INFORMATION SYSTEMS RESEARCH

Data-Driven Auditing of Black-Box AI Systems: An Application to Political Campaigns

Journal:	*Information Systems Research*
Manuscript ID	ISRE-2025-2100
Manuscript Type:	Research Note
Manuscript Category:	Regular Issue
Keywords:	AI auditing, Agent-based modeling, Content personalization, Political messaging
Abstract:	AI-mediated content creation and distribution systems are pervasive across digital platforms and increasingly central to online communication. However, little is known about how these "black-box" systems operate—the methods they use, the data they rely on, and the goals they optimize. This study introduces a data-driven auditing framework to analyze and infer the decision-making processes behind AI-driven content distribution. Using agent-based causal experimentation, we develop an approach to estimate how effectively these systems personalize outreach based on explicit user traits and implicit engagement signals. We validate the approach through simulation and apply it to a field experiment during the 2024 U.S. elections. Digital agents, designed to mimic human engagement, subscribed to campaign mailing lists across presidential and Senate races. The resulting data allow us to assess campaign messaging strategies. We find notable differences between Democratic and Republican campaigns in email frequency and personalization. Some Democratic campaigns dynamically adjust content based on engagement. Machine learning models trained using the data accurately predict messaging behavior, enabling inference of underlying policy functions. Beyond politics, this framework offers a scalable method for auditing opaque AI systems in domains like e-commerce, recommendation, and advertising.

https://pubsonline.informs.org/journal/isre

INFORMATION SYSTEMS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 1047-7047, EISSN 1526-5536

Submitted to *Information Systems Research*

Data-Driven Auditing of Black-Box AI Systems: An Application to Political Campaigns

(Authors' names are not included for peer review)

Abstract. AI-mediated content creation and distribution systems are pervasive across digital platforms and increasingly central to online communication. However, little is known about how these “black-box" systems operate—the methods they use, the data they rely on, and the goals they optimize. This study introduces a data-driven auditing framework to analyze and infer the decision-making processes behind AI-driven content distribution. Using agent-based causal experimentation, we develop an approach to estimate how effectively these systems personalize outreach based on explicit user traits and implicit engagement signals. We validate the approach through simulation and apply it to a field experiment during the 2024 U.S. elections. Digital agents, designed to mimic human engagement, subscribed to campaign mailing lists across presidential and Senate races. The resulting data allow us to assess campaign messaging strategies. We find notable differences between Democratic and Republican campaigns in email frequency and personalization. Some Democratic campaigns dynamically adjust content based on engagement. Machine learning models trained using the data accurately predict messaging behavior, enabling inference of underlying policy functions. Beyond politics, this framework offers a scalable method for auditing opaque AI systems in domains like e-commerce, recommendation, and advertising.

Key words: AI auditing, Agent-based modeling, Content personalization, Telemetry Infrastructure, Political messaging

1. Introduction

Algorithmic curation, driven by AI-mediated content creation and distribution systems, is ubiquitous across digital platforms. AI-based technologies are at the heart of recommender systems (Zhang et al. 2021), targeted advertising (Davenport et al. 2020), and personalized messaging (Wattal et al. 2012) in business, media, and politics. Despite these advances, we have a limited understanding of the inner workings of the “black boxes" that power these systems, specifically the methods they employ, the data they leverage, and the objectives they optimize. The opacity of these systems raises significant concerns for the public, businesses, and policymakers. For example, political institutions may seek to audit campaign strategies for transparency. Prior to deploying mission critical systems, firms would like to know to what extent a competitor could learn about the decision rules implemented in its systems. Red team/blue team approaches are often used to evaluate such vulnerabilities (Mathur et al. 2023).

In this paper, we develop data-driven auditing methods designed to systematically analyze, infer, and evaluate latent decision-making processes within AI-powered content creation and distribution systems. Our study has two key objectives: (1) designing a data-driven approach to estimate how AI systems distribute content and determine the frequency of outreach based on both explicit user features and implicit engagement signals, and (2) applying this framework to empirically investigate the methods used by political campaigns to create and distribute content to their followers based on their engagement behaviors.

We develop a modular simulation “meta model" of an AI-driven content creation and distribution system to support a systematic evaluation of how its decision-making could be a function of potential user engagement behaviors. The modular design of the simulation meta model permits alternative model instances to be created with different user engagement models (e.g., stochastic vs. deterministic response models) as well as alternative policy functions. This capacity to model content distribution policies, engagement behaviors, and personalization strategies permits the meta model to create a digital twin of a campaign management system. The simulation model enables the design and evaluation of a statistically well-founded approach to learn the “latent" policy function used by the campaign. Using simulation, for different parametric choices, alternative configurations of agents that mimic different user engagement behaviors can be investigated. This statistical exploration is used to determine the design of the agent system that will produce the best approximation of the policy function of interest. We use this approach to develop a decision-making grid to help auditors with agent designs. Our results show that both rule-based and neural network–based policy functions interacting with stochastic and deterministic user response functions can be closely approximated. This simulation framework provides a rigorous method for determining the cost effective agent design, both in terms of population size and engagement behaviors, required to approximate the latent policy function within acceptable error bounds or under potential budget constraints.

Political campaigns offer a pertinent application context for our framework, where there is significant potential for AI-driven messaging. AI development enables political campaigns to tailor messaging at scale, using features such as voter demographics, online behavior, and engagement patterns to optimize outreach (Lorenzo-Dus and Blitvich 2013, Guzman et al. 2020). Political messaging through emails has become increasingly popular among politicians (Chen et al. 2024, Mathur et al. 2023), enabling controlled dissemination of candidate messaging to potential supporters and donors.[1] However, personalization at this level requires extensive tracking and data collection, often leveraging browsing behavior, location, and inferred demographic characteristics (DeLuca and Curiel 2023, Moser 2024). Recent advancements in large language models (LLMs), such as ChatGPT, have also dramatically reduced the cost associated with content personalization, making dynamic personalization more feasible than ever (Simchon et al. 2024). Therefore, the 2024 U.S. elections marked a significant milestone in this evolution, being the first major election cycle after the widespread adoption of generative AI technologies.

During the 2024 election cycle, we deployed an experimental testbed composed of autonomous agent personas (AAPs). Each AAP was assigned static features (e.g., name, gender, zip code) and browsing

[1] See this NPR article from 2015 where the role of emails as a key communication strategy is discussed `https://www.npr.org/sections/itsallpolitics/2015/07/28/426022093/as-political-campaigns-go-digital-and-social-email-is-still-king`.

behavior cookies (e.g., engagement with liberal or conservative websites) and programmed to exhibit dynamic engagement behaviors (e.g., opening emails, clicking on links). The systematic variation of agent behaviors was used to observe how campaigns adapt their outreach strategies in response to varying levels of voter interaction and engagement signals. The objective was to audit/estimate the policy function that the campaign used to create and distribute emails to campaign followers based on their engagement behaviors. We used the insights from our simulation modeling and statistical approach to evaluate the agent design we deployed in the November 2024 elections.

The results of our field experiment show that Democratic campaigns sent significantly more emails than Republican campaigns, with 65% of Republican campaigns sending no emails to our AAPs. There was significant personalization by Democrats based on engagement with clicks on links embedded in the body of emails, leading to an increase of approximately seven emails received relative to the control group. We do not find any variation in the volume of emails sent based on observable characteristics such as gender and location (e.g., state, rural vs. urban). In terms of the content of the emails, we found limited personalization across observable characteristics of AAPs in the aggregate. Despite limited personalization in the aggregate, six Democratic campaigns were sophisticated in their content personalization. These campaigns dynamically created and distributed content, with clickers receiving more sustained and specific topic coverage in the body of the email. To quantify different types of personalization in the subject line, we develop a framework identifying Level 1 to Level 5 personalization. These six sophisticated candidates dynamically adapt their overall messaging (like reproductive rights, healthcare, and the economy) and also specific instances of topics within messages (like IVF and Roe vs. Wade within reproductive rights) based on prior engagement. Furthermore, our results indicate that the survival of the topics presented by each campaign depends on engagement. Finally, we estimate predictive models (as part of our audit) to understand the mechanics of different campaigns' messaging strategies, based on data collected through our experimental infrastructure. Models like AdaBoost (Favaro and Vedaldi 2021) and CNN (Gu et al. 2018) effectively predicted engagement, the generation of personalized content, and the selective targeting of content to specific AAPs.

These results suggest a technological divide between campaigns, with only a handful embracing sophisticated personalization techniques in a relatively low-cost messaging channel. Such techniques have been demonstrated to be effective in a different but comparable context of e-commerce and non-profit marketing. For a set of sophisticated Democratic campaigns, our results indicate the use of AI in content generation and distribution, suggesting a path forward for future political campaigns. More generally, our flexible infrastructure can serve as a foundation for auditing messaging strategies across various channels and opening up the “black-box" in different ways. Variations of the experimental infrastructure can also be used to understand the use of personalization in other contexts, such as pricing and promotions.

Our study contributes to multiple strands of literature. First, we contribute methodologically to the literature on agent-based causal experimentation, which explores how varying agent behaviors influence algorithmic decision-making (Valogianni et al. 2023). Although prior research has studied agent-based modeling in customer dynamics and market behavior, including product adoption (Rand and Rust 2011), diffusion of innovations (Garcia 2005, Goldenberg et al. 2001), decision making in retail locations (Heppenstall et al. 2006), targeted marketing campaign (Delre et al. 2007, North et al. 2010), and recommendation systems (Zhang et al. 2020), there remains a critical need for models capable of more holistically capturing complex interdependencies between diverse actors and their decisions. Our proposed meta-model framework simulates interactions between distinct classes of actors: individual agents (representing voters) and campaign systems (representing political organizations), enabling researchers and practitioners to systematically examine how decisions made by one set of actors influence and constrain the behaviors of others. This holistic and modular modeling approach supports the study of strategic decision-making over time, enabling repeated simulations, comparative analyses, and robust auditing across multiple domains such as content delivery, digital marketplaces, and political campaign ecosystems.

Next, we extend research on email messaging and political donations. Prior work has examined the use of manipulative tactics in campaign emails, such as dark patterns that prompt unintended donations (Posner et al. 2023), the evolution of small-dollar campaign fundraising via platforms like ActBlue and WinRed (Bouton et al. 2022) and the content strategies employed by the 2020 Trump and Biden campaigns (Chen et al. 2024, Mathur et al. 2023). By designing an infrastructure that actively engages with campaign emails, our study reveals how digital and AI-driven techniques shape political messaging at scale, providing a foundation for auditing political campaigns through data-driven methodologies. Second, we contribute to the broader literature on political communication strategies. While past work has examined how social media influences candidate visibility and fundraising (Petrova et al. 2021), the impact of TV ads on voter behavior (Spenkuch and Toniatti 2018), and the role of ad tone and consistency in political messaging (Gordon et al. 2023, Fossen et al. 2022), our study uniquely focuses on email—a direct, persistent, and highly trackable channel of engagement. Prior research has begun to scrutinize political ads on platforms like Facebook (Baviera et al. 2022, Beraldo et al. 2021), but we take this further by examining how campaigns dynamically personalize outreach through email tracking and AI-generated content. Third, we contribute to the literature addressing tracking technologies and regulatory concerns surrounding digital personalization. Studies have shown that restricting cookie tracking reduces the effectiveness of ad targeting and platform recommendations (Sun et al. 2024, Wernerfelt et al. 2024, Peukert et al. 2024), yet little is known about how political campaigns employ these same tracking mechanisms. Our work sheds light on the extent to which campaigns leverage explicit user features and implicit engagement signals to optimize their outreach, revealing the increasing sophistication of AI-driven political messaging and its implications for transparency and regulation.

2. Modeling the “System"

The development of a principled approach to auditing an AI-driven content creation and distribution system requires access to both the decision logic governing the campaign system as well as the behavioral logic governing the engagement behavior of agents interacting with the content delivered to them by the AI system. A dynamic feedback loop links user interactions and engagement to the decision logic, shaping subsequent content creation and distribution decisions. We develop a modular meta model of a digital AI driven campaign management system – a digital twin – consisting of two interconnected subsystems: the *Policy System*, which governs topic selection, targeting, and frequency of email dispatches, and the *User System*, which models recipient behavior.

2.1. Policy System

The Policy System determines the topics to include in an email to an agent (as a function of agent engagement) and when the emails are sent. The key elements of the model used by the campaign to make these decisions are described below. An agent refers to a user subscribed to the campaign's mailing list. Each campaign, indexed by $c \in C$, employs its own policy function $\Pi^{(c)}(x \mid \beta_t^n)$ to decide the content and scheduling. Each agent can be subscribed to one or more campaigns.

1. **Users and Time Steps:** Each agent has a state that evolves as a function of time-invariant demographic features, such as gender and location, as well as time-varying features, such as historical engagement. Each agent is denoted as $AAP^1, AAP^2, \ldots, AAP^N$, each observed over T^n time steps indexed by $t = 1, \ldots, T^n$.
2. **User State (β_t^n):** For each user n at time t, the state β_t^n encapsulates features of aggregate engagement score (ES) computed over the last m emails from the historical record H, and demographic attributes such as gender and location. The engagement score is defined as $ES_t = \frac{1}{m}\sum_{i=t-m+1}^{t}\sum_{y\in Y}\alpha_y\, \mathbb{1}\{Y_i = y\}$, where $\mathbb{1}\{Y_i = y\}$ is the indicator function that equals 1 if $Y_i = y$ and 0 otherwise; α reflects the intensity of engagement for the outcome y. For example, opening an email might have $\alpha = 1$, while donating, being a stronger signal, could be assigned $\alpha = 4$.
3. **Actions (x_t^n):** At each time step, each campaign $c \in C$ selects an action x_t^n from a predefined topic space $\mathcal{A}$ (e.g., 'Healthcare', 'Economy'), treated as a hyperparameter.
4. **Email Scheduling ($S(\beta, ES_t)$):** The waiting time ω until the next email is a function of the current state β and the computed engagement score ES_t, defined as $\omega = \omega_0 \cdot \exp\Big(-\lambda \cdot ES_t\Big)$ where ω_0 is a baseline interval and λ controls the sensitivity of the waiting time to the engagement score.

We consider two scenarios for the policy function: (1) rule-based policy function, and a (2) learned policy function. The rule-based policy function follows a structured statistical model in which topic selection is governed initially by a pre-specified prior $\pi_0(x)$ and updated as engagement data accumulate. The final distribution over topics is given by:

$$\Pi^{(c)}_{\text{rule-based}}(x \mid \beta_t^n; \gamma) = \frac{\pi_0(x)\ \exp\{\gamma\ \cdot \phi(x, \beta_t^n)\}}{\sum_{x' \in \mathcal{A}} \pi_0(x')\ \exp\{\gamma \cdot \phi(x', \beta_t^n)\}}$$

where $\phi(x, ES_t)$ quantifies the alignment between topic x and the current engagement states, and γ is a sensitivity parameter.

The learned policy function leverages a neural network to adaptively map the user state to a probability distribution over topics. At each decision step, the updated topic prior $\pi(x \mid \beta^n_{t+1})$ is concatenated with the user state, which includes the updated ES and encoded demographics to form a combined feature vector. This vector is then passed through a three-layer fully connected network. The network first maps the input feature vector to a hidden representation using two ReLU-activated layers, and then projects this representation onto the topic space via a final linear layer that outputs logits. A softmax is applied to the logits to ensure the valid probability distribution over topics:

$$\Pi^{(c)}_{\text{learned}}(x \mid \beta^n_t; \theta) = \text{softmax}\Big(\text{PolicyNetwork}\big([\pi(x \mid \beta^n_t), \beta^n_t]; \theta\big)\Big)$$

Both types of policy functions are updated periodically (e.g., daily or weekly) using newly observed engagement behavior. By selecting these two modules, our simulation framework incorporates two representative approaches to modeling decision-making: one that emphasizes interpretability and a principled Bayesian update (the rule-based function) and another that leverages data-driven adaptivity via neural networks (the learned policy function). In this modular infrastructure, each campaign can explore alternative policy functions while competing for the attention of agents, allowing researchers to explore alternative configurations.

2.2. User System

The User System comprises agents that mimic human behavior. While the policy functions used by the policy systems are latent to the agents, similarly, the behavior of the agents in the user system is unobserved and unknown to the policy system. In this sense, the twin models the information environment present in a real-world campaign.

The agents in the user system employ alternative behavioral models to respond to the content of the received email and are subject to limited attention capacity. When an email is received, containing a set of topics, user's response function generates a probability distribution over engagement outcomes, $Y = \{1, 2, 3, 4\}$. These outcomes represent a spectrum of engagement, ranging from complete inattention to increasingly costly actions such as opening, clicking, or donating. Each outcome is associated with a cost $O(y)$: $O(1) = 1,\ O(2) = 2,\ O(3) = 3,\ O(4) = 5$. We consider these four outcomes with a partial order based on the cost incurred; however, the framework can be extended to include additional outcomes organized by a total order.

At each time step, an agent follows a three-stage decision process to determine their engagement behavior, subject to the constraint of limited attention capacity:

1. **Email Processing Decision:** agent n decides whether or not to process emails in a given time step. Otherwise, all emails are ignored.
2. **Campaign Selection:** If multiple campaign emails arrive simultaneously (i.e., from campaigns $c \in C^n$, where C^n denotes the campaigns to which user n is subscribed), the agent selects one campaign to

process. This selection decision is determined via a logit model based on the aggregated engagement score. Specifically, let $ES_t^{n,c}$ denote the engagement score of agent n with campaign c up to time t. Then, the probability of selecting campaign c is given by $S_t^n(c) = \frac{\exp\left\{\delta ES_t^{n,c}\right\}}{\sum_{c' \in C^n} \exp\left\{\delta ES_t^{n,c'}\right\}}$ where δ is a sensitivity parameter.

3. **Engagement Outcome Decision:** Conditional on processing an email from campaign c, the agent then selects an engagement action. We consider two alternative approaches to model the response decision:
   (a) **Deterministic Response Function:** The probabilities $f(y \mid \beta, x)$ are fixed and defined via a linear interpolation between a low-engagement vector $\mathbf{P}_{\text{low}} = [0.45,\ 0.20,\ 0.15,\ 0.20]$ and a high-engagement vector $\mathbf{P}_{\text{high}} = [0.30,\ 0.20,\ 0.25,\ 0.25]$, modulated by the user's propensity score ρ: $f(y \mid \rho, x) = (1-\rho)\, P_{\text{low}}[y] + \rho\, P_{\text{high}}[y]$.
   (b) **Stochastic Response Function:** Here, the response function $f(y \mid \beta, x)$ is unknown and modeled nonparametrically. For each topic x, logit parameters are sampled to compute the categorical distribution via a softmax transformation. For each email received, an outcome Y_t^n is sampled as $P(Y_t^n = y \mid \beta_t^n, x_t^n) = f(y \mid \beta_t^n, x_t^n)$.

The observed user engagement behavior is then used to update the user's state β_{t+1}^n and the full historical engagement record H, which in turn informs the Bayesian update of the topic priors for each campaign: $\pi^{(c)}(x \mid ES) \propto \pi_0^{(c)}(x)\, L(ES \mid x, c)$, where $L(ES \mid x, c)$ is the likelihood of the observed engagement history given topic x for campaign c. For the learned policy function, after each email interaction, the updated user state, comprising the updated ES and encoded demographic attributes is fed into the neural network, allowing the network to adapt its mapping from state to topic distribution continually.

2.3. Estimation via Maximum Likelihood

Our objective is to estimate the latent policy $\Pi^{(c)}$, which maps any given user state β to a probability distribution over topics $\mathcal{A}$, i.e., $\Pi^{(c)}(x \mid \beta) \geq 0,\ \sum_{x \in \mathcal{A}} \Pi^{(c)}(x \mid \beta) = 1,\ \forall \beta$. We consider two scenarios for the estimation procedure:

1. **Deterministic Response Function:** When the response function f is known, the probability of observing an outcome Y_t^n =y for a user in state β_t^n is given by $P(Y_t^n = y \mid \beta_t^n) = \sum_{x \in \mathcal{A}} \Pi^{(c)}(x \mid \beta_t^n)\, f(y \mid \beta_t^n, x, c)$. The cost-weighted likelihood and log-likelihood are

$$L^{(c)}(\Pi^{(c)}) = \prod_{n=1}^{N} \prod_{t \in \mathcal{T}^{n,c}} \left(\sum_{x \in \mathcal{A}} \Pi^{(c)}(x \mid \beta_t^n)\, f(Y_t^n \mid \beta_t^n, x, c) \right)^{O(Y_t^n)},$$

$$\ell^{(c)}(\Pi^{(c)}) = \sum_{n=1}^{N} \sum_{t \in \mathcal{T}^{n,c}} O(Y_t^n) \cdot \log\left(\sum_{x \in \mathcal{A}} \Pi^{(c)}(x \mid \beta_t^n)\, f(Y_t^n \mid \beta_t^n, x, c) \right).$$

The estimation involves maximizing $\ell(\Pi)$ subject to the constraints on Π.

2. **Stochastic Response Function:** When f is unknown, we treat $f(y \mid \beta, x)$ as an arbitrary but valid probability distribution. The cost-weighted likelihood and log-likelihood are then

$$L^{(c)}(\Pi^{(c)}, f) = \prod_{n=1}^{N} \prod_{t \in \mathcal{T}^{n,c}} \left(\sum_{x \in \mathcal{A}} \Pi^{(c)}(x \mid \beta_t^n)\, f(Y_t^n \mid \beta_t^n, x, c) \right)^{O(Y_t^n)},$$

$$\ell^{(c)}(\Pi^{(c)}, f) = \sum_{n=1}^{N} \sum_{t \in \mathcal{T}^{n,c}} O(Y_t^n) \cdot \log\left(\sum_{x \in \mathcal{A}} \Pi^{(c)}(x \mid \beta_t^n)\, f(Y_t^n \mid \beta_t^n, x, c) \right).$$

In both cases, the following constraints must hold for all campaigns c, user states β, and pairs (β, x) $\Pi^{(c)}(x \mid \beta) \geq 0$, $\sum_{x \in \mathcal{A}} \Pi^{(c)}(x \mid \beta) = 1$, $f(y \mid \beta, x, c) \geq 0$, $\sum_{y \in Y} f(y \mid \beta, x, c) = 1$. We evaluate the estimated policy $\hat{\Pi}^{(c)}$ against the true policy $\Pi^{(c)}$ using divergence measures, the Kullback–Leibler divergence (Kullback and Leibler 1951) with the objective of minimizing divergence. We provide a summary of the notation used in Appendix A.

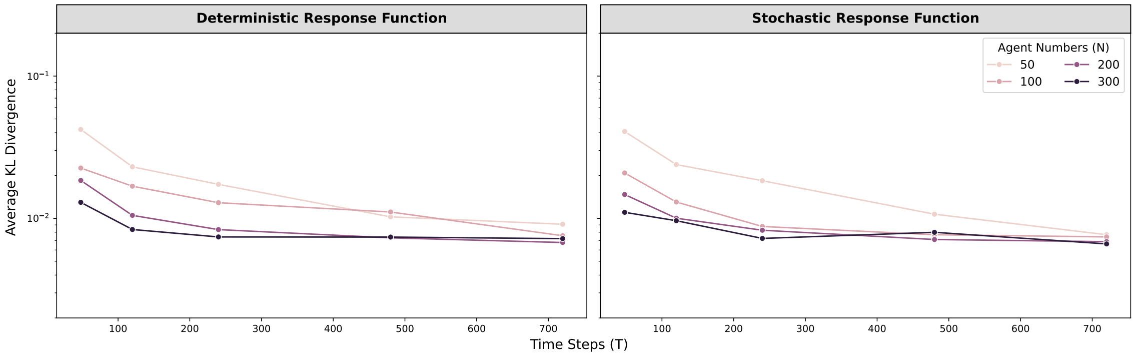
3. Simulations

We use our simulation model of both the policy system and the user system to systematically investigate and audit the campaign system's policy function. More specifically, we use the simulation to examine the design of the agent system that we use to gather data on content received and distributed as a function of engagement behaviors. The simulation is modular and permits model instances consisting of system designs with alternative policy function and user engagement models to be investigated. These modules (shown below) allow representative simulation instances to be created and investigated.

- **Policy Function Modules:** Rule-based and neural network-based approaches.
- **Response Function Modules:** Deterministic and stochastic response functions.
- **Agent Features:** Static (e.g., demographics, location) and dynamic behavioral (e.g., engagement history) components.
- **Hyperparameters:** The topics and the number of topics per email ($k_{\min}$ to $k_{\max}$), engagement-score look-back window m, scheduling baseline interval ω_0, the sensitivity parameter λ controlling frequency, the agent's propensity score ρ, and the user-response cost weights $O(y)$.

Since we are interested in the number of agents and the behaviors they expose to acquire the data required to estimate the latent policy function, we vary the number of agents N ∈ {50, 100, 200, 300}and simulation time steps T ∈ {48, 120, 240, 480, 720} over which data will be collected about content creation, distribution and frequency based on the engagement behaviors of agents. For each configuration, we apply maximum likelihood estimation to evaluate how good an approximation of the true policy function can be recovered by a given agent design and time step combination. The simulation results are shown in Figure 1, which presents the divergence between the estimated and true policy functions over time T, stratified by agent count N. We observe that the number of agents increases (represented by darker lines), the divergence consistently decreases, indicating improved estimation accuracy. Similarly, as T grows, the divergence declines, with the gains beyond approximately N = 300 and T = 480 beginning to plateau. We provide rule-based policy function estimation in Appendix F with the gains also around N = 300 and T = 480 beginning to plateau. This demonstrates that our MLE-based estimation framework robustly approximates the underlying policy functions across varying decision-making scenarios and user engagement patterns.

Figure 1 Evaluation of NN-based Policy Function Across Agent Counts and Time Steps



4. Empirical Setting

To validate our simulation-based findings, we conducted a field experiment during the 2024 U.S. elections, deploying our model in a real-world environment and recording both actual campaign emails and user interactions. We operationalize the parameters and modules described in Section 2 in this context. Our study focused on the presidential race and 13 Senate races. Among the Senate contests, three solid/likely Democratic races (NY, MA, MN), three solid/likely Republican races (MO, TX, TN), and seven toss-up races (AZ, NV, PA, WI, MI, MT, OH) were selected based on the Cook Political Report ratings.[2] This yielded a balanced pool of 28 campaigns (14 Republican and 14 Democratic), enabling us to compare email content generation and distribution processes between competitive (toss-up or lean) and non-competitive (solid Republican or solid Democratic) races.

4.1. Construction of Autonomous Agent Personas

To simulate potential voters, we construct AAPs that reflect varied demographic and political profiles, with explicit signals provided along multiple dimensions. Additional details on the relationship between simulated agents and AAPs are provided in Appendix B.

1. **Gender:** Names were chosen to represent gender signals, drawing from the top white names based on the analysis in Fryer Jr and Levitt (2004). Although only 53% of campaigns required name entry, our subsequent analysis (see Table E6) indicated no statistically significant difference in emailing strategies based on name requirements.

[2] `https://www.cookpolitical.com/ratings/senate-race-ratings`

2. **Major Geographical Location (State):** AAPs were based in the seven Senate battleground states (AZ, NV, PA, WI, MI, OH, MT), as our pilot studies indicated that campaigns in competitive states were more proactive in emailing likely voters compared to solid Red or solid Blue states.
3. **Minor Geographical Location (Zip Code):** For each AAP, zip codes were randomly selected based on income levels (high/low), geographical settings (urban, suburban, rural), and racial composition (Black, Non-White/Black, White) [3].
4. **Personal Political Affiliation:** Each AAP was pre-assigned a political affiliation (Democratic or Republican) to ensure that emails were routed within the appropriate party context and to prevent cross-party interference.

4.2. Experimental Conditions

Our experimental design manipulates two primary dimensions: website browsing (cookie-based tracking) and behavioral engagement. Each condition is applied to the AAPs to isolate its effect on campaign email personalization and engagement outcomes.

4.2.1. Website Browsing Conditions Each AAP is assigned a specific browsing history that corresponds to one of four cookie conditions. This assignment was implemented using four different laptops with unique cookie profiles: (1) Democratic Browsing: Visits only left-wing websites, (2) Republican Browsing: Visits only right-wing websites, (3) Mixed Browsing: Visits both left- and right-wing websites and (4) No Browsing History: Maintains a blank cookie profile. The websites were selected based on Media Bias Fact Check ratings, and to ensure comparability, each AAP is replicated across the four browsing conditions.

4.2.2. Behavioral Engagement Conditions We design three distinct behavioral engagement conditions to evaluate the impact of different campaign strategies: (1) Control (No Interaction): AAPs do not interact with the emails. (2) Treatment 1 (Open Only): AAPs check their email and decide whether to open it. (3) Treatment 2 (Open and Click): AAPs check their email, decide whether to open it, and, if opened, decide whether to click on embedded links. For each AAP, parameters governing email-checking frequency and the likelihood to open or click links [4] are drawn from distributions that incorporate local factors, which we determine using the congressional district competitiveness measured via the Cook Political Report's PVI score. This design allows us to assess whether differential engagement behavior affects both the volume and the nature of campaign emails delivered.

Combining dimensions of demographic, zip code characteristics, state, browsing history, and engagement behavior results in a structured matrix of AAPs. In our experiment, 4,032 unique AAPs (2 genders × 12 zip

[3] Zip codes were categorized into 12 groups based on geographical setting (rural, suburban, urban), income level (high, low), and racial composition (Black, Non-White/Black, White). Classifications were made using Census Bureau data on mean income, racial demographics, and Rural-Urban Commuting Area codes, labeling as "high," "medium," or "low" based on national percentiles.

[4] It is pertinent to note that we did not donate to any campaign since only humans, who are US citizens, are legally allowed to donate money to a political candidate.

codes × 2 political affiliations × 7 states × 3 engagement conditions × 4 browsing conditions) signed up for 5 distinct races, yielding a total of 20,160 signups. Each signup is associated with a unique email ID visible to the campaign, facilitating detailed comparisons. For instance, we can isolate the effect of political affiliation by comparing AAPs with identical demographic and browsing profiles but different party alignments.

4.3. Operational Infrastructure and Email Interaction Simulation

Once the AAPs were constructed and assigned to the experimental conditions, we developed an automated operational algorithm to simulate realistic email interactions that consist of the following key components:

Email-Checking Frequency: Each AAP is assigned an email-checking frequency drawn from a normal distribution (with a mean of 3.5 hours and a range of 0 to 17 hours Adobe Blog (2019)). Figure C1 shows the corresponding distribution. The algorithm compares this frequency (normalized to a 0–1 scale) with a random threshold to decide whether an AAP checks their email during each hourly cycle.

Decision to Open/Click: If an AAP checks their email, a decision is made on whether to open it based on a beta distribution defined at the ZIP code level (shown in Figure C2). This beta distribution is modulated by the competitiveness of the local congressional district (derived from the PVI score), with more competitive districts yielding higher probabilities of opening. For AAPs in the "open and click" condition (Treatment 2), an additional beta draw determines whether the agent clicks on a link within the email. Priority is given to donation links (e.g., via ActBlue or WinRed), followed by other calls to action.

Operational Algorithm: The algorithm operates in a cyclic manner, where it first simulate email-checking frequency for each AAP then for those who check their email, simulate the decision to open the email. Finally, for emails that are opened, simulate the decision to click on a link. Time stamps are recorded to prevent multiple actions on the same email, and a "last acted upon" (LTA) log is maintained to ensure that only new emails are processed in subsequent cycles. VPNs are also aligned with each AAP's designated home state to reinforce location consistency.

The signup process was implemented using our automated infrastructure across four laptops (each representing a distinct cookie profile). To maintain data integrity, VPN settings were switched for each AAP to match their designated home state. A sequential sign-up process was enforced, in which each AAP was registered for their designated races (one home Senate race, one presidential race, and three out-of-state Senate races) with random delays between 20 and 100 seconds to mimic organic sign-up behavior. Cookie and browser consistency was maintained throughout each AAP's signup activity, and each signup event was recorded and monitored to confirm adherence to the experimental design.

5. Quantity Analysis: Baseline and Heterogeneity

We start by analyzing the quantity of emails received across treatment conditions and then look at potential heterogeneity across different personas.

5.1. Quantity Analysis: Baseline Results

Figure 2 illustrates the volume of emails sent by campaigns over time, showing a clear upward trend as the elections approach, particularly among Democratic candidates. In contrast, Republican candidates maintain a relatively stable email frequency throughout the sample period. Notably, Democrats send emails at a rate eight times higher than their Republican counterparts. Figure 3, which summarizes the average number of emails sent across parties and treatments, reveals a striking disparity: 65% of Republican candidates in our sample (9 out of 14) do not send any emails to any AAP, highlighting a significant difference in campaign outreach strategies between the two parties. The heatmap also highlights that Kamala Harris sent the highest number of emails across treatment conditions, as indicated by the darker shades. The horizontal axis represents the number of emails received over the experimental period across conditions.

Figure 2 **Total Email Received by Date**

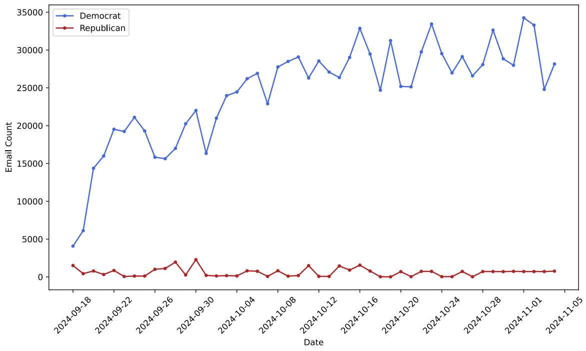


Table 1 **Email Volume by Treatment Conditions**

	Democrats	Republican	Democrats	Republican
DV: email count	(1)	(2)	(3)	(4)
Dem_cookie		-0.059**		
		(0.024)		
None_cookie	-1.026***	-0.037		
	(0.349)	(0.024)		
Rep_cookie	-0.694**			
	(0.350)			
Mix_cookie	-2.024***	-0.102***		
	(0.321)	(0.022)		
Open Only			-0.525***	-0.011
			(0.165)	(0.019)
Open + Click			6.873***	0.015
			(0.241)	(0.020)
Time FE	Yes	Yes	Yes	Yes
Observations	96,768	96,768	96,768	96,768
R^2	0.014	0.005	0.272	0.001

Clustered (persona_id) standard-errors in parentheses. ***: 0.01; **: 0.05; *: 0.1

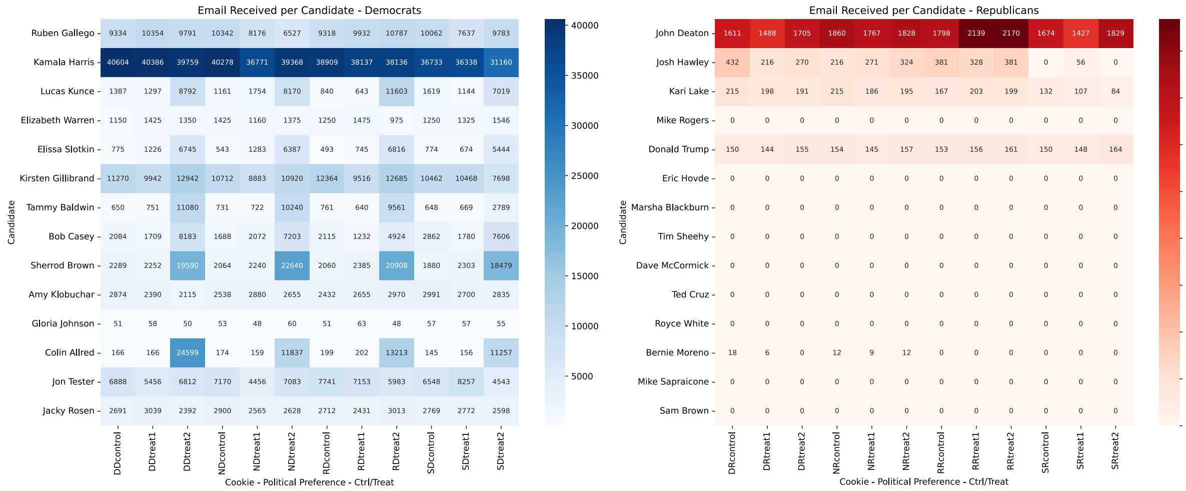
To analyze these trends quantitatively, we aggregate these values and estimate regressions based on Equation 1, using AAP-day as the unit of analysis.

$$\begin{aligned} \text{Email Count}_{it} = \beta_0 + \beta_1 \cdot \text{Cookie Condition/Engagement Level}_i + \beta_2 \cdot \text{Area Type}_i \\ + \beta_3 \cdot \text{Income}_i + \beta_4 \cdot \text{Race}_i + Date_t + \epsilon_i \end{aligned} \tag{1}$$

where Email Count$_{it}$ represents the number of emails received by AAP i on date t. Cookie Condition/Engagement Level$_i$ represents treatment assignment based on browsing behavior (e.g., left-leaning, right-leaning, mixed, or no website) or engagement level (e.g., never open, open only, open + click). Area Type$_i$, Income$_i$ and Race$_i$ are AAP-specific covariates capturing demographic information where the AAP is located. Date fixed effects are also included, and standard errors are clustered at the persona level.

Table 1 presents the effects of AAPs browsing left or right-leaning websites, a mixture of the two, or no website on campaign email engagement. Column (1) shows that AAPs browsing non-left-leaning sites

Figure 3 Heatmap of Email Received by Group by Candidate



Note. DDcontrol, NRtreat1, and similar labels in the horizontal axis represent different experimental conditions based on browsing history, political affiliation, and treatment assignment. For example, DDcontrol refers to an AAP assigned a left-leaning browsing history with a Democratic affiliation in the control group.

receive fewer Democratic campaign emails, while Column (2) indicates that AAPs with left-leaning or mixed browsing histories receive fewer Republican emails compared to those browsing right-leaning sites. This suggests campaigns track and target like-minded AAPs using cookie-based information.

Next, we examine how email engagement influences email volume. The control group never opens emails, while two treatment groups either always open but never click or click on at least one link. Column (3) shows that clickers receive about seven more emails than the control group, reinforcing the idea that campaigns personalize outreach based on engagement. AAPs who only open emails receive about half an email fewer than the control group, suggesting campaigns adjust expectations about engagement potential. For Republican campaigns (Column 4), neither treatment condition significantly affects email volume, where the campaigns treat all the AAPs the same and do not incorporate the signals based on tracking engagement with the email (or lack thereof), consistent with findings that email plays a lesser role in Republican campaign strategies. Interaction effects further support these trends. Democratic campaigns send more emails to left-leaning cookie clickers, while Republican campaigns show no significant variation by cookie type or engagement level. Detailed results are in Table E1.

5.2. Quantity Analysis: Heterogeneity Results

The baseline results above suggest some personalization based on cookie and engagement tracking. Campaigns may want to better target their message across different demographics and locations as interest may vary significantly. To understand the prevalence of such finer targeting, we analyze heterogeneous treatment effects across income, gender, and location (urban, suburban vs. rural) for the two experiment conditions.

We analyze the degree of targeting by the gender of the individual signing up. As a reminder, about half of the campaigns ask for names at signup, and email marketing software can enable gender prediction at a low cost. Focusing on male versus female comparisons (see Table E3), we analyze the interaction of the cookie conditions with the gender of the recipient. As in the case of income, we find null results across all conditions for both Democrats and Republicans. The statistically and economically null effect is also observed for the engagement experimental conditions (never open, open, open + click). Finally, in Table E2, we also see statistically and economically insignificant results on the interaction between the experimental conditions and whether the zip code sign-up is from an urban, suburban, or rural zip code. Together, these results suggest that there is no variation based on the explicit signals provided by the subscriber at sign-up.

In general, these results on the heterogeneity of treatment effects suggest that tracking appears to occur only at the aggregate cookie and engagement levels. But does the content of the emails vary by persona?

6. Content Analysis: Baseline and Heterogeneity

In this section, we analyze the content of emails received focusing on the subject, body text and the topics sent to different personas.

6.1. Content Analysis: Baseline Results

To capture the nuances of personalization in each email message, we break down the body of each email into individual sentences, subject lines and topics. To determine the extent to which the same content is sent to multiple personas, we derive a measure *AAP reached percentage* that calculates the percentage of AAPs reached per text type $\tau \in \{sentence, subject, topic\}$ for each candidate c on each date t specified as the following,

$$\text{AAP Reached Percentage}_{ct\tau} = \left(\frac{\text{AAPs Reached}_{ct\tau}}{\text{Total AAPs Reached}_{ct\tau}}\right) \times 100 \tag{2}$$

A higher AAP reached percentage means that the exact particular text type was sent to a larger percentage of AAPs on that date, implying a broader or less targeted (personalized) message distribution strategy. These statistics provide a birds-eye view of the content strategy and the potential degree of targeted messaging.

Table 2 Summary Statistics of AAP Reached Percentage across Content Types

	Mean	std	Min	25%	50%	75%	Max
Democrat							
Average Sent-Persona-Pct	71.91	37.33	0.05	41.53	99.05	100.00	100.00
Average Topic-Persona-Pct	85.12	28.45	0.05	87.74	100.00	100.00	100.00
Average Subject-Persona-Pct	42.35	43.47	0.05	1.19	15.56	98.35	100.00
Republican							
Average Sent-Persona-Pct	92.29	21.02	3.85	100.00	100.00	100.00	100.00
Average Topic-Persona-Pct	97.12	14.57	3.85	100.00	100.00	100.00	100.00
Average Subject-Persona-Pct	78.32	36.16	3.85	50.00	100.00	100.00	100.00

Republican campaign emails show minimal personalization, with both sentence- and topic-persona percentages exceeding 90% on average and reaching 100% at the median, indicating highly uniform messaging. Subject lines also show little variation with some differences in subject lines sent out at the bottom of the distribution, aligning with the limited personalization observed in email quantity analysis. Democratic campaigns, in contrast, exhibit more variation, with sentence- and topic-persona percentages averaging 72 and 85, respectively, though still high. While median campaigns use broad messaging, subject lines show greater targeting, with a mean of 42.35 and a median of 15.56.

Table 3 AAP Text Reached pct of Democrats Candidates

	sentence-persona-pct		*subject-persona-pct*		*topic-persona-pct*	
	Six Candidates	Non-Six Candidates	Six Candidates	Non-Six Candidates	Six Candidates	Non-Six Candidates
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.0890	0.0338	-0.4983***	0.0664	0.2532***	0.0152
	(0.0747)	(0.0405)	(0.0746)	(0.0701)	(0.0504)	(0.0217)
Suburban	-0.0802	-0.0386	-0.1743	-0.0368	0.0153	-0.0444
	(0.1498)	(0.0810)	(0.1499)	(0.1401)	(0.1011)	(0.0435)
Urban	-0.3580***	0.0116	-0.5058***	0.2086**	-0.0128	-0.0090
	(0.1061)	(0.0571)	(0.1064)	(0.0996)	(0.0718)	(0.0307)
Low	0.1535	0.0225	0.2229**	0.0619	0.0376	0.0038
	(0.1072)	(0.0556)	(0.1081)	(0.0963)	(0.0718)	(0.0302)
White	-0.6048***	-0.0019	-0.4255***	-0.0514	-0.2557***	0.0161
	(0.0914)	(0.0491)	(0.0921)	(0.0852)	(0.0616)	(0.0267)
Non-white/black	-0.3407***	-0.0004	-0.3468***	-0.2737***	-0.2606***	0.0188
	(0.0916)	(0.0498)	(0.0918)	(0.0859)	(0.0617)	(0.0267)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Treatment FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,633,254	9,334,279	671,679	505,194	3,171,905	3,442,918
R^2	0.17375	0.12843	0.12521	0.07081	0.09851	0.06580

Clustered (persona_id) standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

The distributions highlighted above are aggregate numbers that may obscure nuances in campaign strategies. Specifically, campaigns might focus their outreach on one or two specific demographic groups while sending more generic messages to others. Thus, we analyze text variation across demographics, focusing on six candidates: Bob Casey, Collin Allred, Elissa Slotkin, Kamala Harris, Amy Klobuchar, and Lucas Kunce, who have lower AAP Reached Percentages, indicating higher personalization potential (Table 3).

For the other eight candidates (labeled non-six candidates) (Columns 2, 4, 6), we observed limited personalization efforts across demographics where the estimates are statistically insignificant for body and topics. In contrast, the six targeted candidates (Columns 1, 3, 5) show greater personalization, particularly across gender, urban-rural divides, and racial demographics.

These findings highlight a clear divergence in email strategies. Republican candidates appear to deprioritize email as a tool for personalization, while Democratic candidates employ more nuanced messaging, especially in subject lines. Additionally, 43% of candidates (6 out of 14) engage in more personalized communication, with engagement levels influencing content distribution.

6.2. Tailored Messaging: Insights from Key Candidates

To further explore campaign personalization, we focus on Bob Casey, Collin Allred, Elissa Slotkin, Kamala Harris, Amy Klobuchar, and Lucas Kunce, who show noticeable levels of personalized messaging.

6.2.1. Subject Line Analysis To assess the extent of subject line personalization, we classify subject lines into five distinct types, each reflecting varying dimensions of tailored messaging designed to align with recipient preferences and engagement behaviors. This classification provides a framework to uncover the strategic intent behind subject line personalization.

- Level 1: Generic, non-personalized lines (e.g., "Last time you'll hear from me").
- Level 2: Basic personalization with details like the recipient's name or city (e.g., "Connor, is there anything – ANYTHING – we can say?")
- Level 3: References to time-sensitive events, such as recent debates, without action-oriented language (e.g., "re: Fox News" or "Check out what CNN just reported").
- Level 4: Combines current events with persona details and action-oriented phrasing (e.g., "[INVITATION] Join Vice President Harris in La Crosse on Thursday").
- Level 5: Leveraging in-depth data like zip codes and neighborhood characteristics (e.g., "Our campaign is not on track to reach the 277-donation goal we set for Henderson").

The results in Table 4 suggest higher engagement personas (captured by open + click) are more likely to receive action-oriented subject lines such as invitations to join local rallies or events (Level 3 and 4). On the other hand, lower engagement personas that never open emails and serve as the control group are targeted with hyper-personalized content emphasizing personal reference or local details (Level 2 and 5).

Table 4 Engagement Level on Personalized Subject Types

	Dependent variable: subject count				
	Level 1	Level 2	Level 3	Level 4	Level 5
Model:	(1)	(2)	(3)	(4)	(5)
Intercept	4.878***	2.032***	1.200***	0.097***	0.120***
	(0.083)	(0.014)	(0.011)	(0.022)	(0.022)
Open Only	0.008	-0.007	-0.006	0.005	-0.121***
	(0.017)	(0.008)	(0.005)	(0.012)	(0.015)
Open + Click	1.091***	-0.084***	0.150***	0.103***	-0.151***
	(0.046)	(0.008)	(0.007)	(0.013)	(0.013)
Observations	536,614	105,598	49,622	12,683	12,721
R^2	0.059	0.001	0.010	0.011	0.066

Clustered (persona_id) standard-errors in parentheses. ***: 0.01, **: 0.05, *: 0.1

6.2.2. Email Body Content Analysis Building on subject line analysis, we examine how email topics evolve with engagement using OpenAI's GPT-4o model to extract and categorize topics and nested subtopics. We employ multiple metrics (detailed in Appendix D), including topic recurrence, topic entropy (to assess focus), cosine similarity (to measure semantic continuity), topic introduction, survival duration, and co-occurrence probabilities. These measures help determine how prior engagement shapes content in subsequent periods. The results are presented in Table 5.

At the user level, Column (1) shows that clicking behavior increases topic recurrence, while Column (2) indicates that higher engagement reduces topic entropy, suggesting a narrowing of content around specific themes that signal interest. Columns (3) and (4) reveal that engagement enhances semantic continuity and increases the introduction of new topic instances (e.g., discussing IVF or Roe v. Wade under reproductive rights), inferring the evolution of topics over time. At the campaign level, engagement significantly influences content strategy. Column (5) shows that clickers receive topics with longer survival durations, reflecting campaigns' focus on resonant themes. Column (6) indicates greater topic diversification with increased clicks, while Column (7) shows stronger alignment between recipient exposure and campaign themes. Finally, Column (8) reveals higher topic co-occurrence for engaged users, reinforcing a more cohesive narrative.

Table 5 Impact of Engagement on Campaign Content Generation Strategy

	Persona-Level Personalization		**Topic Evolution**		**Campaign-Level Content Strategy**			
DV	*Same Topic Received*	*Persona Topic Entropy*	*Cosine Similarity*	*New Instances Pct*	*Topic Survival Duration*	*Candidate Topic Entropy*	*Topic Cluster Similarity*	*Topic Co-occurrence*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Click (1/0)	0.719***							
	(0.048)							
Click pct$_{it-1}$		-0.304***	0.394***	0.464***	0.005***	1.244**	0.010***	
		(0.051)	(0.005)	(0.005)	(0.001)	(0.575)	(0.000)	
Topic Entropy$_{it-1}$		0.036				0.152		
		(0.016)				(0.134)		
is_T2								0.1701***
								(0.068)
Observations	17,264	9,005	100,362	100,362	2,803	42	90,291	4,052
R^2	0.227	0.809	0.052	0.047	0.009	0.218	0.010	0.557

Clustered (persona_id) standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1. *is_T2* is a binary indicator denoting whether a user is in treatment group 2 (open and click)

The findings suggest important implications for campaign strategy and messaging optimization. The persistence of topics with high survival durations highlights the strategic value campaigns place on resonant themes, suggesting that click data not only refines content but also guide long-term messaging strategies. The alignment between candidate topic clusters and AAP-specific topics shows how campaigns balance overarching narratives with individualized exposure. This balance is critical for maintaining message consistency while catering to diverse recipient preferences. The co-occurrence analysis further shows the thematic coherence within emails, suggesting that campaigns intentionally group related topics to enhance message salience and boost recipient engagement.

Overall, the analysis demonstrates that campaigns dynamically adapt their content based on recipient engagement, leveraging sophisticated personalization mechanisms to balance content focus, diversity, and alignment. Doing this at scale is strongly indicative of the use of AI technology in managing the closed loop system that uses user feedback to govern content generation and content distribution.

7. Estimating and Evaluating Predictive Models for Campaign Messaging

To understand the mechanisms driving the observed behaviors used by campaigns and motivated by our auditing approach, we hypothesize a family of models that campaigns are likely to use to predict engagement outcomes, with the aim of identifying models that closely approximate the mechanisms campaigns use to predict and optimize engagement outcomes. While in a real world audit, the the true policy function used by the campaign is unavailable for comparison, the observed performance of predictive models provides insights into the infrastructure and processes campaigns employ.

We test a range of predictive models, including logistic regression, ensemble models, and deep learning models, using all available features we have collected and simulated such as persona demographics, email delivery attributes, and topic distribution. This evaluation covers the entire campaign messaging process: (1) predicting email delivery, (2) modeling AAP engagement, and (3) forecasting topics based on prior interactions[5]. Performance results are summarized in Table 6.

Table 6 Performance Metrics for Campaign Messaging Predictions Across Different Models

	Email Delivery Classification	**Email Engagement Classification**		**Multi-Label Topic Prediction**
Model/Metrics	*Recall*	*Recall*	**Model/Metrics**	*Hamming Accuracy*
Logistic Regression	0.8497 ± 0.0008	0.7072 ± 0.0045	CNN	**0.9448 ± 0.0003**
AdaBoost	0.8560 ± 0.0009	**0.9660 ± 0.0099**	LSTM	0.9122 ± 0.0590
XGBoost	0.8838 ± 0.0006	0.9276 ± 0.0058	Transformer	0.9380 ± 0.0019
SGD	0.8459 ± 0.0015	0.7936 ± 0.0237		
Decision Tree	**0.9414 ± 0.0010**	0.7838 ± 0.0065		
Random Forest	0.9382 ± 0.0014	0.5443 ± 0.0040		
Neural Network	0.9186 ± 0.0073	0.9251 ± 0.0103		

Note: mean ± standard deviation across 5 experimental repeats

The results, presented in Table 6, demonstrates the predictive accuracy of various machine learning models across different tasks. Ensemble models, particularly AdaBoost and XGBoost, excel at predicting engagement outcomes, with AdaBoost achieving the highest recall (0.966). For email delivery, decision tree models perform best, achieving a recall of 0.941, indicating their robustness in identifying successful delivery cases. For multi-label topic prediction, convolutional neural networks (CNNs) yield the highest

[5] Models are trained on 80% of the data and validated on a 20% holdout set. Recall is prioritized for engagement tasks to minimize missed opportunities while Hamming accuracy is used for multi-label topic prediction to assess the model's ability to identify relevant topics and exclude irrelevant ones, reflecting the multi-thematic nature of campaign messaging.

Hamming accuracy (0.945), showing abilities to capture nuanced relationships between multiple concurrent topics within emails.

These findings suggest that although our infrastructure does not record the universe of all campaign messaging and user engagement, under reasonable assumptions, the predictive models we estimate using our telemetry data from our agent system capture critical components in campaign's decision-making process. The success of these models in predicting engagement and topic relevance demonstrates that campaigns likely use similar models as policy functions to optimize their messaging strategies.

The implications of this analysis extend to both academic research and practical applications. For researchers, the ability to simulate campaign behavior using predictive models highlights the value of data-driven approaches in exploring complex adaptive systems. These models provide a lens into the infrastructure of campaign decision-making, offering a foundation for future explorations into personalization and targeted communication strategies.

8. Discussion and Conclusion

This study introduces a comprehensive framework for auditing AI-driven campaign systems by modeling and estimating the latent policy functions that govern content distribution. We introduce a novel analytical infrastructure designed to uncover the “black box" of campaign content management and distribution systems. Our framework is structured as a digital twin, a modular meta model, that simulates interactions between political campaigns (policy systems) and voter responses (user systems). By modeling components such as the policy function (rule-based vs. neural network-based), response function (deterministic vs. stochastic), and the connections between static and dynamic agent features, we provide a flexible and scalable approach to auditing diverse campaign strategies across domains.

Through grid-based simulation experiments, we systematically varied key hyperparameters, including the number of agents, observation duration, and topic selection granularity to establish benchmarks for the minimal agent infrastructure needed to approximate underlying policy functions. Our findings indicate that approximately 300 agents observed over 480 hours are sufficient to recover both rule-based and learned policy dynamics with low divergence, laying the groundwork for real-world deployment of our auditing framework. To mirror this decision-making grid in practice, we constructed over 4,000 Autonomous Agent Personas (AAPs), extending simulated agents' basic demographic and engagement features to include political affiliation, state, and browsing history. This expansion ensured coverage above the simulation-identified threshold while enabling fine-grained descriptive analyses across demographic and market strata.

Empirically, our field experiment during the 2024 U.S. elections both validates and extends our simulation results. By deploying Autonomous Agent Personas (AAPs) with varying demographic attributes and dynamic engagement patterns, we captured how actual campaigns respond to different voter signals. Our findings show that while most candidates employ limited personalization, nearly half of Democratic campaigns distinguish themselves through robust use of dynamic content adaptation and engagement tracking.

These campaigns effectively leverage engagement signals, such as link clicks, to craft more cohesive and personalized narratives. In contrast, other Democratic and Republican campaigns demonstrate a generalized approach, with limited adoption of advanced techniques. Our predictive modeling exercise further demonstrates the potential of machine learning models like AdaBoost and CNN in predicting granular engagement behaviors and dynamically adapting content. Such infrastructure and predictive frameworks would become more valuable as content strategies increasingly leverage AI.

Practically, our data-driven audit infrastructure opens the "black box" of campaign messaging and provides a method to measure the extent of targeting employed. Given the widespread use of targeted messaging, recommendations, and pricing across industries and organizations, our approach serves as a proof of concept for evaluating opaque algorithmic behavior. In this context, our study aligns with ongoing research on auditing algorithmic systems like YouTube and Google Search (Srba et al. 2023, Hannak et al. 2013).

Applying our audit framework, we find that a significant proportion of campaigns do not personalize the messaging based on observable demographics, highlighting an opportunity for AI-driven improvements in targeting. Some sophisticated Democratic campaigns do leverage granular engagement data to dynamically adapt content. However, there remains substantial potential for more advanced algorithms and Generative AI to further refine audience segmentation and content personalization. Finally, our findings suggest that predictive modeling frameworks, even when constrained by observational data, can reveal key aspects of adaptive communication systems. Such infrastructures pave the way for greater transparency and optimization in communication, both in politics and beyond.

Our study is not without limitations and caveats. First, AAPs are unable to perform high-cost behaviors such as donations, which may be significant triggers for content adaptation, as only human U.S. citizens can legally contribute to political candidates. Second, because AAPs subscribe to campaign emails by design, our analysis focuses on users who are already somewhat engaged; the findings may not generalize to efforts aimed at mobilizing passive or disengaged individuals. Moreover, as a next step, it would be interesting to compare the messaging in emails to other communication channels such political ads on TV or online. A final caveat is that we do not aim to uncover dark patterns in email messaging. We take a neutral stance to understand the role of technology and personalization in political emails. Our analysis does not preclude the use of dark patterns as previously highlighted in Posner et al. (2023).

In conclusion, our research advances a principled, modular, and systematic approach to auditing AI-driven political messaging, bridging the gap between rigorous simulation-based analysis and real-world campaign dynamics. By explicitly connecting simulated agent systems and empirical validation through predictive modeling, our methodology provides a robust foundation for evaluating and optimizing agent infrastructures. This enables more transparent and accountable communication strategies, not only in political campaigns but across diverse domains where AI personalization plays a critical role.

References

Adobe Blog (2019) If you think email is dead, think again. URL `https://business.adobe.com/blog/perspectives/if-you-think-email-is-dead-think-again`, accessed: 2024-10-21.

Baviera T, Sánchez-Junqueras J, Rosso P (2022) Political advertising on social media: Issues sponsored on facebook ads during the 2019 general elections in spain. *Communication & Society* 35(3):39–49.

Beraldo D, Milan S, de Vos J, Agosti C, Sotic BN, Vliegenthart R, Kruikemeier S, Otto LP, Vermeer SA, Chu X, et al. (2021) Political advertising exposed: Tracking facebook ads in the 2021 dutch elections. *Internet Policy Review* 11.

Bouton L, Cagé J, Dewitte E, Pons V (2022) Small campaign donors. Technical report, National Bureau of Economic Research.

Chen B, Borah P, Dahlke R, Lukito J (2024) Battle for inbox and bucks: Comparing email fundraising strategies of donald trump and joe biden in the 2020 us presidential election. *Journal of Quantitative Description: Digital Media* 4.

Davenport T, Guha A, Grewal D, Bressgott T (2020) How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science* 48:24–42.

Delre SA, Jager W, Bijmolt TH, Janssen MA (2007) Targeting and timing promotional activities: An agent-based model for the takeoff of new products. *Journal of business research* 60(8):826–835.

DeLuca K, Curiel JA (2023) Validating the applicability of bayesian inference with surname and geocoding to congressional redistricting. *Political Analysis* 31(3):465–471.

Favaro P, Vedaldi A (2021) Adaboost. *Computer Vision: A Reference Guide*, 36–40 (Springer).

Fossen BL, Kim D, Schweidel DA, Thomadsen R (2022) The role of slant and message consistency in political advertising effectiveness: evidence from the 2016 presidential election. *Quantitative Marketing and Economics* 20(1):1–37.

Fryer Jr RG, Levitt SD (2004) The causes and consequences of distinctively black names. *The Quarterly Journal of Economics* 119(3):767–805.

Garcia R (2005) Uses of agent-based modeling in innovation/new product development research. *Journal of product innovation management* 22(5):380–398.

Goldenberg J, Libai B, Muller E (2001) Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review* 9(3):1–18.

Gordon BR, Lovett MJ, Luo B, Reeder III JC (2023) Disentangling the effects of ad tone on voter turnout and candidate choice in presidential elections. *Management Science* 69(1):220–243.

Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, et al. (2018) Recent advances in convolutional neural networks. *Pattern recognition* 77:354–377.

Guzman J, Oh JJ, Sen A (2020) What motivates innovative entrepreneurs? evidence from a global field experiment. *Management science* 66(10):4808–4819.

Hannak A, Sapiezynski P, Molavi Kakhki A, Krishnamurthy B, Lazer D, Mislove A, Wilson C (2013) Measuring personalization of web search. *Proceedings of the 22nd international conference on World Wide Web*, 527–538.

Heppenstall A, Evans A, Birkin M (2006) Using hybrid agent-based systems to model spatially-influenced retail markets. *Journal of Artificial Societies and Social Simulation* 9(3).

Kullback S, Leibler RA (1951) On information and sufficiency. *The annals of mathematical statistics* 22(1):79–86.

Lorenzo-Dus N, Blitvich PGC (2013) Get involved! communication and engagement in the 2008 obama presidential e-campaign. *Media talk and political elections in Europe and America*, 229–251 (Springer).

Mathur A, Wang A, Schwemmer C, Hamin M, Stewart BM, Narayanan A (2023) Manipulative tactics are the norm in political emails: Evidence from 300k emails from the 2020 us election cycle. *Big Data & Society* 10(1):20539517221145371.

Moser E (2024) How u.s. presidential campaigns are targeting digital ads by zip code. Accessed: 2024-12-10.

North MJ, Macal CM, Aubin JS, Thimmapuram P, Bragen M, Hahn J, Karr J, Brigham N, Lacy ME, Hampton D (2010) Multiscale agent-based consumer market modeling. *Complexity* 15(5):37–47.

Petrova M, Sen A, Yildirim P (2021) Social media and political contributions: The impact of new technology on political competition. *Management Science* 67(5):2997–3021.

Peukert C, Sen A, Claussen J (2024) The editor and the algorithm: Recommendation technology in online news. *Management science* 70(9):5816–5831.

Posner N, Simonov A, Mrkva K, Johnson EJ (2023) Dark defaults: How choice architecture steers political campaign donations. *Proceedings of the National Academy of Sciences* 120(40):e2218385120.

Rand W, Rust RT (2011) Agent-based modeling in marketing: Guidelines for rigor. *International Journal of research in Marketing* 28(3):181–193.

Simchon A, Edwards M, Lewandowsky S (2024) The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS nexus* 3(2):pgae035.

Spenkuch JL, Toniatti D (2018) Political advertising and election results. *The Quarterly Journal of Economics* 133(4):1981–2036.

Srba I, Moro R, Tomlein M, Pecher B, Simko J, Stefancova E, Kompan M, Hrckova A, Podrouzek J, Gavornik A, et al. (2023) Auditing youtube's recommendation algorithm for misinformation filter bubbles. *ACM Transactions on Recommender Systems* 1(1):1–33.

Sun T, Yuan Z, Li C, Zhang K, Xu J (2024) The value of personal data in internet commerce: A high-stakes field experiment on data regulation policy. *Management Science* 70(4):2645–2660.

Valogianni K, Padmanabhan B, Qiu L (2023) Causal abm: a methodology for learning plausible causal models using agent-based modeling. *Available at SSRN 4343647* .

Wattal S, Telang R, Mukhopadhyay T, Boatwright P (2012) What's in a "name"? impact of use of customer information in e-mail advertisements. *Information Systems Research* 23(3-part-1):679–697.

Wernerfelt N, Tuchman A, Shapiro BT, Moakler R (2024) Estimating the value of offsite tracking data to advertisers: Evidence from meta. *Marketing Science* .

Zhang J, Adomavicius G, Gupta A, Ketter W (2020) Consumption and performance: Understanding longitudinal dynamics of recommender systems via an agent-based simulation framework. *Information Systems Research* 31(1):76–101.

Zhang Q, Lu J, Jin Y (2021) Artificial intelligence in recommender systems. *Complex & Intelligent Systems* 7(1):439–457.

Appendix A: Notation

Table A1 Notation and Descriptions for the System Model

Notation	Description
C	The set of all campaigns in the policy system.
AAP^n	The nth agent (subscribed user) in the campaign system, where $n = 1, \ldots, N$.
T^n	The total number of discrete time steps during which agent n is observed.
t	Discrete time index for each agent, $t = 1, \ldots, T^n$.
β_t^n	The state of agent n at time t, which encapsulates the engagement score and demographic attributes (e.g., gender, location).
H	The full historical record of engagement data.
ES_t	Engagement score computed over the previous m emails, defined as $ES_t = \frac{1}{m}\sum_{i=t-m+1}^{t}\sum_{y\in Y}\alpha_y\,\mathbb{1}\{Y_i = y\}$, where $\mathbb{1}\{Y_i = y\}$ is the indicator function and α_y reflects its engagement intensity for outcome y.
$S(\beta, ES_t)$	The email scheduling function that determines the waiting time until the next email based on the current state β and the engagement score ES_t.
ω	The waiting time until the next email, defined as $\omega = \omega_0 \cdot \exp\{-\lambda \cdot ES_t\}$.
ω_0	Baseline waiting time parameter.
λ	Sensitivity parameter controlling how ES_t influences the waiting time.
Y_i	Engagement outcome (e.g., 1 to 4) for the ith email.
x_t^n	The action (selected topic) for agent n at time t, drawn from the topic set $\mathcal{A}$.
$\mathcal{A}$	The predefined topic space (e.g., *Healthcare*, *Economy*).
$\pi_0(x)$	The baseline (prior) probability of topic x.
$\phi(x, \beta_t^n)$	The alignment function that quantifies the match between topic x and the user's state β_t^n.
γ	Sensitivity parameter governing the influence of $\phi(x, \beta_t^n)$ in the rule-based policy update.
θ	The parameters of the learned policy network.
$f(y \mid \beta, x, c)$	The response function giving the probability of observing outcome y for a user in state β when receiving topic x from campaign c.
$O(y)$	The cost associated with outcome y.
δ	Sensitivity parameter in the campaign selection (logit) model.
ρ	The scalar between 0 and 1 that quantifies a user's tendency to exhibit higher engagement.
$S_t^n(c)$	The probability that agent n selects campaign c at time t, given by $S_t^n(c) = \frac{\exp\{\delta\cdot ES_t^{n,c}\}}{\sum_{c'\in C^n}\exp\{\delta\cdot ES_t^{n,c'}\}}$.
$\Pi^{(c)}_{\text{rule-based}}(x \mid \beta_t^n; \gamma)$	The rule-based policy distribution over topics for campaign c, given by $\frac{\pi_0(x)\ \exp\{\gamma\cdot\phi(x,\beta_t^n)\}}{\sum_{x'\in\mathcal{A}}\pi_0(x')\ \exp\{\gamma\cdot\phi(x',\beta_t^n)\}}$.
$\Pi^{(c)}_{\text{learned}}(x \mid \beta_t^n; \theta)$	The learned policy distribution for campaign c, where a neural network maps the combined feature vector $[\pi(x \mid \beta_t^n), \beta_t^n]$ to topic probabilities via a softmax layer.

Appendix B: Agents and AAP's: Using the simulation to guide the field experiment

Section 2 introduced a meta-model and a methodological approach to approximate the latent policy function used by the policy system to distribute content to agents based on their static features and dynamic engagement behavior. The auditor of a real-world AI content distribution system can use the simulation model to design or evaluate an agent system, starting with the pilot data collection to identify which static and dynamic user features the policy system appears to leverage when adaptively distributing content to users. For example, if the real-world campaign system is addressing users with two static features (gender and location) and four dynamic engagement behaviors, the simulation model instance described in the paper shows that approximately 300 agents are required to approximate the latent policy function. Since the auditor is also interested in conducting a descriptive analysis of how agents representing different user types receive and respond to contents, we convert agents into autonomous agent personas (AAPs), where each combination of dynamic and static features is represented as a distinct AAP. In the simulation example in Section 2, each agent is characterized by two static features (gender: male or female and location: urban, suburban, rural) and one of four dynamic engagement features. Thus, there are 24 combinations (2×3×4), yielding 24 AAPs. Hence, the 300 agents recommended by the simulation would result in 1,200 AAPs being deployed in the field experiment.

The political campaign setting described in the field experiment in Section 4 is richer than the simulation example developed in Section 3. Each agent has multiple static features (gender, political affiliation, zip codes, state, and browsing condition) and one of three dynamic engagement behaviors. As described above, for the purposes of conducting descriptive analyses of how the campaign engages with users with different combinations of static and dynamic features, we deployed 4,032 AAPs (the product of the combinations of static and dynamic features described in Section 4.1). The simulation model can be used to estimate the minimum number of AAPs necessary for approximating the policy function, possibly fewer than the full set, depending on the policy system's use of features. As discussed in Section 4, while the 4,032 AAPs and the 20,160 signups are essential to understand the campaign strategy, the accuracy of the predictive model is determined by the dynamic engagement features, which is a subset of the AAP features.

In summary, AAPs enable the collection of rich and granular data that provides the basis for the auditor to understand the behavior of the campaign with respect to dynamic and static features of the agent. It also provides the data required to recover the latent policy function that the campaign is using and to demonstrate whether it is using all or only a subset of the features that the agent exposes.

Appendix C: Distribution of User Email Behavior

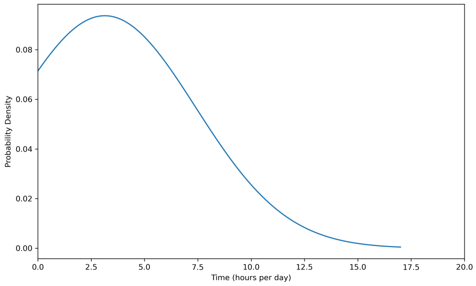


Figure C1 Email Checking Frequency Distribution

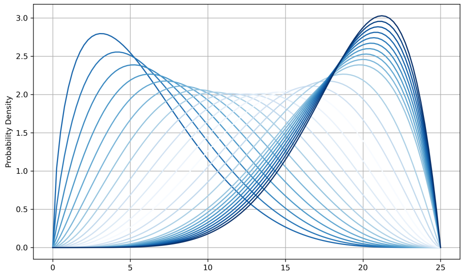


Figure C2 Beta Distribution of Engagement

Appendix D: Metrics Specification

To capture topic evolution, we calculate both the Cosine Similarity and the Percentage of New Instances between subtopics within a given topic. Cosine Similarity measures the semantic similarity of subtopic distributions across consecutive periods, while the Percentage of New Instances quantifies the introduction of new subtopics over time. The cosine similarity for a AAP i , topic τ, and time period t is given by:

$$\text{Cosine Similarity}_{i\tau} = \frac{\mathbf{v}_{i\tau t} \cdot \mathbf{v}_{i\tau(t-1)}}{\|\mathbf{v}_{i\tau t}\| \|\mathbf{v}_{i\tau(t-1)}\|} \tag{D.1}$$

where $\mathbf{v}_{i\tau t}$ is the embedding vector [6] for the subtopics of topic τ AAP i during period t, $\mathbf{v}_{i\tau(t-1)}$ is the embedding vector for the subtopics of topic τ received by APP i during period $t-1$.

The percentage of new instances for a AAP i , topic τ, and time period t is given by:

$$\text{New Instances Percentage}_{i\tau t} = \frac{\left|\mathbf{C}_{i\tau t} \setminus \mathbf{C}_{i\tau(t-1)}\right|}{|\mathbf{C}_{i\tau t}|} \times 100, \tag{D.2}$$

$\mathbf{C}_{i\tau t}$ is the set of unique subtopics in topic τ for AAP i during time t and $\mathbf{C}_{i\tau(t-1)}$ is the set of unique subtopics in topic τ for AAP i during time $t-1$. $|\mathbf{C}_{i\tau t} \setminus \mathbf{C}_{i\tau(t-1)}|$ represents the count of subtopics in $\mathbf{C}_{i\tau t}$ that are not present in $\mathbf{C}_{i\tau(t-1)}$, and $|\mathbf{C}_{i\tau t}|$ is the total count of subtopics in $\mathbf{C}_{i\tau t}$.

A high Cosine Similarity indicates that subtopics are converging and becoming more focused on a central theme over time while a high Percentage of New Instances suggests that within the same topic, more novel subtopics are being introduced over time, reflecting greater depth development in the messaging strategy.

To estimate the content distribution from the candidate perspective, we quantify the persistence of topics where we calculate the topic survival duration for candidate c and topic τ from candidate c as follows:

$$\begin{aligned} &\text{Survival Duration}_{c\tau}(t) = (t - \operatorname*{argmin}_t\{\mathbb{1}(I_{c\tau t} = 1)\} \quad \text{if} \quad \mathbb{1}(I_{c\tau t} = 1), \\ &I_{c\tau t} = \begin{cases} 1 & \text{if topic } \tau \text{ was sent by candidate } c \text{ at time } t \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{D.3}$$

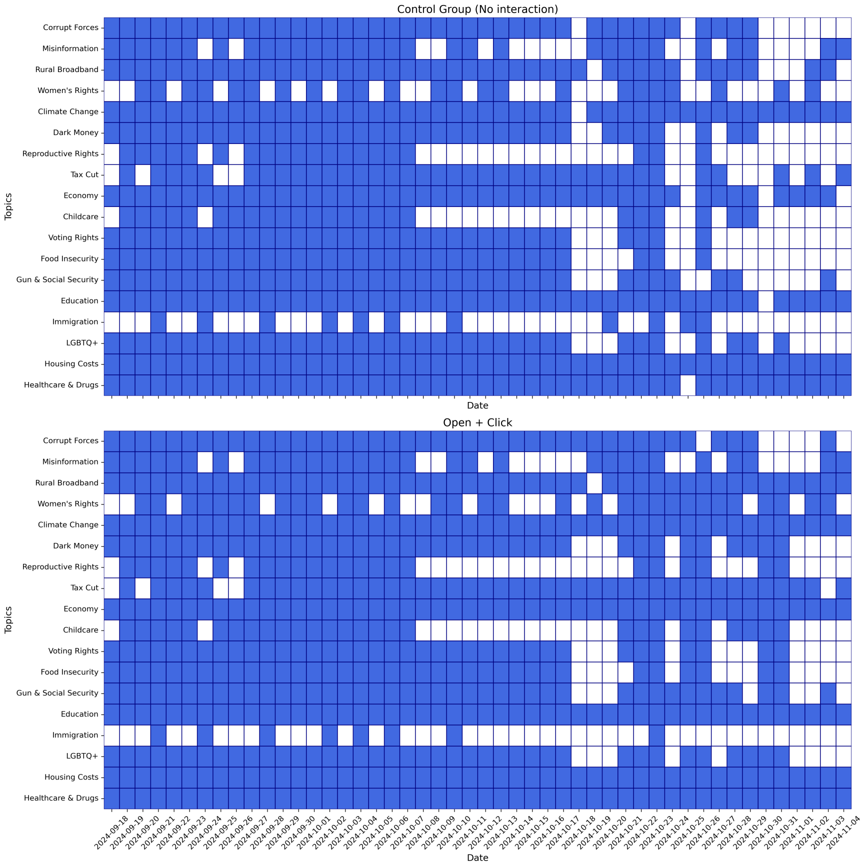
here t represents the current time point and $\operatorname*{argmin}_t\{\mathbb{1}(I_{c\tau t} = 1)\}$ represents the earliest time topic τ was sent by candidate c. We visualize the survival of topics over time for the Control Group and Open + Click treatment in Figure D1. The heatmap illustrates how topic survival duration varies across time and treatment groups.

To measure the alignment of topics between AAPs and the candidate, we compute the cosine similarity between the topic embeddings for all topics τ sent by candidate c on day t and those received by AAP i on the same day. The unique set of topics sent by the candidate on day t is referred to as the *Topic Clusters*. This metric captures the semantic alignment between the overall focus of the candidate's messaging and the topics actually received by the persona. A higher cosine similarity value indicates that the AAP's experience closely mirrors the candidate's central thematic emphasis, reflecting stronger alignment between the campaign's overarching strategy and the individual's exposure. The cosine similarity follows the logic in Model D.1, specified as

$$\text{Cosine Similarity}_{cit} = \frac{\mathbf{v}_{ct} \cdot \mathbf{v}_{it}}{\|\mathbf{v}_{ct}\| \|\mathbf{v}_{it}\|} \tag{D.4}$$

[6] We extracted all embedding vectors using the Sentence-Transformer model, specifically `all-MiniLM-L6-v2`, , which produces embeddings with a dimensionality of 384.

Figure D1 Topic Survival Heatmap



Note. Note: Each row represents a distinct topic, while each column corresponds to a date within the analysis period. Blue cells indicate topic survival (1), while white cells indicate the absence of the topic (0).

where $\mathbf{v}_{ct}$ represents the embedding vector of all the topics sent by candidate c on day t and $\mathbf{v}_{i,t}$ represents the embedding vector of all the topics received by AAP i on day t.

Lastly, to analyze thematic relationships within campaign messages, we construct a topic co-occurrence probability that quantifies the likelihood of pairs of topic pairs appear together in the same email. For candidate c, AAP i on date t, the co-occurrence of topics τ_1 and τ_2 is defined as:

$$\begin{aligned} P(\tau_1, \tau_2) &= \frac{\text{Co-occurrence Count}(\tau_1, \tau_2)}{\sum_{\tau_1'} \sum_{\tau_2'} \text{Co-occurrence Count}(\tau_1', \tau_2')} \\ \text{where} \quad \text{Co-occurrence}(\tau_1, \tau_2) &= \sum t \sum_i \mathbb{1}(\tau_1 \in \mathbf{T}_{cit} \wedge \tau_2 \in \mathbf{T}_{cit}) \end{aligned} \tag{D.5}$$

where $\mathbf{T}_{cit}$ is the set of topics included in the email sent to AAP i by candidate c on date t, $\mathbb{1}(\cdot)$ is an indicator function that equals 1 if both τ_1 and τ_2 are present in the email, and 0 otherwise. $\sum_{\tau_1'} \sum_{\tau_2'} \text{Co-occurrence Count}(\tau_1', \tau_2')$ represents the total number of topic co-occurrences across all topic pairs. A high co-occurrence probability indicates that this pair is relatively more frequent to appear together compared to others.

We utilize these metrics as dependent variables to examine how engagement levels influence topic diversity, alignment, and persistence:

$$Y_{it} = \beta_0 + \beta_1 \cdot \text{click pct}_{it-1} + X_{it} + \text{AAP}_i + \text{period}_t + \epsilon_{it} \tag{D.6}$$

$Y_{it} \in$ [topic entropy, cosine similarity, percentage of new instances, survival duration, or co-occurrence probability] is one of the dependent variables, representing a specific aspect of campaign messaging. Click pct_{it-1} captures the engagement level of persona i from the previous period. X_{it} are controls including topic entropy_{t-1} that controls for baseline topic diversity in the preceding period and, email counts_{t-1} controls for the number of emails sent in the preceding period. AAP i and Date t fixed effects are also included.

Appendix E: Heterogeneous Effects

Table E1 Impact of Cookie Type and Treatment Interaction on Email Counts

	Dependent variable: Email Counts	
	Democrat	Republican
	(1)	(2)
Dem_cookie		-0.009
		(0.042)
Dem_cookie x Open Only		-0.087
		(0.058)
Dem_cookie x Open + Click		-0.064
		(0.060)
None_cookie	-0.054	-0.005
	(0.333)	(0.039)
None_cookie x Open Only	-0.849*	-0.050
	(0.455)	(0.057)
None_cookie x Open + Click	-2.068***	-0.044
	(0.675)	(0.058)
Rep_cookie	-0.120	
	(0.337)	
Rep_cookie x Open Only	-0.282	
	(0.463)	
Rep_cookie x Open + Click	-1.440**	
	(0.664)	
Mix_cookie	-0.423	-0.067*
	(0.338)	(0.037)
Mix_cookie x Open Only	-0.094	-0.068
	(0.472)	(0.053)
Mix_cookie x Open + Click	-4.709***	-0.036
	(0.646)	(0.055)
Open Only	-0.219	0.041
	(0.328)	(0.042)
Open + Click	8.927***	0.051
	(0.485)	(0.043)
Time Effect	Yes	Yes
Observations	96,768	96,768
R^2	0.301	0.006

Clustered (persona_id) standard-errors in parentheses.
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table E2 Impact on Email Counts by Location

	Dependent variable: Email Counts			
Model:	Democrat	Republican	Democrat	Republican
	(1)	(2)	(3)	(4)
Constant	13.07***	0.3373***	9.738***	0.2948***
	(0.5252)	(0.0340)	(0.2296)	(0.0265)
None_cookie	-1.407**	0.0025		
	(0.7070)	(0.0479)		
Rep_cookie	-0.5491			
	(0.7118)			
Mix_cookie	-2.248***	-0.0881**		
	(0.6279)	(0.0436)		
Urban	0.2697	-0.0111	0.6652**	-0.0155
	(0.6417)	(0.0417)	(0.2865)	(0.0336)
Suburban	-0.4137	0.0347	-0.0977	0.0099
	(0.7340)	(0.0514)	(0.3267)	(0.0373)
None_cookie × Urban	0.3763	-0.0108		
	(0.8580)	(0.0586)		
Rep_cookie × Urban	-0.3346			
	(0.8658)			
Mix_cookie × Urban	0.0769	-0.0297		
	(0.7788)	(0.0530)		
None_cookie × Suburban	0.7727	-0.1346**		
	(1.002)	(0.0673)		
Rep_cookie × Suburban	0.0898			
	(0.9966)			
Mix_cookie × Suburban	0.7397	0.0043		
	(0.8936)	(0.0650)		
Dem_cookie		-0.0463		
		(0.0482)		
Dem_cookie × Urban		0.0080		
		(0.0592)		
Dem_cookie × Suburban		-0.0684		
		(0.0678)		
Open Only			-0.3445	-0.0048
			(0.3178)	(0.0389)
Open + Click			7.178***	0.0335
			(0.4770)	(0.0397)
Open Only × Urban			-0.4453	-0.0055
			(0.3951)	(0.0480)
Open + Click × Urban			-0.6525	-0.0056
			(0.5891)	(0.0488)
Open Only × Suburban			0.1693	-0.0123
			(0.4589)	(0.0540)
Open + Click × Suburban			0.0844	-0.0624
			(0.6719)	(0.0554)
Observations	96,768	96,768	96,768	96,768
R^2	0.01104	0.00551	0.21777	0.00078

Clustered (persona_id) standard-errors in parentheses.
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table E3 Impact on Email Counts by Gender

	Dependent variable: Email Counts			
Model:	Democrat	Republican	Democrat	Republican
	(1)	(2)	(3)	(4)
Constant	12.83***	0.3273***	9.824***	0.2675***
	(0.3568)	(0.0253)	(0.1587)	(0.0183)
None_cookie	-0.6619	-0.0446		
	(0.4896)	(0.0340)		
Rep_cookie	-0.5394			
	(0.4817)			
Mix_cookie	-2.010***	-0.0737**		
	(0.4368)	(0.0320)		
Male	0.5271	0.0263	0.4429*	0.0440
	(0.5203)	(0.0353)	(0.2377)	(0.0277)
None_cookie × Male	-0.7284	0.0161		
	(0.6984)	(0.0478)		
Rep_cookie × Male	-0.3092			
	(0.7010)			
Mix_cookie × Male	-0.0276	-0.0565		
	(0.6416)	(0.0445)		
Dem_cookie		-0.0332		
		(0.0334)		
Dem_cookie × Male		-0.0523		
		(0.0482)		
Open Only			-0.2918	0.0243
			(0.2255)	(0.0272)
Open + Click			6.913***	0.0415
			(0.3312)	(0.0271)
Open Only × Male			-0.4660	-0.0699*
			(0.3291)	(0.0390)
Open + Click × Male			-0.0802	-0.0528
			(0.4823)	(0.0397)
Observations	96,768	96,768	96,768	96,768
R^2	0.01089	0.00418	0.21742	0.00087

Clustered (persona_id) standard-errors in parentheses.
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table E4 Impact on Email Counts by Income

	Dependent variable: Email Counts			
Model:	Democrat	Republican	Democrat	Republican
	(1)	(2)	(3)	(4)
Constant	12.91***	0.3472***	9.970***	0.2840***
	(0.3645)	(0.0247)	(0.1704)	(0.0191)
None_cookie	-0.8112*	-0.0655**		
	(0.4876)	(0.0331)		
Rep_cookie	-0.6744			
	(0.4930)			
Mix_cookie	-1.739***	-0.1005***		
	(0.4527)	(0.0314)		
Low Income	0.3697	-0.0136	0.1511	0.0109
	(0.5206)	(0.0353)	(0.2383)	(0.0277)
None_cookie × Low Income	-0.4298	0.0578		
	(0.6987)	(0.0478)		
Rep_cookie × Low Income	-0.0392			
	(0.7011)			
Mix_cookie × Low Income	-0.5709	-0.0027		
	(0.6421)	(0.0445)		
Dem_cookie		-0.0711**		
		(0.0334)		
Dem_cookie × Low Income		0.0234		
		(0.0483)		
Open Only			-0.3313	-0.0035
			(0.2347)	(0.0268)
Open + Click			6.741***	0.0152
			(0.3446)	(0.0278)
Open Only × Low Income			-0.3870	-0.0143
			(0.3294)	(0.0390)
Open + Click × Low Income			0.2629	-0.0003
			(0.4825)	(0.0397)
Observations	96,768	96,768	96,768	96,768
R^2	0.01050	0.00392	0.21729	0.00034

Clustered (persona_id) standard-errors in parentheses.
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table E5 Impact on Email Counts by Race

	Dependent variable: Email Counts			
Model:	Democrat	Republican	Democrat	Republican
	(1)	(2)	(3)	(4)
Constant	12.84***	0.2968***	9.940***	0.2604***
	(0.4440)	(0.0270)	(0.2049)	(0.0224)
None_cookie	-0.8070	-0.0067		
	(0.5892)	(0.0381)		
Rep_cookie	-0.3566			
	(0.6096)			
Mix_cookie	-1.661***	-0.0624*		
	(0.5519)	(0.0357)		
White	0.2758	0.0900**	0.1690	0.0920***
	(0.6106)	(0.0428)	(0.3002)	(0.0344)
Non-white/black	0.5063	0.0411	0.1489	-0.0047
	(0.6587)	(0.0410)	(0.2821)	(0.0320)
None_cookie × White	-0.1834	-0.0476		
	(0.8317)	(0.0578)		
Rep_cookie × White	-0.6633			
	(0.8292)			
Mix_cookie × White	-0.4733	-0.0839		
	(0.7601)	(0.0539)		
None_cookie × Non-white/black	-0.4737	-0.0420		
	(0.8680)	(0.0567)		
Rep_cookie × Non-white/black	-0.3486			
	(0.8885)			
Mix_cookie × Non-white/black	-0.6171	-0.0346		
	(0.8091)	(0.0529)		
Dem_cookie		-0.0069		
		(0.0395)		
Dem_cookie × White		-0.0882		
		(0.0600)		
Dem_cookie × Non-white/black		-0.0692		
		(0.0564)		
Open Only			-0.3002	0.0157
			(0.2907)	(0.0324)
Open + Click			6.874***	0.0363
			(0.4116)	(0.0324)
Open Only × White			-0.3238	-0.0889*
			(0.4055)	(0.0482)
Open + Click × White			-0.3458	-0.0818*
			(0.5767)	(0.0488)
Open Only × Non-white/black			-0.3499	0.0098
			(0.4054)	(0.0460)
Open + Click × Non-white/black			0.3426	0.0182
			(0.6002)	(0.0469)
Observations	96,768	96,768	96,768	96,768
R^2	0.01065	0.00479	0.21749	0.00202

Clustered (persona_id) standard-errors in parentheses.
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table E6 Impact of Name on Email Counts

	Dependent variable: Email Counts	
	Democract	Republican
	(1)	(2)
Name Field Required	1093.286	-37.404
	(1173.236)	(186.944)
Observations	574	77
R^2	0.035	0.001

Clustered (persona_id) standard-errors in parentheses.
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table E7 Experiment Impact on Subject Counts

	Dependent variable: Subject Counts			
Model:	Democrat	Republican	Democrat	Republican
	(1)	(2)	(3)	(4)
Constant	13.15***	1.344***	10.14***	1.322***
	(0.2602)	(0.0185)	(0.1160)	(0.0149)
None_cookie	-0.8474**	-0.0210		
	(0.3483)	(0.0246)		
Rep_cookie	-0.6642*			
	(0.3494)			
Mix_cookie	-1.914***	-0.1070***		
	(0.3201)	(0.0191)		
Dem_cookie		0.0071		
		(0.0271)		
Open Only			-0.4464***	-0.0091
			(0.1622)	(0.0206)
Open + Click			6.893***	-0.0068
			(0.2386)	(0.0211)
Observations	95,565	21,388	95,565	21,388
R^2	0.00928	0.00467	0.22201	3.63×10^{-5}

Clustered (persona_id) standard-errors in parentheses.
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table E8 Impact on Subject Counts by Gender

	Dependent variable: Subject Counts			
Model:	Democrat	Republican	Democrat	Republican
	(1)	(2)	(3)	(4)
Constant	12.90***	1.345***	9.892***	1.302***
	(0.3567)	(0.0287)	(0.1568)	(0.0176)
None_cookie	-0.5091	0.0002		
	(0.4887)	(0.0381)		
Rep_cookie	-0.5153			
	(0.4801)			
Mix_cookie	-1.935***	-0.1046***		
	(0.4366)	(0.0297)		
Male	0.5002	-0.0023	0.4957**	0.0369
	(0.5198)	(0.0373)	(0.2312)	(0.0291)
None_cookie × Male	-0.6762	-0.0390		
	(0.6962)	(0.0496)		
Rep_cookie × Male	-0.2975			
	(0.6984)			
Mix_cookie × Male	0.0422	-0.0053		
	(0.6393)	(0.0383)		
Dem_cookie		-0.0317		
		(0.0350)		
Dem_cookie × Male		0.0837		
		(0.0548)		
Open Only			-0.1760	0.0258
			(0.2235)	(0.0279)
Open + Click			6.967***	0.0050
			(0.3283)	(0.0268)
Open Only × Male			-0.5418*	-0.0686*
			(0.3239)	(0.0404)
Open + Click × Male			-0.1482	-0.0211
			(0.4766)	(0.0418)
Observations	95,565	21,388	95,565	21,388
R^2	0.01004	0.00591	0.22261	0.00057

Clustered (persona_id) standard-errors in parentheses.
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table E9 Impact on Subject Counts by Location

	Dependent variable: Subject Counts			
Model:	Democrat	Republican	Democrat	Republican
	(1)	(2)	(3)	(4)
Constant	13.07***	0.3373***	9.738***	0.2948***
	(0.5252)	(0.0340)	(0.2296)	(0.0265)
Rep_cookie	-0.5491			
	(0.7118)			
None_cookie	-1.407**	0.0025		
	(0.7070)	(0.0479)		
Mix_cookie	-2.248***	-0.0881**		
	(0.6279)	(0.0436)		
Urban	0.2697	-0.0111	0.6652**	-0.0155
	(0.6417)	(0.0417)	(0.2865)	(0.0336)
Suburban	-0.4137	0.0347	-0.0977	0.0099
	(0.7340)	(0.0514)	(0.3267)	(0.0373)
Rep_cookie × Urban	-0.3346			
	(0.8658)			
None_cookie × Urban	0.3763	-0.0108		
	(0.8580)	(0.0586)		
Mix_cookie × Urban	0.0769	-0.0297		
	(0.7788)	(0.0530)		
Rep_cookie × Suburban	0.0898			
	(0.9966)			
None_cookie × Suburban	0.7727	-0.1346**		
	(1.002)	(0.0673)		
Mix_cookie × Suburban	0.7397	0.0043		
	(0.8936)	(0.0650)		
Dem_cookie		-0.0463		
		(0.0482)		
Dem_cookie × Urban		0.0080		
		(0.0592)		
Dem_cookie × Suburban		-0.0684		
		(0.0678)		
Open Only			-0.3445	-0.0048
			(0.3178)	(0.0389)
Open + Click			7.178***	0.0335
			(0.4770)	(0.0397)
Open Only × Urban			-0.4453	-0.0055
			(0.3951)	(0.0480)
Open + Click × Urban			-0.6525	-0.0056
			(0.5891)	(0.0488)
Open Only × Suburban			0.1693	-0.0123
			(0.4589)	(0.0540)
Open + Click × Suburban			0.0844	-0.0624
			(0.6719)	(0.0554)
Observations	96,768	96,768	96,768	96,768
R^2	0.01104	0.00551	0.21777	0.00078

Clustered (persona_id) standard-errors in parentheses.
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table E10 Impact on Subject Counts by Income

	Dependent variable: Subject Counts			
Model:	Democrat (1)	Republican (2)	Democrat (3)	Republican (4)
Constant	12.97***	1.346***	10.07***	1.323***
	(0.3632)	(0.0247)	(0.1658)	(0.0202)
Rep_cookie	-0.6348			
	(0.4903)			
None_cookie	-0.6364	-0.0289		
	(0.4855)	(0.0341)		
Mix_cookie	-1.638***	-0.1050***		
	(0.4499)	(0.0261)		
Low Income	0.3652	-0.0044	0.1402	-0.0029
	(0.5201)	(0.0371)	(0.2320)	(0.0298)
Rep_cookie × Low Income	-0.0591			
	(0.6985)			
None_cookie × Low Income	-0.4218	0.0152		
	(0.6965)	(0.0494)		
Mix_cookie × Low Income	-0.5529	-0.0041		
	(0.6399)	(0.0382)		
Dem_cookie		-0.0076		
		(0.0356)		
Dem_cookie × Low Income		0.0292		
		(0.0541)		
Open Only			-0.2522	-0.0266
			(0.2312)	(0.0263)
Open + Click			6.743***	-0.0040
			(0.3401)	(0.0298)
Open Only × Low Income			-0.3882	0.0356
			(0.3243)	(0.0411)
Open + Click × Low Income			0.3017	-0.0055
			(0.4768)	(0.0423)
Observations	95,565	21,388	95,565	21,388
R^2	0.00961	0.00479	0.22246	0.00026

Clustered (persona_id) standard-errors in parentheses.
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

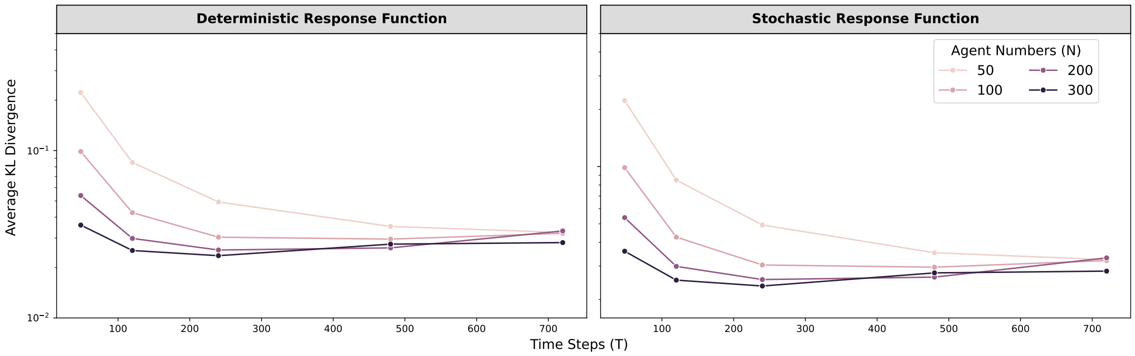
Table E11 Impact on Subject Counts by Race

	Dependent variable: Subject Counts			
Model:	Democrat (1)	Republican (2)	Democrat (3)	Republican (4)
Constant	12.91***	1.297***	10.06***	1.308***
	(0.4439)	(0.0237)	(0.1973)	(0.0219)
Rep_cookie	-0.3160			
	(0.6065)			
None_cookie	-0.6346	0.0090		
	(0.5861)	(0.0340)		
Mix_cookie	-1.561***	-0.0625***		
	(0.5508)	(0.0240)		
White	0.2420	0.0806*	0.1621	0.0392
	(0.6106)	(0.0417)	(0.2903)	(0.0351)
Non-white/black	0.4867	0.0508	0.0898	-0.0056
	(0.6578)	(0.0413)	(0.2753)	(0.0335)
Rep_cookie × White	-0.6596			
	(0.8258)			
None_cookie × White	-0.1630	-0.0501		
	(0.8287)	(0.0550)		
Mix_cookie × White	-0.4318	-0.0831**		
	(0.7570)	(0.0420)		
Rep_cookie × Non-white/black	-0.3851			
	(0.8850)			
None_cookie × Non-white/black	-0.4749	-0.0331		
	(0.8649)	(0.0584)		
Mix_cookie × Non-white/black	-0.6282	-0.0420		
	(0.8072)	(0.0433)		
Dem_cookie		0.0649		
		(0.0411)		
Dem_cookie × White		-0.0590		
		(0.0670)		
Dem_cookie × Non-white/black		-0.1086*		
		(0.0588)		
Open Only			-0.2405	-0.0135
			(0.2848)	(0.0315)
Open + Click			6.897***	-0.0047
			(0.4048)	(0.0307)
Open Only × White			-0.3321	-0.0002
			(0.3977)	(0.0499)
Open + Click × White			-0.3671	-0.0111
			(0.5677)	(0.0498)
Open Only × Non-white/black			-0.2854	0.0223
			(0.3997)	(0.0467)
Open + Click × Non-white/black			0.3566	0.0129
			(0.5939)	(0.0483)
Observations	95,565	21,388	95,565	21,388
R^2	0.00976	0.00620	0.22257	0.00071

Clustered (persona_id) standard-errors in parentheses.
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Appendix F: Estimation Accuracy Evaluation for the Rule-based Policy Function

Figure F1 Evaluation of Estimation Accuracy Across Agent Counts and Time Steps


Deterministic Response Function
Stochastic Response Function
Agent Numbers (N)
Average KL Divergence
10^{-1}
10^{-2}
Time Steps (T)