# From Bytes to Booth: Exploring Privacy Concerns in Voting Advice Applications

Sampsa Rauti
Department of Computing
University of Turku
Turku, Finland
sjprau@utu.fi

Timi Heino
Department of Computing
University of Turku
Turku, Finland
tdhein@utu.fi

Panu Puhtila
Department of Computing
University of Turku
Turku, Finland
papuht@utu.fi

## Abstract

In this study, we conduct a network traffic analysis of Finnish voting advice applications. We analyze the potential third-party data leaks in 8 voting advice applications first during the 2023 parliamentary election and then during the 2024 presidential election. The goal is to study whether sensitive information, such as individuals' political opinions, leak to third parties. Furthermore, the paper also describes the change in privacy of voting advice applications after the leaks found in 2023 were covered in the media. The findings indicate that media attention and public discourse has had a clear positive impact on the privacy of Finnish voting advice applications, but a few data leaks and privacy concerns still remain. We conclude by discussing how to protect political opinions and democratic processes in increasingly digital society.

## CCS Concepts

• **Security and privacy → Web application security**.

## Keywords

Voting advice applications, third-party services, online privacy

**ACM Reference Format:**
Sampsa Rauti, Timi Heino, and Panu Puhtila. 2024. From Bytes to Booth: Exploring Privacy Concerns in Voting Advice Applications. In *The 8th International Conference on Computer Science and Application Engineering (CSAE 2024), November 28–29, 2024, Shanghai, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3704814.3705969

## 1 Introduction

The digital world has an important role in today's society, also when it comes to politics. People can participate in politics and access information in completely new ways. Political campaigns have also evolved due to digitalization, and voters can now reach candidates and parties on the web. Information about candidates is now available for voter through many different web services, which makes more informed decisions in elections possible. Voting advice applications and political party websites are examples of such resources.



*CSAE 2024, November 28–29, 2024, Shanghai, China*

ACM ISBN 979-8-4007-1809-0/24/11
https://doi.org/10.1145/3704814.3705969

Since voting advice applications (VAAs) were adopted in Europe in the early 2000s, the have become increasingly popular [14]. The idea of these applications is the ask users questions about important topics in politics and match them with candidates who have previously given similar answers. This makes it easier for a voter to decide on a suitable candidate and compare their own opinions to the candidate's views. Therefore, the VAAs aim to expedite the objective analysis of political positions and deliberate decision-making.

VAAs have continuously become more popular in many European countries. In Finland, for instance, nearly half of voters reportedly use VAAs during election season [19]. Modern websites regularly use third-party services in order to track user's behavior and improve user experience. While users benefit from this in many ways, it also often causes serious privacy concerns [16, 23, 24]. The challenge here is that personal data may be sent to third parties without website maintainers realizing, which is a threat to confidentiality and privacy. This means that for VAAs specifically, there is a risk of sharing sensitive personal data on users' political opinions.

Privacy can be considered as an important condition for democracy [7]. Democracy needs citizens with critical thinking, willingness to consider diverse viewpoints, and courage to stand up for their beliefs. Privacy is required to build these qualities, because it supports the development of informed and empowered citizens who create a healthy democracy. It is also worth noting the GDPR prohibits the processing of sensitive personal data – including political opinions – without explicit consent, unless there is a legal basis for processing it.

The existing research on VAAs primarily focuses on their impact and algorithms. There is not much research on online privacy considerations, a gap this paper aims to fill. We present a network traffic analysis of 8 Finnish VAAs, first during the 2023 parliamentary election, and then during the 2024 presidential election. The goal is to study whether sensitive political opinions leak to third parties. Our findings reveal that VAAs had significant privacy leaks in 2023, but the situation has improved in 2024 after the maintainers of the VAAs were informed of our findings. The current study is a continuation to our earlier study on VAAs in 2023 [10]. It gives insights into the positive influence of public discourse and media coverage on the privacy of VAAs.

The subsequent sections of this paper detail the selection and analysis of the studied applications, present our findings on data leaks in VAAs, discuss implications from political and technical perspectives, and conclude the study.

## 2 Related Work

On modern websites, it is commonplace that website users' personal data is sent to these third parties integrated in the website, such as web analytics [1]. Time and time again, this has been demonstrated to be the case even in many web services and public sector websites that are essential in nature and handle sensitive data [17, 20]. Privacy of VAAs [10] is only one example of such critical topic areas.

Most of the studies published on VAAs seem to revolve around the impact, algorithms, and design of these applications [2, 8, 13, 15, 18, 22]. While this subject is important, it seems that the privacy concerns associated with VAAs are largely neglected from this field of research. In their study, Kaskina et al. [12] propose a privacy framework used to protect the information on users' political affiliations from other users of VAAs. However, in their study a VAA is considered as a social platform where users can discuss political matters with each other.

Therefore, these studies are very different from our topic of analyzing and preventing leaks of sensitive political data to third parties such as web analytics providers. We specifically address the privacy challenges of VAAs from a software engineering point of view. The current study addresses the research gap concerning privacy issues caused by third-party analytics in VAAs and by using network traffic analysis, reveals leaks of sensitive data concerning political opinions in these applications.

## 3 Study Setting and Method

The current study examines the privacy of 8 widely-used Finnish VAAs. In Finland, almost all VAAs are provided by media outlets. We have selected VAAs offered by major media outlets in Finland and those with high rankings in Google search results. Also, only VAAs available both in parliamentary election in 2023 and presidential election in 2024 are included in the current study. The providers of VAAs are anonymously outlined in Table 1. Our aim was not to encompass all Finnish VAAs. Rather, we want to highlight the privacy risks associated with third-party services with a smaller representative snapshot of popular VAAs.

In our experiment, we tested the selected VAAs manually. This involved selecting an electoral district or municipality, answering the application's questionnaire, and finally viewing the recommended candidate's page. In this experiment, all cookies were accepted. Network traffic was recorded using Google Chrome's Developer Tools with caching disabled. We filtered network requests to only cover traffic to third parties. The recorded data was saved for further examination. In analyzing the collected data, we focused on third-party web requests' payloads. We read the payloads carefully to see whether they contained potential political opinions, for example in leaked URLs.

We also examined the privacy policies of the websites hosting the studied VAAs. If an application had its own privacy policy, that document was analyzed. We identified any indications of sharing data concerning political opinions with third parties. We also looked into whether the privacy policies clearly specified the third parties who, based on our earlier analysis of network traffic, were shown to be recipients of potentially sensitive personal data from the website.

## 4 Results

While the VAAs were not found to leak the user's answers to the political questions, VAAs often leaked the candidates the user displayed after getting the list of best fitting candidates at the end of the questionnaire. When the user clicks on a candidate on this list, a new web page inside the VAA is usually opened, and this page's URL address leaks to third parties through analytics services. In many cases, these kinds of URL leaks may reveal the top candidates the VAA is recommending to the user. Consequently, information on the user's political opinions can be leaked.

Figure 1 illustrates this situation. After answering questions in VAA, the user gets the best fitting candidates as results. From here, candidate pages can be opened, but at the same time, a third-party analytics script on the page leaks the data on the candidate and the user's identity (e.g. IP address) to the third party. After this, there is a possibility that the sensitive personal data is given or sold to some fourth party, or disclosed in a data breach.

Table 2 shows the URL leaks found in the studied VAAs in spring 2023 and spring 2024. We can see that in spring 2023, 6 out of 8 studied VAAs leaked the URLs of opened candidate pages. What is more, the 6 VAAs each had at least two third-party services that collected candidate URLs. VAA7 leaked this information to 6 different third-party services.

Not surprisingly, Google Analytics was the most common third-party service with 4 instances in our experiments in spring 2023. This means that half of the studied websites had Google receiving information about accessed candidate pages. UserReport (a user feedback tool) and Giosg (a live chat service) had 3 instances. While these are not necessarily data-gathering analytics services per se, they are still third parties receiving potentially sensitive data. Meta Pixel (a tracking pixel used for advertising and analytics) and Chart-Beat (an analytics tool to improve audience engagement) both had 2 instances.

Other miscellaneous third-party services only appeared once, all in VAA7. These services included Rubiconproject (an advertising exchange platform), AppNexus (an advertising technology company) Adform (a digital advertising platform), Smartadserver (an ad serving platform), and OneTag (a tag management and analytics platform).

The situation is noticeably better in 2024. The number of total data leaks in the 8 VAAs has gone down from 19 to 5, when we count a leaking candidate page URL once for each receiving third party (for example, Google often receives the same URL in several HTTP requests). This means that between spring 2023 and spring 2024, there was a decrease in the average number of URL leaks per VAA, dropping from 2.38 to 0.63.

Most media outlets providing the studied VAAs have either completely removed all third parties (VAA2, VAA3, VAA4) or the number of third-parties has been toned down (VAA5, VAA6, VAA7). VAA1 is a peculiar exception to this rule, as the maintainers have added two third-parties since spring 2023. VAA8, on the other hand, is the only studied VAA that had no leaks in either of our analyses. Curiously, VAA7 does not contain any of the 6 third parties it had before, but has now added one completely new third-party service (New Relic, an application performance monitoring platform).

**Table 1: The providers of the studied voting advice applications.**

| | |
|---|---|
| VAA1 | Broadcasting company |
| VAA2 | Finnish daily newspaper |
| VAA3 | Finnish tabloid newspaper |
| VAA4 | Regional newspaper |
| VAA5 | Finnish tabloid newspaper |
| VAA6 | Regional newspaper |
| VAA7 | Regional newspaper |
| VAA8 | Broadcasting company |

**Table 2: The data leaks found in voting advice applications in spring 2023 and spring 2024.**

| Application | Spring 2023 | Spring 2024 |
|---|---|---|
| VAA1 | | Chartbeat, UserReport |
| VAA2 | Google, UserReport, Giosg | |
| VAA3 | Google, UserReport, Giosg | |
| VAA4 | Google, UserReport, Giosg | |
| VAA5 | ChartBeat, Google | Google |
| VAA6 | ChartBeat, Meta | ChartBeat |
| VAA7 | Smartadserver, Rubiconproject, Adform, OneTag, Meta, AppNexus | New Relic |
| VAA8 | | |

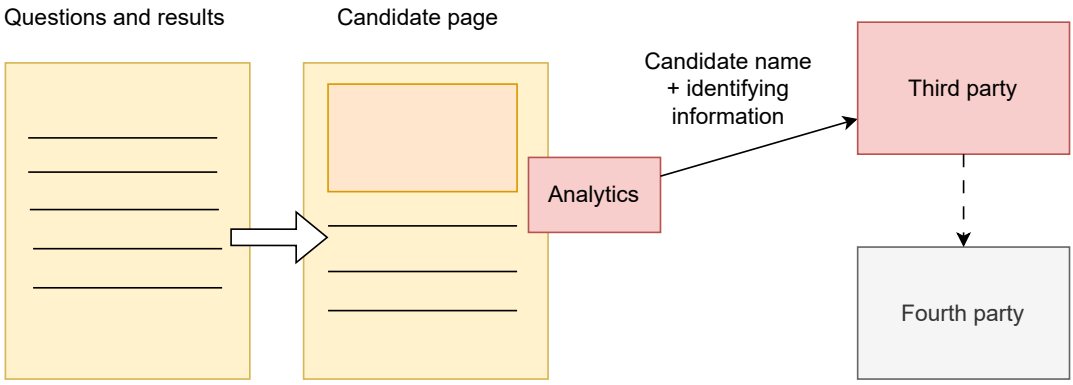


**Figure 1: Data leaks in VAAs.**

What is very noteworthy is that Google and Meta are absent in 2024 with the exception of VAA5, where Google Analytics still persists as the only third party. The exclusion of Google and Meta's services from VAAs is important, considering the capability of these technology giants to link identifiable personal data (e.g. IP addresses) to the names of individual users. This omission can be seen as a positive step towards protecting user privacy in VAAs. When a user uses the same device to log into services run by Google or Meta, as well as to use a VAA, for example, their real name can often be discerned based on their account name, potentially linking them to sensitive political data.

In general, the privacy improvements of VAAs also illustrate the positive impact and constructive influence that public attention and media coverage can have on data protection of online services. Our initial findings on several data leaks in Finnish VAAs received media attention in Suomen Kuvalehti[1], a weekly Finnish news magazine, in summer 2023. The data leaks were also reported by several other new outlets, and a Member of Parliament of Finland demanded an

[1]https://suomenkuvalehti.fi/politiikka-ja-talous/useiden-vaalikoneiden-tietosuojassa-aukkoja-kayttajan-tarkastelemia-tietoja-vuotanut-ulkopuolisille-tahoille-kuten-googlelle-metalle-ja-tiktokille/

improvement to the data protection in VAAs, noting that "information regarding democratic participation must be protected in all circumstances."[2]

Lastly, we also inspected the privacy policy documents of the studied VAAs to see how they justify the use of third-party tools that collect information on visited URLs. The privacy policies mentioned reasons such as measuring performance and website functionality, tracking browser sessions and site navigation, and gathering data and insights about the behavior. Although these may be valid reasons for data collection on many websites, it is hard to see why this should also be done inside a VAA. If this really is necessary, at least the collected URLs should be anonymized so that they cannot be connected to candidates, or analytics tools storing data locally should be employed. It is clear that even if data collection needs to be carried out, data that links a certain candidate or political party to a specific user should never be sent to external data collectors. None of the studied privacy policies directly mentioned that data on the displayed candidates leaks to third parties. The users were not adequately informed of the data collection taking place.

## 5 Discussion

Many of the studied VAAs were found to leak their results, i.e. the information on candidates that the user had been displayed by the application based on their answers, to third parties. Consequently, this can be argued to compromise the privacy of users' political opinions. The definition of "political opinion" in the GDPR is unfortunately vague. It is not clear if it refers to a political ideology (e.g. liberal, conservative), partisanship, or even more specific stances on various policies (such as economical or environmental matters) [3]. However, it is still quite clear that leaking the candidates a user has displayed in a VAA to third parties cannot be a good idea.

Privacy and secrecy of political opinions and affiliations is a fundamental principle in the currently prevailing representative democracies, and is often enforced by the letter of law. Reasons for this are based on historical examples of what could happen when the privacy of political opinions is breached. For instance, if the supporters of political opposition can be identified and based on data leaks, they may be persecuted by those in power [4]. While such oppressive actions may feel distant in the modern day, it should be remembered that they were quite common even in the western world just one human lifetime ago, and are still happening in less democratic parts of the world. If voters feel that their political opinions are no longer private, they may be pressured to not express their beliefs in elections either. This can distort the election outcome and have grave and adverse effects on democracy.

If the scenarios considered above seem unlikely dystopian, as they hopefully should, one should remember that the potential for political oppression is hardly the only concern when evaluating the consequences of these leakages. Different types of political data leaked by the VAAs, including ideology and party affiliation, can be used in politically profiling the user. In the past decade, an analytics company named Cambridge Analytica harvested psychological profiles and personal data from millions of Facebook users without their consent and used the collected information for targeted political advertising in the USA presidential campaigns of 2016. [6, 11, 21] In the light of this scandal, and research that has shown how this kind of political profiling in general is increasing [3, 5], indicating the privacy of VAAs is important in order to mitigate this kind of political microtargeting online. The leaking of political opinions by these applications can lead to political profiling, and subsequently to voter manipulation, which has the potential to jeopardize the fairness and freedom of elections.

Based on our findings, it appears that many actors among the media corporations make no distinction to treat VAAs separately from their main websites, which reflects on the use of web analytics as well. While the case for the necessity of the web analytics can be argued for on some parts of media websites, the sensitivity of data concerning political opinions should not be overlooked by including these third-party analytics services in the subpages of VAAs as well. This lack of attention to the specific political context of the VAAs can be seen as one of the contributing factors to the unintentional leaks of sensitive data.

To help protect against these kinds of breaches of the user privacy, we strongly recommend the web application developers and the maintainers of these services to adopt certain practices which should mitigate this phenomenon. First, all third-party web analytics tools should be reviewed and their use should be justified. Moreover, a thorough network traffic analysis should be conducted in order to study whether they leak the user data to third parties, similar to what we have done in this research. Since this kind of traffic analysis requires neither special expertise nor large temporal investment, there are hardly any good reasons to not undertake it. Secondly, analytics services that are found to leak data should be removed, and if it is decreed absolutely necessary to use analytics tools in the first place, solutions that are possible to deploy locally should be favored, as this mitigates the risk of the data falling into wrong hands [9]. Ultimately, serious consideration should be given to the reasons for deploying any kinds of web analytics tools in the first place within VAAs because of the sensitive nature of these applications.

Our study also discovered deficiencies in the privacy policies presented by the VAA providers – specifically when it came to the clarity and transparency of these documents. As noted previously, the need for privacy is much more stringent in the case of VAAs than in the media websites in general. This is also why the website maintainers and data protection officers should consider composing separate privacy policies for these applications. These documents should be transparent and comprehensible enough to adequately inform the users of any data collections activity taking place. Due to the political nature of these applications, there is a greater need to provide justifications and disclosure of this kind of data processing.

On the brighter side, our results highlight the positive impact public discourse and media attention has had on the overall privacy situation of the studied VAAs. On a less positive note, it is discouraging to witness the continued use of certain leaky web analytics tools still in some of the VAAs, such as Google Analytics and New Relic even after all of the public attention and discourse. At this point all of the VAAs' providers must be cognizant of the potential threats to the user privacy these analytics services pose, and thus

[2] https://www.sttinfo.fi/tiedote/69991701/kansanedustaja-peltonen-vaatii-ryhtiliiketta-vaalikoneiden-tietosuojan-parantamiseksi:-%22demokraattista-osallistumista-koskevan-tiedon-on-oltava-kaikissa-tilanteissa-suojattua%22?publisherId=66784162

their continued usage can not be justified by claiming ignorance of this issue. Therefore, all the remaining VAA providers with leaks in their applications should carefully fix the privacy issues.

## 6 Conclusions

We have presented an overview of data leaks in Finnish VAAs, and explored how their privacy improved after the initial leaks were made public. While the dataset does not cover all Finnish VAAs, the finding that several studied applications leaked sensitive political and some even continued to do so after the leaks were made public is concerning. Our findings serve as a reminder to service providers and software developers for keeping the users' personal data safe and informing users appropriately of the data processing practices and the third parties receiving data. In web applications that deal with sensitive political data, it is challenging to find legitimate reasons for using third-party web analytics at all. In future, a further comparison between VAAs in Finland and other countries should also be conducted in order to get a better understanding of how privacy situation varies in different regions.

## Acknowledgments

This research has been funded by Academy of Finland project 327397, IDA – Intimacy in Data-Driven Culture.

## References

[1] Gunes Acar, Steven Englehardt, and Arvind Narayanan. 2020. No boundaries: data exfiltration by third parties embedded on web pages. *Proceedings on Privacy Enhancing Technologies* 2020, 4 (2020), 220–238.

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

[3] Colin Bennett. 2013. The politics of privacy and the privacy of politics: Parties, elections and voter surveillance in Western democracies. *Elections and Voter Surveillance in Western Democracies (June 15, 2013)* (2013).

[4] Colin Bennett and Smith Oduro Marfo. 2019. Privacy, voter surveillance and democratic engagement: challenges for data protection authorities. In *International Conference of Data Protection and Privacy Commissioners (ICDPPC)*.

[5] Colin J. Bennett. 2016. Voter databases, micro-targeting, and data protection law: can political parties campaign in Europe as they do in North America? *International Data Privacy Law* 6, 4 (12 2016), 261–275. https://doi.org/10.1093/idpl/ipw021 arXiv:https://academic.oup.com/idpl/article-pdf/6/4/261/9598014/ipw021.pdf

[6] Hal Berghel. 2018. Malice domestic: The Cambridge analytica dystopia. *Computer* 51, 05 (2018), 84–89.

[7] Volker Boehme-Nessler. 2016. Privacy: a matter of democracy. Why democracy needs privacy and data protection. *International Data Privacy Law* 6, 3 (2016), 222–229.

[8] 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.

[9] Jonas Gamalielsson, Björn Lundell, Simon Butler, Christoffer Brax, Tomas Persson, Anders Mattsson, Tomas Gustavsson, Jonas Feist, and Erik Lönroth. 2021. Towards open government through open source software for web analytics: The case of Matomo. *JeDEM-eJournal of eDemocracy and Open Government* 13, 2 (2021), 133–153.

[10] Timi Heino, Sampsa Rauti, Samuli Laato, Robin Carlsson, and Ville Leppänen. 2024. Leaky Democracy: Third Parties in Voting Advice Applications. Accepted to the 8th International Conference on Smart Trends in Computing and Communications.

[11] Harshil Kanakia, Giridhar Shenoy, and Jimit Shah. 2019. Cambridge Analytica–a case study. *Indian Journal of Science and Technology* 12, 29 (2019), 1–5.

[12] Aigul Kaskina and Andreas Meier. 2016. Integrating privacy and trust in voting advice applications. In *2016 Third International Conference on eDemocracy & eGovernment (ICEDEG)*. IEEE, 20–25.

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

[14] Valérie-Anne Mahéo. 2017. Information campaigns and (under) privileged citizens: An experiment on the differential effects of a voting advice application. *Political Communication* 34, 4 (2017), 511–529.

[15] Simon Otjes and Tom Louwerse. 2014. Spatial models in voting advice applications. *Electoral Studies* 36 (2014), 263–271.

[16] Denise Quintel and Robert Wilson. 2020. Analytics and privacy. *Information Technology and Libraries* 39, 3 (2020).

[17] Sampsa Rauti, Robin Carlsson, Sini Mickelsson, Tuomas Mäkilä, Timi Heino, Elina Pirjatanniemi, and Ville Leppänen. 2024. Analyzing third-party data leaks on online pharmacy websites. *Health and Technology* 14 (2024), 1–18.

[18] Martin Schultze. 2014. Effects of voting advice applications (VAAs) on political knowledge about party positions. *Policy & Internet* 6, 1 (2014), 46–68.

[19] Statistics Finland. 2019. Puolet äänioikeutetuista käytti vaalikonetta ennen eduskuntavaaleja. https://www.stat.fi/til/sutivi/2019/sutivi_2019_2019-11-07_kat_002_fi.html.

[20] Nik Thompson, Ravi Ravindran, and Salvatore Nicosia. 2015. Government data does not mean data governance: Lessons learned from a public sector application audit. *Government information quarterly* 32, 3 (2015), 316–322.

[21] Ikhlaq ur Rehman. 2019. Facebook-Cambridge Analytica data harvesting: What you need to know. *Library Philosophy and Practice* (2019), 1–11.

[22] Markus Wagner and Outi Ruusuvirta. 2009. Faulty recommendations? Party positions in online voting advice applications. *Party Positions in Online Voting Advice Applications* (2009).

[23] Tim Wambach and Katharina Bräunlich. 2017. The Evolution of Third-Party Web Tracking. In *Information Systems Security and Privacy*, Olivier Camp, Steven Furnell, and Paolo Mori (Eds.). Springer International Publishing, 130–147.

[24] Alexander R Zheutlin, Joshua D Niforatos, and Jeremy B Sussman. 2021. Data-tracking on government, non-profit, and commercial health-related websites. *Journal of general internal medicine* (2021), 1–3.