Towards a Typology of User Engagement in Conversational Agent Voting Advice Applications

Daan van der Weijden*
Department of Informatics
University of Zurich
Zurich, Switzerland
weijden@ifi.uzh.ch

Thilo Ignaz Dieing*
TU Darmstadt
Institute of Political Science
Darmstadt, Hessen, Germany
thilo.dieing@tu-darmstadt.de

Fynn Bachmann*
Department of Informatics
University of Zurich
Zurich, Switzerland
fynn.bachmann@uzh.ch

Abstract

Voting Advice Applications (VAAs) help citizens align with political parties, but are limited by frequent comprehension problems. Conversational Agent VAAs (CAVAAs) address this by integrating chatbot-based support. Yet, user interaction patterns and their effects on completing the CAVAA remain underexplored. This study identifies behavior-based CAVAA user types and examines their interaction with chatbot personas. Using interaction data from 189 users of an GPT-driven CAVAA during the 2024 European Parliament elections, a Latent Class Analysis reveals three types: Checkers (low interaction), Seekers (high engagement and uncertainty), and Testers (system probing rather than advice seeking). While user types do not predict completion, the chatbot personas significantly did. We find that the more active chatbot (asking follow-up questions) increased dropout rates. Our analysis introduces a novel behavioral typology and highlights the importance of conversational design for reducing dropout and improving CAVAA effectiveness.

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools**; **User studies**.

Keywords

CAVAA, voting advice applications, conversational agents, dialogue acts, latent class analysis, user typology

ACM Reference Format:
Daan van der Weijden, Thilo Ignaz Dieing, and Fynn Bachmann. 2026. Towards a Typology of User Engagement in Conversational Agent Voting Advice Applications. In *ACM Conversational User Interfaces 2026 (CUI '26), July 21–24, 2026, Bremen, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3816046.3816272

1 Introduction

Voting Advice Applications (VAAs) are online tools that match citizens with parties based on policy responses. Widely used across Europe and beyond, VAAs attract millions of users during election cycles [6, 13, 14]. In doing so, VAAs can increase political knowledge [18, 31, 43], political interest [11, 24], and in some cases even influence vote choice, particularly among users who experience the largest knowledge gains [18, 28, 36].

Despite these positive effects, roughly 20% of VAA statements cause comprehension problems. These lead to neutral or "don't know" answers [17], thereby effectively reducing knowledge gains [19] and distorting final voting advice. In order to address this limitation, Kamoen and Liebrecht [19, 20, 23] introduced Conversational Agent VAAs (CAVAAs), combining VAAs with conversational agents to support comprehension, semantic clarification, and user questions about parties and issues. While early CAVAAs were rule-based and limited in scope [20, 27], recent work extends them with Large Language Models (LLMs) [7, 46]. These systems were shown to improve political knowledge [40], outperform traditional VAAs [19], and foster users' "exploration, reflection and rationalization" of the political landscape [46, p. 12].

In this work, we focus on three gaps that remain in understanding CAVAA usage. To our knowledge, no study has identified *types* of conversational engagement in CAVAA dialogues. While Van De Pol et al. [38] and Dieing [8] proposed a VAA user typology relying on self-reported survey data, dialogue logs of CAVAAs could help researchers to tailor dialogue strategies effectively. As non-completion rates in CAVAAs are high [46], such a strategy could adapt to user types with specific early-turn dialogue behaviors, such as information-seeking questions or expressions of conviction. While there is an emerging body of literature on chatbot personas [35] and adaptive VAA systems [2, 3], the interaction of different user types with chatbot personas is less explored. These gaps lead us to our three research questions:

RQ1 Which latent classes of CAVAA users emerge based on conversational engagement and interaction features?
RQ2A To what extent do CAVAA user types exhibit different completion rates?
RQ2B To what extent does the chatbot behavior influence users' completion rates?

Utilizing data from 189 users of ChatEP2024, an LLM-driven CAVAA for the 2024 EU elections [46], the Latent Class Analysis (LCA) reveals three distinct CAVAA user types based on interactional features. The largest group, called Checkers, shows low interaction with the chatbot and uses the CAVAA like a traditional VAA; the Seekers, depicting political uncertainty, ask more questions in order to get information. Extending the typology identified by Van De Pol et al. [38] and Dieing [8], we then discover a third CAVAA-specific type called the Testers. These users are politically interested and convicted but prefer to *probe* the chatbot rather than receiving voting advice from it. Furthermore, our results show that the *chatbot behavior* significantly affects users' completion

*These authors contributed equally to this work

CUI '26, Bremen, Germany

ACM ISBN 979-8-4007-2741-2/26/07
https://doi.org/10.1145/3816046.3816272

rate. Here, the active condition significantly reduced the number of turns users interacted with the CAVAA.

Establishing such a first user typology of CAVAAs and providing a pipeline for dialogue act classification during CAVAA interaction, this study gives practitioners further design insights for the development of type specific solutions that can help prevent user drop-out.

2 Related Work

2.1 (CA)VAA user typology

As electoral information shortcuts, VAAs suffer from an inherent self-selection bias [15]. Users tend to be highly educated [30, 41], politically interested [25, 26], young [10, 11], and male [25, 26], though some differences are diminishing [1]. Van De Pol et al. [38], analyzing 52,999 Kieskompas users via LCA identified three classes, distinguishing users by political interest, efficacy, vote certainty, and motivation: *doubters* (9.9%), *seekers* (31.8%), and *checkers* (58.3%). Seekers access the tool earlier in the campaign and use it for genuine guidance, while checkers—the majority—use it mainly to confirm pre-existing views. This typology has been reaffirmed for second-order elections [39] and was extended by Dieing [8] by profiling users who consult multiple VAAs within the same election cycle.

Since their introduction by Kamoen and Liebrecht [19, 20], CAVAAs have shown potential to improve VAA comprehension and enhance users' political knowledge. Beyond LLM-based extensions [7] and safeguards [9], studies on different modalities indicate that structured chatbots with clickable buttons are preferred and encourage more information-seeking than unstructured designs, though knowledge gains are similar across formats [19, 20]. However, a gap remains: no typology exists for CAVAA users, who may differ from traditional VAA user types.

While existing VAA typologies rely on self-reported survey data, Stenmark [33] showed that interpretable user typologies can be derived solely from interaction features, producing predictively useful profiles. CAVAAs, as conversational tools that save chat histories, offer an ideal context to develop such log-based typologies.

VAAs can be seen as online surveys [37] capturing users' political stances, and CAVAAs enrich this by adding interactive features. Xiao et al. [44] found that chatbot surveys ($N \approx 600$) increased engagement and elicited responses that were more informative, relevant, specific, and clear. This suggests that the conversational format of CAVAAs may elicit richer, more authentic opinions than static VAAs, motivating our exploration of whether free-dialogue engagement patterns can inform user typology. Such a typology could then be used to better explain CAVAA usage and effects, and to design customized CAVAAs tailored to the requirements of each type.

2.2 Completion & chatbot personas

For CAVAAs and VAAs alike, users can only maximize the advantages of the tools if they complete them in order to receive voting advice. Studies show break-off rates are similar across CAVAA designs, indicating that non-completion is driven more by user-level factors than interface features [19, 20]. LLM-driven CAVAAs exhibit similar behavior: Zhu et al. [46] reported roughly 61% non-completion, highlighting that dropout is a persistent issue. High dropout rates limit CAVAAs' effectiveness, as incomplete users cannot receive voting advice or learn about the political landscape.

Recent research on chatbot interaction highlights how personas and anthropomorphism shape user engagement. A chatbot persona encompasses human-like traits—personality, tone, communication style, and social role—that influence interactions [35]. With LLMs, persona creation via prompting has become common [4]. Studies show extraverted chatbots boost engagement [21], assertive or anthropomorphized bots increase trust [5, 45], and agreeable personalities enhance satisfaction [32]. Conversely, frustrating interaction loops reduce usage and engagement [12], and users may experience fatigue when responding to question-focused bots [47]. Thus, careful design of chatbot personas and behaviors is crucial for effective user interactions.

For CAVAAs, modality studies examined whether chatbot personas affect tool use, comparing formal and anthropomorphic bots. While tool evaluations were similar, users of anthropomorphic bots showed greater factual knowledge gains [16]. Further studies found that the renownedness of embodied CAVAAs did not affect user behavior, and users generally preferred classic disembodied chatbots [42].

While early studies highlight how different chatbot personas affect CAVAA users, many chatbot types remain untested. Although prior research shows that more engaging, question-asking chatbots can have positive [44] or negative effects [47], it is unclear if these findings apply to CAVAAs or whether engagement impacts completion rates. Additionally, personality congruence between users and chatbots can boost engagement [22], suggesting that CAVAA user types might moderate possible chatbot persona effects.

3 Methods

3.1 Dataset

We analyze the available chats from ChatEP2024, a LLM-based CAVAA for the 2024 European Parliament elections [46]. The dataset contains 189 German-language conversations, with a total of 3,189 conversation turns, and an average of 17 turns per user. Each conversation is structured into a fixed five-phase protocol: (1) *greeting*, in which the chatbot introduces itself and the VAA task; (2) *free dialogue*, in which users discuss topics before committing to a structured evaluation; (3) *topic and party selection*, in which the chatbot presents all available topics and parties for the user to choose from; (4) *statement evaluation*, in which the chatbot presents up to ten numbered policy statements, and the user responds with their stance; and (5) *party-alignment recommendation*, in which the chatbot summarizes the user's positions and recommends parties that best match their expressed views.

During the statement evaluation phase, the participants were split into two conditions: in one condition, the users were interviewed by an *active* bot, which asked follow-up questions after each of the user's answers; in the other condition, the chatbot was *passive*, i.e. merely recorded the user's response and went on to the next statement. This between-subjects design was intended by Zhu et al. [46]; however, not reported in the original paper. We are not aware of any other configuration differences across the two conditions.

Table 1: Dialogue act label set used for user-turn annotation, adapted from Stolcke et al. [34] and Prahallad and Mamidi [29]

Label	Description
Statement	Factual claim without explicit opinion.
Opinion	Subjective view, evaluation, or position.
Agreement	Explicit agreement with a position of the bot.
Reject	Explicit disagreement with the bot.
Hedge	Uncertain, conditional, or qualified response.
Question	Question directed at the assistant.
Backchannel	Minimal acknowledgment (e.g., "Ok").
Challenge	Questioning a claim of the bot.
Rebuttal	Counter-argument with explicit justification.
Explanation	Seeking elaboration or reasoning.
Correction	Explicit correction of a misunderstanding.
Request	Request an action of the bot.

3.2 Text Classification

3.2.1 Dialogue Acts. For Dialogue Act Classification, we combine two dialogue-act (DA) schemes to capture the full range of moves in political conversation. The Switchboard Dialog Act corpus (SWBD-DAMSL) provides categories for cooperative dialogue-statements [34]. However, designed for casual speech, it lacks categories for confrontational or persuasive moves. BEADS [29] extends DAMSL for political discourse, adding challenge, rebuttal, and explanation-seeking acts. The combined annotation scheme from elements of SWBD-DAMSL and BEADS resulted in a set of 12 labels tailored to the political chatbot dialogue context (see Table 1).

3.2.2 Prompt Types. In addition to dialogue act coding, each user turn was automatically assigned a *prompt type*, reflecting CAVAA-relevant inquiries. The label set is adapted from Dieing [7], who identified five prompt types in rule-based CAVAAs and extended these with categories covering broader political questions and content-filter-relevant prompts (see Table 2). Based on initial analysis, we added two categories to the existing prompt types: Answer and Off_Topic. This extension addresses user interactions in a conversational interviewing setup.

3.2.3 Conviction Scoring. In a standard VAA, stance is directly encoded by Likert-scale responses to structured policy statements; in the free-dialogue phase, by contrast, the intensity of any expressed political view must be measured separately. To characterize the intensity of users' expressed political positions during the free-dialogue phase, we use an *ideological conviction* score. Each Phase 2 turn is assigned a value on a three-point scale (0 = no stance, 1 = moderate, 2 = strong). Turn-level scores were aggregated into a *mean conviction* per conversation.

3.2.4 LLM classification. In order to classify each turn for the three interaction features, we leveraged a locally hosted open-source LLM. The choice of an LLM-based annotation approach was preferred over rule-based classifiers as the tags are sensitive to context. Three human annotators independently labeled a sample of 200 turns across all three tasks and therefore established a gold standard for model evaluation. Inter-annotator agreement was substantial across all dimensions (Krippendorff's α = 0.80, 0.68, and 0.63 for dialogue act, prompt type, and conviction scoring, respectively) with a cross task mean of $\alpha = 0.71$. We benchmarked eight models across four families (Mistral, Qwen, Llama, GPT-5.4) and selected Llama 3.3-70B based on its top-tier agreement with the gold standard (Cohen's κ = 0.69, 0.58, and 0.68 for dialogue act, prompt type, and conviction scoring, respectively; mean $\kappa = 0.65$) while being fully open-weight and locally reproducible. All classifiers used zero-shot prompting at temperature 0.0, with German-language system prompts providing the full label set and disambiguation rules, and each turn was classified in context with the preceding bot message (up to 1 000 characters).

Table 2: Prompt type label set used for user-turn classification, adapted from Dieing [7].

Label	Description
Semantic_Meaning	Question about the meaning of a word or concept in a VAA statement.
Statement_Explanation	Request for elaboration or background on a VAA statement.
Current_Affairs	Question about current political events relevant to a statement.
Pros_And_Cons	Request for arguments for and against a policy position.
Party_Position	Question about a specific party's stance on a topic.
Political_Knowledge	General political question extending beyond the VAA statements.
Party_Ideology	Question about a party's broader political orientation or ideology.
Critical	Sensitive, offensive, or discriminating prompt requiring content filtering.
LLM_Opinion	Prompt attempting to elicit the chatbot's own political opinion.
Vote_Intention	Prompt asking the chatbot for a voting recommendation.
Answer	Direct response or Likert-scale answer to a VAA statement.
Off_Topic	Prompt unrelated to the VAA context or politics.

3.3 Experimental Setup

3.3.1 RQ1: Latent Class Analysis. To identify latent engagement typologies, we apply LCA to per-conversation behavioral profiles, following the methodology of van de Pol et al. [38]. For each conversation, we construct a feature vector from three sources: dialogue act rates, prompt type rates, and early conviction score. Only turns from Phase 1 and Phase 2 (the free dialogue phases) are included. The feature set comprises 25 features (12 dialogue act rates, 12 prompt type rates, and early conviction score). All features are standardized (z-scored) prior to fitting to ensure that differences in absolute scale do not distort the latent class solution. The number of classes is selected from a sweep of $K = 2$–8. A BIC/AIC sweep showed decreasing values from $K = 2$ (BIC = 6209) through $K = 7$ (BIC = −4965); however, solutions beyond $K = 3$ produced classes with fewer than 15 observations, which are too small for reliable interpretation given the total $N = 189$.

3.3.2 RQ2: Completion Prediction. We operationalize completion as the fraction of statements a user responded to. This count captures variation in engagement depth that a binary measure would conceal: two users who both drop off may have engaged with very different numbers of statements. We then test whether completion differs across the latent classes identified in RQ1. As a supplementary analysis, we test which specific features correlate with completion independent of class membership. This is achieved with an OLS regression on the completion per user. Lastly, we also compare completion across experiment conditions (active and passive bot).

4 Results

4.1 RQ1: Engagement Typology

4.1.1 Class identification. Of the initial 25 features, six were removed due to near-zero variance (raw standard deviation < 0.05): Challenge, Agreement, Critical, Reject, Correction, and Rebuttal, leaving 19 features for the LCA. We selected $K = 3$ based on interpretability, balanced class sizes, and clear separation; higher values of K yielded increasingly fragmented classes with less distinct profiles. The solution shows excellent classification certainty: average normalized entropy was 0.01, average maximum posterior probability was 0.99, and all 189 conversations were assigned to their modal class with posterior probability above 0.80. The three resulting types are *Checkers* ($n = 124$, 65.6%), *Seekers* ($n = 39$, 20.6%), and *Testers* ($n = 26$, 13.8%). Figure 1 displays the ten most discriminating features per class, and Figure 2 shows clear separation in a two-dimensional Linear Discriminant Analysis projection. Kruskal–Wallis tests with Benjamini-Hochberg correction confirmed that 7 of the 19 features differ significantly across classes ($\alpha = 0.05$); the strongest discriminators are LLM_Opinion rate ($H = 101.4$, $\eta^2 = 0.53$), Party_Ideology rate ($H = 69.0$, $\eta^2 = 0.36$), and Pros_And_Cons rate ($H = 52.2$, $\eta^2 = 0.27$).

4.1.2 Checkers. The Checkers, as the biggest group, are mainly defined by their more traditional use of the CAVAA. As such, they mainly exhibit dialogue acts such as Answer, Statement, and Opinion, while they also show below the cross-class mean characteristics of asking fewer Questions, Pros_And_Cons, and Party_Ideology items. In comparison with the other identified CAVAA user types, the Checkers therefore appear to be more passive users who mostly use the CAVAA like a normal VAA, giving the conversational agent their positions on statements and then moving on, rather than asking specific questions. With this behavior, they come closest to the largest VAA user type, also named Checker, who are already politically interested users, utilizing the VAA predominantly to check whether their chosen party actually matches with them [38]. This core utilization of a VAA as a "checking tool" appears to also be the case for these CAVAA users, as they too exhibit a higher mean for asking the CA for voting advice, i.e., checking if the CAVAA recommends the party they intend to vote for.

4.1.3 Seekers. As the second largest CAVAA user group, the Seekers exhibit more dialogue acts in which the conversational agent is asked Questions such as about general Political_Knowledge, Party_Ideology, and Pros_And_Cons. Simultaneously, the users in the Seeker class ask fewer Statement_Explanation, Semantic_Meaning, have fewer Request and Backchannel acts, and display an overall lower political conviction. These distinctive class features draw a picture of more politically uncertain users who are in need of voting advice as the main focus and ask more questions about parties and general political topics to receive such advice, instead of asking more detailed questions about statements and their meaning or about the chatbot.

Therefore, this group is more or less the reversal of the Checker and is not politically informed and certain of whom to vote for, but actually utilizes the CAVAA as an information shortcut for vote choice. Like the Checkers, the Seekers also appear in normal VAA user typologies [8, 38]. Here, the Seekers distinguish themselves from the other types via low vote certainty and political interest. Hence, Seekers engage more with the chatbot compared to the more traditional Checkers and might benefit more from the interactive conversational offer made by CAVAAs.

4.1.4 Testers. The Testers is the smallest group of the three class CAVAA typology. They display above average usage of LLM_Opinion, Off_Topic, Backchannel, and Request acts and ask more about semantic meanings. Furthermore, their inquiries display fewer uncertainties (Hedge), and they ask fewer questions about parties or whom to vote for. These dialogue-acts present the Tester as a unique group that neither uses the CAVAA to check their party choice nor find out whom to vote for, but rather to test the conversational agent as an addition to the VAA design. As such, they mainly focus on testing the chatbot's limits and capabilities instead of asking questions to gain more political knowledge. When compared to the existing VAA typology, the Tester appears to be a distinct group in the context of CAVAAs and therefore can not directly be found in the types of normal (non-CA) VAAs. Hence, the Tester is a unique group that arises because of the chatbot, which can help developers find the limits of new CAVAAs, but might also bias effect studies, as they do not use the CAVAA as the intended information tool per se.

4.2 RQ2: Completion prediction

4.2.1 RQ2A: User types. With a full completion rate of only 51.9% only roughly half of the 189 CAVAA users answered all ten statements and subsequently received their voting advice. The result of an OLS regression with the derived three-class CAVAA typology as an independent variable shows that neither type is statistically more significant than the other in completing the CAVAA. Testers, nevertheless, exhibit the lowest full completion rate (34.6%), likely because of their lower interest in a voting advice.

While the user types do not seem to influence the number of answered statements, several individual features do significantly correlate with completion. An OLS regression revealed that, for example, Question ($\beta = -0.47$, $p < 0.001$), Off_Topic ($\beta = -2.86$, $p < 0.001$), and Political_Knowledge ($\beta = -0.93$, $p < 0.001$) are among the strongest negative predictors. Conversely, Answer ($\beta = 0.64$, $p < 0.05$) and mean word count per user message ($\beta = 0.07$, $p < 0.01$) are positively associated with completion, indicating that users who write longer responses progress further through the VAA.

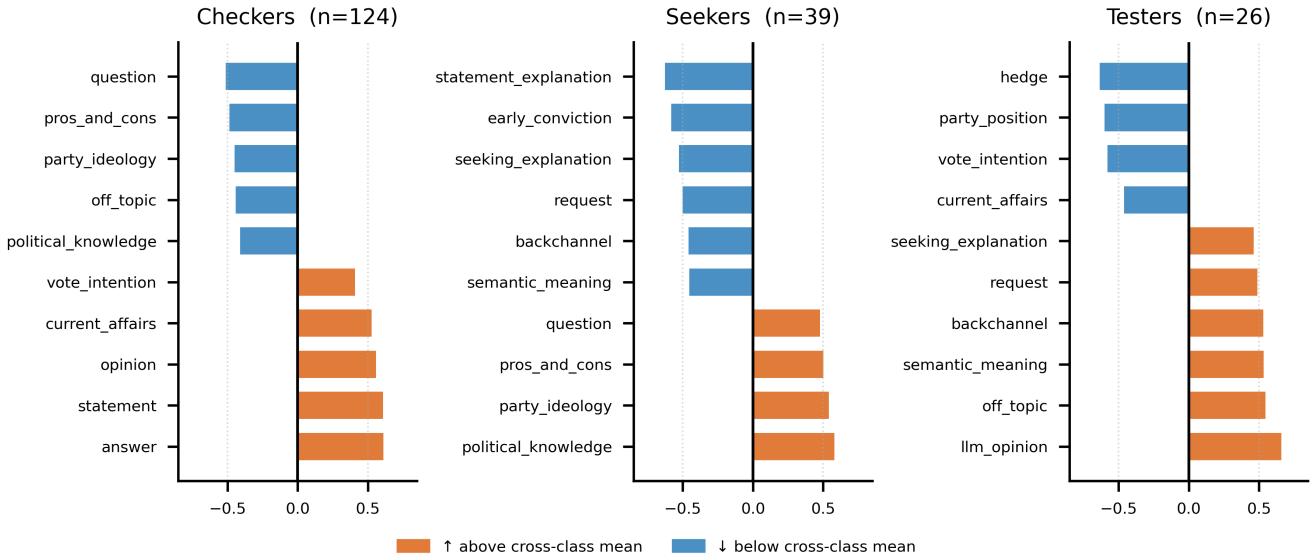

Figure 1: Top-10 most discriminating features per user type shown as a diverging bar chart (standardized mean difference from the grand mean).

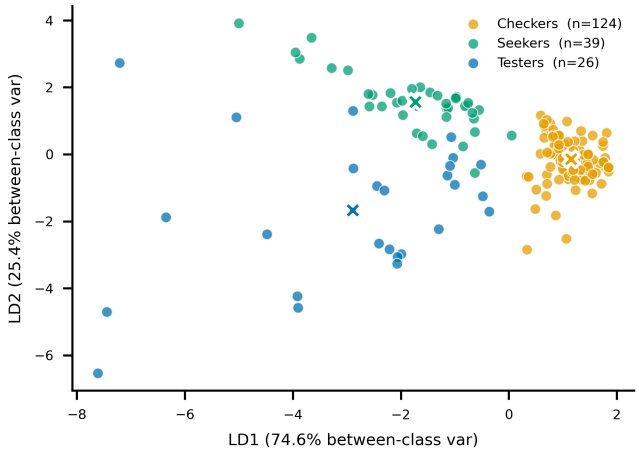

Figure 2: Linear Discriminant Analysis (LDA) projection of conversational feature vectors, showing separation of the three latent classes.

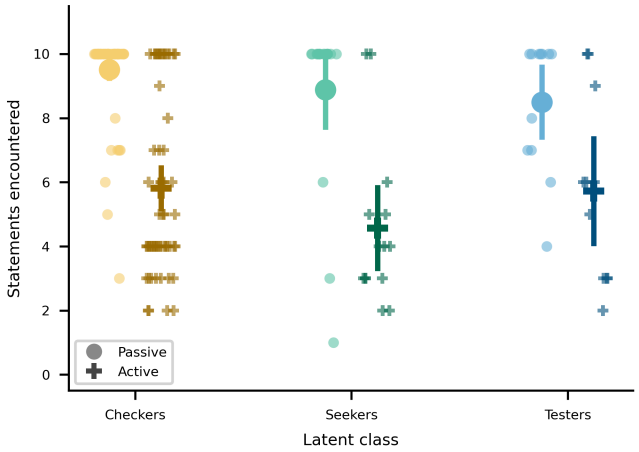

Figure 3: Number of CAVAA statements encountered per participant, split by user type and chatbot behavior.

4.2.2 RQ2B: Chatbot Engagement. The chatbot persona had a strong significant effect on the completion rate. The *active* bot, asking follow-up questions, resulted in significantly lower completion rates than the *passive* chatbot ($M = 5.60$, $SD = 2.79$ vs. $M = 9.25$, $SD = 1.82$; Welch's $t = -10.16$, $p < 0.001$). This finding is also in line with previous modality studies that show that users prefer clickable, less interactive CAVAAs [20]. As such, it seems that CAVAA users are more inclined to use them as information resources, not as an interactive discussion partner.

In a last step, we analyze whether the findings of RQ2A persist if controlling for the strong effect of the chatbot persona. As shown by the results in Figure 3, we find that Seekers exhibit a stronger reaction to the chatbot behavior, as such, demonstrate a lower likelihood of completion when the chatbot is *active*. Moreover, we confirm that several individual features retain their predictive power even after controlling for the chatbot persona. An OLS regression with condition as a covariate reveals that Question ($\beta = -0.20$, $p = 0.023$) and Political_Knowledge ($\beta = -0.39$, $p = 0.042$) remain significant predictors of completion. Notably, the mean word count of the chatbot ($\beta = -0.007$, $p = 0.024$) emerges as a new significant predictor that was not among the strongest effects in the uncontrolled model. This aligns with the negative effect of the *active* chatbot persona: longer bot responses appear to overwhelm

users and reduce completion. This finding underscores that verbose chatbot behavior, whether by design or by prompt type, detracts from the user's progression through the CAVAA.

5 Conclusion

Driven by a missing typology of CAVAA users, this study investigates whether distinct users could be established based on their interactions with a chatbot. A LCA reveals three distinct CAVAA user types: While the Checkers and the Seekers appear similar to the two largest VAA user groups, the Testers emerge as a unique CAVAA user type. Additional analyses indicate that user types do not help predict CAVAA completion, while chatbot personas (active vs. passive) significantly do—with users of a more active chatbot dropping out 3-4 turns (out of 10) earlier on average. As such, this study contributes to the ongoing research on CAVAAs and general chatbot systems by presenting an original typology, possible effects, and design recommendations to mitigate user dropout.

Meanwhile, we acknowledge several limitations of this study. First, the quality of the utilized data is limited, as the dataset is very small and only stems from an experimental German CAVAA. The 189 conversations represent all interactions recorded during the ChatEP2024 deployment, with the sample size constrained by the scope of that single study. This limits the stability of the LCA solution, particularly for the smallest class (Testers, $n = 26$). Furthermore, while the LLM-based categorization was benchmarked against expert coders, using LLMs could still have biased the results. Lastly, the derived typology based on interaction features can only approximate real demographic markers of the types. In particular, the Tester type may be specific to research settings, where curiosity about the LLM inflates testing behaviour that may not occur in real-world deployments. We recommend that future research reinvestigates the established three-class typology on larger and more diverse datasets, accounting for socio-cultural differences as well as adding additional survey items [8, 38] to the interactional features for a clearer interpretation of the types. Such validation would also allow researchers to assess whether the typology can serve as a practical tool for tailoring CAVAA design to specific user needs, for instance by adapting chatbot personas or information depth to the predominant user type. Despite these limitations, this study establishes a novel typology for electoral information chatbot systems, as well as giving developers further insights into chatbot persona effects. These results can help better understand CAVAAs effects on users, design tailor-made solutions and personas for type-specific needs, thereby increasing the usability of CAVAAs and their completion.

References

[1] Andreas Albertsen. 2022. How do the characteristics of voting advice application users change over time? Evidence from the German election studies. *German Politics* 31, 3 (2022), 399–419.

[2] Fynn Bachmann, Cristina Sarasua, and Abraham Bernstein. 2024. Fast and adaptive questionnaires for voting advice applications. In *Joint European Conference on machine learning and knowledge discovery in databases*. Springer, 365–380.

[3] Fynn Bachmann, Daan van der Weijden, Cristina Sarasua, and Abraham Bernstein. 2026. Estimating the Recommendation Certainty in Candidate-Based Voting Advice Applications. *Politics and Governance* 14 (2026).

[4] Ting-Chi Chang, Yu-Jou Chen, Sheng Hung, Ning-Hsuan Chang, Chih-Hao Ku, Szu-Yin Lin, and Shih-Yi Chien. 2025. A Review on Shaping Chatbot Personalities via Large Language Models. In *58th Hawaii International Conference on System Sciences, HICSS 2025*.

[5] Xusen Cheng, Xiaoping Zhang, Jason Cohen, and Jian Mou. 2022. Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. *Information Processing & Management* 59, 3 (2022), 102940.

[6] Frederico Ferreira Da Silva and Diego Garzia. 2023. Voting advice applications. In *The Routledge handbook of political parties*, Neil Carter, Daniel Keith, Gyda M. Sindre, and Sofia Vasilopoulou (Eds.). Routledge, New York, 450–465. doi:10.4324/9780429263859-47

[7] Thilo I Dieing. 2024. AI-Driven Dialogue: Leveraging Generative AI in Conversational Agent Voting Advice Applications (CAVAAs). In *International Symposium on Chatbots and Human-Centered AI*. Springer, 161–180.

[8] Thilo I. Dieing. 2025. Just can't get enough – profiling users of multiple Voting Advice Applications. *Journal of Information Technology & Politics* 0, 0 (2025), 1–14. doi:10.1080/19331681.2025.2496259

[9] Thilo Ignaz Dieing. 2025. Traditional and Deep Learning Methods for Trustworthy Intent Classification in Conversational Agent Voting Advice Applications (CAVAAs). In *Proceedings of the Mensch Und Computer 2025 (MuC '25)*. Association for Computing Machinery, New York, NY, USA, 749–754. doi:10.1145/3743049.3748572

[10] Patrick Dumont and Raphaël Kies. 2012. Smartvote. lu: usage and impact of the first VAA in Luxembourg. *International Journal of Electronic Governance* 5, 3-4 (2012), 388–410.

[11] Jan Fivaz and Giorgio Nadig. 2010. Impact of voting advice applications (VAAs) on voter turnout and their potential use for civic education. *Policy & Internet* 2, 4 (2010), 167–200.

[12] Asbjørn Følstad and Petter Bae Brandtzaeg. 2020. Users' experiences with chatbots: findings from a questionnaire study. *Quality and User Experience* 5, 1 (2020), 3.

[13] Diego Garzia and Stefan Marschall. 2019. Voting Advice Applications. In *Oxford research encyclopedia of politics*, William R. Thompson (Ed.). Oxford University Press, New York, NY. doi:10.1093/acrefore/9780190228637.013.620

[14] Kostas Gemenis. 2018. The Impact of Voting Advice Applications on Electoral Turnout: Evidence from Greece. *Statistics, Politics and Policy* 9, 2 (2018), 161–179. doi:10.1515/spp-2018-0011

[15] Micha Germann and Kostas Gemenis. 2018. Getting Out the Vote With Voting Advice Applications. *Political Communication* 36, 1 (2018), 149–170. doi:10.1080/10584609.2018.1526237

[16] Stef Hankel, Christine Liebrecht, and Naomi Kamoen. 2024. 'Hi Chatbot, let's Talk about Politics!'Examining the Impact of Verbal Anthropomorphism in Conversational Agent Voting Advice Applications (CAVAAs) on Higher and Lower Politically Sophisticated Users. *Interacting with Computers* (2024), iwae031.

[17] Naomi Kamoen and Bregje Holleman. 2017. I Don't Get It. Response Difficulties in Answering Political Attitude Statements in Voting Advice Applications. *Survey Research Methods* 11, 2 (2017). doi:10.18148/srm/2017.v11i2.6728

[18] Naomi Kamoen, Bregje Holleman, André Krouwel, Jasper van de Pol, and Claes de Vreese. 2015. The Effect of Voting Advice Applications on Political Knowledge and Vote Choice. *Irish Political Studies* 30, 4 (2015), 595–618. doi:10.1080/07907184.2015.1099096

[19] Naomi Kamoen and Christine Liebrecht. 2022. I Need a CAVAA: How Conversational Agent Voting Advice Applications (CAVAAs) Affect Users' Political Knowledge and Tool Experience. *Frontiers in Artificial Intelligence* 5 (2022). doi:10.3389/frai.2022.835505

[20] Naomi Kamoen, Tessa McCartan, and Christine Liebrecht. 2022. Conversational Agent Voting Advice Applications: A Comparison Between a Structured, Semi-structured, and Non-structured Chatbot Design for Communicating with Voters About Political Issues. doi:10.1007/978-3-030-94890-0_10

[21] Leon OH Kroczek, Alexander May, Selina Hettenkofer, Andreas Ruider, Bernd Ludwig, and Andreas Mühlberger. 2025. The influence of persona and conversational task on social interactions with a LLM-controlled embodied conversational agent. *Computers in Human Behavior* (2025), 108759.

[22] Mohammad Amin Kuhail, Mohamed Bahja, Ons Al-Shamaileh, Justin Thomas, Amina Alkazemi, and Joao Negreiros. 2024. Assessing the impact of chatbot-human personality congruence on user behavior: A chatbot-based advising system case. *IEEE Access* 12 (2024), 71761–71782.

[23] Christine Liebrecht, Naomi Kamoen, and Celine Aerts. 2023. Voice Your Opinion! Young Voters' Usage and Perceptions of a Text-Based, Voice-Based and Text-Voice Combined Conversational Agent Voting Advice Application (CAVAA). In *Chatbot Research and Design*, Asbjørn Følstad, Theo Araujo, Symeon Papadopoulos, Effie L.-C. Law, Ewa Luger, Morten Goodwin, and Petter Bae Brandtzaeg (Eds.). Springer International Publishing, Cham, 34–49.

[24] Vasilis Manavopoulos, Vasiliki Triga, Stefan Marschall, and Lucas Constantin Wurthmann. 2018. The impact of VAAs on (non-voting) aspects of political participation: Insights from panel data collected during the 2017 German federal elections campaign. *Statistics, Politics and Policy* 9, 2 (2018), 105–134.

[25] Marschall. 2014. "Profiling users,". In *Matching Voters with Parties and Candidates: Voting Advice Applications in Comparative Perspective*, Garzia Diego and Marschall Stefan (Eds.). ECPR Press, 93–104.

[26] Stefan Marschall and Martin Schultze. 2012. The emergence of the "Voter 2.0". In *Proceedings of the XXVI Convegno SISP: VAA Users in a Changing Political Communication Sphere*. Rome, Italy, 13–15.

[27] Tessa McCartan and CC Liebrecht. 2021. Conversational Agent Voting Advice Applications (CAVAAs). *Master's thesis* (2021).

[28] Joëlle Pianzola, Alexander H Trechsel, Kristjan Vassil, Guido Schwerdt, and R Michael Alvarez. 2019. The impact of personalized information on vote intention: Evidence from a randomized field experiment. *The Journal of Politics* 81, 3 (2019), 833–847.

[29] Lavanya Prahallad and Radhika Mamidi. 2025. Analyzing Biases in Political Dialogue: Tagging US Presidential Debates with an Extended DAMSL Framework. *arXiv preprint arXiv:2505.19515* (2025).

[30] Ain Ramonaite. 2010. Voting advice applications in Lithuania: promoting programmatic competition or breeding populism? *Policy & Internet* 2, 1 (2010), 117–147.

[31] Martin Schultze. 2013. Effekte des Wahl-O-Mat auf politisches Wissen über Parteipositionen. *ZPol Zeitschrift für Politikwissenschaft* 22, 3 (2013), 367–391.

[32] Tuva Lunde Smestad and Frode Volden. 2018. Chatbot personalities matters: improving the user experience of chatbot interfaces. In *International conference on internet science*. Springer, 170–181.

[33] Dick Stenmark. 2008. Identifying Clusters of User Behavior in Intranet Search Engine Log Files. *Journal of the American Society for Information Science and Technology* 59, 14 (2008), 2232–2243. doi:10.1002/asi.20931

[34] Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics* 26, 3 (2000), 339–373. doi:10.1162/089120100561737

[35] Richard Sutcliffe. 2023. A survey of personality, persona, and profile in conversational agents and chatbots. *arXiv preprint arXiv:2401.00609* (2023).

[36] Mathias Wessel Tromborg and Andreas Albertsen. 2023. Candidates, voters, and voting advice applications. *European Political Science Review* 15, 4 (2023), 582–599.

[37] Kirsten van Camp, Jonas Lefevere, and Stefaan Walgrave. 2014. The content and formulation of statements in voting advice applications: a comparative analysis of 26 VAAs. In *Matching Voters with Parties and Candidates: Voting Advice Applications in Comparative Perspective*, Garzia Diego and Marschall Stefan (Eds.). ECPR Press, 11–31.

[38] Jasper van de Pol, Bregje Holleman, Naomi Kamoen, André Krouwel, and Claes de Vreese. 2014. Beyond young, highly educated males: a typology of VAA users. *Journal of Information Technology & Politics* 11, 4 (2014), 397–411.

[39] Jasper van de Pol, Naomi Kamoen, André Krouwel, Claes de Vreese, and Bregje Holleman. 2019. Same but different: A typology of Voting Advice Application users in first- and second-order elections. *Acta Politica* 54, 2 (2019), 225–244. doi:10.1057/s41269-018-0083-3

[40] Nina van Zanten and Roel Boumans. 2024. Voting Assistant Chatbot for Increasing Voter Turnout at Local Elections: An Exploratory Study. In *Chatbot Research and Design*, Asbjørn Følstad, Theo Araujo, Symeon Papadopoulos, Effie L.-C. Law, Ewa Luger, Morten Goodwin, Sebastian Hobert, and Petter Bae Brandtzaeg (Eds.). Springer Nature Switzerland, Cham, 3–22.

[41] Kristjan Vassil. 2011. Voting smarter? The impact of voting advice applications on political behavior. *European University Institute* (2011).

[42] Elke van Veggel, Christine Liebrecht, and Naomi Kamoen. 2025. How Life-like Digital Humans in Voting Advice Applications Can Stimulate Young Voters to Inform Themselves About Politics: The Role of Familiarity and Expertise. *International Journal of Human–Computer Interaction* (2025), 1–15.

[43] Bettina Westle, Christian Begemann, and Astrid Rütter. 2014. The "Wahl-O-Mat" in the course of the German Federal Election 2013 – Effects of a German VAA on users' election-relevant political knowledge. *Zeitschrift für Politikwissenschaft* 24, 4 (2014), 389–426. doi:10.5771/1430-6387-2014-4-389

[44] Ziang Xiao, Michelle X Zhou, Q Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020. Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 3 (2020), 1–37.

[45] Michelle X Zhou, Gloria Mark, Jingyi Li, and Huahai Yang. 2019. Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 9, 2-3 (2019), 1–36.

[46] Jianlong Zhu, Manon Kempermann, Vikram Kamath Cannanure, Alexander Hartland, Rosa M. Navarrete, Giuseppe Carteny, Daniela Braun, and Ingmar Weber. 2025. Learn, Explore and Reflect by Chatting: Understanding the Value of an LLM-Based Voting Advice Application Chatbot. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 58, 19 pages. doi:10.1145/3719160.3736611

[47] Jie Zou, Evangelos Kanoulas, and Yiqun Liu. 2020. An empirical study on clarifying question-based systems. In *Proceedings of the 29th ACM international conference on information & knowledge management*. 2361–2364.