Brennstoff|Dekarbonisierung|Erneuerbare.*Energie|Solarenergie|Windkraft|Windrad|Windräd|Windenergie|Wasserkraft|Geothermie|Energiewende|logisch.*Fußabdruck|Emissionshandel|Emissionsreduktion|Emissionsfrei|Kohlenstoffbudget|Pariser.*Abkommen|Kyoto-Protokoll|IPCC|Fridays.*for.*Future|Thunberg|Letzte.*Generation|Extinction.*rebellion"
```

Translated Version (not used for filtering, added for readability only)

```
r"(?i)climate|global warming|global.*warming|greenhouse effect|greenhouse gas|CO2 emission|carbon dioxide emission|methane emission|fossil.*fuel|decarbonization|renewable.*energy|solar energy|wind power|wind turbine|wind turbin|wind energy|hydropower|geothermal energy|energy transition|ecological.*footprint|emissions trading|emission reduction|emission-free|carbon budget|Paris.*Agreement|Kyoto Protocol|IPCC|Fridays.*for.*Future|Thunberg|Last.*Generation|Extinction.*Rebellion"
```

Table D.2: Description of Datasets: validation and full dataset

Category	Full Validation Dataset	Full Analysis Dataset
Total Tweets	586	64223
Total conversations	202	1251
Avg. conversation length	3.01	51.54
Min conversation length	1	5
Max conversation length	26	350
Avg. tweet length (word)	21.89	20.40
Min. tweet length (word)	2	1
Max. tweet length (words)	57	350
Avg. tweet length (character)	165.74	156.01
Min tweet length (character)	7	6
Max tweet length (character)	465	654

Note: The table summarizes the main characteristics of the human validation, LLM validation and the full analysis dataset sampled from the SMO large-scale Twitter dataset.

Table D.3: Results automated measurement of type I and type II deliberative quality in large dataset

Category	All users (n=64,223)			Public Speaker (n=3,447)			Non-Public speaker (n=60,776)		
	Absent (0)	Present (1)		Absent (0)	Present (1)		Absent (0)	Present (1)	
External Justification	55584 (86.55)	8639 (13.45)		1902 (55.18)	1545 (44.82)		53682 (88.33)	7094 (11.67)	
Follow up questions	61416 (95.63)	2807 (4.37)		3352 (97.24)	95 (2.76)		58064 (95.54)	2712 (4.46)	
Internal Justification	34815 (54.21)	29408 (45.79)		1563 (45.34)	1884 (54.66)		33252 (54.71)	27524 (45.29)	
Formal Dialogic Exchange	29701 (46.25)	34522 (53.75)		2078 (60.28)	1369 (39.72)		27623 (45.45)	33153 (54.55)	
Substantial Dialogic Exchange	2780 (4.33)	61443 (95.67)		1311 (38.03)	2136 (61.97)		1469 (2.42)	59307 (97.58)	
Civility	13415 (20.89)	50808 (79.11)		50 (1.45)	3397 (98.55)		13365 (21.99)	47411 (78.01)	
Constructiveness	45716 (71.18)	18507 (28.82)		572 (16.59)	2875 (83.41)		45144 (74.28)	15632 (25.72)	
Common Good Orientation	42588 (66.31)	21635 (33.69)		1103 (32.0)	2344 (68.0)		41485 (68.26)	19291 (31.74)	
Empathy	61889 (96.37)	2334 (3.63)		3256 (94.46)	191 (5.54)		58633 (96.47)	2143 (3.53)	
Storytelling	58323 (90.81)	5900 (9.19)		2113 (61.3)	1334 (38.7)		56210 (92.49)	4566 (7.51)	
Reflexivity	63991 (99.64)	232 (0.36)		3412 (98.98)	35 (1.02)		60579 (99.68)	197 (0.32)	
Humour	53500 (83.3)	10723 (16.7)		3275 (95.01)	172 (4.99)		50225 (82.64)	10551 (17.36)	
Emotions	Absent (0)	Negative (1)	Positive (2)	Absent (0)	Negative (1)	Positive (2)	Absent (0)	Negative (1)	Positive (2)
	59641 (92.87)	2805 (4.37)	1777 (2.77)	3215 (93.27)	21 (0.61)	211 (6.12)	56426 (92.84)	2784 (4.58)	1566 (2.58)

Note: All categories are dichotomous (absent/present) except emotionality, which includes the labels "Absent: no emotion or unclassifiable emotion" (0), or negative (1), and positive emotion (2).