A Dataset of Generalizable Election-Related Questions for AI Tools Compiled from Leading Global Democracies

Kausik Lakkaraju, Sara Elizabeth Jones, Bharath Muppasani, Biplav Srivastava

AI Institute, University of South Carolina, Columbia, USA

Abstract

Elections represent the unified voice of citizens in a country, empowering the voters to shape the government and the future. Understanding the concerns of different stakeholders involved in voting is pivotal and it fosters better engagement between these stakeholders. We collected 181 questions concerning different stakeholders - voters, candidates, journalists, and election administration - across different geopolitical contexts. We extracted 79 questions that can be generalized and adapted to various regions. With 57 parameterizable questions, we hypothesize that addressing these concerns could promote voter participation. We observed that there is a significant disparity between the questions we collected and the questions provided in the official FAQs. We categorize and identify the gaps, and pave the way for future research in this area, offering a way to enhance the current electoral processes.

Introduction

Elections are crucial for a society and it serves as a collective voice for citizens in shaping the government thereby, shaping their future. Voters have the power to enact change making their participation in the electoral process pivotal. Hence, it is paramount for the election administration and the candidates contesting the elections to know the concerns of the voters to improve the current polling methods to engage with the voters effectively. Similarly, the potential candidates themselves might have some questions concerning the election procedure and their eligibility to contest. Moreover, the Electoral Management Board (EMB) needs to foster close cooperation with various stakeholders to ensure their support for its policies and to strengthen their confidence in EMB's performance (AKPAN 2022). To organize the elections effectively, it is very important to know the various concerns different stakeholders might have.

We collected questions that different stakeholders (voters, candidates, journalists, and election administration) might have concerning the elections. A total of 181 questions were collected from various sources out of which 79 can be generalized or applicable to other regions in the world. 57 of these questions have parameters that can be changed to augment

the data or paraphrase the questions. Our contributions are that:

- We provide a common set of parameterized election questions.
- Furthermore, we hypothesize that if they are answered they will help drive voter participation in any given region.
- We categorize these questions into various categories (based on the stakeholders involved) and various subcategories with identified parameters for each so that they can be generalized to other regions.
- We identify further gaps and pave the way for future work in this area.

Related Work

In political sciences, there is a lot of literature on election poll performance (Kenett, Pfeffermann, and Steinberg 2018; Chen and Körtner 2022; Kennedy et al. 2018; Hillygus 2011) and internet or e-voting, or more specifically how internet voting can improve the polling process and make it more accessible to everyone (Benabdallah et al. 2022; Ehin et al. 2022). Notably, recent literature assessed the impact of social media posts (eg., posts on X (formerly called Twitter)) on voting results (Kofi Frimpong et al. 2022; Carney 2022).

Despite the vast research in these areas, there is still a gap in realizing the concerns of various stakeholders such as voters, candidates, journalists, and election administration involved in the electoral process. This gap paves the way for an opportunity to explore electoral concerns across different geopolitical contexts.

Dataset Overview: Collection, Description, and Applications

In this section, we describe the data collection process and potential applications of the data.

Data Collection

We looked for each country's 'Most googled/asked election-related questions'. Table 1 details the number of collected questions per country, the corresponding country name, and the sources from which we collected these questions. From the set of questions collected for each country, we extracted

the questions that can be generalized and applied to different regions or periods. In each question, we identified adaptable parameters (or variables), enabling the query to be paraphrased in different forms by adjusting these parameters. For instance, one of the questions we came across for Ireland, "What happens if I cannot attend the polling station?" underwent paraphrasing and parameter identification, resulting in the modified query: "What happens if I cannot come to the polling station (in $<region>$)?". Here, $<region>$ serves as an optional parameter, allowing for customization based on the voter's region. This parameter can also be omitted, providing flexibility in the question formulation.

Data Description

Each of the datasets created for different countries has the following columns: Query Number, Question-General, Parameters, Question-specific, and Source. The final dataset that was created after extracting the generalized questions from each country has all the columns present in the individual country datasets. In addition to those columns, there are three additional columns: Question Category (stakeholder), Question Sub-category, Region(s) Applicable - Assumption as well as from source, and Reference to Country-Specific Tab. 'Question-General' refers to the generalized version of 'Question-Specific' with the identified parameters. We considered three question categories based on the stakeholders involved: voters, candidates, and Election Administration / Journalists. Within each of these categories, we considered different sub-categories ('Question Sub-category'): 'What', 'How', 'When', and 'Who' questions. The 'Region(s) Applicable' column has the list of all the region(s) to which the corresponding query is applicable. Table 2 shows some sample queries from each category (stakeholders). The parameters present in each query, the applicable region(s), and the country data from which the question is extracted are also provided along with the query number. The complete list of queries along with a Python notebook for data analysis can be found in https://github.com/ai4society/election-dataset

Potential Applications

The dataset we created can have several applications:

1. **Chatbot development:** The data can be used to train the chatbots that can answer diverse election-related queries accurately.

2. **Understanding stakeholder concerns:** This dataset can be analyzed by interested parties to understand the concerns and queries different stakeholders have.

3. **Policy formulation:** This dataset can be used by policymakers to better align their agendas with public interests.

4. **Understanding intercultural contrasts:** Such a diverse dataset helps us understand the differences between the election-related queries each stakeholder from different countries may have.

5. **Improve polling procedure:** The concerns of the voters and candidates can help the election administration to improve the polling procedure they are following.

6. **Candidate-voter engagement:** The candidates contesting the elections can use the dataset to get insights into the voter concerns so that they can address the issues the voters are facing more effectively.

Discussion

While our study shed light on stakeholder concerns, there are still critical gaps that need to be addressed. Notably, the data we collected for India and Indonesia, based on most googled questions, presents valuable insights, reflecting the actual concerns the voters have regarding the voting procedure. However, similar comprehensive studies were not found for many other countries, indicating a gap that needs to be addressed. The following are some more critical gaps in this area that need to be addressed:

- **Disparity between official FAQs and independent surveys:** A discrepancy still exists between the election FAQs provided on official websites and the queries that resulted from different independent surveys that were conducted. This gap can be filled if the government takes the initiative to conduct such surveys to align the official resources with real-time concerns.

- **Lack of different stakeholder perspectives:** Existing studies explore the voter concerns, neglecting the concerns of election administration or potential candidates who would like to contest elections. These gaps can be addressed by conducting inclusive surveys that take into consideration various stakeholders' voices.

- **Poor and non-inclusive data:** In (Muppasani et al. 2023), the authors trained a chatbot on the FAQs data provided on the official election websites of South Carolina and Mississippi. Such efforts when combined with more inclusive and vast data help researchers build better chatbots or AI tools to promote voter participation.

Conclusion

Elections serve as the primary platform through which citizens across multiple countries and regions can articulate their voices and opinions, thereby weaving each constituent into the fabric of their respective governments. Recognizing that governments are intended to be reflections of the collective needs, desires, and opinions of the voting populace, the expansion of the voter pool in a given region or country is deemed desirable in the context of election processes. As such, our objective is to contribute to a global increase in voter participation by consolidating pertinent queries on election processes. To accomplish this, we gathered a total of 181 specific queries and 79 generalized queries, of which 59 may be parameterized for tailoring to a specific demographic or region. We hold the belief that by centralizing common election questions, we can alleviate some of the confusion surrounding election processes. Moreover, we have identified a notable gap in Frequently Asked Questions (FAQs), particularly in addressing common queries that electoral candidates or government administrations may themselves pose. Recognizing the importance of providing a comprehensive query list, we have structured our dataset to

Country	No. of Queries	Sources
India (IND)	11	https://www.vice.com/en/article/gy4x3m/10-most-googled-questions-about-indian-election
United States of America (USA)	39	https://www.usa.gov/how-to-vote; https://www.cnn.com/interactive/2020/politics/voting-questions-answers/
Indonesia (IDN)	6	https://www.vice.com/en/article/neawv8/most-googled-questions-about-indonesian-election-2019
Ireland (IRE)	11	https://www.citizensinformation.ie/en/government-in-ireland/elections-and-referenda/types-of-elections-and-referendums/
United Kingdom (UK)	7	https://fullfact.org/election-2019/election-questions-answered/
Estonia (EST)	50	https://www.valimised.ee/en/internet-voting/frequently-asked-questions/
South Africa (SA)	47	https://www.elections.org.za/content/for-voters/FAQ-For-Voters/
Canada (CA)	11	https://www.tvo.org/article/your-federal-election-questions-answered

Table 1: The number of queries collected from each country with the corresponding country name and source.

Stakeholders	Sub-category	Question - general	Parameter(s)	Region(s) Applicable	Reference to country-specific tab
Voter	What	What counts as a "fair" election?	None	Any	IDN-Q6
	How	How would the voters of advanced age vote as actively as the young?	None	Any	EST-Q36
	When	When can the <region> begin to count absentee & mail-in ballots?	<region>	Any	USA-Q35
	Who	Who can run for <Head-of-state> of <region>?	<Head-of-state>, <region>	<Head-of-state>, <region> = King | Queen, Canada, President, USA, India, South Africa, Indonesia	IRE-Q2
Candidate	What	Do <poll officers> have any guidelines about conditions when a candidate can request a recount in a close result?	<poll officers>	Any	UK-Q3
	How	How is it ensured that each voter votes themselves (independently) (in <region>)?	<region>	Any	EST-Q20, EST-Q43
Election Administration / Journalist questions	What	What goes into the cost of an election in <region>?	<region>	Any	CA-Q7
	How	How does <region> handle voter abstention?	<region>	Any	IDN-Q3
	Who	Who are Millennial voters?	None	Any	IDN-Q1

Table 2: Sample generalized queries extracted from different sources with corresponding stakeholders, sub-categories, parameters, regions applicable, and reference to the country-specific tab. The country codes for each country are shown in Table 1.

include these perspectives. This positions our dataset to provide a holistic view of election processes by ensuring that it not only encompasses voter queries but also those pertinent to electoral candidates and government administrations. This approach serves to enhance the depth of our dataset, as well as provide a more nuanced examination of the electoral landscape. In light of this, we explore various applications for our dataset, such as developing an election-related chatbot, conducting research on stakeholder concerns and intercultural differences using our dataset, and enhancing polling procedures to increase accessibility for all constituents.

References

AKPAN, E. E. 2022. The Roles of Stakeholders Relationship in the Electoral Process. *International Journal of Research Education and Management Science*, 3(3): 9–21.

Benabdallah, A.; Audras, A.; Coudert, L.; El Madhoun, N.; and Badra, M. 2022. Analysis of blockchain solutions for E-voting: A systematic literature review. *IEEE Access*.

Carney, K. 2022. The effect of social media on voters: experimental evidence from an Indian election. *Job Market Paper*, 1–44.

Chen, S.; and Körtner, J. 2022. Did "Trumpists" perform worse in US Senate election polls? *Significance*, 19(3): 28–31.

Ehin, P.; Solvak, M.; Willemson, J.; and Vinkel, P. 2022. Internet voting in Estonia 2005–2019: Evidence from eleven elections. *Government Information Quarterly*, 39(4): 101718.

Hillygus, D. S. 2011. The evolution of election polling in the United States. *Public opinion quarterly*, 75(5): 962–981.

Kenett, R. S.; Pfeffermann, D.; and Steinberg, D. M. 2018. Election Polls—A Survey, A Critique, and Proposals. *Annual Review of Statistics and Its Application*, 5(1): 1–24.

Kennedy, C.; Blumenthal, M.; Clement, S.; Clinton, J. D.; Durand, C.; Franklin, C.; McGeeney, K.; Miringoff, L.; Olson, K.; Rivers, D.; et al. 2018. An evaluation of the 2016 election polls in the United States. *Public Opinion Quarterly*, 82(1): 1–33.

Kofi Frimpong, A. N.; Li, P.; Nyame, G.; and Hossin, M. A. 2022. The impact of social media political activists on voting patterns. *Political Behavior*, 1–54.

Muppasani, B.; Pallagani, V.; Lakkaraju, K.; Lei, S.; Srivastava, B.; Robertson, B.; Hickerson, A.; and Narayanan, V. 2023. On safe and usable chatbots for promoting voter participation. *AI Magazine*, 44(3): 240–247.