Ballotbot: Can Chatbots Strengthen Direct Democracy?

Pre-Analisys Plan

Elliott Ash, Sergio Galletta, and Giacomo Opocher

October 10, 2024

1 Introduction

In this study, we investigate the impact of new information technologies on American voters' political knowledge through a two-wave online survey experiment. We developed BallotBot, an interactive chatbot powered by state-of-the-art large language models (LLMs) trained to provide official information about statewide ballot measures. A random sample of California voters is assigned to either use BallotBot or the digital version of the official voter guide to answer questions about upcoming ballot initiatives. We want to compare participants' short- and medium-term political knowledge, as well as the perceived and objective costs of acquiring that knowledge, across the two experimental groups. Additionally, we collect data on voting intentions, turnout behavior, and the reasoning behind participants' vote choices to assess how these factors differ between the groups.

2 Experimental Design

2.1 Experiment Overview

The experiment takes the form of a two-wave survey experiment where participants receive information about upcoming ballots. The treatment varies the method through which information about the ballot initiatives is communicated—specifically, through an interactive chatbot or a static voter guide. Following the treatment, participants will answer a series of questions related to the ballot content. Responses will be collected via Qualtrics, and participants (around 2,400 Californian residents) will be recruited in October 2024 through Prolific.

2.2 Experiment Procedure

In the first survey experiment, we start by collecting baseline information from all respondents (i.e., demographic, baseline attitudes, and attention check questions). Next, participants will be randomly assigned to one of three upcoming ballot initiatives.[1] At this point, participants will be encouraged

[1]We select Proposition 32 (raises minimum wage), Proposition 33 (expands local governments' authority to enact rent control on residential property), and Proposition 36 (allows felony charges and increases sentences for certain drug and theft crimes).

to voluntarily submit up to three questions they would like answered to help them make an informed voting decision. Following this, within each subgroup, participants will be randomly assigned to one of two treatment groups, each having access to information through a different communication method:

- **T1 Group:** Participants in this group will be provided access to the digital version of the official voter guide that is distributed to households in California.
- **T2 Group:** Participants in this group will interact with a chatbot (via a web application) that is trained on the same official information as the voter guide. The chatbot allows users to interactively ask questions specific to the referendum.

Next, we will ask the participants to answer a set of questions:

- Five compulsory closed-ended questions about the content of the assigned proposition. We generated a pool of 12 questions for each proposition, equally distributed over four difficulty levels. Respondents receive one randomly selected very easy, easy, hard, and two very hard questions, in random order. Participants may use either the chatbot (T2) or the voter guide (T1), depending on their assigned group, to find the information to answer the questions. In addition to providing their answers, participants will be asked to indicate their confidence level for each response.
- Half of the participants will be informed of the actual accuracy of their responses, while the other half will not. Next, participants will be asked to state the fee (compensation for correct answers) they would be willing to accept for answering an additional question (between 0 and $1), following the Becker-DeGroot-Marschak (BDM) method. If their bid is selected, they will be presented with an additional closed-ended question about the proposition's content, randomly selected from a pool of three very difficult questions.
- We remind participants of the questions they raised at the beginning of the study and suggest they use the provided information tool to find answers. We then ask whether they were able to find the relevant information.

- We conclude with a final set of questions regarding participants' perceived usefulness of the information tool they were provided.

At the end of the survey, we inform all participants about their scores in answering the questions and the total compensation. Also, we inform them that they will have continued access to their assigned information tool for a certain period, up until they are recalled for the second wave of the survey. Additionally, we provide two links, each directing them to a version of the same information tool, but specifically tailored to provide information about the two propositions they were not initially assigned.

The second wave will take place in the following weeks. We will recontact participants, deactivate their access to all previously provided information, and present them with a new set of questions:

- Three compulsory closed-ended questions about the content of the previously assigned proposition.
- Questions about whether they have already voted or if they plan to vote in the upcoming election, including whether they intend to vote "yes" or "no" (or have already voted "yes" or "no") on the proposition.[2]
- For all respondents—those who have already voted, those who plan to vote, and those who do not expect to vote—we will ask an open-ended question requesting the reasons behind their decision to vote or not vote, and, if applicable, their reasons for voting "yes" or "no."

2.3 Treatment Details

As already discussed in the procedure section, participants in this experiment are assigned to one of two information tools:

- **Voter Guide**: This treatment is the official California voter guide, which we have converted into an HTML format while maintaining the same layout, structure, and content as the PDF version available on the California state website. This guide is identical to the physical voter guide that participants will receive at home. Participants in this group

[2]Respondents who have not yet voted at the time of this survey will be contacted a few days after the election to collect this information.

are encouraged to consult this guide to answer the quiz questions on ballot measures.

- **BallotBot**: BallotBot is an interactive chatbot powered by a state-of-the-art large language model (GPT-4o) and utilizes an information retrieval tool from Azure. BallotBot is specifically designed to answer questions about upcoming statewide ballot measures based solely on official information. Participants can interact with BallotBot in real-time, asking questions about the ballot initiatives and receiving tailored, accurate responses drawn from the same official sources as the voter guide.

2.4 Outcome Variables

We are interested in several outcomes, as reflected by respondents' choices during the survey. Specifically, we are focused on the following key outcomes:

1. **Political Knowledge:** We will assess respondents' political knowledge by measuring the proportion of correct answers they provide. We will differentiate between instantaneous knowledge (measured during the first wave of the survey) and retained knowledge (measured during the second wave). Additionally, for the open-ended question, we will generate various text-based metrics, including the number of words, number of arguments, complexity, and measures of argument quality, which will be evaluated using large language models (LLMs).[3]

2. **Cost of Acquiring Knowledge:** We will assess the cost of acquiring knowledge using two distinct measures. The first is an objective measure, which captures the mean or total time spent answering the questions. The second is a subjective measure, which will be derived from the participants' bids in the BDM task, reflecting the compensation they would accept for answering additional questions, indicating their perceived effort or cost associated with acquiring knowledge.

[3]We plan to use gpt-4o or a LLM model with similar or higher capabilities. The expected prompt is: "Evaluate the strength and persuasiveness of this response explaining why someone chose to vote 'yes' or 'no,' decided not to vote, or remained undecided on a specific ballot proposition. Rank the argument on a scale from 1 to 10, and provide a brief explanation for the score, focusing on how well the response justifies the decision, the clarity of reasoning, and the use of relevant evidence or information about the proposition."

3. **Turnout Behavior:** We will measure self-reported turnout (whether participants voted or plan to vote) and their self-reported vote choice (e.g., "yes" or "no" on the ballot proposition).

We expect to assess the effect of the treatment on other secondary outcomes, which we believe will provide additional insights alongside the primary outcomes of interest.

1. **Confidence:** We will measure respondents' confidence by analyzing the self-reported confidence levels they provide after answering each question.

2. **Information Tool Usage:** We will evaluate the actual usage of the information tools during the period between the two survey waves, measuring both the number of times the tools were accessed and the duration of their usage.

Finally, we plan to generate text-based metrics for participants in the experimental group exposed to the chatbot. These metrics will provide non-experimental insights into how respondents interact with and use the tool.

2.5 Economic Incentives

We incentivize participants to answer the closed-ended questions correctly by adding $0.20 to their participation fee for each correct answer. Similarly, they will receive the amount they requested in the BDM task if the randomization works in their favor and if they reply correctly to the question.

2.6 Dealing with Experimenter Demand Effect

We take several steps to minimize the experimenter demand effect. We emphasize non-partisanship in the beginning of the survey right after screening. We do not collect name or any contact information. This provides maximum anonymity and privacy and could potentially increase truthful reporting.

3 Data

3.1 Sample characteristics

We will assess the representativeness of the sample in comparison to the target population, evaluate the balance of covariates between experimental groups, and document attrition both within and between rounds. These analyses will be conducted using standard econometric techniques commonly employed in experimental research.

4 Empirical Strategy and Results

4.1 Empirical Strategy

Let's define Y_i as an outcome of interest measured for participant i, T_i as the treatment group indicator, and X_i as a vector of covariates. We estimate the effect of using BallotBot compared to the voter guide by means of the following linear OLS regression model:[4]

$$Y_i = \tau T_i + \beta X_i + \epsilon_i$$

We control for a range of covariates, including age, gender, race, education, income, attitudes toward A.I., political views, baseline (self-reported) political knowledge, and usual consumption habits of politically relevant information. Most of the controls will be discretized according to standard econometric procedures.

To study heterogeneous treatment effects, we will estimate the effects separately for different levels of covariates or by including interaction terms. We plan to analyze heterogeneity across at least four dimensions: question difficulty, baseline knowledge, baseline political preferences, and socio-economic background. While these are the dimensions of primary interest, we may also explore additional dimensions, such as whether the respondent was informed about the accuracy of their responses before the BDM task.

[4]Given that some of the outcomes are binary or discrete, we will also estimate Probit or Logit non-linear models.

4.2 Results

All figures and tables will refer to the analysis sample, although sample sizes may vary across outcomes due to missing data. We will present the estimated average treatment effects in standard tables, along with additional tables displaying the estimated heterogeneous treatment effects by covariates of interest. Corresponding figures will be included to visually represent these estimates.

While these results will form the core of our analysis, we may also conduct additional empirical exercises to better describe the data and explore the mechanisms behind the observed treatment effects. For instance, we can analyze chatbot interaction data to examine how the topics participants inquire about vary by party affiliation or baseline knowledge. Additionally, we can explore variations in time and intensity of chatbot usage based on socio-economic status and education level.