Recommender Systems for Democracy: Toward Adversarial Robustness in Voting Advice Applications

Frédéric Berdoz[1] , **Dustin Brunner** , **Yann Vonlanthen**[1] and **Roger Wattenhofer**
ETH Zurich, Switzerland
fberdoz@ethz.ch, yvonlanthen@ethz.ch, wattenhofer@ethz.ch

Abstract

Voting advice applications (VAAs) help millions of voters understand which political parties or candidates best align with their views. This paper explores the potential risks these applications pose to the democratic process when targeted by adversarial entities. In particular, we expose 11 manipulation strategies and measure their impact using data from Switzerland's primary VAA, Smartvote, collected during the last two national elections. We find that altering application parameters, such as the matching method, can shift a party's recommendation frequency by up to 105%. Cherry-picking questionnaire items can increase party recommendation frequency by over 261%, while subtle changes to parties' or candidates' responses can lead to a 248% increase. To address these vulnerabilities, we propose adversarial robustness properties VAAs should satisfy, introduce empirical metrics for assessing the resilience of various matching methods, and suggest possible avenues for research toward mitigating the effect of manipulation. Our framework is key to ensuring secure and reliable AI-based VAAs poised to emerge in the near future.

1 Introduction

Recent advances in information technology have significantly transformed our daily lives. One area that remains relatively underexplored is digital democracy, which integrates digital innovations into the political system. Among the most notable developments in this field is the emergence of Voting Advice Applications (VAAs). VAAs provide voters with personalized recommendations on which parties or candidates best align with their preferences and policy stances. VAAs exist in as many as 30 countries across the world, including the USA, Canada, Australia, as well as many European countries [Terán, 2020]. Interestingly, the legal basis of VAAs varies widely from country to country, ranging from publicly governed and regulated entities to loosely controlled private associations [Garzia and Marschall, 2012]. Strikingly, almost every country chooses a different method to match voters to candidates [Louwerse and Rosema, 2014]. In countries where VAAs are currently in use, they are often consulted by 10-50% of voters [Terán, 2020], making them a highly popular source of information. On top of that, the advice provided by these applications has been shown to significantly influence both voter turnout and voter decisions [Munzert and Ramirez-Ruiz, 2021]. In Switzerland specifically, Germann and Gemenis [2019] showed that the VAA mobilized 58,000 additional voters in 2007, while Ladner and Pianzola [2010] reported that 67% of the users had stated that the VAA had influenced their voting behavior. The profound impact of VAAs has gone as far as triggering a shift from representative to promissory democracy, in which VAA profiles are interpreted as electoral promises [Ladner, 2016]. This transition occurred without requiring any changes to constitutional or legal frameworks. While the benefits of VAAs are undeniable and well-documented [Munzert and Ramirez-Ruiz, 2021], for the first time, this study aims to shed light on their potential vulnerabilities. Specifically, we seek to quantify the impact that a hypothetical adversarial actor could have on the recommendations. Toward this goal, we focus our analysis on Smartvote, Switzerland's primary VAA. In 2023, Smartvote was used by up to 20% of eligible Swiss voters, up from 17% in 2011. In 2023, a total of 2.1 million voting advice reports were created [Politools, 2024a]. Our contributions:

1. We propose three adversarial robustness properties for VAAs. Namely, robustness against manipulation by (i) candidates and parties, (ii) platform operators, and (iii) question designers (Section 2).
2. We empirically demonstrate the importance of these robustness properties by leveraging two comprehensive datasets collected by Smartvote during the Swiss national elections of 2019 and 2023. We uncover a total of 11 vulnerabilities through which adversaries could manipulate the recommendations (Table 1 and App. C).
3. Based on the highest-risk vulnerabilities, we suggest 9 metrics to compare the adversarial robustness of existing and newly proposed matching methods (Section 5).
4. Finally, with input from Politools, the non-profit organization behind Smartvote, we propose research directions to mitigate these vulnerabilities, enabling the development of more robust VAAs in the future (Section 6).

[1]Corresponding authors.

Vulnerability	Adversary Type	Code	Section	Data	Benefactors	Visibility Gain	Likelihood	Impact
Answer Optimization	Candidates	AO	4.1	✓	👤	259%	*Low*	*High*
Answer Calibration	Candidates	AC	4.1	✗	👤 ☰ 🏴 ⚖	248%	*High*	*High*
Diversification	Candidates (Party)	DIV	4.1	✗	🏴	345%	*Medium*	*High*
List Centralization	Candidates (Party)	LC	App. C	✗	☰	-	*Low*	*Low*
Matching Method	Platform operator	MM	4.2	✗	⚖	105%	*Medium*	*Medium*
Question Ordering	Platform operator	QO	App. C	✓	🏴	6%	*Low*	*Low*
Weight Selection	Platform operator	WS	4.2	✗	⚖	≈15%	*Low*	*Low*
Similarity Score	Platform operator	SS	4.2	✗	🏴 ⚖	-	*Medium*	*Low*
Tie-breaking	Platform operator	TB	App. C	✗	👤	210%	*Medium*	*Medium*
Question Favoritism	Question designer	QF	4.3	✓	🏴	261%	*Low*	*High*
Question Correlation	Question designer	QC	4.3	✗	🏴	-	*Medium*	*Medium*

Table 1: Overview of the main vulnerabilities associated with each type of adversary, with type-specific color codes for reference in the paper. The *Data* column indicates whether a strategy exploiting that vulnerability requires knowledge of the voters' or candidates' answers. For *Benefactors*, 👤 denotes single candidates, ☰ denotes lists, 🏴 denotes parties, and ⚖ denotes party coalitions (i.e., left, center, right as shown in Figure 5). The primary benefactor is highlighted in **black**, secondary benefactors are shown in gray. The *Visibility Gain* factor indicates its best-case potential relative increase in visibility in the VAA if that vulnerability is exploited, as estimated by our experiments throughout the paper (left blank if no experiment was conducted). The table also includes a subjective assessment of the *Likelihood* and *Impact* of each strategy.

2 Background

Most popular VAAs use a set of questions $Q = \{q_t\}_{t=1}^{N_q}$ to position both candidates $C = \{c_j\}_{j=1}^{N_c}$ and voters $V = \{v_i\}_{i=1}^{N_v}$ within the high-dimensional Euclidean space $\mathbb{R}^{N_q}$. Formally, each question $q_t : V \cup C \to A_t$ assigns an answer to a given voter or candidate, with $A_t \subseteq \mathbb{R}$ being the set of allowable answers for that question (generally discrete and bounded). For example, the question "Are VAAs robust?" might map the answers "No", "Rather no", "Rather yes", and "Yes" to the numerical values 0, 25, 75, and 100, respectively. Additionally, for each question q_t, voters can typically choose a numerical weight within a set of allowable values $W_t \subseteq \mathbb{R}$ to reflect how important each question is to them. This weight is formally represented as a mapping $w_t : V \to W_t$. Given a voter-candidate pair (v_i, c_j) and their respective answer and weight vectors $\mathbf{v}_i = [q_1(v_i), ..., q_{N_q}(v_i)]^T$, $\mathbf{c}_j = [q_1(c_j), ..., q_{N_q}(c_j)]^T$ and $\mathbf{w}_i = [w_1(v_i), ..., w_{N_q}(v_i)]^T$, the VAA computes a similarity score $s(v_i, c_j)$ between v_i's and c_j's opinions using a predefined weighted distance function $d(\mathbf{v}_i, \mathbf{w}_i, \mathbf{c}_j)$, with $d : \mathbb{R}^{N_q} \times \mathbb{R}^{N_q} \times \mathbb{R}^{N_q} \to \mathbb{R}_+$. Lastly, for each voter v_i, the VAA provides a ranking $\mathbf{r}_i \in \mathcal{R}(C)$ based on these similarity scores, with $\mathcal{R}(C)$ the set of total orders on C. See Table 3 in the Appendix for a summary of how the most popular VAAs align with this framework. As some of our analysis will concern parties and lists, we also account for the fact that candidates can belong to exactly one party $p \in P$ and one list $l \in L$, with P and L being the set of all parties and lists, respectively. In Swiss National Council elections, lists are party- or coalition-specific slates of candidates from which voters choose or modify their preferred selections (see Appendix A.1 for more details). A canonical set of properties that any safe VAA must satisfy commonly includes [Garzia and Marschall, 2014]:

(R) *Reproducibility*: The VAA produces reproducible recommendations, enabling users to verify the system's reliability.

(I) *Interpretability*: The rationale behind the VAA's recommendations is easily understandable and intuitive to users, including those with less technical expertise.

(T) *Transparency*: The VAA's matching algorithm and all factors influencing recommendations are open-source.

(F) *Fairness*: The VAA is purely issue-based and does not consider any other characteristics of voters or candidates.

(E) *Explainability*: Voters receive clear and intuitive explanations for candidate or list recommendations.

(P) *Privacy*: The VAA ensures the privacy and anonymity of users' responses and preferences.

Although the importance of these properties is clear, they do not offer protection against malicious actors (i.e., adversaries) aiming to manipulate the recommendations to favor a particular candidate or party. From the above definitions, one can identify three potential types of such adversaries: (i) The **candidates** providing their answer vectors, (ii) the **platform operator** in charge of choosing d, $\{A_t\}_{t=1}^{N_q}$, $\{W_t\}_{t=1}^{N_q}$ and all other aspects related to VAA's interface (such as question ordering, tie-breaking, etc.), and (iii) the **question designers** writing the questions Q.[2] In Section 4, we analyze the primary dangers associated with each type of adversary, grounding our analysis in the two datasets from Smartvote presented in Section 3. Then, in Section 6, we propose solutions to mitigate these risks.

[2]For Smartvote, the non-profit association Politools is responsible for selecting the questions and operating the platform [Politools, 2024a].

3 Dataset

We empirically evaluate our claims using two comprehensive datasets collected by Smartvote [Politools, 2024a], which include questionnaire responses and metadata from both voters and candidates in the 2019 and 2023 Swiss National Council elections. In both elections, approximately 85% of electable candidates participated by completing the questionnaire, and around 20% of eligible Swiss voters used Smartvote for voting recommendations. These recent datasets provide a solid foundation for analyzing VAA robustness, capturing a significant portion of both voters and candidates. Smartvote contains $N_q = 75$ questions with $A_t = \{0, 25, 75, 100\}$ for questions $1 \leq t \leq 60$ (policy questions), $A_t = \{0, 17, 33, 50, 67, 83, 100\}$ for $61 \leq t \leq 67$ (value questions) and $A_t = \{0, 25, 50, 75, 100\}$ for $68 \leq t \leq 75$ (budget questions). For all questions, the allowable values for the weights are $W_t = \{0, 0.5, 1, 2\}$, with 1 being the default value for answered questions and 0 the value automatically assigned to any unanswered question. The distance metric used in Smartvote is the L2 distance

$$d_{\text{L2}}(\mathbf{v}_i, \mathbf{w}_i, \mathbf{c}_j) = \sqrt{\sum_{t=1}^{N_q} (\mathbf{w}_{i,t}(\mathbf{v}_{i,t} - \mathbf{c}_{j,t}))^2}, \quad (1)$$

which is used to compute the normalized similarity scores

$$s(v_i, c_j) = 100 \cdot \left(1 - \frac{d_{\text{L2}}(\mathbf{v}_i, \mathbf{w}_i, \mathbf{c}_j)}{d_{\text{L2}}(100 \cdot \mathbf{1}_{N_q}, \mathbf{w}_i, \mathbf{0}_{N_q})}\right), \quad (2)$$

where $\mathbf{1}_{N_q}$ (respectively $\mathbf{0}_{N_q}$) denote the one-valued (respectively zero-valued) N_q dimensional vector. In addition to candidate rankings, Smartvote also provides a list ranking by averaging the similarity scores of all candidates on each list $l \in L$, i.e., $s(v_i, l) = \frac{1}{|l|} \sum_{c \in l} s(v_i, c)$. For a more detailed description of the Swiss political system and Smartvote, we refer the reader to Appendix A. In Appendix B, we provide a comprehensive description of the preprocessing applied to the two datasets, as well as an exploratory data analysis. We conducted all analyses and experiments on both datasets, but present results from the more recent 2023 dataset, as the overall findings are consistent across both elections.

4 Vulnerabilities

While Smartvote satisfies in large part[3] all the safety properties listed in Section 2, its robustness to adversarial entities remains unclear. In this section, we analyze the key strategies that the different types of adversaries might use to increase the visibility of a particular candidate or party. Given a set of candidates C and a set of recommendations (i.e., rankings) $R_C = \{\mathbf{r}_i \in \mathcal{R}(C) \mid v_i \in V\}$, we define the k-**visibility of a candidate** $\nu_k(c \mid C)$ as the frequency with which candidate c appears in the top k positions of the rankings R_C. Additionally, we define the k-**visibility of a party** $\nu_k(p \mid P)$ as the fraction of the top k recommendations that are occupied by members of that party. Finally, we define the k-**visibility of a list** $\nu_k(l \mid L)$ as the frequency with which l appears in the top k positions of the list rankings $R_L = \{\mathbf{r}_i \in \mathcal{R}(L) \mid v_i \in V\}$. Throughout this work, unless specified otherwise, we set k to the number of seats allocated to the candidate's state[4] in the National Council, for both candidate and party visibility. For lists, we use $k = 1$ by default, as voters can only vote for one list. These default values also correspond to the number of candidates and lists visually put forward by Smartvote. Due to their specificity, we discuss the list centralization (LC), the question ordering (QO), and the tie-breaking (TB) vulnerabilities in Appendix C.

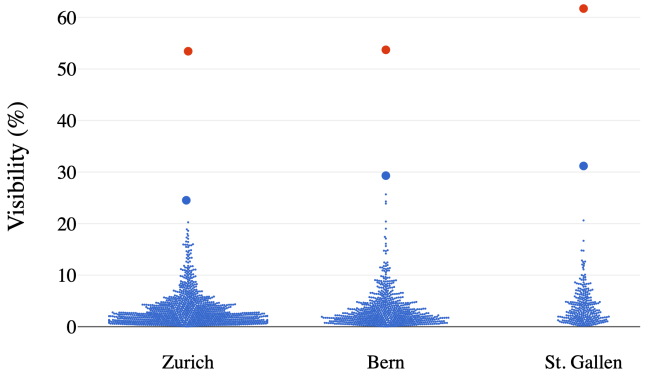


Figure 1: Visibility of crafted candidates (red) compared to all other candidates (blue) in the states of Zurich ($k = 36$), Bern ($k = 24$), and St. Gallen ($k = 12$). The larger dots highlight the crafted and actual most visible candidates.

4.1 Candidates and Parties

Answer Optimization (AO)

We start by investigating the potential for a single candidate to manipulate their answers to increase their popularity. The computation of the provably optimal candidate is of combinatorial complexity and thus infeasible, as pointed out by Etter *et al.* [2014]. However, we can find an approximate solution through randomized optimization. For each state, we craft an artificial candidate c^* using *simulated annealing* [Kirkpatrick *et al.*, 1983] and optimizing $\nu_k(c^* \mid C \cup \{c^*\})$. In almost all states, the crafted candidate appears in more than 50% of top k recommendations, significantly outperforming the previously crafted candidate by Etter *et al.* [2014], as well as any actual candidate.[5] Figure 1 shows that the crafted candidates in the states of Zurich, Bern, and St. Gallen easily surpass their competition in terms of visibility. Table 7 in Appendix C contains the popularity of our best crafted candidate for each state, as well as a comparison with other optimization strategies. Specifically, it demonstrates that the visibility of candidates crafted using only 1% of the voters' data is nearly as high as those optimized with the full dataset, achieving 51.70%, 50.66%, and 52.55% in Zurich, Bern, and

[3]The *fairness* and *reproducibility* properties of Smartvote are not fully met, as they break ties using last names and allowed some candidates to overwrite their initial answers on a few questions.

[4]Usually referred to as a *canton* in Switzerland

[5]Note that Etter *et al.* [2014] set $k = 50$ in the popularity metric, while for us $k \in \{1, \cdots, 36\}$. Our result is thus strictly stronger.

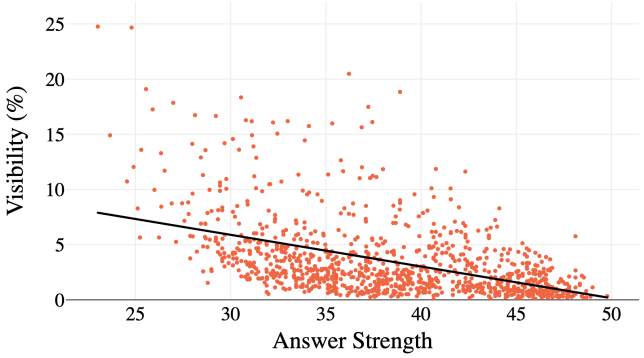


Figure 2: Relationship between the answer strength of candidates, as defined in Eq. (3), and their visibility in the state of Zurich ($k = 36$). Each dot shows a candidate and the black line represents an ordinary least squares trend line.

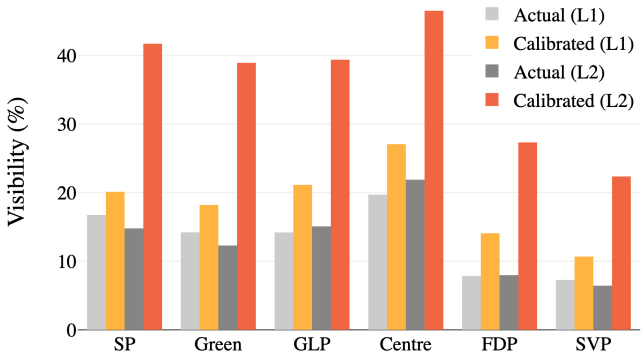


Figure 3: Comparison of actual and calibrated party visibility using the L1 and L2 distance metrics. To simulate this scenario, the answer profiles of all candidates in the party were adjusted to weaken their responses (e.g., changing all "Yes" to "Rather yes"), and the recommendations were recalculated using the L1 and L2 distance metrics.

St. Gallen, respectively. The analysis of the crafted candidate's profile reveals that almost no questions are answered on the answer spectrum's extremities (e.g., only 2 out of 75 answers for the crafted candidate in Zurich). This points to a systematic bias toward candidates with moderate positions. We investigate this lead next.

Answer Calibration (AC)

In Smartvote, candidates are provided with four or more response options. They can deliberately choose to respond "strongly" by selecting answers at the poles (0 or 100) or "moderately" by choosing options closer to the middle of the answer spectrum (25 or 75).[6] We define the **strength** σ of an answer $\mathbf{c}_j$ by its deviation from the neutral position in absolute value, i.e.,

$$\sigma(\mathbf{c}_j) = \frac{1}{N_q} \sum_{t=1}^{N_q} |\mathbf{c}_{j,t} - \frac{1}{2}(\max A_t + \min A_t)|. \quad (3)$$

In Figure 2, we find that in Smartvote, candidates with moderate answers (i.e., lower answer strength) are recommended significantly more often. This concerning trend suggests that candidates can artificially boost their visibility by providing moderate answers to all questions. This strategy is particularly problematic because it can be executed with minimal deviation from the true candidate's position, making it difficult to detect. Figure 3 reveals that with the current distance metric used in Smartvote (d_{L2}), some parties can increase their visibility fourfold by unilaterally adopting this strategy.

Diversification (DIV)

Figure 4 shows that parties with more candidates relative to their vote share tend to receive disproportionately more recommendations on Smartvote. This significant correlation suggests that having more candidates can skew recommendations, thereby providing an artificial advantage in voter outreach and potentially electoral success.

[6] Answering moderately can be used to indicate a nuanced position, openness to compromise, or ambivalence. As such, the added expressivity is regarded to be beneficial [Batterton and Hale, 2017].

4.2 Platform Designers

Matching Method (MM)

Louwerse and Rosema [2014] show how sensitive recommendations are to changes in the matching method. We extend these findings by quantitatively evaluating the bias and accuracy of each distance function in Table 2. Additionally, in Table 8 of the Appendix, we show that some methods can disproportionately favor candidates at either end of the political spectrum.

Weight Selection (WS)

In Smartvote, voters have the option to decrease or increase the weight of each question q_t, but without knowing the actual numerical weights $W_t = \{0, \frac{1}{2}, 1, 2\}$ corresponding to these actions.[7] Figure 5 displays the relative change of the main parties' visibility (among voters that have weighted at least one question) if these values are changed.

Similarity Score (SS)

Apart from determining the ranking $\mathbf{r}_i$, the similarity scores $s(v_i, c_j)$ can also be displayed to provide voters with a sense of their relative proximity to different candidates. The exact calculation of such a score is mostly arbitrary. In Smartvote, the Euclidean distance between the voter and candidate is scaled by the maximum possible distance between two answers, as specified in Eq. (2). Figure 6 shows that the similarity scores of the best-matching candidate vary by party and are generally quite low, which is in large part a consequence of the curse of dimensionality [Thirey and Hickman, 2015]. This disparity could ultimately influence voters from different parties in different ways.

4.3 Question Designers

Question Favoritism (QF)

Certain questions can significantly benefit specific parties by aligning closely with their popular stances. Figure 7 shows

[7] These values are available on the *About* page on Smartvote, but they are not displayed directly alongside the questions.

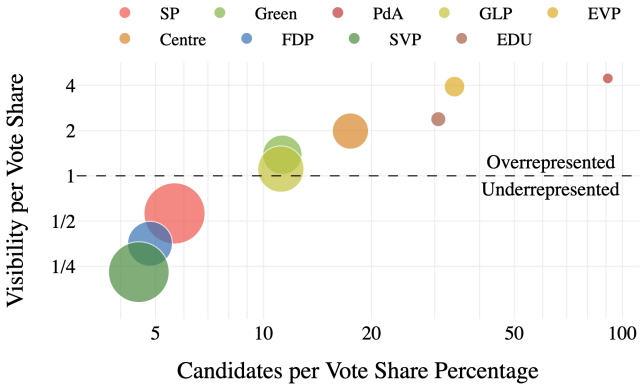


Figure 4: Relationship between the number of candidates per percent of vote share and the ratio of visibility to vote share for parties in the state of Zurich. The size of each dot represents the vote share of the corresponding party. Vote shares are calculated based on the votes received by candidates participating in Smartvote for the 2023 National Council election. Exact values can be found in the column *Vote Share (adjusted)* of Table 4 in the Appendix.

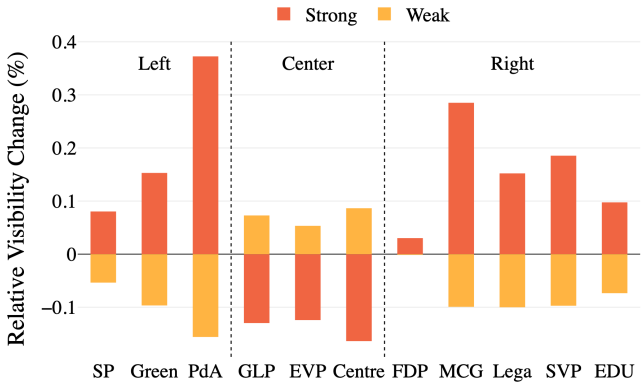


Figure 5: Relative visibility change of all parties if the available question weights are set to $W_t = \{0, \frac{1}{10}, 1, 10\}$ (strong) or $W_t = \{0, \frac{9}{10}, 1, \frac{10}{9}\}$ (weak). The visibility of each party is computed using only the voters that have weighted at least one question. Parties are listed according to their parliamentary seating arrangement, with traditional larger coalitions (left, center, right) shown at the top. As observed, the actual numerical value of the weights can significantly favor certain coalitions, with center parties benefiting from weak weights and left- and right-wing parties from strong weights.

the relative change in party visibility based on the size of alternative questionnaires. These questionnaires consist of a subset of questions from the original set, selected to benefit the respective parties the most during the elections in the state of St. Gallen. With this knowledge, an adversarial question designer could favor questions that benefit their preferred party.

Question Correlation (QC)
If a question is advantageous for a particular party, introducing additional questions with answers highly correlated to this question (among voters and candidates) implicitly increases its weight. For instance, asking the negation of a question effectively doubles the original question's weight. Although this strategy is inherently associated with question favoritism, it has the potential to magnify its impact.

5 Measuring Robustness

From Table 1, we note that three high-risk vulnerabilities, namely AC, AO, and MM, are highly dependent on the matching method. To assess the impact of matching methods on robustness, we compare the five most commonly used distance functions and two novel proposals using various key robustness metrics. A formal definition of these distance functions is provided in Appendix D.3.

Party Bias (BIA).
We assess the deviations in party visibility for each matching method relative to the median visibility observed across all other evaluated methods (see Appendix D.1 for a detailed discussion). Here we consider the mean absolute deviation (BIA1) and max deviation (BIA2) over the eight largest parties.

Calibration Potential (CP).
For each matching method, we repeat the analysis of Figure 3 and measure the average relative visibility gain or loss that results from a party employing the moderate answering strategy (CP-M) or the strong answering strategy (CP-S) weighted by the adjusted voter shares of the parties in the 2023 election (see Table 4 in Appendix A for the exact values).

Answer Strength Correlation (ASC).
This metric addresses the answer calibration manipulation strategy. It is defined as the Pearson correlation between the answer strength (defined in Eq. 3) and the expectation-normalized visibility of candidates. The expectation-normalized visibility adjusts for the varying number of candidates in each state by multiplying the visibility by the ratio of the number of candidates to the number of available seats in the state, ensuring comparability across different states. To minimize the effectiveness of any answer calibration strategy regarding the answer strength, this metric should ideally be close to zero, indicating no systematic bias toward candidates with moderate or strong answers.

Gini Coefficient (GIN).
This metric measures the Gini coefficient of the expectation-normalized visibilities over all candidates, indicating how evenly distributed the recommendations are among them. A Gini coefficient of 0 represents a perfectly even distribution, and a coefficient of 1 indicates a completely uneven distribution. While there is no ideal Gini coefficient for a distance method, and actual election votes are typically less evenly distributed than Smartvote recommendations (see Figure 20 in the Appendix), the Gini coefficient offers insight into the differences in recommendation diversity between matching methods.

Party Match Accuracy (ACC1).
This metric measures the proportion of voters whose top list recommendation matches their preferred party. As manual accuracy checks are impractical, comparing the voter's stated

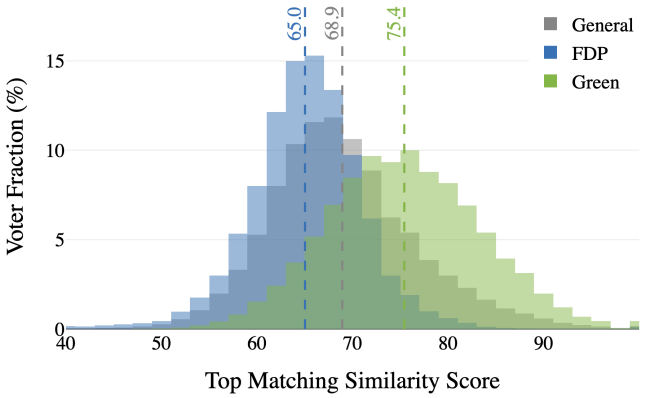


Figure 6: Distribution of similarity scores between voters and their top matching candidate, with colored histograms isolating voters whose top candidate is from a specific party. This histogram reveals that the matching percentages vary significantly based on the party of the top matching candidate. It also shows that for many voters, their top matching candidate is surprisingly low (below 70%).

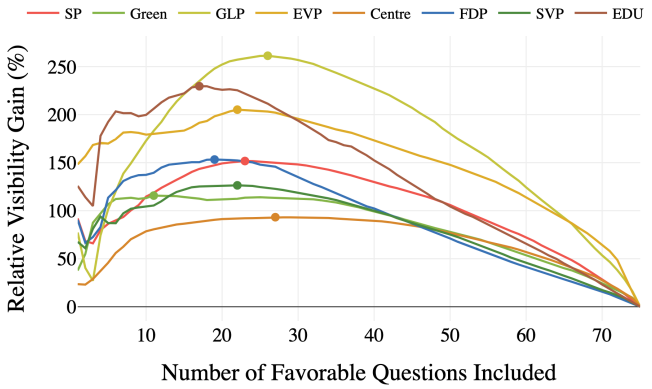


Figure 7: Relative visibility gain for each party using a set of greedily selected optimal questions to generate voting advice in the state of St. Gallen. Each line represents a political party and shows its increase in visibility as more and more favorable questions are included, compared to the baseline scenario with the full questionnaire. Circles indicate each party's maximum attainable visibility (e.g., when only choosing the best-aligned 12 questions, the Green party can increase its visibility by 120%).

preferred party with the party recommended by the algorithm is common for assessing the accuracy of VAAs [Garzia and Marschall, 2014]. For Smartvote, which does not directly recommend parties, we use the party from the best-matching list as a proxy. While this metric is appealing for its simplicity, it assumes that voters know the party that best represents them, which may not always be true.

Normalized Party Rank (ACC2).
This metric provides deeper insight into the rankings of lists associated with voters' preferred parties. It measures the average normalized rank of the top list for the preferred party, with normalization adjusting for the number of lists per state. A normalized rank of 0 means the list is recommended first, while a value of 1 means it is recommended last.

Strong Disagreement Accuracy (ACC3).
This metric measures the disagreements between voters and their recommended candidates. However, it specifically focuses on questions that voters weighted more strongly, indicating their greater importance. This metric should ideally be low, as voters likely expect their recommended candidates to align with them on these high-priority questions.

6 Future Work on Mitigation Approaches

Below, we present a series of possible mitigation strategies, specifying the vulnerabilities they aim to address. We also provide mitigations for TB, QO and LC in Appendix E. We emphasize that these strategies have not been extensively tested and may introduce unintended harms. We introduce them here as a foundation for future work, aiming to facilitate systematic research in this direction. Mitigation strategies currently under Politools review are marked with 🔍.

🔍 L1 or Angular instead of L2 (AC, AO, MM).
While each distance metric has its trade-offs, we find in Table 2 that L1 and Angular consistently offer better robustness than L2 without sacrificing accuracy. Specifically, L1 outperforms L2 in ACC1 and ACC2, while Angular excels in ACC3 with only minor reductions in ACC1 and ACC2. Therefore, we argue that any of these two methods is a viable robust substitution for L2. Alternatively, the Hybrid method appears to offer strong robustness properties with only a slight decrease in accuracy across all three metrics.

Lower Expressivity (AC, AO).
Reducing the number of allowable answers can reduce the impact of many vulnerabilities by limiting opportunities for fine-grained manipulation. Since expressivity is important for voters, it could be reduced specifically for candidates. For example, candidates could be restricted to answering "Yes" or "No" for each question, while voters still have "Rather yes" and "Rather no" as options. This would effectively mitigate the answer calibration strategy, which, based on our subjective assessment in Table 1, poses the greatest risk.

🔍 Deal-breaker Filtering (WS).
As demonstrated by the vulnerability to weight selection, voters could easily misunderstand the effect of weighting questions. To address this issue, we propose to allow only the weights to $W_t = \{0, 1, \infty\}$ for each question q_t. Assigning a weight ∞ to a question effectively treats it as a deal-breaker [Isotalo, 2021], directly excluding all candidates who answered differently from the voter on that question. To avoid leaving voters without candidates due to excessive filtering, the matching algorithm could consider the number of disagreements on deal-breakers as the primary factor in determining the similarity scores. Alternatively, one could also allow voters to exclude all candidates not aligned with their chosen side of the answer spectrum relative to the neutral response.

🔍 Selective Answering (QF, QC).
Voters should be informed that answering more questions does not necessarily lead to a more accurate recommendation

Distance Function (See App. D.3)	BIA1 |↓|	BIA2 |↓|	CP-M ↓	CP-S ↓	ASC |↓|	GIN	ACC1 ↑	ACC2 ↓	ACC3 ↓	Used By
L2	23.0%	+40.7% (EVP)	+207%	-71%	-0.470	0.475	41.0%	0.103	7.8%	*Smartvote*
L1	14.3%	+24.4% (Centre)	+46%	-50%	-0.280	0.373	**41.8%**	**0.101**	11.0%	Wahl-O-Mat
Angular	4.1%	-12.2% (GLP)	-27%	-13%	0.190	0.349	40.2%	0.109	**7.0%**	-
Agreement Count	**3.6%**	+7.8% (SVP)	-36%	-4%	0.256	0.317	35.7%	0.111	15.0%	Stemwijzer
Mahalanobis	29.2%	-47.2% (EDU)	+305%	-69%	**0.044**	0.523	29.0%	0.142	21.6%	-
L1 Bonus	15.5%	-27.1% (GLP)	-81%	+27%	0.583	0.387	37.9%	0.109	11.3%	Smartvote (old)
Hybrid	5.3%	-15.7% (GLP)	-55%	-12%	0.292	0.349	40.2%	0.106	10.1%	EUVox

Table 2: Comparison of alternative distance functions based on various metrics defined in Section 5. The arrows indicate what is desired from the metric (↑: Higher is better, ↓: Lower is better, |↓|: Closer to 0 is better). The best value for each metric is highlighted in bold, and the worst value is underlined. The GIN metric is purely informational, with no suggestion that higher or lower values are better. *Smartvote (old)* refers to the Smartvote VAA until 2010. A detailed discussion about the counterintuitive CP-M and ASC value for Mahalanobis is provided in Appendix D.2.

and may even distort the results. The user interface could instead promote a more selective approach to question selection by each voter.

Distance to Party Mean (AC, AO).
Voters often lack tools to assess a candidate's honesty and determine if they have answered truthfully or exploited VAA vulnerabilities to boost their visibility. One solution is to display the distance between each candidate's answers and their party's mean answers. A large distance might prompt voters to scrutinize the candidate's responses more closely. However, this metric would only be a proxy for honesty, as some candidates may naturally deviate from their party's position [Schwarz *et al.*, 2010].

Limiting the Number of Candidates (DIV).
To prevent parties from disproportionately boosting their visibility by increasing candidate numbers, we propose limiting the number of candidates from the same party that can be recommended to any voter. This limit could be based on the similarity score between the voter's position and each party's average position. For instance, if two parties have the same similarity score with a given voter but one has more candidates, the top k recommendations should be evenly distributed between the parties, minimizing the risk of biased recommendations arising from the diversification strategy.

Fair Answer Normalization (SS).
To avoid presenting varying similarity scores to voters from different parties (as shown in Figure 6), we propose normalizing similarity scores relative to the top candidate for each voter (who would always be considered a 100% match). While this would change the score's meaning and might reduce its overall usefulness, it would also eliminate bias.

7 Related Work

VAAs emerged around 30 years ago and have quickly gained popularity since then. Garzia and Marschall [2012] provide a comprehensive overview of existing VAAs, summarized in Table 3 in the Appendix. The voter data collected by VAAs are a treasure trove, for political, social, and computer scientists alike. Etter *et al.* [2014] for example, extract valuable data on the Swiss political landscape. An extended related work discussion on the influence of VAAs on democratic institutions and their development is detailed in Appendix F.

VAAs under Scrutiny.
Walgrave *et al.* [2009] show that the question selection has a substantial impact on the voting advice. Louwerse and Rosema [2014] highlight the significant impact matching methods (mainly L1 and L2) have on recommendations, using *StemWijzer* as an example. We corroborate this finding but crucially demonstrate that these matching methods behave differently in the presence of an adversary. Van der Linden and Dufresne [2017] critically analyze current methods to visualize aggregate results, and propose a technique based on learned dimensions to correct shortcomings. Finally, Isotalo [2021] identifies several issues with Finnish VAAs, including lack of transparency, user interactivity, and problems in statement structure. Our work supports the effectiveness of their suggested filtering method.

Adversarial Robustness of Recommender Systems.
Other applications have recognized the importance of adversarial robustness [Hurley, 2011; Tang *et al.*, 2019] and the challenges of questionnaire design [Pasek and Krosnick, 2010]. Ovaisi *et al.* [2022] provide a toolkit to compare the robustness of learning-based recommender systems. Given the much stricter requirements of recommender systems for democracy (see Section 2), while our introduced metrics apply to all methods, we restrict our evaluation to non-learning-based methods for now.

8 Outlook

This study highlights critical vulnerabilities in voting advice applications (VAAs), providing empirical evidence that malicious actors could pose a risk to democratic processes. Crucially, many vulnerabilities also uncover the existence of strong biases in VAAs, even in the absence of adversarial entities. We are convinced that VAAs are a highly desirable addition to the political landscape and believe that our proposed comparative metrics and mitigations can help guide future VAA development toward more robust designs. As VAAs continue to evolve in the era of AI, future work should also aspire to extend our results to other types of political recommender systems that fall outside our formalism.

Ethical Statement

The dataset has been collected and anonymized by Politools in accordance with the new Swiss Federal Act on Data Protection (nFADP), the Telecommunications Act (TCA), and other applicable data protection regulations [Politools, 2024b]. Further, we strictly follow the platform's terms of use for data. As such, we do not publish results that may be attributed to specific individuals. In accordance with the terms of use for research, the dataset is kept private, and we adhere to established best practices for dealing with sensitive data. While the dataset cannot be made accessible directly, it might be made available to researchers by Politools upon request [Politools, 2024a]. Given access to the data, all numerical results and figures can be easily reproduced using the code in the supplementary material.

In the absence of established Ethics guidelines, we follow Menlo's report on Computer Science research principles [Kenneally and Dittrich, 2012]. Publicly disclosing all found vulnerabilities presents a risk, as various actors might benefit from exploiting them. We mitigate these risks by publishing our results after Switzerland's national election, leaving enough time to implement potential mitigation for the 2027 elections. To the best of our knowledge, no countries with popular VAAs that could be affected by our research will hold national elections in the months following the publication of this work. Thus, we believe that this is the right time to shed light on these vulnerabilities. Overall, we believe that despite some inherent risks, this work will have a clear net positive social impact by providing tools to enhance the robustness of VAAs, and consequently, democracies.

Acknowledgments

We thank Michael Erne and Daniel Schwarz from Politools for the uncomplicated and fruitful collaboration, as well as their helpful feedback. Our gratitude extends to Roger Germann, Douglas Orsini-Rosenberg, Leon Plath, and Gina Stoffel, whose Bachelor and Master thesis contributed to our understanding of VAAs and the success of this project. Finally, we thank Judith Beestermöller and Robin Fritsch for their valuable input and guidance.

References

R Michael Alvarez, Ines Levin, Peter Mair, and Alexander Trechsel. Party preferences in the digital age: The impact of voting advice applications. *Party Politics*, 20(2):227–236, 2014.

R Michael Alvarez, Ines Levin, Alexander H Trechsel, and Kristjan Vassil. Voting advice applications: How useful and for whom? *Journal of Information Technology & Politics*, 11(1):82–101, 2014.

Solomon E Asch. Forming impressions of personality. *The journal of abnormal and social psychology*, 41(3):258, 1946.

Katherine A Batterton and Kimberly N Hale. The likert scale what it is and how to use it. *Phalanx*, 50(2):32–39, 2017.

Vincent Etter, Julien Herzen, Matthias Grossglauser, and Patrick Thiran. Mining democracy. In *Proceedings of the second ACM conference on Online social networks*, pages 1–12, 2014.

Diego Garzia and Stefan Marschall. Voting advice applications under review: the state of research. *International Journal of Electronic Governance*, 5(3-4):203–222, 2012.

Diego Garzia and Stefan Marschall. *Matching Voters With Parties and Candidates. Voting Advice Applications in a Comparative Perspective*. ECPR Press, 01 2014.

Diego Garzia and Stefan Marschall. *Voting advice applications*. Oxford University Press, 2019.

Micha Germann and Kostas Gemenis. Getting out the vote with voting advice applications. *Political Communication*, 36(1):149–170, 2019.

Neil J Hurley. Robustness of recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 9–10, 2011.

Veikko Isotalo. Improving candidate-based voting advice application design: The case of finland. *Informaatiotutkimus*, 40(3):85–109, 2021.

Dahyeon Jeong, Shilpa Aggarwal, Jonathan Robinson, Naresh Kumar, Alan Spearot, and David Sungho Park. Exhaustive or exhausting? evidence on respondent fatigue in long surveys. *Journal of Development Economics*, 161:102992, 2023.

Erin Kenneally and David Dittrich. The menlo report: Ethical principles guiding information and communication technology research. *Available at SSRN 2445102*, 2012.

Scott Kirkpatrick, C Daniel Gelatt Jr, and Mario P Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

Jan Kleinnijenhuis and Andre Krouwel. Dimensionality of the european issue space. In *Political Representation*, pages 99–116. Routledge, 2016.

Yakov Kuzyakov, JK Friedel, and Karl Stahr. Review of mechanisms and quantification of priming effects. *Soil Biology and Biochemistry*, 32(11-12):1485–1498, 2000.

Andreas Ladner and Joëlle Pianzola. Do voting advice applications have an effect on electoral participation and voter turnout? evidence from the 2007 swiss federal elections. In *Electronic Participation: Second IFIP WG 8.5 International Conference, ePart 2010, Lausanne, Switzerland, August 29–September 2, 2010. Proceedings 2*, pages 211–224. Springer, 2010.

Andreas Ladner, Jan Fivaz, and Joëlle Pianzola. Voting advice applications and party choice: evidence from smartvote users in switzerland. *International Journal of Electronic Governance*, 5(3-4):367–387, 2012.

Andreas Ladner. Do vaas encourage issue voting and promissory representation? evidence from the swiss smartvote. *Policy & Internet*, 8(4):412–430, 2016.

Tom Louwerse and Martin Rosema. The design effects of voting advice applications: Comparing methods of calculating matches. *Acta politica*, 49:286–312, 2014.

Fernando Mendez. Matching voters with political parties and candidates: an empirical test of four algorithms. *International Journal of Electronic Governance*, 5(3-4):264–278, 2012.

Fernando Mendez. Modeling proximity and directional decisional logic: What can we learn from applying statistical learning techniques to vaa-generated data? *Journal of Elections, Public Opinion and Parties*, 27(1):31–55, 2017.

Simon Munzert and Sebastian Ramirez-Ruiz. Meta-analysis of the effects of voting advice applications. *Political Communication*, 38(6):691–706, 2021.

Zohreh Ovaisi, Shelby Heinecke, Jia Li, Yongfeng Zhang, Elena Zheleva, and Caiming Xiong. Rgrecsys: A toolkit for robustness evaluation of recommender systems. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 1597–1600, 2022.

Josh Pasek and Jon A Krosnick. Optimizing survey questionnaire design in political science: Insights from psychology. *The Oxford Handbook of American Elections and Political Behavior*, 2010.

Politools. Smartvote. smartvote.ch, 2024.

Politools. Terms of use and data protection. https://www.smartvote.ch/en/wiki/anb-privacy, 2024.

Daniel Schwarz, Lisa Schädel, and Andreas Ladner. Preelection positions and voting behaviour in parliament: Consistency among swiss mps. *Swiss Political Science Review*, 16(3):533–564, 2010.

Jinhui Tang, Xiaoyu Du, Xiangnan He, Fajie Yuan, Qi Tian, and Tat-Seng Chua. Adversarial training towards robust multimedia recommender system. *IEEE Transactions on Knowledge and Data Engineering*, 32(5):855–867, 2019.

Luis Terán. Voting advice applications. *Dynamic Profiles for Voting Advice Applications: An Implementation for the 2017 Ecuador National Elections*, pages 15–26, 2020.

Benjamin Thirey and Randal Hickman. Distribution of euclidean distances between randomly distributed gaussian points in n-space. *arXiv preprint arXiv:1508.02238*, 2015.

Clifton Van der Linden and Yannick Dufresne. The curse of dimensionality in voting advice applications: reliability and validity in algorithm design. *Journal of Elections, Public Opinion and Parties*, 27(1):9–30, 2017.

Stefaan Walgrave, Michiel Nuytemans, and Koen Pepermans. Voting aid applications and the effect of statement selection. *West European Politics*, 32(6):1161–1180, 2009.

A Smartvote and Swiss Politics

A.1 Swiss National Elections

The Swiss parliament is composed of two chambers: the National Council and the Council of States. Together they hold the legislative power in Switzerland, meaning they are responsible for passing laws and approving the federal budget. The National Council is constituted of 200 seats, which are assigned proportionally to the states' populations (approximately). This proportional distribution of seats to the states ensures that each seat in the National Council represents approximately the same number of voters.

The distribution of seats in the National Council following the 2023 elections is illustrated in Figure 8. The largest party is the Swiss People's Party (SVP), which holds 31% of the seats, followed by the Social Democratic Party of Switzerland (SP) with 20.5%. Parliamentary elections are held every four years, and the Swiss parliament is elected on a cantonal basis. This means voters can only cast votes for candidates who stand for election in the same state as their place of residence. The number of candidates a voter can vote for is determined by the number of seats allocated to the state of residence of the voter in the National Council.

In most states, candidates are required to form a coalition, known as a candidate list, to be eligible to stand for election. Voters are then presented with the option of either voting for a predefined list of candidates or composing their custom list of individual candidates. The number of candidates on a list corresponds to the number of seats available in the state. Furthermore, voters have the option to give more weight to specific candidates by placing a candidate's name twice on the list, effectively giving them two votes. The political landscape of Switzerland's federal government comprises 11 major parties whose name, abbreviation, current number of seats, and vote share in the 2023 National Council elections are summarized in Table 4.

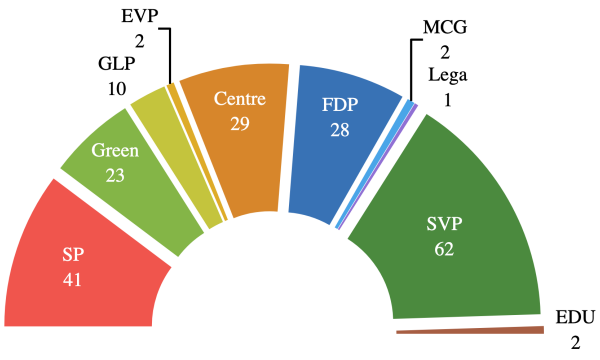


Figure 8: Seat distribution of the National Council after the 2023 elections.

A.2 Questionnaire Design

The questionnaire is the core element for capturing political positions and is newly created for each election. The creation process starts months before the elections and involves the collection of question proposals from official party or government websites, media analysis, and public submissions. Over 1,500 proposals were gathered for the 2019 Swiss federal elections. The Smartvote team independently refines these proposals into a 45 to 75-question survey, ensuring clarity, neutrality, and relevance to current and future political discussions. Finally, the draft questionnaire is quality-checked by experts from the scientific community and selected Smartvote users.

The final questionnaire consists of 75 questions split into three question types: 60 policy questions with $A_t = \{0, 25, 75, 100\}$, 7 value questions with $A_t = \{0, 17, 33, 50, 67, 83, 100\}$, and 8 budget questions with $A_t = \{0, 25, 50, 75, 100\}$. Policy questions capture positions on specific political issues (e.g., *Do you support an increase in the retirement age?*). Value questions capture agreement or disagreement with broad political principles (e.g., *What is your position on the following statement: "In the long term, everyone benefits from a free market economy."?*). Budget questions capture how the spending on key federal budget items should be adjusted (e.g., *Should the federal government spend more or less in the area of social services?*).

Voters also have the option to complete a rapid version of the questionnaire, which consists of the 30 policy questions deemed most important by the questionnaire designers, as an alternative to the full 75-question deluxe version.

Party	Code	Vote Share	Vote Share (adjusted)
Swiss People's Party	SVP	27.9%	27.1%
Social Democratic Party	SP	18.2%	19.1%
The Centre	Centre	14.3%	14.7%
FDP. The Liberals	FDP	14.1%	13.8%
Green Party	Green	9.8%	10.4%
Green Liberal Party	GLP	7.6%	8.0%
Evangelical People's Party	EVP	2.0%	2.0%
Federal Democratic Union	EDU	1.2%	1.1%
Party of Labour	PdA	0.7%	0.4%
Lega dei Ticinesi	Lega	0.6%	0.2%
Geneva Citizens Movement	MCG	0.5%	0.3%

Table 4: Overview of the major parties in the 2023 Swiss National Council elections. The table displays the party names, their codes, their vote shares as a percentage of the total national vote, and their adjusted vote shares (calculated by excluding votes cast for candidates not registered in Smartvote). Tiny parties and candidates without party affiliations are excluded, explaining why percentages may not add up to 100%.

A.3 Smartvote Usage

Before each National Council election cycle, candidates are invited to participate by filling out a detailed questionnaire. This invitation process typically starts a few months before the election, allowing candidates enough time to complete their profiles. All candidates are required to answer all questions in the questionnaire. Candidates then have time to complete their profiles until the questionnaire is made public to the voters. After the publications candidates are generally not

VAA Name	Region	Type	$\lvert A_t \rvert$	W_t	Distance Metric
Smartvote		👤 ☰	4, 5, 7	$\{0, 0.5, 1, 2\}$	L2
Wahl-O-Mat		⚑	3	$\{0, 1, 2\}$	L1
StemWijzer (VoteMatch)		⚑	3	$\{0, 1, 2\}$	Agreement Count
Wahlkabine		⚑	2	$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$	Agreement Count
VoteCompass		👤 ⚑	5	$\{0, 1\}$	L1
Volkskabin, Politikkabine		👤 ⚑	2	$\{0, 0.5, 1, 2\}^8$	Agreement Count (adj.)
Manobalsas		👤 ⚑	5	$\{0, 1\}$	L2 (low-dim.)
EU&I (Euandi)		⚑	5	$\{0, 0.5, 1, 2\}$	L1
EUVox 2014		⚑	3	$\{0, 1\}$	Hybrid
iSideWith		👤	2	5 weights (undisclosed)	Agreement Count (adj.)

Table 3: The table includes the name, region (flags), type of recommendation (👤: candidates, ☰: lists, ⚑: parties), number of answer options, question weights, and distance metric used by some of the most popular VAAs globally.

allowed to adjust their answers anymore. In individual, justified exceptional cases, however, corrections have been made after publication in the case of clear errors.

Once the candidate profiles are finalized, Smartvote is made available to the public. Voters can then access the platform, fill out the questionnaire, and receive voting advice based on the candidate's responses. Additionally, voters have the option to provide metadata such as demographic details, political preferences, and their interest in politics before completing the questionnaire. Voters are free to use Smartvote as often as they like until the election date, allowing them to make informed decisions based on the most up-to-date information available. It is important to note that the voters using Smartvote represent a biased sample of the Swiss voting population. Our analysis indicates that Smartvote users are generally younger, more likely to reside in urban areas and have a preference for left-leaning parties compared to the overall voting population. This bias may influence the results and should be considered when interpreting the data.

A.4 Smartvote UI

Figure 9a illustrates how candidate recommendations are displayed on Smartvote. It shows how the top-ranked candidates are prominently featured based on the number of seats available in the state. This also demonstrates our definition of candidate visibility, determined by the proportion of voters for whom a candidate is prominently displayed. Similarly, Figure 9b illustrates how list recommendations are presented to voters, effectively demonstrating the concept of list visibility, which is defined as the proportion of voters for whom a given list is prominently displayed.

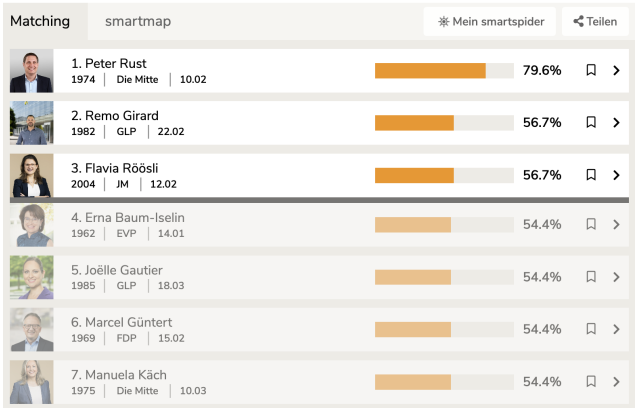
A.5 Limitations

To validate our implementation of the matching algorithm used by Smartvote and to ensure the integrity of the dataset, we tried to reproduce the recommendations from the 2023 Smartvote elections dataset, representing the exact recommendations voters were given on the platform. Despite our best efforts, we encountered certain limitations. A small number of candidates deleted their profiles after the publication of Smartvote, which meant that recommendations involving them were not fully reproducible. Additionally, 17 candidates were allowed to correct their responses after the publication due to clear errors, making the recommendations involving them before the corrections not fully reproducible.

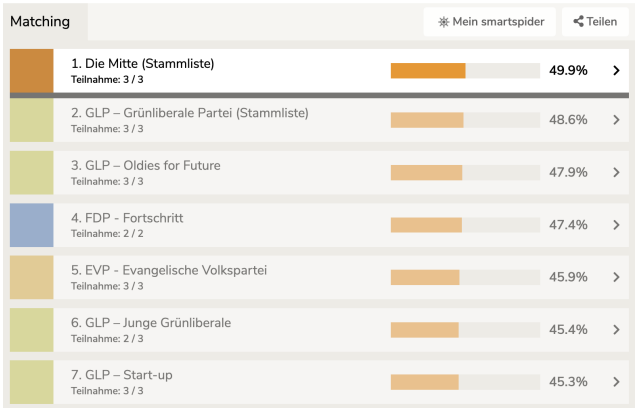
B Dataset (continued)

B.1 Preprocessing & Cleaning

To enhance the quality of our data and prepare it for analysis, we conducted thorough preprocessing on both the voter and candidate datasets from both elections. For the voter datasets, we first addressed corrupt entries by identifying and removing those lacking essential information, such as the election ID. To ensure the reliability of our analysis, we excluded recommendations where voters answered fewer than 15 questions, which accounted for less than 1.3% of the dataset. Further refinement involved filtering out recommendations both before the questionnaire went live and after the election date. This step aimed to exclude recommendations intended for testing purposes and those no longer relevant to voting decisions. To streamline our dataset, we implemented a deduplication process. We retained only the most recent recommendations with the highest number of answered questions for each unique voter ID, ensuring data consistency. Lastly, to eliminate recommendations resulting from random clicking behavior, we filtered out those with more than 14 consecutive identical answers, which accounted for about 0.1% of the dataset. As for the candidate datasets, our preprocessing primarily involved filtering out candidates who did not participate in Smartvote as the dataset included all candidates up for election. As part of the preprocessing, we ensured that all variables used in our analysis contained valid and sensible values for all voters and candidates. Additionally, we created a consolidated party variable by merging the youth wings of political parties with their respective main parties, consistent with the preferred



(a) Smartvote UI for candidate recommendations.



(b) Smartvote UI for list recommendations.

Figure 9: Screenshots of candidate and list recommendations on Smartvote for a voter in the state of Zug. Candidates and lists are ranked by their similarity score in descending order. Given that the state of Zug has three seats available in the National Council (which is also the list size for that state), the top three candidates and the first list are prominently displayed. This example also highlights how Smartvote resolves ties between candidates with identical similarity scores using their last names.

party variable provided by users. For instance, JUSO (Young Socialists) was merged with SP (Social Democratic Party), and JSVP (Young Swiss People's Party) was combined with SVP (Swiss People's Party). Table 5 provides an overview of how youth parties were mapped to the main parties.

Youth Party	Main Party
JUSO	SP
JG	Green
JGLP	GLP
JEVP	EVP
JM	Centre
JFS	FDP
JSVP	SVP

Table 5: Party combinations showing how youth parties are combined into main parties.

B.2 Exploratory Data Analysis

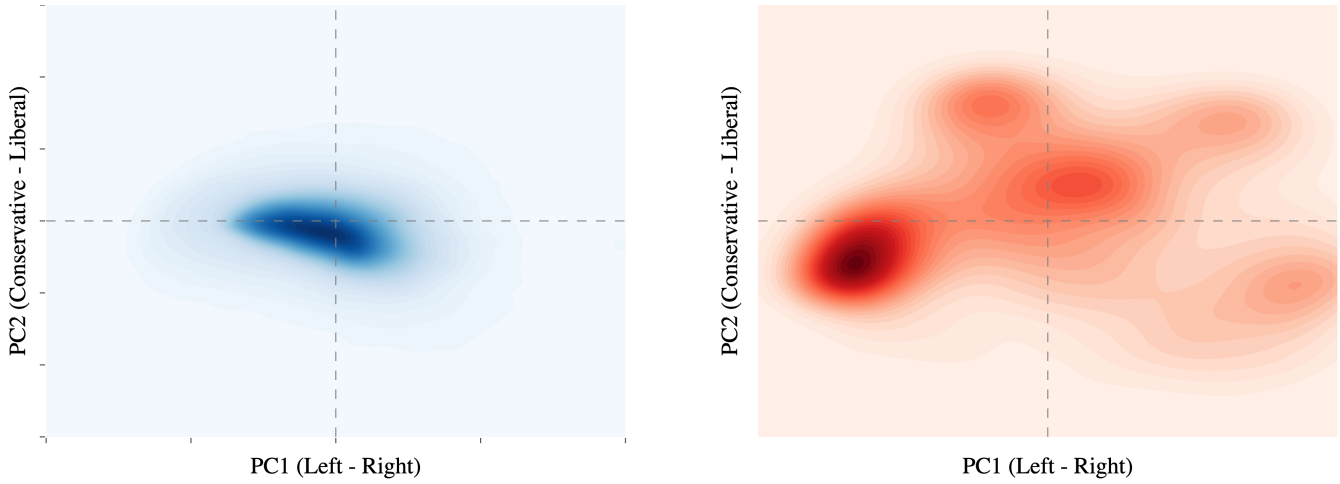
Dataset	# Rec.	N_v	N_c	Lists	N_q
2019	427,572	389,881	3,926	508	75
2023	1,662,683	485,838	4,983	623	75

Table 6: Dataset overview for the 2019 and 2023 National Council elections. The table includes the number of voting recommendations before cleaning (# Rec.), the number of recommendations after cleaning (taken as the number of unique voters N_v), the number of participating candidates (N_c), the number of lists, and the number of questions in each dataset (N_q).

To gain a deeper understanding of the datasets and provide intuition to the reader, we performed an exploratory data analysis built in large part around a Principal Component Analysis (PCA) on the voter and candidate answer profiles. These analyses help to uncover underlying patterns and structures in the data, which are crucial for evaluating and improving the matching algorithm.

One of the key analyses involves visualizing the density distributions of voters and candidates in a two-dimensional political space defined by the two first principal components of the candidate answer profiles. These dimensions can be roughly interpreted as Left-Right and Conservative-Liberal. The first plot (Figure 10a) depicts the voter density for the 2023 election, showing how voters are distributed across this political landscape. The second plot (Figure 10b) illustrates the candidate density, highlighting where candidates position themselves in the same political space. Comparing these plots reveals the alignment or gaps between voter preferences and candidate positions, providing valuable insights into the effectiveness of the VAA's recommendations.

Additionally, Figure 11 shows the cumulative explained variance of answers as a function of the number of principal components, for both voters and candidates. From this plot, we observe that while the first three principal components capture more than 50% of the variance in the candidate data, they only capture around 30% of the variance in the voter data. This suggests that candidate responses are more structured and consistent, likely due to alignment with party platforms, whereas voters' responses are more varied, reflecting a broader range of opinions and requiring more principal components to capture the same variance. This observation aligns with findings by Kleinnijenhuis and Krouwel [2016], who noted that voters' policy preferences are not strongly structured and cannot be easily captured by low-dimensional spatial models. We also visualize the positions of individual candidates in a two-dimensional space. Figure 12 displays the PCA plot of the candidates' positions, colored by their respective parties. This visualization highlights the distribution and clustering of candidates along the primary axes of political orientation, which can be interpreted as Left-Right and Conservative-Liberal [Garzia and Marschall, 2014; Politools, 2024a]. From the PCA plot, we can observe distinct cluster-



(a) Voters' answer distribution.

(b) Candidates' answer distribution.

Figure 10: Density plots of voters and candidates in the 2D political space for the 2023 election, showing the distribution of their answers along the two principal components dimensions of the candidate answers. These dimensions can be interpreted as economic Left-Right and social Conservative-Liberal [Garzia and Marschall, 2014; Politools, 2024a].

ing patterns among different parties. For instance, the SP and PdA parties, represented in red, are predominantly situated in the left-wing quadrant. Conversely, the SVP, shown in green, occupies the right-wing, conservative area of the plot. The

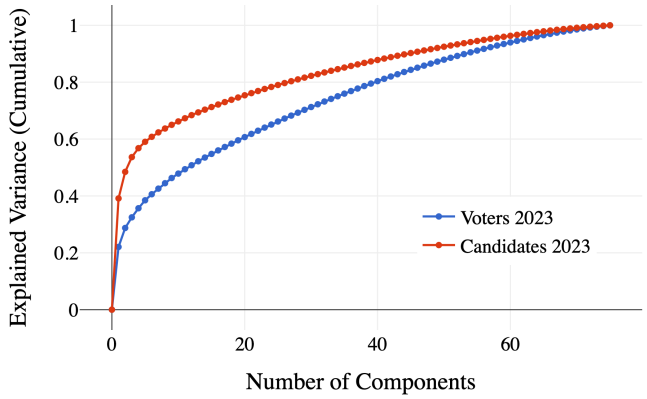


Figure 11: PCA explained variance for voter and candidate answer profiles in the 2023 dataset. The plot shows the cumulative explained variance as a function of the number of principal components for both voters and candidates.

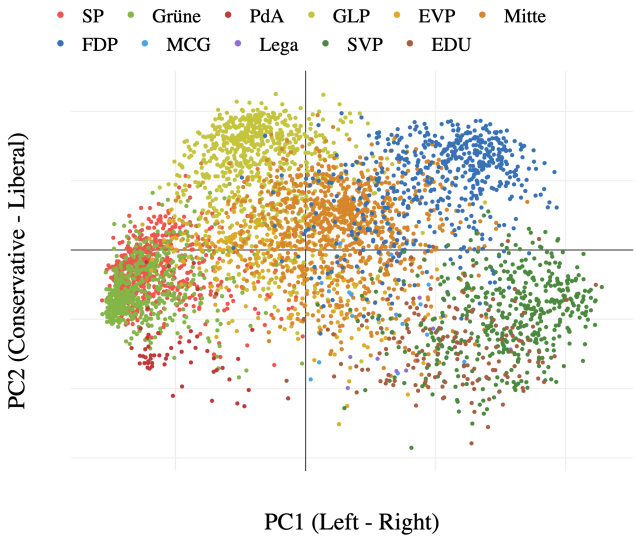


Figure 12: PCA plot of candidate positions by party in the 2023 dataset, highlighting candidate clustering and each party's political orientation.

FDP party, colored in blue, is more dispersed but generally aligns with liberal positions. This differentiation underscores the varying political ideologies and strategies of each party.

C Vulnerabilities (continued)

Supplementary Material

As additional material for the AO vulnerability, Table 7 provides the visibility of the crafted candidates for all states and different optimization strategies. Concerning the MM vulnerability, Table 8 provided detailed visibility measurement for each party using several matching methods. Lastly, Figure 15 displays the correlation between the voters' answers, supporting the fact that QC can have a non-negligible impact on recommendations. We now expose three additional vulnerabilities, namely LC, QO and TB.

List Centralization (LC)

This danger pertains to the safety of the list recommendation functionality of VAAs. We show that there exists a bias toward favoring lists that group many candidates with similar answer vectors compared to lists with more diversity. Figure 13 quantitatively demonstrates the impact of this bias for some states. If parties are limited in their number of lists, this creates a trade-off with the diversification strategy (see Section 4.1 in the main paper). However, in Smartvote, par-

State (canton)	N_v	N_c	Seats	Highest Visibility	Visibility Optimized 1% Data	Visibility Optimized 100% Data
Zurich	104,826	1,029	36	24.77%	51.70%	**53.43%**
Bern	89,378	685	24	29.87%	50.66%	**53.71%**
Aargau	47,442	568	16	20.44%	51.29%	53.99%
Lucerne	32,378	329	9	31.12%	49.70%	53.66%
St. Gallen	27,421	288	12	32.19%	52.55%	**61.71%**
Vaud	25,591	337	19	37.66%	54.51%	64.92%
Valais	20,586	199	8	25.70%	55.27%	61.82%
Fribourg	18,982	137	7	35.07%	59.83%	63.84%
Solothurn	16,364	163	6	19.29%	52.71%	59.84%
Basel-Landschaft	15,872	163	7	20.60%	51.08%	58.05%
Thurgau	14,644	187	6	16.58%	44.52%	54.73%
Basel-Stadt	13,233	107	4	18.07%	46.91%	57.10%
Graubünden	10,666	109	5	36.27%	55.93%	61.92%
Geneva	9,892	217	12	35.05%	43.13%	63.92%
Schwyz	8,593	98	4	25.54%	44.13%	58.01%
Zug	7,287	84	3	22.73%	34.49%	55.11%
Neuchâtel	6,689	56	4	46.17%	32.74%	70.82%
Ticino	5,023	144	8	29.94%	25.22%	58.21%
Schaffhausen	3,257	36	2	33.93%	41.63%	64.05%
Jura	3,039	34	2	16.75%	33.30%	60.09%
Appenzell Ausserrhoden	1,263	2	1	79.41%	53.13%	96.75%
Glarus	892	3	1	51.79%	21.19%	76.01%
Nidwalden	806	3	1	72.58%	51.36%	88.46%
Uri	699	2	1	72.53%	66.95%	95.42%
Obwalden	662	2	1	75.83%	77.49%	96.22%
Appenzell Innerrhoden	353	1	1	100.0%	79.89%	100.0%

Table 7: List of states along with their corresponding number of unique voters (N_v), registered candidates (N_c), and the number of available seats for the 2023 National Council elections. The *highest visibility* denotes the most visible candidate in the dataset. The last two columns show the visibility of candidates generated using the *simulated annealing* method [Kirkpatrick *et al.*, 1983] with the first 1% of the data and with 100% of the data, respectively. The values in **bold** correspond to the red dots shown in Figure 1 of the main paper.

Distance Metric	SP	Green	PdA	GLP	EVP	Centre	FDP	MCG	Lega	SVP	EDU
Vote Share	18.2%	9.8%	0.7%	7.6%	2.0%	14.3%	14.1%	0.5%	0.6%	27.9%	1.2%
Preferred Party	27.60%	14.28%	0.74%	18.27%	1.92%	11.01%	13.52%	0.02%	0.07%	9.79%	0.57%
L2	14.79%	12.29%	0.69%	15.07%	8.49%	21.88%	7.97%	0.14%	0.06%	6.43%	2.85%
L1	16.73%	14.21%	0.83%	14.2%	7.47%	19.69%	7.86%	0.16%	0.06%	7.27%	2.78%
Agreement Count	18.84%	16.54%	0.95%	12.97%	6.07%	16.1%	7.79%	0.19%	0.06%	8.32%	2.72%
Angular	19.38%	17.29%	1.2%	11.92%	6.46%	15.83%	8.2%	0.19%	0.07%	8.26%	2.7%
Mahalanobis	23.17%	23.26%	0.58%	15.75%	5.4%	14.7%	5.37%	0.05%	0.04%	4.38%	1.43%
L1 Bonus	21.78%	20.09%	1.17%	9.9%	4.87%	13.06%	8.08%	0.21%	0.07%	8.97%	2.56%
Hybrid	19.74%	18.25%	1.12%	11.44%	6.0%	15.82%	8.07%	0.19%	0.07%	8.16%	2.62%

Table 8: Party visibilities when using different matching methods. The first two rows serve as references: the first row displays the parties' vote shares in the 2023 Swiss National Council elections, while the second row shows the distribution of preferred parties as indicated by voters. The cell colors indicate deviations from these preferences: higher values are green and lower values are red, with the saturation representing the magnitude of the deviation.

ties are allowed to create many lists, making both strategies exploitable simultaneously.

Tie-breaking (TB)

With the extensive number of voters and candidates using Smartvote, and considering that some voters answer only a

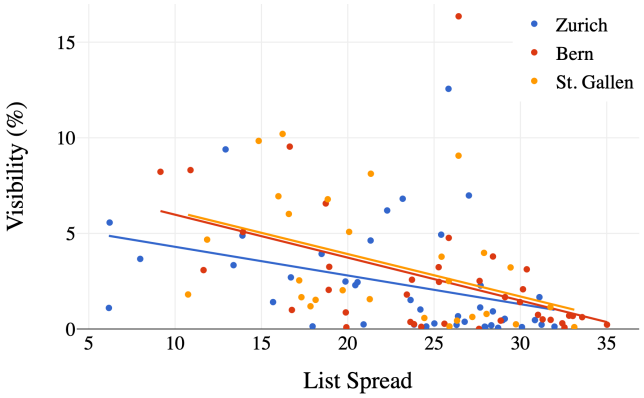


Figure 13: Relationship between the spread and visibility of lists in the states of Zurich, Bern, and St. Gallen. The list spread is defined as the average (over questions) of the answers' standard deviation (across the candidates on the list).

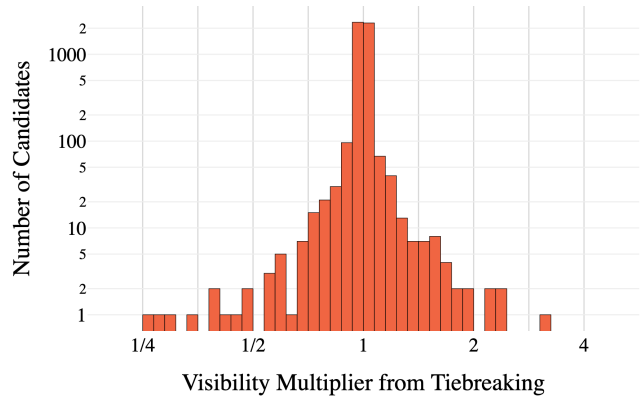


Figure 14: Distribution of relative changes in the visibility of candidates when switching from a fair proportional distribution of ties to Smartvote's tie-breaking method. The x-axis represents the relative visibility change, while the logarithmic y-axis indicates the fraction of candidates affected.

limited number of questions, multiple candidates may end up at the same distance from a given voter. In such cases, Smartvote resolves ties by using the alphabetical order of the candidates' last names. This approach is problematic, as it systematically favors candidates with last names starting with letters from the beginning of the alphabet over those with last names starting with letters from the end. Figure 14 shows how the visibility of candidates differs between a fair distribution of ties and the Smartvote tie-breaking, revealing that certain candidates are significantly affected by this tie-breaking methodology.

Question Ordering (QO)

The order in which questions are presented can introduce biases such as the primacy effect [Asch, 1946], the priming effect [Kuzyakov *et al.*, 2000], and survey fatigue [Jeong *et al.*, 2023]. The primacy effect indicates that early questions tend to be given more weight, while the priming effect suggests that responses may vary depending on the order of questions. Finally, survey fatigue indicates that later questions are often skipped or answered less thoughtfully, as illustrated in Figure 16. These effects could easily be exploited by an adversarial platform designer. To quantitatively understand the potential impact of such a strategy, we perform the following analysis. First, for each $q_t \in Q$, we compute the fraction f_t of answer vectors for which q_t was answered (the dots on Figure 16) and we fit a function $f(t)$ to these values (the line on Figure 16). Then, for each party p, we rearranged all complete answer vectors (i.e., the answer vectors of voters in $V^c = \{v_i \mid \mathbf{w}_{i,t} \neq 0, \forall t\}$) following the order shown in Figure 7. Let $\mathbf{v}_i^p$ be the answer vector of voter v_i rearranged according to p's favorable ordering. Then, for each $v_i \in V^c$, we drop with probability $1 - f(t)$ the answer to question t in $\mathbf{v}_i^p$ (i.e., by setting $\mathbf{w}_{i,t} = 0$). Finally, we compare the visibility gain of each party with respect to the party's popularity resulting from the original ordering, only considering voters in V^c. The results of this analysis are displayed in Figure 17.

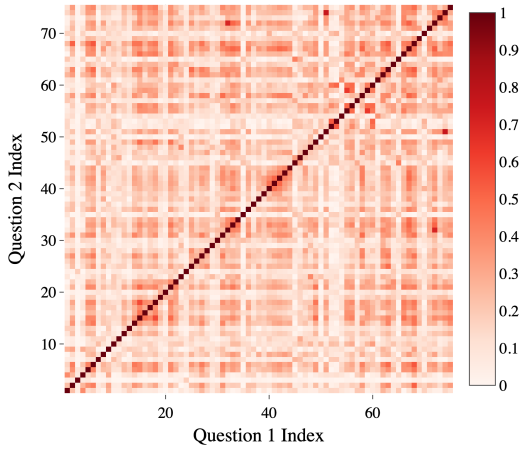


Figure 15: Absolute correlation between question responses in the voter dataset. The heatmap reveals clusters of highly correlated questions (≈ 0.8), suggesting that topics covered by these questions are implicitly weighted more heavily for the recommendation calculation. This provides an advantage to parties that are typically favored by these questions.

D Measuring Robustness (continued)

D.1 Party Bias (BIA)

To quantify the bias of a distance function, we evaluate relative changes in party visibilities when using this distance function compared to the median party visibility across all other evaluated distance functions. BIA1 represents the average of the absolute values of these relative changes in visibility for the 8 largest parties, ensuring that smaller parties' noisy relative changes do not disproportionally affect the metric. BIA2 captures the maximum absolute value of relative change across the 8 largest parties, highlighting the most significant deviation.

We define the visibility of a party p when using a specific distance function d as $\nu(p \mid P, d)$. Let $\mathcal{D}$ represent the set of all evaluated distance functions, which includes L2, L1, Agreement Count, Angular, Mahalanobis, L1 Bonus, and Hybrid. Let P_8 represent the set of the 8 largest parties, which

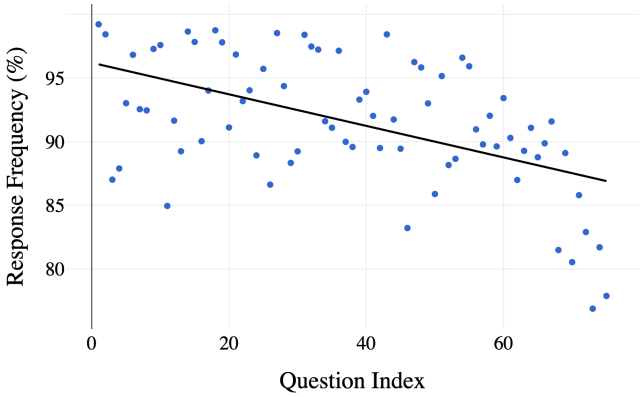


Figure 16: Relationship between response frequency and question position in the deluxe questionnaire (all 75 questions). The black line represents an Ordinary Least Squares (OLS) trend line with equation $f(t) = 96\% - t \cdot 0.12\%$, where t is the index of the question. The plot shows a pattern where later questions get answered less frequently.

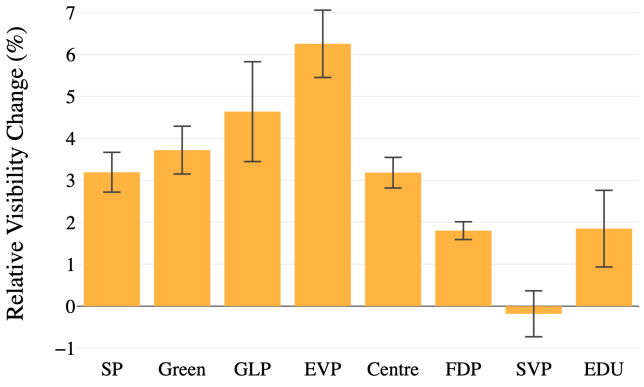


Figure 17: Relative visibility change for each political party resulting from ordering the questionnaire according to the most favorable questions for each party. The bars represent the average visibility change across 10 randomized trials, with error bars indicating the standard deviation. The figure highlights how certain parties, like EVP, could potentially benefit from favorable question orderings.

includes SP, Green, GLP, EVP, Centre, FDP, SVP, and EDU. BIA1 and BIA2 are then defined as follows:

$$\begin{aligned}
\text{Median}(p \mid d) &= \text{med}\{\nu(p \mid P, d') \mid d' \in \mathcal{D} \setminus \{d\}\}, \\
\Delta_{\text{rel}}(p, d) &= \frac{\nu(p \mid P, d) - \text{Median}(p \mid d)}{\text{Median}(p \mid d)}, \\
p_{\max} &= \arg\max_{p \in P_8} |\Delta_{\text{rel}}(p, d)|, \\
\text{BIA1}(d) &= \frac{1}{|P_8|} \sum_{p \in P_8} |\Delta_{\text{rel}}(p, d)|, \\
\text{BIA2}(d) &= \Delta_{\text{rel}}(p_{\max}, d).
\end{aligned}$$

Table 8 provides a detailed overview of how party visibilities change based on the matching methods used for calculating recommendations. When comparing these party visibilities with the references, it's important to note that the Smartvote user base is inherently biased, with only about 85% of candidates participating in Smartvote and only around 20% of eligible voters using the platform, not all of whom disclosed their preferred party. A key indicator of this bias is the notable difference between election vote shares and the distribution of preferred parties among Smartvote users.

D.2 Answer Calibration Metrics (ASC, CP)

To assess the vulnerability of each matching method to an answer calibration strategy, we employed the Answer Strength Correlation (ASC) and Calibration Potential (CP) metrics. These metrics generally align, meaning that if one suggests a matching method is vulnerable, the other typically indicates the same. However, a notable exception is the Mahalanobis matching method. As shown in Table 2, the ASC for Mahalanobis is very low at 0.044, indicating a weak correlation between candidates' answer strengths and their expectation-normalized visibilities. This suggests that an answer calibration strategy would likely be ineffective. In contrast, the CP metric reveals that parties could increase their visibility by over 300% when all candidates used the moderate answering strategy. Although this result may initially appear counterintuitive, further analysis offers an explanation: When analyzing the ASC in scenarios where a party's candidates adopt a moderate answering strategy, the correlation becomes quite negative, with values dropping below -0.6, varying by party. We hypothesize that this is due to the alteration of the precision matrix (inverse covariance matrix), which plays a crucial role in calculating the Mahalanobis distance. This alteration leads to candidates using the neutral answering strategy receiving a significantly increased number of recommendations.

D.3 Distance Metrics

L1

The L1 distance, also known as the City-Block or Manhattan distance (see Figure 18b), is defined as the sum of the absolute differences across all dimensions, making it a simple and intuitive metric:

$$d_{\text{L1}}(\mathbf{v}_i, \mathbf{w}_i, \mathbf{c}_j) = \sum_{t=1}^{N_q} \mathbf{w}_{i,t} \cdot |\mathbf{v}_{i,t} - \mathbf{c}_{j,t}|.$$

Agreement Count

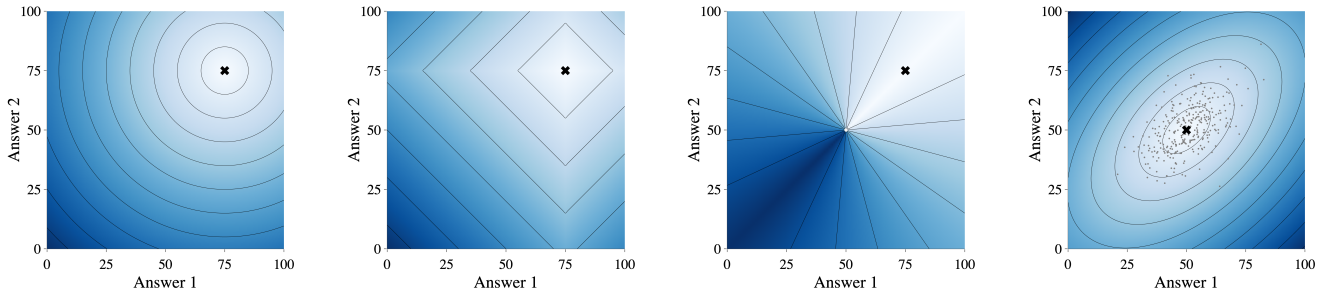
The Agreement Count metric simply counts the agreements between voter and candidate responses. To convert it into a distance metric, we take the negative value of the count and incorporate voter weights.

$$d_{\text{AC}}(\mathbf{v}_i, \mathbf{w}_i, \mathbf{c}_j) = \sum_{t=1}^{N_q} -\mathbf{w}_{i,t} \cdot \delta_{\mathbf{v}_{i,t}\mathbf{c}_{j,t}}$$

where

$$\delta_{xy} = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases}$$

A bias can be added to make the distance positive.



(a) L2 (b) L1 (c) Angular (d) Mahalanobis

Figure 18: Visualizations of the L2, L1, Angular, and Mahalanobis distance metrics in a simplified answer space with two questions. The contour plot illustrates how the distance of a candidate behaves based on their relative position to the voter, who is represented as ✖. For the Mahalanobis distance, the positions of the candidates used for calculating the precision matrix are additionally visualized as small gray dots.

Angular

The Angular distance measures the angle between the dimension-weighted voter and candidate answer vectors. This metric captures the directional similarity between the voter's and candidate's responses, putting less emphasis on the magnitude of their deviation from the neutral point in comparison to other distance metrics. Assuming that the neutral response vector $50 \cdot \mathbf{1}_{N_q}$ serves as the origin of the answer space, it is defined as

$$d_{\text{Angular}}(\mathbf{v}_i, \mathbf{w}_i, \mathbf{c}_j) = \arccos\left(\frac{\tilde{\mathbf{v}}_i^\top \tilde{\mathbf{c}}_j}{\|\tilde{\mathbf{v}}_i\| \cdot \|\tilde{\mathbf{c}}_j\|}\right),$$

where

$$\tilde{\mathbf{v}}_i = \mathbf{w}_i \odot (\mathbf{v}_i - 50), \qquad \tilde{\mathbf{c}}_j = \mathbf{w}_i \odot (\mathbf{c}_j - 50).$$

Here, $\odot$ denotes element-wise multiplication. See Figure 18c for a visualization.

Mahalanobis

The Mahalanobis distance is a covariance-rescaled version of the Euclidean distance. It accounts for the correlations between questions based on the candidate answer vectors and adjusts the importance of each dimension accordingly. Unlike the other distance metrics, the Mahalanobis distance does not take into account the question weights of voters. Figure 18d illustrates how the Mahalanobis distance behaves in a simplified answer space.

$$d_{\text{Mahalanobis}}(\mathbf{v}_i, \mathbf{c}_j) = \sqrt{(\mathbf{v}_i - \mathbf{c}_j)^\top \text{Cov}(\mathbf{C})^{-1} (\mathbf{v}_i - \mathbf{c}_j)},$$

where $\mathbf{C}$ is the candidate answer matrix, with $\mathbf{C}_j$ representing the j-th row of $\mathbf{C}$, corresponding to the answer vector $\mathbf{c}_j$ of candidate j. The covariance matrix $\text{Cov}(\mathbf{C})$ is defined as

$$\text{Cov}(\mathbf{C}) = \frac{1}{N_q - 1}(\mathbf{C} - \bar{\mathbf{C}})^\top (\mathbf{C} - \bar{\mathbf{C}}),$$

with $\bar{\mathbf{C}}$ the matrix containing the column means of the candidate answer matrix $\mathbf{C}$.

$q(v)$ \ $q(c)$	0	25	50	75	100
0	0	125	150	175	200
25	125	75	125	150	175
50	150	125	100	125	150
75	175	150	125	75	125
100	200	175	150	125	0

Table 9: The L1 Bonus distance matrix used by Smartvote until 2010. The table shows how the distance between a voter and candidate is determined for all pairs of voter $q(v)$ and candidate $q(c)$ responses to a question with five options.

L1 Bonus

Distance matrices offer an alternative approach to defining distances by specifying the distance between every possible pair of responses in a matrix format. Until 2010, Smartvote utilized a modified version of the L1 distance, known as L1 with Bonus, which employed such a distance matrix for calculating recommendations. This method ensures that the row and column sums of the distance matrix are equal by adding a bonus when voters and candidates strongly agree on a question. The distance for a voter i and candidate j under this method is given by:

$$d_{\text{L1 Bonus}}(\mathbf{v}_i, \mathbf{w}_i, \mathbf{c}_j) = \sum_{t=1}^{N_q} \mathbf{w}_{i,t} \cdot \mathbf{D}_{\text{L1 Bonus}}(\mathbf{v}_{i,t}, \mathbf{c}_{j,t}),$$

where $\mathbf{D}_{\text{L1 Bonus}}(\mathbf{v}_{i,t}, \mathbf{c}_{j,t})$ is the entry from the distance matrix $\mathbf{D}_{\text{L1 Bonus}}$ corresponding to the voter's response $\mathbf{v}_{i,t}$ and the candidate's response $\mathbf{c}_{j,t}$ to question t. Table 9 illustrates $\mathbf{D}_{\text{L1 Bonus}}$ for questions with five answer options.

Hybrid

The Hybrid distance method, used by the EUVox VAA in 2014 [Mendez, 2012], balances the proximity voting logic model, which generally focuses on how close candidate and voter answer vectors are, and the directional voting logic model, which emphasizes their similarity in direction over the exact strength of the deviations from the neutral point. The Hybrid distance is calculated as the average of the L1

and scalar distance matrices. The distance for a voter i and candidate j under this method is given by

$$d_{\text{Hybrid}}(\mathbf{v}_i, \mathbf{w}_i, \mathbf{c}_j) = \sum_{t=1}^{N_q} \mathbf{w}_{i,t} \cdot \mathbf{D}_{\text{Hybrid}}(\mathbf{v}_{i,t}, \mathbf{c}_{j,t}),$$

where $\mathbf{D}_{\text{Hybrid}}(\mathbf{v}_{i,t}, \mathbf{c}_{j,t})$ is the entry from the Hybrid distance matrix $\mathbf{D}_{\text{Hybrid}}$ corresponding to the voter's response $\mathbf{v}_{i,t}$ and the candidate's response $\mathbf{c}_{j,t}$ to question t. Table 10 illustrates $\mathbf{D}_{\text{Hybrid}}$ for questions with five answer options.

$q(v)$ \ $q(c)$	0	25	50	75	100
0	0	50	100	150	200
25	50	37.5	75	112.5	150
50	100	75	50	75	100
75	150	112.5	75	37.5	50
100	200	150	100	50	0

Table 10: The Hybrid distance matrix used by the EUVox 2014 VAA. The table shows how the distance between a voter and candidate is determined for all pairs of voter $q(v)$ and candidate $q(c)$ responses to a question with five options.

E Mitigation Strategies (continued)

Question Order Randomization (QO).
The question order vulnerability can be effectively mitigated by randomizing the sequence of questions. Importantly, this randomization does not impact the reproducibility property as long as the matching method is permutation invariant. This is generally true for most VAAs, including Smartvote, as observed in Eq. (1) and (2).

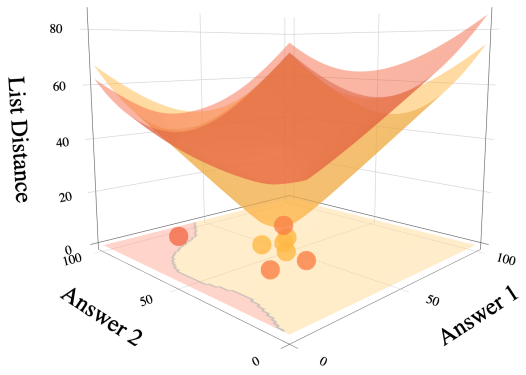


Figure 19: Impact of list centralization on recommendation outcomes. This plot illustrates how a more centralized list, with candidates that have less spread (yellow), is closer to voters in a larger portion of the answer space compared to a list with more spread-out candidates (red).

Fair Tie-Breaking (TB).
To ensure fairness in tie-breaking and prevent systematic bias (for instance based on candidates' last names as in Smartvote), we propose generating random seeds for voters based on their ID (to maintain reproducibility). This introduces variability in tie-breaking, ensuring that ties between candidates are not systematically resolved in favor of the same candidate for all voters.

List Matching Score (LC).
Smartvote ranks lists by computing the average similarity scores between the voter and the list candidates. Figure 19 exposes the mechanism driving the list centralization bias: When candidates on a list have similar answer profiles, the average distance to the voter is minimized, increasing the likelihood of the list being recommended. In contrast, lists with diverse candidate responses incur a higher average distance, reducing their chances of recommendation. Instead of ranking lists by the average similarity scores between the voter and the list candidates, we propose ranking them by the similarity between the voter and the average answer vector of the candidates on the list. This approach would effectively neutralize the impact of list centralization strategies.

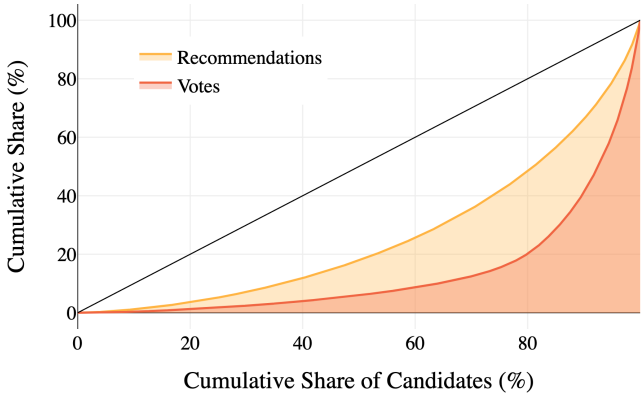


Figure 20: Cumulative share of expectation-normalized visibility and expectation-normalized actual votes over all candidates. The Lorenz curves illustrate that actual votes are distributed less evenly than Smartvote recommendations over all candidates.

F Related Work (continued)

VAA Impact on Democracy.
Post-electoral surveys analyzed by Ladner *et al.* [2012] indicate that 67% of voters are influenced by Smartvote, with 15% of voters stating they had adopted the recommendation in its entirety. Moreover, the candidate recommendations by Smartvote have been shown to be the most decisive factor of electoral choice, ahead of e.g., party membership [Ladner, 2016]. Alvarez *et al.* [2014b] found that 4 out of 5 users of the *EU Profiler* evaluated the VAA as useful. Interestingly, perceived usefulness decreased whenever a user's preferences were less represented. For the same dataset, Alvarez *et al.* [2014a] find that more than 4 in 5 voters were matched with a party they did not initially prefer, with 8% of those voters subsequently changing their vote. Germann and Gemenis [2019] estimate that in 2007 a staggering 58'000 voters were added through the existence of smartvote, making up about 1.2% of the total tally. They compute the cost of each additional voter to be 7.5 USD, exceptionally low compared to costs averaging between 38 and 90 USD in telephone campaigns. Finally,

Munzert and Ramirez-Ruiz [2021] perform a meta-analysis of the effects of VAAs and show that there is significant evidence that VAAs increase voter turnout and influence vote choice.

VAA Development.
Garzia and Marschall [2014] collect an extensive overview of the state-of-the-art on VAA research. Mendez [2017] compares the predictive power of both high- and low-dimensional matching methods, by using the voter's initial preferences as ground truth. Garzia and Marschall [2019] postulates a set of open questions in VAA research, underlining that identifying which matching algorithms are best suited is still an open question. Our study does not give all answers but provides strong evidence that certain matching algorithms should not be used.