What AI Speaks for Your Community: Polling AI Agents for Public Opinion on Data Center Projects

Zhifeng Wu
UC Riverside
zwu178@ucr.edu

Yuelin Han
UC Riverside
yhan116@ucr.edu

Shaolei Ren*
UC Riverside
shaolei@ucr.edu

Abstract

The intense computational demands of AI, especially large foundation models, are driving a global boom in data centers. These facilities bring both tangible benefits and potential environmental burdens to local communities. However, the planning processes for data centers often fail to proactively integrate local public opinion in advance, largely because traditional polling is expensive or is conducted too late to influence decisions. To address this gap, we introduce an AI agent polling framework, leveraging large language models to assess community opinion on data centers and guide responsible development of AI. Our experiments reveal water consumption and utility costs as primary concerns, while tax revenue is a key perceived benefit. Furthermore, our cross-model and cross-regional analyses show opinions vary significantly by LLM and regional context. Finally, agent responses show strong topical alignment with real-world survey data. Our framework can serve as a scalable screening tool, enabling developers to integrate community sentiment into early-stage planning for a more informed and socially responsible AI infrastructure deployment.

1 Introduction

The rapid expansion of artificial intelligence (AI), especially foundation models [5], has fueled an unprecedented demand for computing resources, leading to a surge in the construction of large-scale data centers [11]. These facilities are increasingly woven into the fabric of local communities, bringing both potential benefits and risks. On one hand, data centers can create jobs, generate tax revenues, and position regions at the forefront of the digital economy [32]. On the other, they consume vast amounts of electricity, potentially stress limited water infrastructures if evaporative cooling is used, contribute to carbon emissions and local air pollution, and may strain local infrastructure or alter land use [15, 48, 18].

Public perceptions of these trade-offs are sometimes framed in polarized terms: AI and data centers are either viewed as inherently "bad" or unequivocally "good." Yet, real community sentiment is typically more nuanced and mixed. Many residents may simultaneously acknowledge the economic benefits while expressing concerns over environmental sustainability, health impacts, or long-term resilience.

From the perspective of responsible data center deployment, it is often hard to take local voice into consideration in advance. The primary difficulty lies in capturing the complexity of public sentiment. Traditional polling requires significant financial and human resources [28]. Most community feedback is only collected during public hearings, often after major commitments to a project have already been made. In some cases, the hearing process itself may be shaped by limited participation, sample bias, or incomplete information, resulting in feedback that does not fully represent the broader community.

*Corresponding author

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: ResponsibleFM.

A critical gap exists in mechanisms for obtaining early, scalable and diverse community input for socially responsible AI data center deployment. In this paper, we explore the use of foundational models—in particular, large language models (LLMs)—as AI agents to poll public opinion on data center projects, the physical homes of AI and foundational models. By leveraging the reasoning and world knowledge capabilities of LLMs [6], our approach introduces a scalable and cost-effective framework to approximate the breadth of perspectives that might emerge in community engagement. This framework enables stakeholders to rapidly screen community sentiment across diverse regions, identifying key concerns to inform socially responsible planning.

The AI agent polling framework for public opinion is shown in Figure 1. Our methodology consists of six key stages. First, we establish a data center proposal. Then we generate representative AI agent samples using Iterative Proportional Fitting (IPF) from county-level demographics. Additionally, we model community profiling and project specifications, providing them as system prompts to the sampled agents. After that, we conduct detailed polling by using modern LLMs (e.g., GPT-5, Gemini-2.5-Pro, and Qwen-Max), covering multiple questions across diverse core domains. Results can then be calibrated with a small set of real-world polling data using conformal prediction to provide statistical guarantees. Finally, we perform multi-level analyses including cross-model, cross-regional, and human poll comparisons.

Our experiments in two distinct U.S. counties reveal several key findings. First, overall attitudes vary by region: Taylor County agents show higher support, while Loudoun County agents remain largely neutral. Second, there are some commonalities in specific items: water consumption, tax revenue, and utility bills emerge as top concerns or priorities for AI agents across regions. Third, agent opinions exhibit model-specific patterns. For example, Qwen agents stress economic factors and exhibit higher trust in government than their GPT-5 and Gemini-2.5 counterparts. Finally, while direct quantitative comparison is not feasible, AI agent responses show strong topical alignment with recent national polls on primary concerns and perceived benefits [52].

Disclosure. *We recognize that using AI agents to gauge public opinion may introduce significant challenges including, but not limited to, legitimacy, bias, and representativeness. Synthetic opinions generated by AI agents may not substitute for actual voices from the involved communities, and there is a potential risk that biased training data could skew results. Our goal is not to claim definitive measurements of public sentiment but to use a transparent approach that encourages earlier and more informed public discussions.*

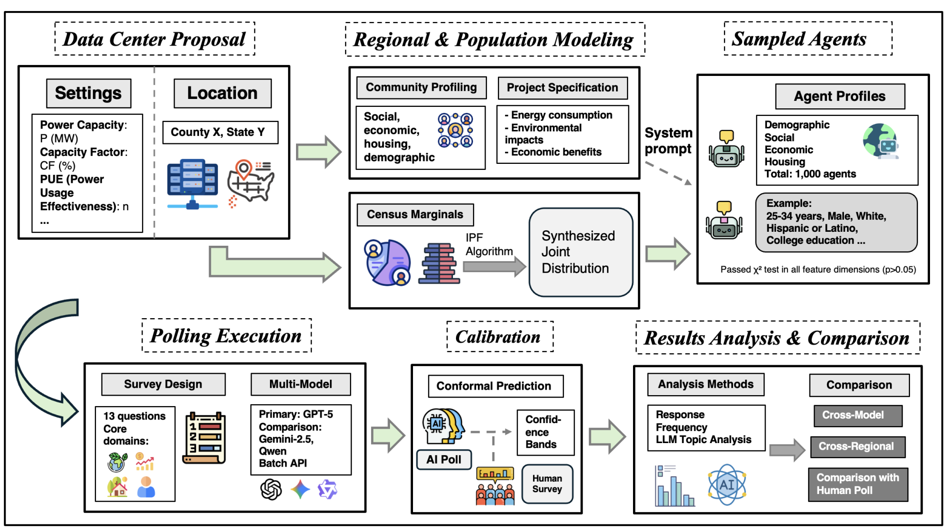


Figure 1: **AI agent polling framework for data center public opinion assessment**. The framework synthesizes county-level demographics with project specifications to generate representative virtual agents, validated through chi-square tests. Multi-model polling across GPT-5, Gemini-2.5-Pro, and Qwen-Max captures responses to 13 questions spanning 5 core domains, enabling cross-model, cross-regional, and human poll comparative analysis.

2 Related Works

The prior research has extensively evaluated the environmental effects of AI data centers, including water consumption, health impact, grid impacts, and life cycle assessments [15, 48, 14, 51]. Studies also examine the economic impacts of data centers [25, 32]. Beyond these macro-level analyses, recent research investigates local impacts of AI data centers [23]. These research efforts provide valuable insights into the diverse consequences of AI infrastructure and establish foundations for understanding the impacts that communities may experience when hosting data center projects.

Public opinion polling traditionally serves as the primary method for gauging local sentiment, involving surveys from representative samples [4]. Currently, the combination of fixed-line and mobile phone surveys remains one of the most widely used methodologies in the field [17]. With the rise of big data, data-augmented approaches like social media analysis have emerged [10, 38]. Currently, data center constructors often rely on traditional polling methods or public hearing to gauge community feedback, typically implementing these efforts only during the late phase of construction.

The widespread adoption of foundation models enables new approaches to opinion research. Researchers can now conduct opinion polling through AI agents that leverage the world knowledge and reasoning capabilities of LLMs [31, 50]. In political research, [53] develops an LLM-based AI agent framework for election simulation. Similarly, prior studies demonstrate that LLM-powered interactive systems can extract themes from public discussions that align with formal reports [49]. Regarding public engagement, [26] suggests that AI could be utilized for evaluating public engagement processes due to its time and cost advantages.

While this research demonstrates the potential of AI agents for public opinion, the focus has seldom turned to the physical infrastructure enabling foundation models and artificial intelligence: the data centers themselves. The traditional engagement used by data center builders faces scalability and timeliness challenges, often failing to provide the early community feedback required for responsible planning. The maturation of LLM-based foundational models for AI agents [7] offers a cost-effective and scalable methodology to fill this specific gap. To our knowledge, this study is the first to apply the AI agent polling framework to assess public opinion on the boom of data centers.

3 Methods

The section presents the methodology for our community-based AI agent polling framework. The overall pipeline comprises four core components: regional context modeling, virtual agent construction, AI agent polling, and conformal prediction calibration. A "community" refers to a U.S. county, matching the level at which census data are provided. More details are provided in Appendix 5.

3.1 Regional Context Modeling

To ensure that AI agents make decisions that closely approximate real-world scenarios, we establish background information for both the data center project and the target community. It contains three components: state-level context, county-level profiling, and proposed data center project description.

For the state in which the target community is located, we incorporate state-level data center electricity consumption data from publicly available reports [11, 39]. The state-level data center electricity consumption provides each agent with a macro-level understanding of the current scale of data center operations in the region.

Since AI agents represent county-level residents, they must possess community relevance comparable to typical residents to improve accuracy. To achieve this local knowledge baseline, we implement a data acquisition and processing module that programmatically interfaces with the U.S. Census Bureau's 2023 American Community Survey 5-year estimates API [42] to obtain detailed population statistics for the target county.

Finally, we construct a standardized project profile for the proposed data center, which is provided to the AI agents as part of the prompt to ensure they are informed about the project. The profile is grounded in publicly available reports and industry standards, encompassing energy consumption, environmental impacts and economic implications.

These three components above, collectively constitute the "Global Context" provided to each AI agent to inform their responses.

3.2 Virtual Agent Construction

To facilitate representative results, we create AI agents that statistically mirror the demographic composition of target communities. This process involves two components: the acquisition of multi-dimensional demographic data and the sampling of the agent population.

Firstly, we extract demographic distributions from the U.S. Census Bureau's 2023 American Community Survey (ACS) [42]. These encompass four categories [43]: social, economic, housing, and demographic characteristics, as detailed in Table 1.

Table 1: Census variables used in agent demographics. All characteristics are based on the U.S. Census Bureau's official categories [44].

Census Category	Agent Attribute	Classification
Social Characteristics	Education Level	No degree, Associate's degree, Bachelor's degree, etc.
	Marital Status	Never Married, Married, Divorced, etc.
	Language at Home	English, Spanish, Other Indo-European languages, etc.
	Citizenship	Native - born in state of residence, Native - born in different state, etc.
Economic Characteristics	Employment Status	Construction, Manufacturing, Wholesale trade, etc
	Household Income	Less than $10K, $10K-$15K, $15K-$25K, etc.
Housing Characteristics	Housing	Owner-occupied: Less than $50K, Rent: Less than $500, etc
	Vehicles	No Vehicle, 1 Vehicle, 2 Vehicles, etc
Demographic Characteristics	Age Group	Under 5 years, 5 to 9 years, 10 to 14 years, etc
	Sex	Male, Female
	Race	Black or African American, Asian Indian, Asian, Japanese, etc.
	Ethnicity	Hispanic, Non-Hispanic

The population statistics acquired previously provide only the marginal distributions for individual attributes. To sample realistic agents with correlated characteristics, we must first synthesize a joint probability distribution. We employ the Iterative Proportional Fitting (IPF) algorithm, a standard method in population synthesis [8]. The resulting agent population's demographic distribution is then verified against the Census data using chi-square goodness-of-fit tests (see Appendix A.2.4).

3.3 AI Agent Polling

This process involves designing a structured questionnaire to capture public sentiment toward the proposed data center and implementing a scalable pipeline to execute surveys and analyze results.

Firstly, to ensure the survey effectively captures community sentiment, the questionnaire is designed to address key aspects of public opinion regarding data center development.

The instrument contains 13 questions, comprising 12 single- and multiple-select items and one open-text question. These questions are structured around five core domains: (1) economic impacts, (2) environmental concerns, (3) community engagement, (4) anticipated personal impacts, and (5) overall project support. The open-text question solicits residents' primary concerns or messages regarding the proposed data center, enabling capture of nuanced perspectives that may not emerge through multi-choice questions. Full questionnaire is available in Appendix C.3.

The subsequent step is to implement the pipeline for executing the survey. Our polling pipeline is an automated workflow that surveys AI agents and processes their responses. It operates in three stages.

First, to optimize for large-scale execution, the prompt is bifurcated. The static regional context, which includes state-level data, the county-level profile, and project specifications, is designated as the system message. The dynamic components, consisting of each agent's unique demographic profile and the survey questionnaire, form the user prompt. This separation enables the API provider to cache the constant system message [29]. Second, we leverage the batch APIs of the selected providers. We choose the latest LLM versions from OpenAI, Google, and Alibaba available as of September

30th, 2025. GPT-5 from OpenAI serves as the primary model throughout all experiments, while Gemini-2.5-Pro from Google and Qwen-Max from Alibaba are used for cross-model comparison to identify model-specific patterns. This asynchronous method facilitates the parallel processing of thousands of agent responses on the provider's infrastructure. Finally, the resulting structured dataset undergoes a two-pronged analysis. For the 12 multiple-choice questions, we perform a quantitative frequency analysis to compute response distributions. For the final open-ended question, we employ LLM-driven topic analysis to identify and quantify the emergent themes in the community's feedback.

3.4 Conformal Prediction Calibration

To obtain results with statistical guarantees and alignment with the real world, we leverage conformal prediction calibration [37]. The workflow proceeds in two phases. First, a calibration phase uses a small set of real-world survey results (y) to compute nonconformity scores ($s = |y - \hat{y}|$) against the corresponding AI agent polling results ($\hat{y}$). This process determines a score threshold $\hat{q}$ based on a target confidence level α. Second, a deployment phase uses this threshold $\hat{q}$ to construct a prediction interval (e.g., $[\hat{y}_{\text{new}} - \hat{q}, \hat{y}_{\text{new}} + \hat{q}]$) for new agent poll results, which is guaranteed to contain the true population probability with at least $1 - \alpha$ probability. This component provides a principled pathway for establishing formal statistical guarantees on agent polling predictions. We note that while included for methodological completeness, this calibration step is not implemented in the current study due to the scope of this preliminary research and the lack of financial resources. However, we will compare our results with an existing national pool (Section 4.3).

4 Experiments

Our experimental analysis comprises two components. First, we conduct a baseline case analyzing AI agent responses to a proposed data center in Taylor County, Texas, with cross-model comparison across three LLMs to identify model-specific variations. Second, we perform cross-regional comparison between Taylor County, Texas and Loudoun County, Virginia to assess how distinct demographic and economic contexts influence agent responses. Finally, we compare our AI agent results with recent human polling data to contextualize our findings.

Implementation scope. This study implements five of the six methodological stages. The calibration component requires resources (e.g., funding, time, personnel) beyond this preliminary study. Validation relies on comparison with human polls (Section 4.3).

Basic settings We select Taylor County, Texas, as our baseline case, which is the home to one of the world's largest data centers. For cross-regional comparison, we contrast this with Loudoun County, Virginia, which hosts the largest concentration of data centers in the world [47]. Each experiment polls 1,000 virtual agents per model or region responding to a hypothetical 100 MW data center proposal through a 13-question survey. In practice, some data centers built in these regions are significantly larger [30].

Model We employ three popular LLMs representing diverse institutional and cultural contexts: OpenAI GPT-5, Google Gemini-2.5-Pro, and Alibaba Qwen-Max[2]. This selection enables cross-model validation and identifying model-specific biases stemming from different training data, institutional backgrounds, and cultural perspectives.

API costs Our large-scale agent polling incurs total API costs of \$36.2 per run . Cost-optimization strategies reduce expenses by at least 50% compared to standard usage. Model-specific costs are \$23.3 for GPT-5 (2,000 agents across baseline and regional comparisons), \$11.2 for Gemini-2.5 (1,000 agents), and \$1.7 for Qwen (1,000 agents). It usually takes more than 24 hours per run.

Detailed specifications, census demographics, data center configurations, and model parameters are provided in Appendix 5.

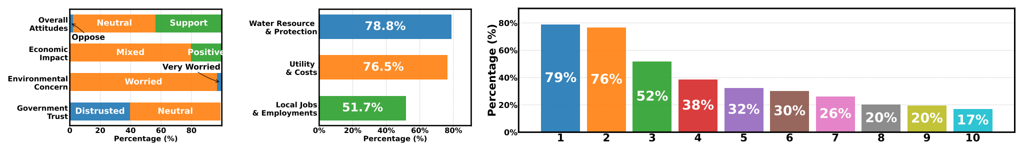
[2]Shortened for readability: GPT-5, Gemini-2.5, and Qwen.

4.1 Baseline Case Analysis

4.1.1 Baseline Results

Our baseline analysis presents an overview of general attitudes from our AI agent polling in Taylor County as shown in Figure 2.

In terms of overall position, slightly over half (54.2%) agents express neutral attitudes toward the data center, while nearly half (43.6%) show positive attitudes. Regarding economic impacts, 80% of agents view the economic effects as mixed, while 20% perceive them as positive. A high level of environmental concern is also evident, with 97% of agents expressing they are worried. On the topic of government oversight, most agents (60%) maintain neutral attitudes, while a 40% express distrust. An analysis of open-ended responses further shows that the top three topics of concern are water resource & protection, utility costs, and local jobs & employments.



(a) Community attitudes (b) LLM topic analysis (c) Detailed topic analysis breakdown

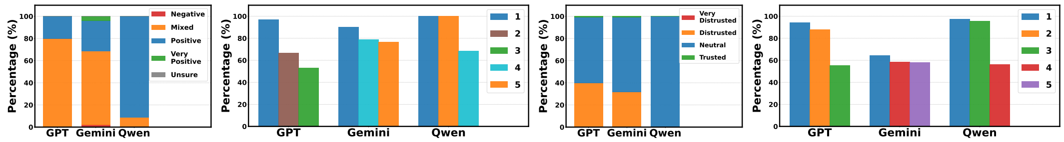
Figure 2: **Taylor County results using GPT-5 (n=1000)**. (a) Community attitude distribution showing neutral-leaning support attitudes, mixed economic impact perception, widespread environmental concerns, and neutral government trust. (b) LLM topic analysis of open-ended community feedback revealing top-three themes. (c) Full topic breakdown (1. Water Resource Protection, 2. Utility Costs, 3. Local Jobs & Employments, 4. Clean Energy, 5. Economic Benefits, 6. Transparency & Public Reporting, 7. Accountability & Enforcement, 8. Grid Impact & Reliability, 9. Housing Costs, 10. Taxes & Public Finance). Charts display only selected response categories, and complete survey options are in Appendix 5.

4.1.2 Cross-Model Comparison

We conduct identical experiments in Taylor County using Gemini and Qwen to assess how polling results vary across different LLMs. While all models identify similar underlying patterns regarding data center development, such as a high level of environmental concern, notable model-specific variations emerge in certain issues as shown in Figure 3:

Economic issues: Agents from GPT-5 and Gemini-2.5 primarily view the project's economic impacts as mixed, whereas Qwen's agents are notably more optimistic, with 91% expressing positive attitudes. This difference is explained by their economic priorities; GPT-5 and Gemini-2.5 agents have relatively diverse preferences, while Qwen agents almost uniformly prioritize tax revenue and job creation. This divergence likely reflects different economic development philosophies embedded in the LLMs' training data. For instance, Qwen's uniform focus on concrete economic outcomes may reflect training data influenced by development-oriented economic paradigms where large-scale infrastructure projects are viewed primarily through the lens of measurable economic advancement. The more varied responses from GPT-5 and Gemini-2.5 may represent training data incorporating diverse concerns.

Governance and preferred information sources: In terms of regulation, both GPT-5 and Gemini-2.5 exhibit substantial distrust in the government's regulatory capacity (40% and 32%, respectively), while nearly no agents from Qwen express distrust. This pattern extends to preferred information sources. Although academic research is a top choice for all models, Qwen's agents show a much stronger preference for local government as a trusted source compared to the other models. These differences may stem from the varying cultural and institutional contexts represented in each model's training data. For example, the high government trust shown by Qwen might be influenced by data from societies with strong state institutions, while the skepticism from GPT-5 and Gemini-2.5 could be more aligned with cultural contexts that emphasize independent oversight and academic validation.



(a) Economic attitudes (b) Top-3 economic benefits (c) Government trust (d) Top-3 information source

Figure 3: **Key differences in cross-model polling results** (n**=1000**). (a) Overall economic attitudes; (b) Top economic benefits (1: Tax Revenue, 2: Infrastructure Upgrades, 3: Business Growth, 4: Economic Diversity, 5: Job Creation); (c) Government trust; (d) Top trustworthy information sources (1: Academic Research, 2: Federal/State Agencies, 3: Local Government, 4: Community Organizations, 5: Local Media). Note: Selected categories; see Appendix 5 for complete data.

4.2 Cross-Regional Analysis

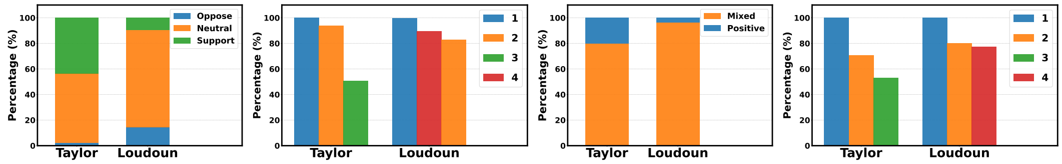
To explore how the regional context influences public attitudes, we conduct a comparative analysis between Loudoun County, Virginia, known as “Data Center Alley,” and Taylor County, Texas. Our analysis focuses on overall attitudes and economic impacts. While some common patterns emerge, we find and explain several notable differences between the two regions. These findings demonstrate that our framework’s outputs are location-specific, highlighting its potential utility in the site-selection process for data center projects.

Agents in Taylor County, Texas, show notably higher support for the proposed data center project than those in Loudoun County, Virginia (43.6% vs 9.7%). Regarding the specific conditions that would make agents more supportive, environmental protection emerges as a top consideration for nearly all agents in both Taylor and Loudoun counties. However, agents in Taylor County show a stronger preference for economic benefits, with 94% selecting lower utility bills and about 51% choosing local job guarantees. This emphasis may be attributable to two factors: Taylor County has a lower median household income, and its hot climate drives up air-conditioning usage, making residents particularly sensitive to the prospect of lower utility bills. Furthermore, as a smaller county with more limited job opportunities compared to Loudoun, the prospect of new employment and corresponding industry upgrades is more attractive to its residents. In contrast, Loudoun County agents prioritize governance, with approximately 90% selecting stricter oversight as a condition for their support. This focus likely stems from the community’s extensive experience with data center development, which may have led to a greater awareness of the importance of refined oversight rules and procedures. Additionally, the higher level of trust in government in Loudoun County may also contribute to this preference, which is a topic discussed in the following section.

Regarding the perceived economic impacts, agents in Taylor County express more positive attitudes, with 20% viewing the project’s effects as positive. This may be because agents representing Taylor County residents attach greater importance to job creation and related industry opportunities than their Loudoun County counterparts. Besides economic benefits, the data center project may lead to some economic issues. Notably, communities in both counties identify higher utility bills as their top economic worry. A finding unique to Loudoun County is a high level of concern about public service strain. This reflects the county’s existing experience, as Loudoun already hosts numerous data centers that have strained local infrastructure and public services. This finding is consistent with real-world scenarios in the region. According to a report of Loudoun County Supervisor[41], the rapid proliferation of data centers has reportedly placed essential public services like the local power grid under severe stress and generated public resistance to new development.

4.3 Comparison with Real-World Human Polls

To contextualize our agent polling results, we compare them with a recent human poll by Heatmap News [52]. However, significant methodological differences (e.g., surveyed population and instrument design) render a direct quantitative comparison inappropriate. Accordingly, this section first details these key differences to acknowledge the inherent limitations of the comparison. Subsequently, we conduct a qualitative analysis, comparing the findings from our AI agent experiments with those of the human poll.



(a) Overall attitudes (b) Top-3 support condition (c) Economic attitudes (d) Top-3 economic concerns

Figure 4: **Cross-regional polling results (n=1000)**. (a) Overall attitudes for proposal data center. (b) The top conditions that would increase AI agents' support for the project (1: Environmental Protections, 2: Lower Utility Bills, 3: Local Job Guarantees, 4: Stricter Oversight). (c) Overall economic attitudes towards the project. (d) The community's top economic concerns regarding the proposed data center (1: Higher Utility Bills, 2: Benefits to Outsiders, 3: Housing Cost Inflation, 4: Public Service Strain). Note: Selected categories; see Appendix 5 for complete data.

4.3.1 Methodological Differences

Firstly, our study focuses on localized public sentiment at the county level, where agent responses are grounded in specific demographic and economic contexts. In contrast, the Heatmap poll has a national scope, surveying voters across all 50 states and Washington, D.C.. Secondly, our research focuses exclusively on a specific data center proposal, whereas the Heatmap survey contextualizes data centers by comparing their public acceptance against other energy infrastructures and analyzes sentiments along political affiliations. A third distinction lies in the questionnaire design. Our survey instrument is more granular, structured around core domains including economic, environmental, and governance issues. The Heatmap survey is more concise, gauging overall support or opposition and evaluating common arguments for each stance. Crucially, the response options differ; for instance, our survey includes "Neutral" as a distinct middle option for overall attitude, an option absent from the Heatmap poll's.

Considering these differences, we conduct the following comparison. We examine common patterns identified in our cross-regional analysis against the national results of the Heatmap poll. Our comparison focuses on similar questions while excluding unrelated survey items. Furthermore, due to differences in response options, we conduct qualitative rather than quantitative analysis.

4.3.2 Topical Comparison and Findings

Table 2 summarizes the key topical alignments and differences. It should be noted that the quantitative figures shown are for illustrative purposes within their respective contexts and are not directly comparable, given the significant methodological distinctions previously discussed.

Regarding overall attitudes toward data centers, our county-level experiments demonstrate substantial geographic heterogeneity, with net support of approximately 41% in Taylor County, Texas, versus -5% in Loudoun County, Virginia. The national Heatmap poll, by comparison, shows net support of approximately 2%. Although the two approaches are not directly comparable due to differences in geographic scope, they may nonetheless reveal a consistent underlying pattern: while national-level polling reflects broadly neutral attitudes toward data centers, actual support appears highly location-dependent, with significant local variation.

Regarding expected benefits brought by data centers, the Heatmap poll identifies tax revenue and job creation as the most popular reasons. Our experimental results rank tax revenue and infrastructure upgrades as the most important economic benefits among AI agents. A similar pattern can be seen: tax revenue is highly selected by both human and AI agents. The divergence lies in the other priority: the human poll indicates a preference for job creation, while the AI agents prioritize infrastructure upgrades. This difference is likely attributable to the survey design. Our study provides a more granular set of choices, which allows agents to prioritize 'infrastructure upgrades' as a distinct benefit. This specific option is not available in the Heatmap poll, whose list of potential benefits focuses on broader arguments such as 'create high-paying jobs' and 'power digital economy'. It is plausible that the human poll's results would have shown greater similarity to our findings had these more detailed options been available.

A strong alignment is also evident regarding the arguments for opposing data centers. The Heatmap poll finds that “water usage” and “electricity usage”, which can lead to higher utility costs, are the most persuasive to the public. Our experiments reveal similar patterns, identifying potential water consumption as a primary environmental worry and higher utility bills as the top economic concern among AI agents. As mentioned by Heatmap, these results are consistent with real-world cases. For example, residents in Tucson, Arizona, have expressed opposition to a proposed data center project due to concerns over water consumption [40].

Table 2: Comparison between human poll and AI agent polling

Metric	**Heatmap Poll (National Voters)**	**AI Agent Polling (County-Level)**	
		Taylor County, TX	**Loudoun County, VA**
Overall Attitudes	Net support +2% (44% support vs. 42% oppose)	Net support +41% (Support 44%, Oppose 2%)	Net support -5% (Support 10%, Oppose 15%)
Benefits	Tax revenue & Create high-paying jobs	Tax revenue & Infrastructure upgrades	
Concerns	Water usage & Electricity usage	Water consumption & Higher utility bills	

Note: Net support is calculated as the sum of all support levels minus the sum of all opposition levels. Methodological differences (Section 4.3.1) mean percentage values reflect different contexts and should be interpreted thematically rather than as direct equivalents. Concerns of AI agent polling represent top environmental (water) and economic (utility bills) issues.

5 Conclusion and Discussion

We propose a novel and scalable method to assess community feedback on data center projects, guiding the siting and design of responsible AI data centers. We introduce an AI agent polling framework that leverages foundation models (i.e., LLM) to gauge nuanced and mixed public attitudes toward these facilities. This approach is beneficial for responsible AI infrastructure deployment because it can allow the local community voice to be taken into consideration in a scalable way. Our results show key concerns and priorities within the targeted community. Cross-regional analysis demonstrates that the framework’s results reflect distinct local contexts, revealing its potential to inform data center site selection. Furthermore, comparison with human polls confirms a clear topical alignment between our experimental results and real-world survey data.

Limitations. While our methodology includes conformal prediction calibration, its implementation requires (small) real-world survey data and financial resources beyond the scope of this study and our budget. Our current validation relies on comparison with existing national polls (Section 4.3). Our analysis focuses on two counties with existing data center infrastructure; applicability to other community contexts warrants further investigation. The framework also has known boundaries. The AI agents model a generalized cognitive response and may not fully capture specific human nuances, such as neurodiverse perspectives or the phenomenon of “engagement silence.” Research also shows that LLMs exhibit inherent bias, such as racial biases [27, 19, 16, 21, 2, 36].

Future work. Our research opens up multiple directions for future work, including implementing the conformal prediction calibration with real-world surveys, extending this framework to other infrastructure domains, developing site selection algorithms that incorporate community sentiment, simulating human deliberative processes and social interactions using multi-agent systems, and establishing hybrid approaches that combine AI polling with traditional engagement methods.

Acknowledgement

This work was supported in part by the U.S. NSF under the grants CCF-2324941, CNS-2326598, and CCF-2324916.

References

[1] Anastasios N. Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification, 2022.

[2] Noel F. Ayoub, Karthik Balakrishnan, Marc S. Ayoub, Thomas F. Barrett, Abel P. David, and Stacey T. Gray. Inherent bias in large language models: A random sampling analysis. *Mayo Clinic Proceedings: Digital Health*, 2(2):186–191, 2024.

[3] Michael Bacharach. Estimating nonnegative matrices from marginal data. *International Economic Review*, 6(3):294–310, 1965.

[4] Adam J. Berinsky. Measuring public opinion with surveys. *Annual Review of Political Science*, 20:309–329, 2017.

[5] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models, 2022.

[6] Shuaihang Chen, Yuanxing Liu, Wei Han, Weinan Zhang, and Ting Liu. A survey on llm-based multi-agent system: Recent advances and new frontiers in application, 2025.

[7] Shuaihang Chen, Yuanxing Liu, Wei Han, Weinan Zhang, and Ting Liu. A survey on llm-based multi-agent system: Recent advances and new frontiers in application, 2025.

[8] Abdoul-Ahad Choupani and Amir Reza Mamdoohi. Population synthesis using iterative proportional fitting (ipf): A review and future research. *Transportation Research Procedia*, 17:223–233, 2016. International Conference on Transportation Planning and Implementation Methodologies for Developing Countries (12th TPMDC) Selected Proceedings, IIT Bombay, Mumbai, India, 10-12 December 2014.

[9] Data Center Map. Amazon aws iad - loudoun county campus, 2025. Data center facility listing; Part of AWS us-east-1 region.

[10] Xuefan Dong and Ying Lian. A review of social media-based public opinion analyses: Challenges and recommendations. *Technology in Society*, 67:101724, 2021.

[11] Electric Power Research Institute. Powering intelligence: Analyzing artificial intelligence and data center energy consumption. Technical Report 3002028905, Electric Power Research Institute, Palo Alto, CA, 2024.

[12] Google. Google's Data Center Efficiency.

[13] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure, 2022.

[14] Fehmi Görkem Üçtuğ and Tayyar Can Ünver. Life cycle assessment-based environmental impact analysis of a tier 4 data center: A case study in turkey. *Sustainable Energy Technologies and Assessments*, 56:103076, 2023.

[15] Yuelin Han, Zhifeng Wu, Pengfei Li, Adam Wierman, and Shaolei Ren. The unpaid toll: Quantifying the public health impact of ai, 2024.

[16] Valentin Hofmann, Pratyusha Ria Kalluri, Dan Jurafsky, and Sharese King. Dialect prejudice predicts ai decisions about people's character, employability, and criminality, 2024.

[17] Gerald Kosicki. Survey methods, traditional, and public opinion polling. pages 1–5, 09 2020.

[18] Pengfei Li, Jianyi Yang, Mohammad A. Islam, and Shaolei Ren. Making AI less "thirsty": Uncovering and addressing the secret water footprint of AI models. *Communications of the ACM*, 2025.

[19] Lin Ling, Fazle Rabbi, Song Wang, and Jinqiu Yang. Bias unveiled: Investigating social bias in llm-generated code. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 27491–27499, 2025.

[20] Loudoun County Government. Loudoun residents survey shows satisfaction with county services, quality of life. County News & Announcements.

[21] Sibo Ma, Alejandro Salinas, Julian Nyarko, and Peter Henderson. Breaking down bias: On the limits of generalizable pruning strategies. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, pages 2437–2450, 2025.

[22] Microsoft Corporation. Northern virginia, 2025. Datacenter Community Pledge and local operations information.

[23] Wacuka M Ngata, Noman Bashir, Michelle Westerlaken, Laurent Liote, Yasra Chandio, and Elsa Olivetti. The cloud next door: Investigating the environmental and socioeconomic strain of datacenters on local communities. In *Proceedings of the 2025 ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*, COMPASS '25. ACM, 2025.

[24] Terry Nguyen and Ben Green. What happens when data centers come to town. Technical report, Science, Technology, and Public Policy Program, Gerald R. Ford School of Public Policy, University of Michigan, July 2025.

[25] Northern Virginia Technology Council. The impact of data centers on virginia's state and local economies: 5th biennial report. Technical Report, prepared by Mangum Economics, April 2024.

[26] Beth Simone Noveck. Governing with ai - learning the how-to's of ai-enhanced public engagement. The GovLab Blog, September 2025.

[27] Abiodun Finbarrs Oketunji, Muhammad Anas, and Deepthi Saina. Large language model (llm) bias index–llmbi. *arXiv preprint arXiv:2312.14769*, 2023.

[28] K. Olson, J. Stevenson, N. Assad, L. Witt-Swanson, C. P. E. Jones, A. Ganshert, and J. Dykema. Examining variation in survey costs across surveys. *Sociological Methods & Research*, 0(0):00491241241298914, 2024.

[29] OpenAI. Prompt caching in the api, 2024. Accessed: September 10, 2025.

[30] OpenAI. Stargate advances with 4.5 gw partnership with oracle, July 2025. Company blog post.

[31] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior, 2023.

[32] Nam Pham. Data centers: Jobs and opportunities in communities nationwide. Technical report, SSRN, May 2017.

[33] Nam D. Pham. Data centers: Jobs and opportunities in communities nationwide. Technical report, U.S. Chamber Technology Engagement Center, May 2017.

[34] Piedmont Environmental Council. Data centers, diesel generators and air quality – PEC web map. https://www.pecva.org/uncategorized/data-centers-diesel-generators-and-air-quality-pec-web-map/, 2024. Accessed: December 2024.

[35] Paul Reig, Tianyi Luo, Eric Christensen, and Julie Sinistore. Guidance for calculating water use embedded in purchased electricity. Working paper, World Resources Institute, August 2022. Collaboration with WSP USA.

[36] Philip Resnik. Large language models are biased because they are large language models. *Computational Linguistics*, pages 1–21, 2025.

[37] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction, 2007.

[38] Thanveer Shaik, Xiaohui Tao, Christopher Dann, Haoran Xie, Yan Li, and Linda Galligan. Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal*, 2:100003, March 2023.

[39] Arman Shehabi, Alex Newkirk, Sarah J. Smith, Alex Hubbard, Nuoa Lei, Md Abu Bakar Siddik, Billie Holecek, Jonathan Koomey, Eric Masanet, and Dale Sartor. 2024 united states data center energy usage report. Technical report, Lawrence Berkeley National Laboratory, Berkeley, CA, December 2024. Report prepared for the U.S. Department of Energy.

[40] The Associated Press. Proposed data center prompts tucson to regulate large water users, require conservation. 2025. Accessed: 2025-09-23.

[41] Mike Turner. Loudoun county, virginia: Data center capital of the world "a strategy for a changing paradigm". Technical report, Loudoun County Board of Supervisors, Loudoun County, VA, August 2025. Original Edition: July 1, 2024.

[42] U.S. Census Bureau. American community survey 5-year data (2009-2023), 2023. Accessed: 2025.

[43] U.S. Census Bureau. American community survey (ACS). https://www.census.gov/programs-surveys/acs.html, 2024. Ongoing survey conducted since 2005.

[44] U.S. Census Bureau. American community survey data, 2024. Accessed: 2024.

[45] U.S. Energy Information Administration. State carbon dioxide emissions data. https://www.eia.gov/environment/emissions/state/, 2025. Accessed: September 10, 2025.

[46] Virginia Department of Environmental Quality. Issued air permits for data centers. https://www.deq.virginia.gov/permits/air/issued-air-permits-for-data-centers, 2024. Accessed: December 2024.

[47] Virginia Joint Legislative Audit and Review Commission. Report to the Governor and the General Assembly of Virginia: Data centers in Virginia (JLARC report 158), December 2024.

[48] Cameron Wade, Mike Blackhurst, Joe DeCarolis, Anderson de Queiroz, Jeremiah Johnson, and Paulina Jaramillo. Electricity grid impacts of rising demand from data centers and cryptocurrency mining operations. Technical report, Carnegie Mellon University, Scott Institute for Energy Innovation and North Carolina State University, June 2025. Open Energy Outlook Initiative.

[49] Maggie Wang, Ella Colby, Jennifer Okwara, Varun Nagaraj Rao, Yuhan Liu, and Andrés Monroy-Hernández. Policypulse: Llm-synthesis tool for policy researchers. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '25, page 1–17. ACM, April 2025.

[50] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2023.

[51] Carole-Jean Wu, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newsha Ardalani, Kiwan Maeng, Gloria Chang, Fiona Aga, Jinshi Huang, Charles Bai, et al. Sustainable AI: Environmental implications, challenges and opportunities. In *Proceedings of Machine Learning and Systems*, volume 4, pages 795–813, 2022.

[52] Matthew Zeitlin. Heatmap poll: Only 44% of americans would welcome a data center nearby, 2025. Accessed: 2025-09-21.

[53] Xinnong Zhang, Jiayu Lin, Libo Sun, Weihong Qi, Yihang Yang, Yue Chen, Hanjia Lyu, Xinyi Mou, Siming Chen, Jiebo Luo, Xuanjing Huang, Shiping Tang, and Zhongyu Wei. Electionsim: Massive population election simulation powered by large language model driven agents, 2024.

Appendix

A Detailed Experimental Settings

A.1 Regional Context Modeling

To construct the region context required for simulation, we integrate data from multiple sources. All demographic and statistical data used are based on 2023 estimates.

A.1.1 State-level Context

The electricity consumption figures, are obtained from publicly available reports by the Electric Power Research Institute (EPRI) and lawrence berkeley national laboratory (LBNL) [11, 39].

A.1.2 County-level Profiling

Detailed profiles of county-level communities are generated programmatically through the American Community Survey (ACS) 5-Year Data Profile API provided by the U.S. Census Bureau [42]. The variable codes we use for the API calls are listed below. Note that the demographic variables here are high-level summaries for a concise community profile; a more granular set of variables is used for the agent construction.

- **Demographic Variables** DP05_0001E (Total population), DP05_0002E (Male), DP05_0003E (Female), DP05_0018E (Median age), DP05_0069E (White), DP05_0070E (Black or African American), DP05_0071E (American Indian and Alaska Native), DP05_0072E (Asian), DP05_0073E (Native Hawaiian and Other Pacific Islander), DP05_0074E (Other race), DP05_0076E (Hispanic or Latino), DP05_0081E (Not Hispanic or Latino).
- **Social Variables** DP02_0001E (Total households), DP02_0016E (Average household size), DP02_0060E to DP02_0066E (Educational attainment), DP02_0153E (Households with a computer).
- **Economic Variables** DP03_0008E (Civilian labor force), DP03_0033E to DP03_0045E (Employment by industry), DP03_0006E (Armed forces), DP03_0062E (Median household income), DP03_0088E (Per capita income).
- **Housing Variables** DP04_0045E (Occupied housing units), DP04_0046E (Owner-occupied units), DP04_0047E (Renter-occupied units), DP04_0089E (Median value of owner-occupied housing units), DP04_0134E (Median gross rent).

A.1.3 Proposed Data Center Project Description

The proposed data center project is designed as a specific and standardized case study, with the following core technical specifications:

- **Power Specifications** Rated capacity of 100 MW, with a Capacity Factor of 70%. The power usage effectiveness (PUE) is set to 1.1, based on values from Google's data center operations [12]. This represents a conservative estimate given the absence of state-specific

PUE data. The annual data center energy consumption is calculated as: rated capacity × capacity factor × annual hours × PUE.

- **Environmental Impacts**
  - Carbon emissions are calculated by multiplying the facility's annual energy consumption by the state-specific emission factor from the U.S. Energy Information Administration [45].
  - Air pollution estimates for nitrogen oxides (NO_x), volatile organic compounds (VOCs), particulate matter ($PM_{2.5}$), and sulfur dioxide (SO_2) are calculated by multiplying data center energy consumption by emission intensities, which are derived from data of diesel backup generators at Virginia data centers [46, 34, 15]. Based on permitted annual emission limits for Northern Virginia data centers, the total allowable emissions are 13,000 tons of NO_x, 1,400 tons of VOCs, 50 tons of SO_2, and 600 tons of $PM_{2.5}$. We assume actual emissions represent 10% of the permitted limits. The resulting emission quantities are then divided by the total annual energy consumption of data centers in the region to derive pollutant-specific emission intensity factors.
  - Water consumption encompasses both on-site and off-site components. For on-site part, the water usage effectiveness (WUE) of 0.36 L/kWh, representing the national average from Lawrence Berkeley National Laboratory [39], is multiplied by the annual IT energy consumption (calculated as total facility energy consumption divided by PUE). For off-site consumption, the electricity water intensity factor (EWIF) of 3.14 L/kWh from the World Resources Institute [35] is multiplied by the total facility energy consumption.
- **Economic Impacts** To estimate the economic impacts of the data center, we reference a case study by the U.S. Chamber of Commerce [33]. During the construction phase (18-24 months), the project supports approximately 1,700 temporary local jobs, and generates around $240 million in local economic activity and $10 million in taxes. Once operational, it supports nearly 160 permanent local jobs annually, with an average salary of about $50k, and contributes over $32 million in local economic activity and $1.1 million in taxes each year. Additionally, we draw upon [24] to ensure objectivity of our prompt design.

A.2 Virtual Agent Construction

To ensure consistency and control across experiments, generated agents are saved and reused for all simulations pertaining to that specific community.

A.2.1 Demographic Data Acquisition

To construct AI agents that statistically mirror real residents, we acquire detailed demographic distributions from the U.S. Census Bureau's American Community Survey (ACS) 5-Year API [42].

The variable codes used for the API calls are listed below.

- **DP05: Demographic Variables** Age distribution (DP05_0005E–DP05_0017E covering 13 age groups from "Under 5 years" to "85 years and over"), sex distribution (DP05_0002E–DP05_0003E for male and female), detailed race categories (DP05_0037E–DP05_0067E including single race and multiracial combinations), and ethnicity classification (DP05_0076E, DP05_0081E for Hispanic/Latino status).
- **DP02: Social Variables** Citizenship status (DP02_0091E–DP02_0097E covering native-born and foreign-born categories), language spoken at home (DP02_0113E–DP02_0122E for English, Spanish, and other language groups), educational attainment (DP02_0060E–DP02_0066E from "Less than 9th grade" to "Graduate or professional degree"), and marital status by gender (DP02_0026E–DP02_0030E for males, DP02_0032E–DP02_0036E for females).
- **DP03: Economic Variables** Employment by industry sectors (DP03_0033E–DP03_0045E covering 13 major industries from agriculture to public administration), unemployment and labor force status (DP03_0005E–DP03_0007E), and household income distribution (DP03_0052E–DP03_0061E spanning 10 income brackets from "Less than $10,000" to "$200,000 or more").

- **DP04: Housing Variables** Housing tenure and value for owner-occupied units (DP04_0081E–DP04_0088E across 8 value ranges), rental costs for renter-occupied units (DP04_0127E–DP04_0135E covering 8 rent brackets), and household vehicle availability (DP04_0058E–DP04_0061E from "No vehicles" to "3 or more vehicles").

A specific preprocessing step is necessary for age data. As our simulated community survey targets only adults, agents must be 18 years or older. Therefore, the ACS age bracket "15-19 years" is partitioned. Assuming a uniform distribution within this bracket, we proportionally allocate the population to isolate an "18-19 years" subgroup, allowing for the accurate sampling of the adult population.

A.2.2 IPF implementation

Key parameters for the IPF implementation are as follows:

- Maximum iterations: 10
- Convergence threshold (ϵ): 10^{-9}

IPF implementation IPF is a numerical method that constructs multidimensional arrays matching specified marginal constraints [3]. The algorithm iteratively adjusts a multidimensional array $\mathbf{M}^{(k)}$ by scaling each dimension to match target marginal distributions. At iteration k, for dimension d, the scaling factor is $\frac{\mathbf{t}_d}{\mathbf{m}_d^{(k)}}$ where $\mathbf{t}_d$ is the target marginal and $\mathbf{m}_d^{(k)}$ is the current marginal. The process continues until convergence: $\|\mathbf{m}_d^{(k)} - \mathbf{t}_d\| < \epsilon$ for all dimensions d. We initialize the process with a uniform distribution and use the ACS marginals as constraints. Then we apply the IPF algorithm to iteratively construct joint distributions. The algorithm iterates until convergence or reaches the maximum iteration limit (set at 10). In all cases, convergence is achieved before reaching the maximum iteration threshold. The resulting joint distribution enables sampling of the desired number of agents with realistic demographic correlations.

Post-sampling adjustments Following the sampling, two post-processing steps are performed. First, any sampled agents under 18 years of age are discarded, ensuring the final population consists solely of adults. Second, since marital status distributions are gender-dependent, we assign marital status after initial agent construction based on each agent's gender characteristics rather than incorporating it directly into the IPF algorithm.

Verification To verify that generated agent demographics conform to Census population distributions, We employ chi-square goodness-of-fit tests. The test compares observed frequencies (generated agents) against expected frequencies (Census data). P-values above 0.05 indicate successful preservation of demographic structure; otherwise, significant deviations require regeneration.

A.2.3 Post-sampling adjustments

Besides excluding agents aged below 18 years, we make two adjustments.

First, we store sex-dependent marital status distributions in advance and assign marital status after IPF implementation, since agents' sex is already allocated. Because marital status is not incorporated in IPF implementation, we may fail to capture correlations between marital status and other variables like age. Therefore, we adjust marital distributions based on age, increasing unmarried probability for young people and widowed probability for elderly.

Second, we assign education levels after agent generation to avoid unrealistic educational assignments. The U.S. Census Bureau provides two types of educational data: general enrollment status for ages 3 and above and detailed educational attainment for ages 25 and above. Using only the general enrollment data could result in unrealistic scenarios (e.g., 30-year-olds in elementary school or 19-year-olds with doctoral degrees). Therefore, we apply age-appropriate educational distributions: for agents aged 19 to 24 years, we use college enrollment data and classify them as "Attending some college or graduate school" or "Not attending any college"; for agents aged 25 and above, we use detailed educational attainment categories from census data.

These adjustments are not perfect but provide reasonable estimates in our cases.

A.2.4 Verification results

To validate that the demographic distribution of the sampled agents is statistically consistent with the U.S. Census data, we perform a Chi-square goodness-of-fit test for each of the 10 demographic dimensions for two regions. Note that marital status and education level are assigned probabilistically after IPF implementation based on Census distributions, ensuring automatic conformity with Census data. The null hypothesis (H_0) for each test is that the observed frequency distribution of the sampled agents is not significantly different from the expected frequency distribution derived from the Census data. We use a significance level of α=0.05.

Table 3: Chi-square Goodness-of-Fit Test Results for Taylor County, Texas

Demographic Attribute	**Chi-square** (χ^2)	**Degree of Freedom**	**P-value**	**Result** (at α=0.05)
Age Group	2.8606	8	0.9428	Fail to reject H_0
Sex	0.3500	1	0.5541	Fail to reject H_0
Race	4.1485	7	0.7625	Fail to reject H_0
Ethnicity	0.3047	1	0.5809	Fail to reject H_0
Citizenship	1.1907	4	0.8796	Fail to reject H_0
Language at Home	7.2699	4	0.1223	Fail to reject H_0
Employment Status	8.0744	15	0.9208	Fail to reject H_0
Household Income	1.6218	9	0.9961	Fail to reject H_0
Housing	7.2419	15	0.9506	Fail to reject H_0
Vehicles	1.3800	3	0.7102	Fail to reject H_0

Table 4: Chi-square Goodness-of-Fit Test Results for Loudoun County, Virginia

Demographic Attribute	**Chi-square** (χ^2)	**Degree of Freedom**	**P-value**	**Result** (at α=0.05)
Age Group	8.3377	8	0.4012	Fail to reject H_0
Sex	0.0081	1	0.9281	Fail to reject H_0
Race	2.3274	7	0.9395	Fail to reject H_0
Ethnicity	0.3392	1	0.5603	Fail to reject H_0
Citizenship	4.6569	4	0.3243	Fail to reject H_0
Language at Home	2.9878	4	0.5599	Fail to reject H_0
Employment Status	16.8381	15	0.3286	Fail to reject H_0
Household Income	7.0488	9	0.6320	Fail to reject H_0
Housing	8.1773	15	0.9165	Fail to reject H_0
Vehicles	0.5687	3	0.9035	Fail to reject H_0

A.3 AI Agent Polling

A.3.1 LLM-Driven Topic Analysis

Our LLM-driven topic analysis modifies traditional BERTopic [13] by replacing vector embedding and clustering steps with LLM operations, leveraging its superior semantic understanding. The LDTA pipeline consists of three stages: (1) an LLM extracts at most three key phrases from each individual response via batch API; (2) these phrases are aggregated by an LLM to generate overarching themes; and (3) the frequency of each theme is quantified across the entire set of responses. This approach enables capture of nuanced perspectives that may not emerge through structured multiple-choice questions alone. The LLM we use here is Gemini-2.5-Pro with default configuration such as temperature and token limit. Since a single response can address multiple themes, the cumulative frequency of all identify themes can exceed 100%.

A.4 Conformal Prediction Calibration

Conformal Prediction Conformal prediction is a practical method that produces statistically valid uncertainty intervals/sets for any model's predictions, regardless of whether the model is interpretable or a black-box [1]. The workflow proceeds in two phases: (1) Calibration: Using held-out data $\{(X_i, Y_i)\}_{i=1}^n$, one would compute nonconformity scores $s_i = s(X_i, Y_i)$ for each example. The larger the nonconformity scores are, the greater the deviation between the model's prediction and the ground truth value, and the greater the prediction uncertainty. The score threshold $\hat{q}$ would then be calculated as the $\lceil (n+1)(1-\alpha) \rceil / n$ empirical quantile of $\{s_i\}_{i=1}^n$, where α is the target error rate (or equivalently, $1-\alpha$ is the desired confidence level). (2) Deployment: For a new input X_{test}, the prediction set would be constructed as $C(X_{\text{test}}) = \{y : s(X_{\text{test}}, y) \leq \hat{q}\}$, including all candidate outputs whose scores fall below the threshold. Under the exchangeability assumption, which is weaker than the independent and identically distributed (i.i.d.) assumption, this provides theoretical guarantees that $\mathbb{P}(Y_{\text{test}} \in C(X_{\text{test}})) \geq 1-\alpha$ [1].

In this setting, each community poll would serve as a single sample. In addition to AI agent polling results, one would collect real-world survey results at a relatively small scale. For each question option, the AI agent polling would produce a selection probability $\hat{y}$ based on the community's context (agent profiles, demographics, socioeconomic features, etc.). The true selection probability y for that option would then be obtained from the collected real-world survey, which serves as the ground truth for calibration. The nonconformity score would be computed as $s = |y - \hat{y}|$, measuring the absolute deviation between the simulated and true probabilities. By repeating this process across n communities, one would obtain a series of calibration scores $\{s_i\}_{i=1}^n$. Given a target confidence level (e.g., 95%, corresponding to $\alpha = 0.05$), the threshold $\hat{q}$ would be computed as the $\lceil (n+1)(1-\alpha) \rceil / n$ quantile of these scores.

During the deployment phase, for each survey question-option pair, let $\hat{y}_{\text{new}}$ denote the agent-predicted probability. The conformal prediction interval $[\hat{y}_{\text{new}} - \hat{q}, \hat{y}_{\text{new}} + \hat{q}]$ would then be guaranteed to contain the true selection probability with at least $(1-\alpha)$ probability under the exchangeability assumption.

Conformal prediction offers three key advantages. First, it establishes theoretically grounded confidence intervals for AI agent polling results, converting black-box outputs into statistically principled uncertainty quantification. Second, by calibrating against real-world survey data, the method aligns synthetic simulations to empirical observations, enhancing prediction reliability. Third, in comparison to conventional large-scale surveys, this approach offers significant economic efficiency, as it requires only a modest number of samples for calibration.

B Supplementary Experimental Analysis and Results

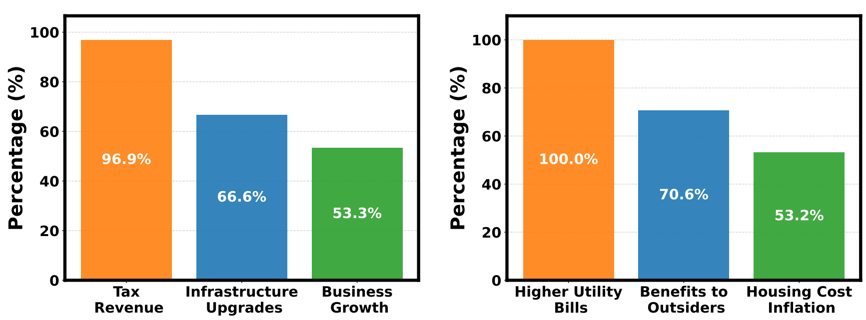
Basic settings We select Taylor County, Texas (Abilene) as our baseline case, which houses the first "Stargate" AI data center [30]. For cross-regional comparison, we contrast this with Loudoun County, Virginia, known as "Data Center Alley," which hosts many data centers including major technology companies [22, 9]. Each experiment polls 1000 virtual agents responding to a hypothetical 100 MW data center proposal through a 13-question survey covering economic, environmental, and community engagement attitudes.

Data collection Demographic data is collected via the U.S. Census Bureau's American Community Survey (ACS) 2023 5-year estimates API [42] based on location of the targeted community, retrieving individual-level characteristics across 12 features and county-level statistics. The data generates representative populations of **1,000 virtual residents** using IPF algorithm, maintaining statistical consistency with regional demographics. Survey responses are collected through batch LLM API calls using standardized prompts incorporating regional context and individual agent profiles.

B.1 Baseline Case Analysis

B.1.1 Baseline Results

A more detailed examination of the survey results is shown in Figures 5, 6, and 7. In the economic domain, agents identify tax revenue (96.9%) as the project's most important benefit, followed by infrastructure upgrades (66.6%) and business growth (53.3%). Conversely, the predominant economic concern shared by all agents is the prospect of higher utility bills, likely due to the immediate and tangible relevance of such costs to residents' household budgets. A strong correspondence exists between the agents' environmental concerns and their requested protections. Water is the paramount issue, with water consumption ranking as the top concern and water conservation as the most demanded protection. Similarly, energy-related impacts are a significant priority, which is understandable given the project's scale. Regarding trusted information sources, AI agents consider academic research the most trustworthy source. Consistent with their previously expressed concerns, the agents' conditions for supporting the project are clear: the environmental protections is the most critical condition (100%), followed by lower utility bills (94.0%), directly mirroring their primary environmental and economic worries.

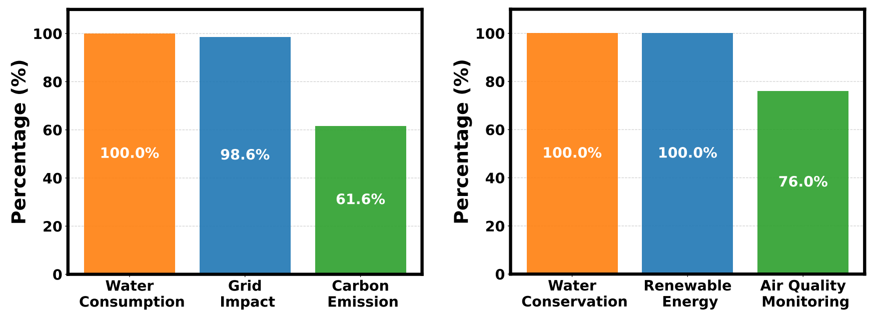


(a) Top-3 economic benefits (b) Top-3 economic concerns

Figure 5: Taylor County results using GPT-5 on economic issues. (a) Community opinions about the most important economic benefits brought by the data center project. (b) Community economic concerns. Note: figures showing distribution for all response options in Appendix 5.

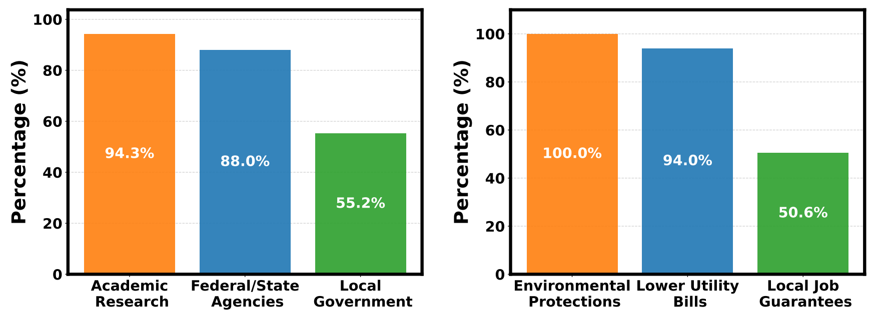
B.1.2 Cross-Model Comparison

Common Results Cross-model analysis reveals several consistent patterns in polling results (see Figure 11). Specifically, a high level of environmental concern is also a consistent finding, with the vast majority of agents in each model simulation reporting being worried about the project's impacts. Furthermore, a strong willingness to participate in community planning discussions is observed, although agents simulated by Gemini-2.5 exhibit a greater degree of neutrality and unwillingness compared to the others. Finally, agents across all models primarily anticipate that the project will have a mixed personal impact on their households.



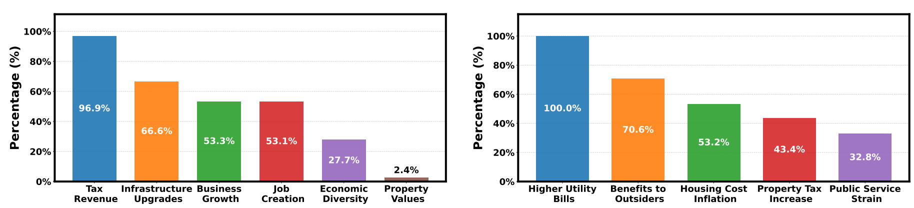
(a) Top-3 environmental concern (b) Top-3 environmental protection

Figure 6: Taylor County results using GPT-5 on environmental issues. (a) The community's top environmental concerns regarding the data center project. (b) The most frequently requested environmental protections for the project. Note: figures showing distribution for all response options in Appendix 5.



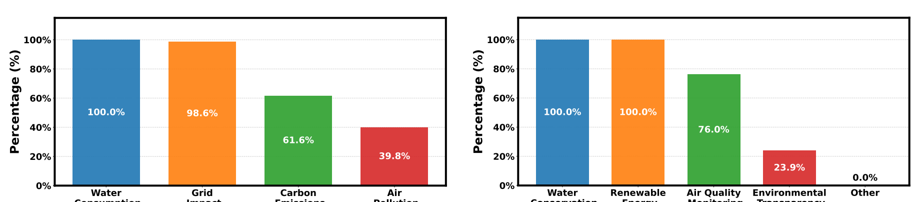
(a) Top-3 information sources (b) Top-3 support condition

Figure 7: Taylor County results using GPT-5 on preferred information sources and support conditions. (a) The community's most trusted sources of information regarding the data center project. (b) The top conditions that would increase community support for the project. Note: figures showing distribution for all response options in Appendix 5.



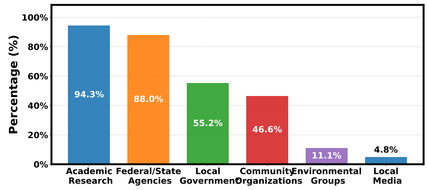
(a) Economic benefits (b) Economic concerns

Figure 8: Taylor County results using GPT-5. (a) Community opinions about the most important economic benefits brought by the data center project. (b) Community economic concerns.

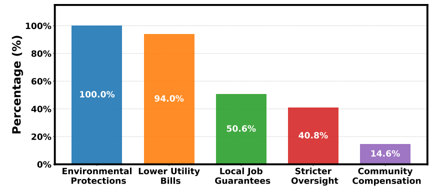


(a) Environmental concern (b) Environmental protection

Figure 9: Taylor County results using GPT-5. (a) The community's top environmental concerns regarding the data center project. (b) The most frequently requested environmental protections for data center project.

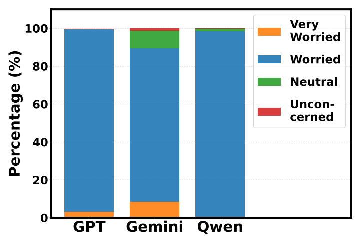


(a) Information sources

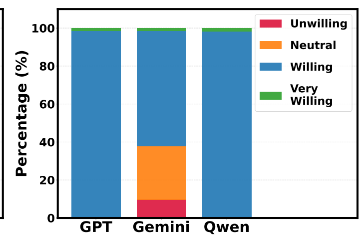


(b) Support condition

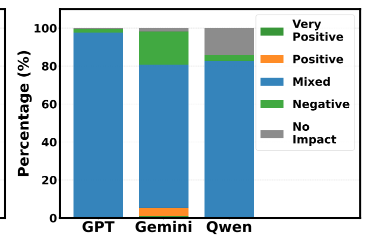
Figure 10: Taylor County results using GPT-5. (a) The community's most trusted sources of information regarding the data center project. (b) The top conditions that would increase community support for the project.



(a) Environmental concern levels

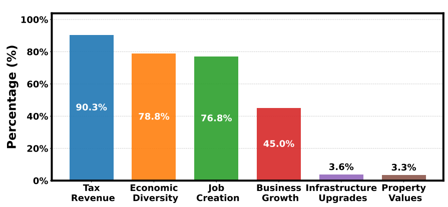


(b) Engagement willingness

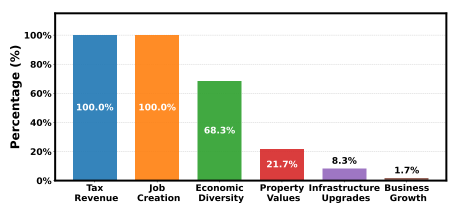


(c) Personal impacts

Figure 11: Common patterns in cross-model polling results. (a) Level of concern expressed by agents regarding the project's potential environmental impacts. (b) Agents' willingness to participate in community planning discussions about the project. (c) Agents' expectations of how the project will personally affect their households. Note: Charts display only selected response categories. Complete survey options are in Appendix 5.

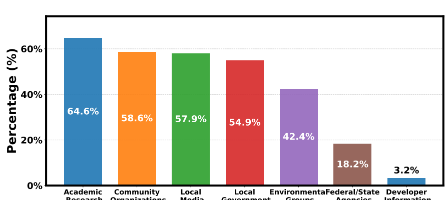


(a) Economic benefits (Gemini)

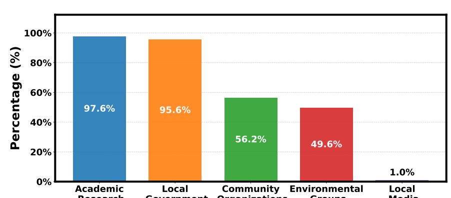


(b) Economic benefits (Qwen)

Figure 12: Taylor County results using Gemini and Qwen. (a) Community opinions about the most important economic benefits brought by the data center project using Gemini. (b) The most important economic benefits using Qwen.

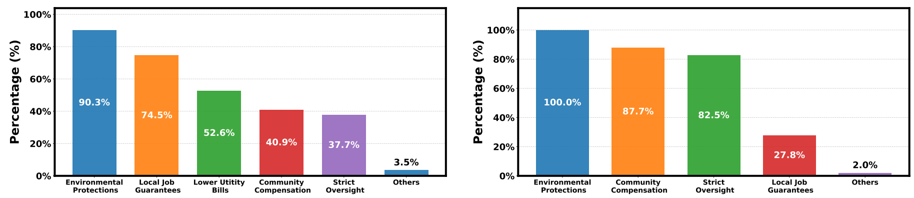


(a) Information sources (Gemini)



(b) Information sources (Qwen)

Figure 13: Taylor County results using Gemini and Qwen. (a) The community's most trusted sources of information regarding the data center project using Gemini. (b) The trusted information source using Qwen.

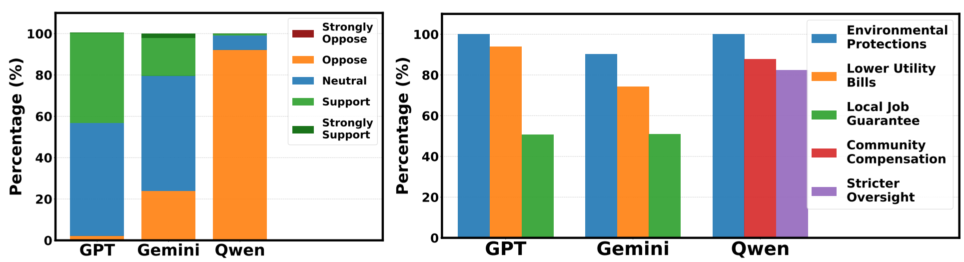


(a) Support conditions (Gemini) (b) Support conditions (Qwen)

Figure 14: Taylor County results using Gemini and Qwen. (a) The conditions that would increase AI agents' support for the project using Gemini. (b) The conditions using Qwen.

Different Results Despite common patterns, significant divergences emerge in key areas:

- **Overall attitudes**: Qwen exhibits notably stronger opposition compared to GPT-5 and Gemini-2.5, despite viewing economic impacts more positively. This apparent contradiction can be explained by examining the conditions under which agents would support the project. Qwen agents strongly demand community compensation and stricter oversight, which are far less frequently mentioned by the other models. Since these conditions are not explicitly guaranteed in the current proposal, it is reasonable that Qwen agents oppose the project despite recognizing its economic benefits. The oversight requirement aligns with Qwen's high government trust, reflecting an expectation that trusted authorities should exercise rigorous regulatory control. This pattern likely stems from training data emphasizing government-led infrastructure development with explicit compensation mechanisms.



(a) Overall attitudes (b) Top-3 support condition

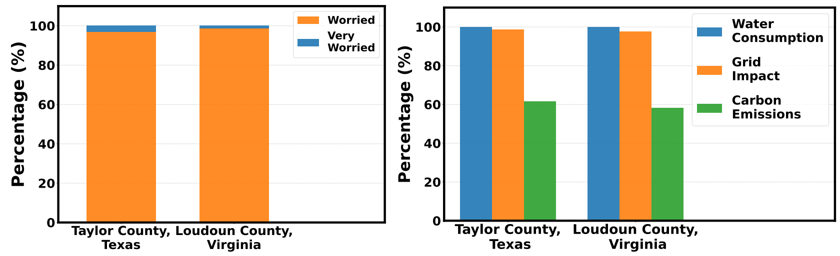
Figure 15: Key differences in cross-model polling results on overall attitudes and support conditions. (a) Overall attitudes for proposed data center. (b) The top conditions that would increase AI agents' support for the project. Note: See Figure 11 regarding response category display conventions.

Key Insights

- **Multi-model analysis reveals complementary viewpoints**: Cross-model comparison reveals that data center project evaluation benefits from incorporating different AI viewpoints. For example, Qwen's emphasis on employment creation and institutional trust offers valuable counterpoints to GPT-5 and Gemini's perspectives.
- **Model-specific biases in agent polling**: The observed variations suggest that different LLMs may produce different responses for certain questions, reflecting distinct cultural, economic, and institutional perspectives embedded in their architectures. These differences demonstrate the sensitivity of polling results to model selection in AI-based social research.

B.2 Cross-Regional Analysis

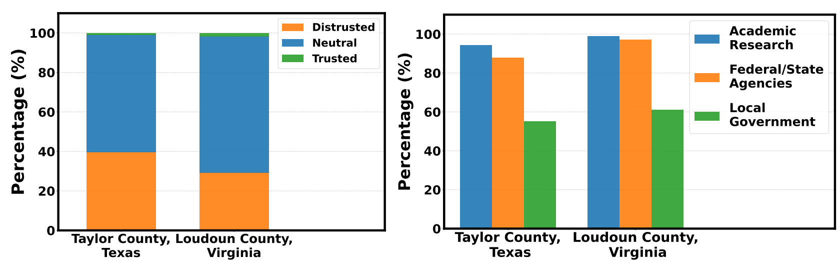
Regarding environmental aspects, agents in the two regions show highly similar trends. When asked about their concern levels toward potential environmental impacts, over 90% of agents in both regions express that they are worried, with a minority being "Very Worried". In terms of specific concerns, water consumption and grid impact are the predominant factors for both communities. The next most common concern is carbon emissions, which is chosen by around half of the agents in both Taylor and Loudoun counties.



(a) Environmental concern levels (b) Top-3 environmental concerns

Figure 16: Cross-regional polling results on environmental aspects. (a) Level of concern expressed by agents regarding the project's potential environmental impacts. (b) The community's top environmental concerns regarding the data center project. Note: See Figure 11 regarding response category display conventions.

Finally, in the domain of governance, we analyze agents' trust in government and their preferred information sources. Regarding trust in government regulation, agents in both counties are predominantly neutral or distrusted. However, a notable difference is that agents in Loudoun County exhibit lower levels of distrust (29%) compared to those in Taylor County (40%). This pattern corresponds with their preferred information sources. While agents in both regions identify academic research as the most trustworthy source, a slightly greater proportion of agents in Loudoun County also express trust in federal and local government. This higher trust in official institutions may be explained by two factors. First, the region's extensive experience with data center regulation has established relatively mature oversight frameworks. Second, Loudoun County's historically high government satisfaction ratings [20] suggest that, on average, residents rate local institutional performance more favorably than in many other jurisdictions.



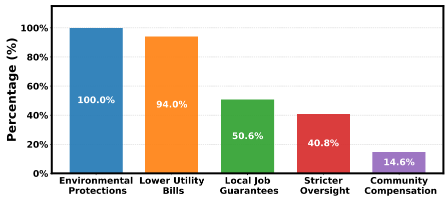
(a) Government trust (b) Top-3 information source

Figure 17: Cross-regional polling results on governance and preferred information sources. (a) Distribution of agents' trust in the government's ability to regulate the data center project. (b) The most trustworthy information sources about the data center project. Note: See Figure 11 regarding response category display conventions.

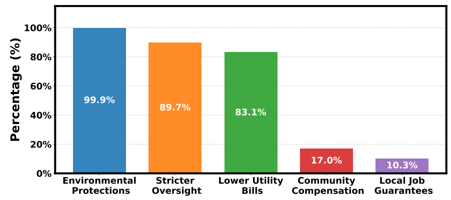
Implications These differences highlight the need for location-specific engagement strategies. For example, communities prioritizing job creation may require stronger employment guarantees and economic development assurances. Varying government trust levels suggest tailored communication approaches are essential, with higher-distrust communities potentially requiring more independent oversight and transparency mechanisms.

B.2.1 Counterfactual Analysis

To disentangle the relative influence of the system prompt versus the agents' demographic features on polling results, we conduct a counterfactual analysis. In this experiment, we utilize the AI agent profiles from Taylor County, Texas (identical to the baseline experiment), but provide them with the regional context (system prompt) of Loudoun County, Virginia. The results exhibit a hybrid pattern:

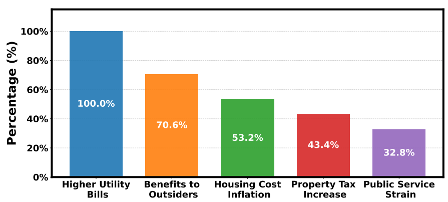


(a) Support condition (Taylor, Texas)

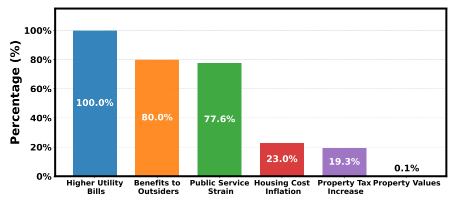


(b) Support condition (Loudoun, Virginia)

Figure 18: The top conditions that would increase community support for the project using GPT-5. (a) Taylor County, Texas. (b) Loudoun County, Virginia.

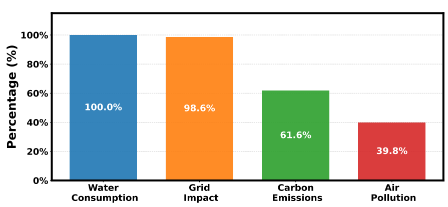


(a) Economic concern (Taylor, Texas)

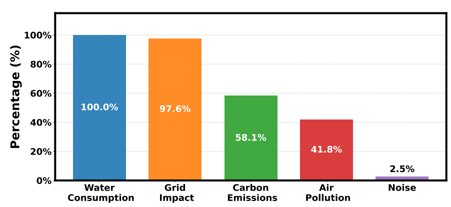


(b) Economic concern (Loudoun, Virginia)

Figure 19: Community economic concerns for the project using GPT-5. (a) Taylor County, Texas. (b) Loudoun County, Virginia.

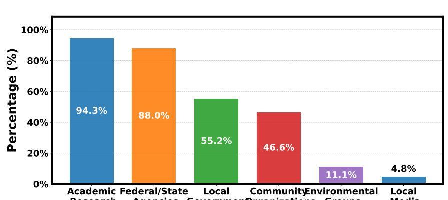


(a) Environmental concern (Taylor, Texas)

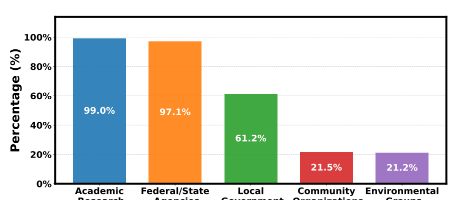


(b) Environmental concern (Loudoun, Virginia)

Figure 20: The community's top environmental concerns regarding the data center project using GPT-5. (a) Taylor County, Texas. (b) Loudoun County, Virginia.



(a) Information sources (Taylor, Texas)



(b) Information sources (Loudoun, Virginia)

Figure 21: The community's most trusted sources of information regarding the data center project using GPT-5. (a) Taylor County, Texas. (b) Loudoun County, Virginia.

while most of attitudes align with the agents' original demographic profiles, specific responses shift to reflect the characteristics of the substituted regional context.

A detailed comparison between this counterfactual experiment (Taylor agents with Loudoun context) and the baseline (Taylor agents with Taylor context) reveals several key divergences. Regarding economic benefits, while tax revenue remains the primary factor in both scenarios, agents in the counterfactual setting demonstrate a significantly higher focus on infrastructure upgrades (97% vs. 68%). Similarly, regarding economic concerns, although higher utility bills remain the top consideration, agents attach considerably more importance to public service strain (45% vs. 29%). These shifts are directly attributable to the Loudoun County system prompt. Because Loudoun County hosts a high density of data centers, the local community experiences strain on public services and infrastructure [41], a factor that directly influences the agents' responses.

In terms of the environmental impacts, agents continue to view water consumption and grid impact as the top two considerations, a pattern consistent across both the Taylor and Loudoun baselines. However, significant differences emerge regarding the conditions required to increase support for the project. In the counterfactual setting, agents demand stricter oversight more frequently (60% vs. 40%) and place less emphasis on local job guarantees (29% vs. 51%). The increased demand for stricter oversight likely stems from the Loudoun context; given the region's extensive experience with data centers and higher government satisfaction metrics, agents may perceive oversight as both a necessary and effective mechanism. Conversely, the reduced demand for local job guarantees is explained by the economic context of Loudoun County, which is characterized by higher income levels and a more robust economy. This enhanced economic backdrop increases the agents' confidence in general economic prospects, thereby reducing the perceived urgency for specific employment guarantees.

B.3 Comparison with Real-World Human Polls

Table 5: Methodological Differences Between Human Poll and AI Agent Polling

Aspect	Heatmap Poll	AI Agent Polling
Scope	National-level	County-level
Context	Data centers vs. other energy infrastructures	Data center projects only
Survey Design	Overall assessment	Granular multi-domain (13 questions, 5 domains)
Population	National voters	Demographically-grounded AI agents
Others	Political affiliation analysis included	No political affiliation analysis

B.3.1 Key Insights

The topical comparison with real-world human polls yields several insights into our AI agent polling framework's performance and characteristics.

From the comparison, we find that agent-based polls exhibit subject-based trends highly consistent with real-world human surveys, suggesting the framework can capture meaningful patterns in public opinion. However, the comparison also reveals limitations: AI agents show less diversity in responses than human participants.

Our framework can serve as a valuable supplement to traditional polling. The topical alignment with human surveys, combined with its scalable and customizable nature, suggests it could help decision-makers predict core community concerns regarding data center development.

C Prompting Framework

C.1 Regional Context Prompt

State Data Center Context
[STATE_NAME] state data center status ([YEAR]): total annual electricity consumption is around [ENERGY_CONSUMPTION] MWh.

Community Profile
[COUNTY_NAME] County, [STATE_NAME] is a community with a total population of [POPULATION]. The population is [FEMALE_PCT]% female and [MALE_PCT]% male, with a median age of [MEDIAN_AGE] years. The racial composition includes [WHITE_PCT]% White residents, [ASIAN_PCT]% Asian residents, [BLACK_PCT]% African American residents, and [HISPANIC_PCT]% Hispanic or Latino residents of any race. The community consists of [TOTAL_HOUSEHOLDS] households with an average household size of [AVG_HOUSEHOLD_SIZE] people per household. In terms of educational attainment, [BACHELOR_OR_HIGHER_PCT]% of adults hold a bachelor's degree or higher, including [GRADUATE_PCT]% with graduate or professional degrees. Additionally, [COMPUTER_PCT]% of households have access to a computer. The economic profile shows a median household income of $[MEDIAN_HOUSEHOLD_INCOME] and a per capita income of $[PER_CAPITA_INCOME]. The major employment sectors include [TOP_INDUSTRIES]. For housing, [HOMEOWNERSHIP_RATE]% of households own their homes, with a median home value of $[MEDIAN_HOME_VALUE]. Rental households pay a median rent of $[MEDIAN_RENT].

Proposed Data Center Project and Its Estimated Impact
The annual electricity consumption of data center project is around [YEARLY_ENERGY_CONSUMPTION] MWh.
During the construction phase ([CONSTRUCTION_DURATION] months), the project is estimated to support approximately [CONSTRUCTION_JOBS] temporary local jobs, and generates around $[CONSTRUCTION_ECONOMIC_ACTIVITY] million in local economic activity and $[CONSTRUCTION_TAX] million in taxes.
Once operational, it is estimated to support nearly [OPERATIONAL_JOBS] permanent local jobs annually, with an average salary of about $[SALARY]k, and generates over $[OPERATIONAL_ECONOMIC_ACTIVITY] million in local economic activity and $[OPERATIONAL_TAX] million in taxes each year.
The annual water consumption includes: [ONSITE_WATER] million liters for on-site water consumption for data center cooling, and [OFFSITE_WATER] million liters for off-site electricity generation.
The annual carbon emissions is [CARBON_EMISSIONS] million short tons.
The annual air pollutants generated by on-site backup generators includes: NO_x [NOX], VOCs [VOCS], $PM_{2.5}$ [PM25], SO_2 [SO2] short tons.

Survey Instructions
You will be asked for your opinions about this proposed data center project. Consider the various impacts of the project on you and your community,including economic factors (such as economic growth, jobs, and tax revenue) and environmental factors (such as energy usage, carbon emissions, water consumption, and air pollution).

C.2 Demographic Agent Prompt

The following is the verbatim prompt used to instruct the LLM agents in our study. The instructional language is designed to be direct and explicit to elicit strong persona adoption from the model.

ASSUME THE ROLE of this resident: [AGE_GROUP], [SEX], [RACE], [ETHNICITY], [EDUCATION_LEVEL] education, [MARITAL_STATUS], [LAN-

GUAGE_SPOKEN_AT_HOME], [CITIZENSHIP], [EMPLOYMENT_STATUS], [HOUSEHOLD_INCOME] household yearly income, [HOUSING], household has [VEHICLES]
Put yourself completely in this person's position. Answer ALL questions from your perspective:
[SURVEY_QUESTIONS]
Give short, clear answers. Be honest and share your real thoughts even if they're critical.
Please respond in JSON format with the following structure: { "question_1_id": "your_answer", "question_2_id": "your_answer", ... }

C.3 Community Survey Questionnaire

1. What do you think will be the overall economic impact of this data center on the local community? Select the one option that best represents your view.

- Very Positive
- Positive
- Mixed
- Negative
- Very Negative
- Unsure

2. Which economic benefits are most important for your community? Select up to three that you consider most important. Separate your answers with a comma only.

- Job Creation
- Tax Revenue
- Infrastructure Upgrades
- Business Growth
- Property Values
- Economic Diversity
- Other (please specify)

3. What economic costs or burdens concern you the most about this data center? Select up to three that you consider most important. Separate your answers with a comma only.

- Higher Utility Bills
- Property Tax Increases
- Job Competition
- Housing Cost Inflation
- Public Service Strain
- Benefits to Outsiders
- No Major Concerns
- Other (please specify)

4. How worried are you about the potential environmental impacts of the data center? Select the one option that best represents your view.

- Very Worried
- Worried
- Neutral
- Unconcerned
- Very Unconcerned

5. Which potential environmental impacts of this data center concern you the most? Select up to three that you consider most important. Separate your answers with a comma only.

- Water Consumption
- Carbon Emissions
- Air Pollution
- Grid Impact
- Heat Generation
- Noise
- No Major Concerns
- Other (please specify)

6. What environmental protections should be required for this data center? Select up to three that you consider most important. Separate your answers with a comma only.

- Water Conservation
- Renewable Energy
- Air Quality Monitoring
- Noise Limits
- Green Building
- Environmental Transparency
- No Special Requirements
- Other (please specify)

7. If given the opportunity to participate in planning discussions for the data center project, would you be willing to participate? Select the one option that best represents your view.

- Very Willing
- Willing
- Neutral
- Unwilling
- Very Unwilling

8. How much do you trust the government and relevant departments' ability to regulate data center operations? Select the one option that best represents your view.

- Very Trusted
- Trusted
- Neutral
- Distrusted
- Very Distrusted

9. Which sources of information would you trust most for this project? Select up to three that you consider most important. Separate your answers with a comma only.

- Environmental Groups
- Local Government
- Community Organizations
- Academic Research
- Developer Information
- Federal/State Agencies
- Local Media
- Other (please specify)

10. How do you expect this data center to personally affect you and your household? Select the one option that best represents your view.

- Very Positive
- Positive
- Mixed
- No Impact
- Negative
- Very Negative
- Other

11. What would make you more supportive of this data center project? Select up to three that you consider most important. Separate your answers with a comma only.

- Already Support
- Lower Utility Bills
- Environmental Protections
- Local Job Guarantees
- Community Compensation
- Stricter Oversight
- Smaller Scale
- Nothing Would Help
- Other (please specify)

12. Would you support or oppose a data center built near your community? Select the one option that best represents your view.

- Strongly Support
- Support
- Neutral
- Oppose
- Strongly Oppose

13. What is the most important thing decision-makers should know about your views on this data center project? Share your key message or main concern.

- No additional thoughts
- Other (please specify)

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claim is to provide a scalable way to gauge public engagement on data centers. We achieve that through Section 4.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We have discussion about limitation, see Section 5.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification: The paper does not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provide details in the main body for understanding results (Section 4), while giving full details in Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Data and code will be released upon publication of the paper.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: API costs of foundation models are listed in Section 4, and the research does not involve traditional training and test processes.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Due to time and resource constraints (Section 4), we do not conduct repeated experiments. We instead establish robustness via multi-model (GPT-5, Gemini-2.5-Pro, Qwen-Max) and cross-regional comparison (2 counties), and alignment with human polls. All sampled agents will be released upon publication to enable reproducibility.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The resources used are mainly the API calls to LLMs, and we provide detailed costs in Section 4.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conform with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss potential positive impacts in the Section 1 and Section 5, and possible negative impacts in disclosure part of Section 1.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All foundation models and data sources used are publicly available and credited throughout the paper.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification: We use LLMs as the backbone of our AI agent system, and we list details in Section 4 and the Appendix.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.