

A Critical Evaluation of LLMs for Analysis of Perspectives Towards Large-Scale Renewable Energy Projects in the U.S.

Anushree Chaudhuri

anuc@mit.edu

Massachusetts Institute of Technology

Jungwoo Chun

Massachusetts Institute of Technology

Lawrence Susskind

Massachusetts Institute of Technology

Article

Keywords:

Posted Date: March 2nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8936990/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Understanding community responses to renewable energy infrastructure siting is essential for accelerating climate action while ensuring a just and equitable energy transition. This study critically examines the capabilities and limitations of large language models (LLMs) for characterizing public sentiment toward renewable energy projects at scale. Drawing on a dataset of 5,095 operational wind and solar projects across the United States, we employed a multi-stage computational methodology to collect and analyze online media coverage, using LLMs to score projects across variables related to opposition types, drivers, and project characteristics. Manually validating a representative sample of the dataset reveals that LLM accuracy varies substantially by variable type: high accuracy (> 95%) for variables with clearly defined, observable indicators but lower accuracy (< 85%) for variables requiring contextual interpretation or narrative synthesis. Analysis across the full dataset indicates that approximately 50% of projects show documented opposition in online media, with threshold effects related to project capacity and differences between wind and solar technologies. However, demographic correlates of opposition documentation appear to reflect digital visibility patterns rather than actual sentiment distributions, raising critical data justice concerns about whose voices become visible in computationally-mediated research. These findings suggest that responsible integration of LLMs into climate and energy social science research requires substantial upfront investment in theoretical grounding, variable operationalization, and validation—meaning effective use of these tools may take considerably longer than anticipated. This performance gradient, where LLMs handle well-defined classification tasks reliably but struggle with tasks requiring contextual judgment and synthesis, has implications beyond research methodology, connecting to broader challenges of scalable oversight in AI systems that will increasingly be asked to assist with the complex, value-laden questions central to climate action. We propose five principles for responsible LLM integration and discuss implications for climate action research methodologies.

Introduction

Communities across the world have mobilized to oppose the siting of wind and solar farms in their local areas, with at least 498 utility-scale projects in 49 U.S. states facing substantial resistance resulting in delays or cancellations as of 2025¹. Understanding the prevalence and underlying drivers of this opposition is essential for policymakers and developers seeking to accelerate the deployment of clean energy infrastructure to mitigate the effects of climate change, while ensuring communities and stakeholders are included in an energy transition that is just and equitable. The urgency of climate action creates intense pressure to rapidly scale renewable energy deployment—the International Energy Agency projects that achieving net-zero emissions by 2050 requires tripling global renewable energy capacity by 2030². Yet this global imperative for speed encounters friction at the local level, where renewable energy projects generate complex responses shaped by community contexts, values, and concerns. The rapid expansion of utility-scale wind and solar projects across the United States—with thousands of projects currently operational or under development—creates both urgency and opportunity for systematic analysis of community responses to renewable energy infrastructure.

Existing research has made substantial contributions to understanding opposition through diverse methodological approaches, each with distinct strengths and trade-offs. Case study research and ethnographic methods provide rich, contextually grounded insights into the complex social dynamics underlying specific siting conflicts, revealing how community values, place attachment, procedural concerns, and distributional justice considerations shape local responses^{3–5}. Prior research has identified numerous factors influencing project acceptance or opposition, including demographics^{6,7}, project characteristics^{8–11}, location factors^{11–14}, economic impacts^{15–19}, environmental concerns^{20–22}, and procedural fairness^{4,22–30}. Petrova (2013) synthesized these factors into the VESPA framework (Visual, Environmental, Socioeconomic, and Procedural Aspects), while Susskind et al. (2022) analyzed 54 case studies to propose seven hypotheses for sources of opposition: environmental concerns about biodiversity impacts; health and safety concerns; intergovernmental disputes over authority; inadequate public participation; violations of tribal sovereignty; perceived property value threats; and financial constraints.

Qualitative case study approaches excel at capturing nuance, context, and the situated meanings communities attach to energy development, though their intensive nature means they typically focus on a limited number of cases. Quantitative approaches, including large-scale surveys and statistical analysis of project characteristics, offer complementary insights by identifying broader patterns and testing hypotheses across larger samples. Stokes et al. (2023) identified demographic predictors of opposition through analysis of newspaper coverage of 1,414 wind projects, manually coding mentions of “anti-wind” actions³¹ across four categories. Rand et al.’s (2024) national survey of 984 residents near solar projects³² revealed overall positive sentiment but dramatic increases in negative attitudes for larger projects. These approaches provide statistical power and generalizability, though surveys are resource-intensive, typically capture sentiment after positions have solidified, and may miss populations less likely to respond. Media analysis can track public discourse at scale but may systematically overrepresent communities and perspectives with greater digital visibility and media access. Each methodological approach thus illuminates certain dimensions of community response while necessarily constraining others.

At the same time, large language models (LLMs) have emerged as potentially powerful tools for analyzing textual data at unprecedented scales, offering new capabilities for identifying patterns across vast corpora of documents. The digitalization of information systems creates both new possibilities and challenges for understanding human-environment interactions, with AI systems increasingly positioned as tools for accelerating sustainability research and climate action^{33,34}. Recent advances in model architectures, including expanded context windows and improved performance on classification tasks, have enabled applications across computational social science³⁵. For renewable energy siting research specifically, Buster et al. (2024) applied LLMs to analyze wind siting ordinances³⁶, demonstrating the potential for computational approaches to process regulatory documents at scale. They integrated a decision tree framework with LLM prompting to achieve 85–90% accuracy in automated extraction—demonstrating both the promise of computational methods for large-scale policy research and the necessity of structured approaches to improve reliability. Systematic reviews document the expanding scope of transformer and LLM applications across the energy sector³⁷, including trajectories toward agentic digital twins and automated policy analysis systems. These developments coincide with institutional pressures on academic research to demonstrate efficiency and impact, making LLMs increasingly attractive for large-scale empirical analysis. These applications also raise questions about what happens when models are deployed on tasks that require not just pattern recognition but contextual judgment—questions that connect to a growing literature on scalable oversight, which examines how human supervision of AI systems can be maintained as model capabilities increase and the complexity of tasks outpaces evaluators' ability to verify outputs directly.³⁸

Recent work demonstrates both the promise and limitations of LLMs in specialized technical domains. Phan et al. (2024) evaluated frontier LLMs on question-answering tasks for environmental review documents under the National Environmental Policy Act, finding that while retrieval-augmented generation approaches achieved 66–69% accuracy, models struggled significantly with divergent and problem-solving questions compared to closed factual questions.³⁹ This pattern—strong performance on factual extraction but weaker performance on complex reasoning—parallels challenges we might expect in renewable energy policy analysis, where understanding community opposition requires synthesizing qualitative judgments rather than extracting discrete facts. Similarly, a systematic review of 99 papers on transformers and LLMs in the energy sector³⁷ found that while these models excel at time-series forecasting and pattern recognition, their application to complex decision-making tasks remains constrained by issues of hallucination, interpretability, and integration with domain-specific knowledge.

However, this technological shift raises important epistemological questions about knowledge production in the social sciences. Critical scholars have examined how AI integration affects both the substance and ethics of research, highlighting tensions between computational efficiency and traditional qualitative methodologies that emphasize interpretation, reflexivity, and meaning-making⁴⁰. These tensions become particularly salient in research domains where understanding community perspectives, place-based knowledge, and cultural context is essential—such as in the study of public responses to renewable energy infrastructure development. Furthermore, the environmental footprint of computationally intensive AI methods presents potential contradictions when applied to sustainability research, as training and operating large language models requires substantial energy consumption and generates significant carbon emissions^{41–43}. Researchers studying climate action pathways thus face an ethical obligation to consider whether the knowledge produced justifies the environmental costs incurred.

Questions of data justice are equally critical when applying AI methods to energy and environmental research. Scholars have documented how algorithmic systems can encode and amplify existing inequities, with particular concern for how digital data practices may reproduce or deepen marginalization of already vulnerable communities^{44–47}. The question of whose voices become visible—and whose remain hidden—in computationally-mediated research has direct implications for environmental justice, as energy infrastructure development often disproportionately affects rural communities, Indigenous nations, and populations with limited political and economic power. While AI has been positioned as a potential tool for advancing both climate action and environmental justice⁴⁸, realizing this potential requires careful attention to how computational methods may inadvertently perpetuate the very inequities they aim to address.

This article critically examines both the capabilities and limitations of LLMs for characterizing public sentiment toward renewable energy projects, and what patterns of opposition and support emerge from large-scale analysis of online media coverage. We address two primary research questions: (1) How accurately can LLMs identify and classify different types of opposition to renewable energy projects based on evidence available from web searches, and what do systematic patterns in LLM performance reveal about the epistemological limits of computational approaches? (2) What patterns of opposition prevalence, drivers, and distribution emerge from LLM-assisted analysis of online media coverage across thousands of operational renewable energy projects in the United States? These results are then discussed within broader debates in computational social science and data justice frameworks, examining how computational methods can be responsibly integrated into energy social science research while maintaining commitments to context, equity, and inclusion. This discussion attends to critical concerns including whose voices become visible in digitally-mediated research, the environmental footprint of computational methods, and the labor required for rigorous validation.

Drawing on a dataset of 5,095 operational wind and solar projects across the United States from the Energy Information Administration's EIA-860 survey⁴⁹, we employed a multi-stage computational methodology to analyze online media coverage and identify mentions of support and opposition (Fig. 1).

First, we collected search engine results for each project using queries combining project name, location, and keywords related to controversy. Second, we retrieved and parsed the full content of relevant webpages, generating aggregated text averaging approximately 10,792 tokens per project. Third, we used an LLM to score search results for relevance to community sentiment, filtering content for subsequent analysis. Fourth, we employed an LLM to analyze aggregated content and generate binary scores for 15 variables related to public perception, including types of opposition (physical protests, legal challenges, policy actions, opinion editorials), drivers of opposition (environmental concerns, health impacts, property values, procedural fairness, tribal sovereignty, intergovernmental disputes), and other project characteristics (support mentions, compensation, delays, co-existing land uses). Variable definitions were developed based on existing literature, particularly the frameworks proposed by Susskind et al. (2022) and Stokes et al. (2023). Full definitions of each variable are included in Table S-3. For each variable, the LLM provided justifications and source citations, and generated narrative summaries synthesizing available information. We then manually validated 501 projects (approximately 10% of the sample) through systematic comparison of LLM-generated scores against human coding. Each project was evaluated across 25 variables, yielding 12,502 individual binary classifications—each representing a human judgment of whether the LLM correctly identified the presence or absence of a specific variable for a specific project. This validation approach allows us to empirically evaluate which aspects of community opposition LLMs can reliably detect and which dimensions require human interpretation.

The findings reveal significant variation in LLM performance across different types of variables, with important implications for computational approaches to studying social responses to energy infrastructure. While LLMs achieve high accuracy (greater than 95%) for certain clearly-defined variables such as mentions of tribal opposition or physical protests, performance decreases substantially (below 85%) for variables requiring narrative understanding, contextual interpretation, or recognition of complex legal and policy dynamics. Narrative summaries—requiring synthesis of information and contextual judgment—achieved only 71.9% accuracy, the lowest among all variables tested. Critically, this performance gradient suggests that variable definition and theoretical grounding may be paramount considerations when employing computational methods. The quality of LLM classification appears to depend fundamentally on how researchers operationalize concepts from social science theory into computationally tractable variables. Variables with clear, observable indicators and explicit textual markers tend to show high accuracy, while those requiring inference, contextual knowledge, or interpretation of implicit meaning tend to show substantially lower performance. This finding suggests that integrating LLMs into social science research may require intensive upfront work by domain experts to critically interrogate theoretical constructs, develop precise operational definitions, and conduct rigorous validation—work that may initially slow rather than accelerate the research process. Furthermore, our analysis suggests potential systematic biases in what LLMs can detect based on the digital visibility of different communities and forms of expression, raising critical questions about whose voices shape our understanding of energy conflicts when research increasingly relies on computational methods and digitally-accessible data sources.

Analysis of LLM-generated scores across the full dataset reveals complex patterns in opposition prevalence and distribution. Overall, 49.68% of operational projects were identified as having at least one mention of opposition in online media, while 34.31% had documented support. Project capacity demonstrates what appears to be a threshold effect, with opposition likelihood increasing markedly above approximately 26 MW. Wind projects face systematically higher opposition rates across nearly all variables compared to solar projects, consistent with prior literature. Regional patterns show higher opposition density in Midwest and Northeastern states. Demographic analysis reveals associations between opposition documentation and community characteristics including population density, racial composition, and energy burden, though these patterns must be interpreted cautiously as they likely reflect digital visibility rather than necessarily actual sentiment distributions.

This research contributes empirical evidence to emerging debates about the role of artificial intelligence in social science research and climate action. Rather than positioning computational and interpretive approaches in opposition, we examine how they might be productively integrated while attending to critical concerns about data justice, epistemic validity, and research ethics. The analysis demonstrates that contrary to popular narratives positioning LLMs as efficiency-enhancing tools that accelerate research, responsible integration may require substantial additional labor: careful theoretical work to define variables, extensive validation against human judgment, critical examination of whose voices are captured in training data and analytical outputs, and reflexive attention to how computational categorization may flatten complex social realities. These findings have implications not only for renewable energy siting research but for broader discussions of computational social science, particularly regarding the pace and process of integrating new technologies into established research methodology.

Results

LLM Accuracy and Evaluation

The validation of 501 projects revealed both the capabilities and limitations of LLM-based classification for renewable energy opposition research. Overall accuracy across all 12,502 individual validations was 92.2%, suggesting generally strong performance. However, this aggregate measure masks significant variation in accuracy across different variable types that has important implications for research applications (Table 1. Validation accuracy for each variable.).

Table 1
Validation accuracy for each variable.

Variable	Accuracy (%)	FPR	FNR	Prevalence	n
Article A	87.6	0.044	0.141	0.820	501
Article B	94.6	0.036	0.061	0.723	501
Article C	92.6	0.065	0.077	0.754	501
Article D	92.0	0.091	0.075	0.693	501
Article E	94.2	0.040	0.067	0.653	501
Article F	92.6	0.056	0.086	0.606	500
Article G	94.6	0.036	0.068	0.556	500
Article H	92.8	0.030	0.108	0.538	498
Article I	93.0	0.049	0.089	0.533	484
Support	90.6	0.010	0.232	0.379	501
Opposition	88.2	0.113	0.120	0.665	501
Physical	98.8	0.002	0.556	0.018	501
Policy	91.2	0.024	0.289	0.242	501
Legal	84.8	0.088	0.254	0.385	501
Opinion-Editorial	90.0	0.035	0.289	0.255	501
Environmental	92.0	0.023	0.298	0.208	501
Participation	92.8	0.023	0.377	0.138	501
Tribal	99.2	0.002	0.150	0.040	501
Health	96.0	0.018	0.231	0.104	501
Intergovernmental	97.8	0.004	0.200	0.090	501
Property	91.4	0.026	0.270	0.244	501
Compensation	96.8	0.009	0.279	0.086	500
Delay	96.2	0.007	0.213	0.150	501
Co- Land Use	93.0	0.015	0.290	0.200	501
Narrative	71.9	0.871	0.147	0.814	501

The distribution of classification outcomes demonstrates the LLM's conservative bias in identifying opposition-related variables. True negatives—cases where the LLM correctly identified a variable as absent—comprised the largest category at 58.21% (7,277 validations). True positives, where the LLM correctly identified a variable as present, accounted for 33.99% (4,249 validations). Error rates remained relatively low: false negatives (the LLM incorrectly marked a variable as absent when it was actually present) occurred in 5.56% of cases (695 validations), while false positives (the LLM incorrectly marked a variable as present when it was actually absent) occurred in only 2.25% (281 validations). This distribution reveals that when the LLM makes errors, it is approximately 2.5 times more likely to miss present variables than to incorrectly identify absent ones—false negatives comprise 71.2% of total errors versus 28.8% for false positives.

Performance varied considerably across variable types, revealing patterns in what aspects of community opposition LLMs may be able to reliably detect. Variables achieving greater than 95% accuracy were predominantly those with clear, observable indicators: tribal opposition (99.2%), physical opposition such as protests (98.8%), intergovernmental disputes (97.8%), compensation issues (96.8%), project delays (96.2%), and health-related opposition (96.0%). These high-performing variables share characteristics of being concrete, typically documented in explicit language, and often involving formal actions or statements. However, it is important to note that for some of these variables—particularly tribal opposition and physical protests—the very low base rates in our sample (4.0% and 1.8% prevalence respectively) mean that the high accuracy figures may not be robust indicators of performance, as there were relatively few positive cases to evaluate.

In contrast, variables requiring interpretive judgment or synthesis of complex information showed notably lower accuracy. Most notably, the narrative summary variable achieved only 71.9% accuracy, the lowest among all variables tested. Legal opposition (84.8%) and opinion editorial (90.0%) also fell below the 90% threshold, despite potentially seeming like clearly documentable phenomena. This performance gradient suggests there may be fundamental differences in how LLMs process explicit versus implicit information in text.

Analysis of LLM scores across the full dataset of 5,095 operational renewable energy projects revealed complex patterns in the prevalence and distribution of community opposition and support. The average content relevance score across all projects was 2.79, with higher capacity projects demonstrating a positive correlation with relevant search results (Pearson's $r = 0.34$). This relationship suggests that larger projects may generate more substantial online media coverage, though notably, operational year showed negligible correlation with content relevance, indicating that temporal proximity does not appear to predict online documentation levels.

Overall, 49.68% of operational projects were identified as having at least one mention of opposition in online media, while 34.31% had documented support. These aggregate figures mask important nuances: 21.65% of projects showed evidence of both opposition and support, suggesting that media coverage often captures multiple perspectives, while only 12.66% had support mentions without corresponding opposition. This asymmetry may align with established media tendencies to prioritize conflict narratives.

Project Size and Opposition Patterns

The relationship between project capacity and opposition likelihood demonstrates a threshold effect. Figure 2 displays the relative frequency distributions of projects with and without documented opposition across the capacity spectrum. Below approximately 26 MW, projects without opposition (blue line) appear at substantially higher frequencies than those with opposition (red line)—indicating that smaller projects rarely generate documented resistance. Around 26 MW, the two distributions converge and cross. Above this threshold, the pattern reverses: projects with documented opposition consistently outnumber those without. The average capacity of projects without documented opposition was 25.4 MW compared to 49.5 MW for those with opposition evidence—nearly twice as high. This inflection point suggests that community responses may be triggered by perceptions of scale and landscape impact that become salient above certain size thresholds.

Temporal Evolution of Public Sentiment

Analysis of opposition and support mentions over time reveals evolving patterns in public discourse around renewable energy projects (Fig. 3). Both opposition and support documentation increased from 2000 to 2022, though this trend likely reflects the expansion of online media coverage as much as or more than changing sentiment. Notably, early projects (2000–2005) showed proportionally higher opposition rates relative to support, which could potentially indicate either greater controversy around initial renewable energy developments or improved public acceptance as technologies became more familiar, though these interpretations remain speculative given the limitations of our data.

Geographic Distribution of Opposition

Spatial analysis reveals distinct regional patterns in opposition documentation. Projects are concentrated in the Northeast, Southwest, and Midwest regions, with notably sparse development in the Pacific Northwest and Southeast. Opposition density shows marked regional variation, with Midwest and Northeastern states demonstrating higher opposition rates, while Western states show marginally lower opposition density (Fig. 4). These patterns may reflect regional differences in renewable energy policy environments, population density, media coverage practices, or cultural attitudes toward energy infrastructure, though our data do not allow us to disentangle these factors.

Technology-Specific Opposition Patterns

Consistent with previous literature, wind projects appear to face systematically higher opposition rates across nearly all perception variables compared to solar projects (Fig. 5). Wind projects showed opposition rates exceeding solar by margins of 10–30 percentage points for most

variables. Particularly notable disparities emerged for property value concerns (wind: 34.6%, solar: 15.1%), health-related opposition (wind: 24.1%, solar: 6.8%), and environmental concerns (wind: 26.1%, solar: 15.2%).

Certain opposition types showed uniformly low prevalence regardless of technology: physical opposition (protests), participation-based opposition, and tribal opposition all registered below 6.7% for both technologies. These low rates may reflect either genuine rarity of these opposition forms or systematic under-documentation in online media, raising questions about whose voices and which forms of resistance become visible in digital discourse. The particularly low rate of tribal opposition documentation warrants careful interpretation, as this may reflect underrepresentation in digitally-accessible media rather than absence of Indigenous concerns about renewable energy development.

Demographic Correlates of Opposition Mentions

Analysis of census tract demographics for project locations reveals associations between community characteristics and opposition documentation (Table 2. Descriptive statistics of projects by demographic characteristics.). Projects with documented opposition were disproportionately located in areas with lower population densities, though this pattern varied by technology type. Solar projects in opposition areas averaged 432.1 people per square mile compared to 714.4 in non-opposition areas ($t = -5.70$, $p < 0.001$), while wind showed even starker contrasts (51.0 vs 149.7, $t = -3.72$, $p < 0.001$).

Racial composition showed statistically significant associations with opposition patterns. Areas with higher Hispanic populations demonstrated lower documented opposition rates (13.5% vs 18.2%, $t = -7.79$, $p < 0.001$), while areas with higher White populations showed increased documented opposition (71.4% vs 64.2%, $t = 9.76$, $p < 0.001$). However, these patterns intersect with other socioeconomic factors in ways that complicate interpretation, and it is critical to emphasize that these are descriptive correlations that may reflect differences in digital visibility, media coverage patterns, or other confounding factors rather than actual differences in community sentiment.

Noteworthy associations also appear with energy burden and environmental quality indicators. Projects with documented opposition were more commonly located in areas with higher energy burden percentiles (66.5% vs 60.4%, $t = 8.84$, $p < 0.001$), which could suggest that communities already facing high electricity costs relative to income may be more likely to express documented opposition to projects that do not directly address their energy affordability challenges—though alternative explanations are equally plausible. Similarly, areas with lower air quality (PM2.5 percentiles) showed higher documented opposition rates, potentially indicating that communities experiencing environmental burdens may view large-scale renewable projects as perpetuating rather than alleviating local environmental impacts, though this interpretation remains speculative.

These demographic patterns must be interpreted with substantial caution, as they reflect only what becomes visible in online media rather than actual community sentiment distributions. The intersection of digital access disparities, media coverage biases, and potential genuine demographic differences in opposition likelihood creates complex challenges for interpretation that underscore the limitations of relying solely on computational analysis of online discourse.

Table 2
Descriptive statistics of projects by demographic characteristics.

	All Projects			Solar			Wind		
Variables	Average With Opposition	Average Without Opposition	t-statistic	Average With Opposition	Average Without Opposition	t-statistic	Average With Opposition	Average Without Opposition	t-statistic
Population Density	300.994	630.052	-8.691***	432.087	714.44	-5.701***	51.015	149.726	-3.715***
Percent Black	0.082	0.101	-4.265***	0.118	0.116	0.263	0.013	0.014	-0.095
Percent Hispanic	0.135	0.182	-7.786***	0.15	0.179	-3.959***	0.107	0.203	-7.921***
Percent White	0.714	0.642	9.763***	0.654	0.626	3.184***	0.828	0.727	7.581***
% >25 without HS Degree	0.127	0.139	-4.099***	0.138	0.14	-0.709	0.107	0.134	-5.016***
% Above 64	0.179	0.163	8.617***	0.169	0.159	4.982***	0.197	0.187	2.668***
% Ages 10–64	0.693	0.707	-7.765***	0.703	0.712	-4.345***	0.674	0.677	-0.92
% of Tract in Tribal Areas	0.151	0.128	0.734	0.059	0.084	-0.897	0.237	0.243	-0.082
Energy Burden Percentile	0.665	0.604	8.842***	0.625	0.587	4.474***	0.743	0.7	3.639***
LMI Percentile From AMI	0.496	0.485	1.567	0.493	0.478	1.700*	0.501	0.52	-1.474
PM2.5 in the Air Percentile	0.297	0.316	-2.538**	0.315	0.332	-1.769*	0.26	0.226	2.266**
Unemployment Percentile	0.473	0.493	-2.441**	0.52	0.513	0.693	0.383	0.374	0.502
*p < 0.1; **p < 0.05; ***p < 0.01.									

Discussion

The validation of LLM performance reveals that classification accuracy depends fundamentally on how researchers operationalize theoretical constructs into computationally tractable variables. While aggregate accuracy of 92.2% suggests generally effective performance, the substantial variation across variable types—from 99.2% to 71.9%—indicates that accuracy emerges from the precision and observability of variable definitions rather than inherent model capabilities. High-performing variables correspond to observable events with explicit textual markers and clear categorical boundaries, while variables requiring synthesis, inference, or contextual interpretation show substantially degraded performance. These patterns align with findings from related domains: Phan et al. (2024) found LLMs performed best on "closed questions" but struggled with "divergent" and "problem-solving" questions requiring complex reasoning,³⁹ and systematic reviews of LLM applications in energy systems document similar patterns where transformers excel at pattern recognition but struggle with complex decision-making requiring contextual understanding³⁷. This suggests current limitations may reflect fundamental architectural constraints rather than implementation-specific issues.

The performance gradient we observe corresponds to a distinction increasingly recognized in the AI alignment literature: the difference between tasks where correct answers are precisely defined and tasks where the correct answer involves judgment, subjectivity, or synthesis across incomplete and conflicting evidence. Scalable oversight research has begun to formalize this challenge; Burns et al. (2023) demonstrate that when weak models supervise stronger ones, generalization is promising for well-defined tasks but degrades for tasks requiring complex reasoning, a pattern structurally analogous to the variable-level accuracy differences we document.³⁸ Efforts to build evaluation benchmarks for such tasks, including datasets of expert-annotated reasoning flaws across multiple domains,⁵⁰ further underscore that the bottleneck for reliable AI evaluation is often the quality and precision of the human judgments against which model outputs are compared—precisely the kind of careful construct definition and validation that social science methodologies are designed to provide.

These findings have important implications for computational social science and climate action research. Contrary to narratives positioning LLMs as efficiency-enhancing tools, responsible integration requires substantial upfront labor that may take considerably longer than anticipated. The sample validation was essential for establishing which variables could be reliably measured, yet required domain expertise and calibration—human coding needed substantial iteration to consistently apply variable definitions. This labor cannot be shortcut without potentially sacrificing epistemic validity.

An overarching principle guides our recommendations: complementarity over replacement. LLMs should augment human interpretive capabilities, not substitute for them. This framing shapes how researchers approach every stage of the research process, from variable development through validation to interpretation of results. Within this frame, we propose five principles for responsible integration of LLMs into energy social science and climate action research, ordered roughly by the sequence of the research process.

Principle 1: Theory-First Variable Development emphasizes that variable operationalization should be driven by deep engagement with domain literature and theoretical frameworks, not by what seems computationally tractable. In our study, variable definitions drew extensively from established frameworks including Petrova's (2013) VESPA model and the opposition typologies developed by Susskind et al. (2022) and Stokes et al. (2023), yet even theoretically grounded variables showed differential performance based on their observability in text.

Principle 2: Proportional Validation requires that the scale of validation match the epistemic claims researchers intend to make. Large-scale claims about patterns across thousands of cases demand substantial validation samples; our 10% validation rate represents one threshold, and variables with lower accuracy may require higher validation rates or should be reported with appropriate uncertainty bounds.

Principle 3: Visibility Auditing mandates explicit documentation of whose voices and which forms of expression are likely captured versus excluded in computationally-analyzed datasets, a point we return to below.

Principle 4: Weighing Computational Costs requires that decisions about when and how to deploy LLMs account for environmental and social impacts. Training and running large language models consumes substantial energy and water, requires significant land use for data centers, and carries associated carbon emissions. The scale of testing, validation, and model deployment should be informed by these trade-offs. Researchers must ask whether the knowledge produced justifies the resources consumed—and seek approaches that minimize computational requirements while maintaining analytical rigor. This principle connects to broader conversations in algorithmic fairness and responsible AI about the uneven distribution of AI's benefits and burdens.

Principle 5: Patience for Rigor acknowledges that responsible LLM integration may require accepting that the research process will take more time, not less, as teams develop robust variable definitions, conduct extensive validation, and document lessons learned. During early stages of integration, the most valuable outputs may not be substantive findings but rather methodological insights from benchmarking—papers reporting "here's what worked, here's what failed, and here's why" may contribute more to advancing computational social science than premature large-scale analyses conducted without adequate validation.

In the longer term, as AI systems become more capable and the tasks they are asked to perform more closely resemble the complex, value-laden judgments that characterize much of social science and policy research, the methodological habits developed during this period of careful integration (precise variable definition, rigorous validation, documentation of systematic biases) will become increasingly important for maintaining meaningful human oversight of these systems.^{38,50}

Our analysis also raises critical data justice concerns extending beyond methodological validity. Demographic correlates of opposition documentation—lower documented opposition in areas with higher Hispanic populations, higher in predominantly White areas—likely reflect digital visibility patterns rather than actual sentiment distributions. Communities with limited internet infrastructure, non-English speakers, and those expressing concerns through non-digital channels will be systematically underrepresented in any analysis relying on online sources^{44,45}. These biases are particularly salient for energy justice, as infrastructure development often disproportionately affects rural communities, Indigenous nations, and populations with limited political power—precisely those most likely to be underrepresented in online discourse⁵¹. The low documented rates of physical protests and participation-based opposition may similarly reflect under-documentation rather than rarity. Critically, our validation methodology cannot assess false negatives for communities and perspectives that never appear in the analyzed dataset, meaning even perfect LLM accuracy would provide only a partial picture of community responses to energy infrastructure.

Despite these constraints, the substantive findings both confirm and extend existing literature. The capacity threshold effect—with opposition increasing markedly above approximately 26 MW—aligns with Rand et al.'s (2024) survey findings showing dramatically increased negative sentiment for larger projects, providing convergent validity across methodological approaches.³² Systematic differences

between wind and solar opposition rates confirm prior research, likely reflecting wind energy's greater visibility, audio impacts, and concentrated rural siting. The relatively high prevalence of documented support (34.31%) challenges framings focused exclusively on conflict. At the same time, some patterns complicate straightforward narratives: the low documented rates of participation-based opposition and intergovernmental disputes—identified as significant drivers in case study literature—may indicate either lower actual prevalence than intensive case studies suggest or reduced media attention to process-oriented concerns compared to dramatic events like protests or lawsuits.

Looking forward, integration of computational approaches with community-based participatory methods could help address the digital visibility biases inherent in media-based analysis, with participatory validation processes involving residents of affected communities assessing whether computational findings align with lived experience. Rather than initially building toward fully agentic AI systems, which face significant security and hallucination risks even in technical domains with clear performance metrics, energy social science may benefit more from bounded approaches where AI handles well-defined extraction tasks while humans maintain control over interpretive work. Alternative approaches to aligning AI outputs with human values, such as constitutional AI methods that replace human labeling with explicit principles and rules,⁵² may reduce some annotation burdens but do not resolve the underlying challenge: the principles themselves must be defined with sufficient precision and domain knowledge to be meaningful for the task at hand. For climate and energy research, where questions about distributional justice, procedural fairness, and community wellbeing are inherently contested and context-dependent, the quality of any alignment approach—whether based on reinforcement learning through human feedback, constitutional principles, or scalable oversight protocols—ultimately depends on the depth of engagement with the substantive domain. Finally, the environmental footprint of computationally intensive LLMs also warrants consideration when applied to sustainability research^{41–43}, reinforcing arguments for comprehensive validation on smaller samples rather than premature deployment across massive datasets.

This research is subject to important limitations. The analysis is constrained to English-language sources and the U.S. context, necessarily excluding voices absent from search engine results due to language barriers, digital access constraints, or algorithmic filtering. The validation was conducted at a single time point using specific LLM models, and model performance may vary as new versions are released. The focus on operational projects may not capture opposition dynamics for proposed projects that were cancelled before construction, potentially underestimating total opposition prevalence. Our decision to provide complete article content rather than implementing retrieval-augmented generation may have affected model performance, particularly for variables requiring information synthesis. Most fundamentally, our findings about what LLMs can and cannot reliably detect are specific to our variable definitions—different operationalizations might yield different performance patterns, underscoring that validation results reflect the relationship between conceptual definitions, operational indicators, and textual manifestation rather than fixed properties of the variables themselves.

The central challenge this research illuminates is not whether to use LLMs in energy social science research but how to do so in ways that advance both knowledge and justice. Uncritical adoption—deploying computational methods without adequate validation or examination of whose voices they privilege—would be epistemologically and ethically problematic, potentially producing findings with unknown validity while amplifying existing exclusions. Equally problematic would be categorical rejection of computational approaches, foregoing genuine opportunities to analyze patterns across scales previously inaccessible to social science. The productive path forward requires what might be termed critical LLM-aided computational social science: approaches that harness LLMs' pattern-detection capabilities for initial screening and large-scale mapping while maintaining commitments to interpretive depth, contextual understanding, and participatory engagement with affected communities. As AI systems are increasingly deployed to assist with the complex, morally and politically consequential questions that climate action demands—questions about who bears the costs of transition, whose concerns are taken seriously, and how competing values are weighed—social scientists have both an opportunity and a responsibility to engage critically with these tools, contributing the methodological rigor and domain expertise needed to ensure that computational methods serve rather than undermine equitable climate action. Success in this integration should be measured not by research speed but by the validity of knowledge produced and its contribution to equitable energy transitions that center the voices and concerns of affected communities in shaping pathways for climate action.

Data and Methods

This analysis uses a federally-maintained dataset of all operational utility-scale power plants in the U.S., updated annually by the Energy Information Administration (EIA). The data is compiled and cleaned from the results of the survey Form EIA-860, which legally requires all project owners or operators to submit generator-level information about existing electric power plants with 1 megawatt (MW) or greater of combined nameplate capacity⁴⁹ ("Form EIA-860" 2023). All utility-scale onshore wind and solar projects that were listed as operational in 2022 were included in the study, with generator-level data—such as information for specific wind turbines within a broader wind energy project—collapsed to the plant-level. There were 5,095 total projects included in the EIA-860 dataset, with summary statistics for the projects included in Table 3. Summary statistics for all plants from EIA-860 dataset (2022).. Based on the latitude and longitude coordinates provided

in the EIA-860 metadata, project locations included 1,240 unique counties in all 50 states. Projects in U.S. territories were excluded from the study due to language differences and translation needs for online media, differing reporting requirements and compliance in the EIA-860 dataset, and a relatively small number of utility-scale projects that are currently operational in territories.

Table 3
Summary statistics for all plants from EIA-860 dataset (2022).

All Projects: n = 5095	Mean	Standard Dev.	Min	50th pct	Max
Capacity (MW)	37.38	70.76	1.1	5.0	1027.0
Operating year	2015.77	4.78	1975	2017	2022
Solar: n = 3841	Mean	Standard Dev.	Min	50th pct	Max
Capacity (MW)	15.51	34.36	1.1	4.0	300.0
Operating year	2017.02	3.12	2002	2017	2022
Wind: n = 1254	Mean	Standard Dev.	Min	50th pct	Max
Capacity (MW)	104.37	103.82	1.2	80.8	1027.0
Operating year	2011.93	6.61	1975	2012	2022

Online evidence related to these projects was collected, cleaned, and analyzed according to the following steps. After collection and cleaning, an LLM was used to score online evidence, first for relevance to the research question and renewable energy project at hand, and second for evidence of public perception variables as well as to generate justifications for its scores, source citations, and an open-ended brief narrative summary of the project if relevant information was identified. A simplified diagram of this process is shown in Fig. 1. A validation process was conducted for a randomly selected sample of approximately 10% of projects to enable accuracy benchmarks.

Collecting Webpage Search Results

First, the BrightData SERP (Search Engine Retrieval Protocol) API was used to collect search engine results as dictionary-style output for each project in the dataset. The SERP API service provides real-user, high-volume results for all major search engines⁵³ (“SERP API Start Guide” 2024); the Google search engine was used for all queries based on the quality of results from initial testing. The search query string entered for each project followed this consistent format: {project name} {project county} {project state} {list of keywords}. The full list of possible keywords provided were “controversy”, “opposition”, “lawsuit”, “conflict”, and “hearing”, separated by the advanced search operator “OR” to ensure that a keyword not being found in the search results did not exclude possibly relevant results.

This list of keywords was finalized based on extensive testing and insights from several stakeholder interviews of renewable energy siting experts. These interviews collected information from experts who regularly use keyword alerts on search engines and newspaper databases to be notified of information about newly proposed or controversial renewable energy projects. Keywords related to positive or neutral public sentiment, such as “support” or “op-ed,” as well as more specific keywords like those denoting reasons for support or opposition (such as “environmental”, “tribal”, or “property value”) were not included after testing because they were rarely indexed by a search engine.

The loaded results were limited to 10 maximum results per search query, imitating the default number of results displayed on a user’s web search page on Google search and ensuring that fewer instances of irrelevant content were included. For any search result that returned fewer than 8 out of the 10 possible results, the list of keywords was reduced or eliminated until the search result output included at least 8 search results. If including zero keywords still resulted in fewer than 8 results, this was the final output saved, as the project name and location were considered required inputs for a valid search query. After following this procedure, the final set of search queries returned an average of 9.58 results per query, with only 229 out of 5,095 projects still returning fewer than 8 search results.

Fetching results for each search query took an average of 67.43 seconds. To reduce runtime, the process was parallelized across the projects in the dataset. This step was run locally, since the server request to the SERP API was the main rate-limiting step. The format of the search result output generated for each project includes the full URL, title, display URL (i.e., the shortened URL that a user would see on a search engine), and short description (i.e., the snippet displayed on a search engine before clicking into a result), with the results ordered based on the rank of the search result on the Google search engine.

Fetching, Parsing, and Cleaning Content for Each Search Result

After saving search results for all projects in the dataset, the URLs of search results were used to fetch the full content of the associated webpage using a combination of the Jina Reader API⁵⁴ and the unstructured library in Python. The Reader API was used specifically to process URLs that load HTML content, simplifying the web scraping process to extract the core content from an HTML document and convert it into clean, LLM-friendly text. For URLs that pointed to non-HTML documents—most commonly PDFs but also including file types like tabular data (.csv, .xlsx) and Word documents—the scraping process used the unstructured library. The unstructured library is an open-source library designed to help preprocess and structure unstructured text/tabular documents for use in downstream machine learning and LLM tasks⁵⁵. The library includes functionality to perform partitioning of documents, as well as chunking, cleaning, and staging before connecting the outputs to a downstream analysis process.

The full text from each URL was saved, as well as an aggregated string containing all of the text content from each search result generated for a specific plant (with separators to indicate the title of each source article). On average, each plant generated aggregated content of about 37,770.55 characters or approximately 10,791.58 tokens. Some requests to fetch content failed, either because of the request timing out from slow loading times from the target server (greater than 30 seconds), or because the content itself could not be accessed. Although only general error messages were logged and all data indexed online was publicly available, reasons for failing to fetch content may include news article paywalls, sophisticated blockers preventing web scrapers from accessing content, and other irregularities in the content that made it unreadable.

Model and Prompt Testing

For steps involving LLM scoring or analysis, model and prompt testing was completed on a smaller sample size of projects, randomly selected with distributional weights for various project variables (like capacity, operating year, technology type, and geographic region) to ensure that a diverse number of projects were represented in the model selection and prompt-refining process. Three sample sizes of 10, 25, or up to 100 projects were used throughout this process.

The first LLM-aided step, scoring search results for relevance to the project and research question, was run using the model Claude 3 Haiku, released by Anthropic AI on March 13, 2024⁵⁶. As the most inexpensive of Claude 3's models and with the fastest runtime, a context window of 200k tokens, and high benchmark accuracy on identification-related tasks, Claude 3 Haiku was an ideal LLM model for this step. Because this step did not strictly require a 200k context window, since the search results being passed in only had a title, display URL, and short description, other models with similar or shorter context windows were also tested. Models tested included GPT 4 Turbo (with a 128k context window), GPT 3.5 Turbo (with a 16k context window), Claude 3 Sonnet, and Claude 3 Opus (both with 200k context windows). However, there was only a maximum of a 6% increase in accuracy when using the most advanced model (Claude 3 Opus) when validated on a sample size of 100 (i.e., 6 out of 100 projects scored by Claude 3 Opus had marginally more accurate scores), with an associated trade-off of more than an order of magnitude in runtime—Claude 3 Haiku required approximately 5 seconds per project to return the output, while Claude 3 Opus required closer to a minute per project.

The second LLM-aided step—scoring each project for a set of public perception variables and generating an overall narrative summary of the project based on all relevant online media identified about the project—was run using the model Claude 3 Opus, released by Anthropic AI on March 3, 2024⁵⁷. As the most expensive but also most advanced of Anthropic's models at the time of release, with benchmarked performance comparable or better than OpenAI's GPT-4 and Google's Gemini, Claude 3 Opus provided an ideal model agent for the identification and text reasoning tasks required for this step. In addition, it was the only model with an accessible developer API at this level of performance with a context window with enough tokens (200k) to incorporate close to the full text of all online media for which content was retrieved. Information about all of the models tested in the development process are included in Table S-1.

For both steps, prompt revision was also completed, with the following major changes for both sets of prompts that resulted in improved accuracy without a great trade-off in cost/runtime: 1) providing a reminder of the name and location of the plant in every line of the scoring criteria instead of just once per prompt, 2) using either binary scores or scores of 1–5 instead of 1–10 to avoid unexplained bunching around odd or even numbers, 3) providing context near the top of the prompt and final instructions near the bottom of the prompt, 4) requiring a brief justification for scores, which can improve model reasoning, and 5) providing additional focus on the model's role and requirements, such as framing the model as “an expert on public perceptions on large renewable energy projects” and including some words, such as “EXTREMELY CONFIDENT,” in capitalization for emphasis.

Scoring Search Results for Relevance (LLM-aided step)

To imitate the process of a search engine user making a judgment on which search results to further click into based on relevance to the query at hand, the content filtering process implemented a search result scoring step aided by an LLM. For all projects for which organic

search results were generated, the LLM received information on each search result's title, display URL, description, and an "article letter," which was assigned in alphabetical order of the search result appearing on the search engine page—i.e. the first article would receive a letter of A, then B ... up to J for 10 search results maximum per project. The following variable structure and prompting were used to instruct the LLM on how to score each search result for relevance to the following system prompt and research question: "You are an expert on public perceptions on large renewable energy projects. Your aim is to take a set of search results from Google corresponding to the following search query: {search_query} and determine whether or not the search results are relevant to our research question. Here are the search results: {search_results}. Based on the title, display link, and description of each URL, we would like to identify which search results are most relevant to this research question: 'What is the narrative surrounding the development of this renewable energy project in this location, and what evidence of opposition or support for the project can be identified?' Score each search result based on the article letter with a number between 1–5, with 1 meaning that the article is least relevant and 5 being the most relevant to the research question." The structure of the grade for each article and rubric used to calibrate the LLM's grading between 1–5 is provided in Table S-2.

Scoring Public Perception Variables (LLM-aided step)

The second step of the LLM-aided analysis and final step in the data collection and scoring process was generating scores for a number of binary variables related to public perception. First, the binary variables were scored either a 1 or 0 depending on whether the LLM identified any evidence in the given texts for project support, opposition (including a number of subcategories based on Susskind et al. (2022) and Stokes et al. (2023)), or other project characteristics like evidence of non-required compensation or a substantial delay. Second, these binary scores were supplemented with short (less than eight word) justifications and a list of sources that the LLM used in justifying its answer. Finally, for each project that had enough relevant content to inform the binary scores, the LLM generated a more open-ended case study-style summary of the project, including details on timeline, location, developer, project characteristics, details of public response, and any evidence of opposition and support, based on all the previous information. A full list of the variables used, and their definitions, are provided in Table S-3. All of the variables used in the analysis are also summarized in Figure S-1.

Validation Protocol

To assess the accuracy of LLM-generated classifications, we conducted comprehensive manual validation of approximately 10% of the total project sample. The validation sample comprised 501 projects selected through stratified random sampling to ensure representative distribution across key project characteristics including technology type (wind/solar), capacity ranges (< 10 MW, 10–50 MW, 50–100 MW, > 100 MW), operational year, and geographic region.

The validation process involved systematic comparison of LLM-generated scores against human coding for each variable. The author independently reviewed the aggregated online content for each project in the validation sample and assessed the accuracy of both article relevance scores (10 article-level validations) and public perception variables (15 binary variables and one short answer variable), resulting in 26 validation points per project.

Each LLM score was assigned one of four categories following standard binary classification metrics:

- True Positive (TP): LLM correctly identified presence of the variable
- True Negative (TN): LLM correctly identified absence of the variable
- False Positive (FP): LLM incorrectly identified presence when variable was absent
- False Negative (FN): LLM incorrectly identified absence when variable was present

The validation dataset encompassed 12,502 individual binary classifications across the 501 projects. This large validation sample enables robust statistical analysis of LLM performance patterns across different variable types and project characteristics. All validation data, coding guidelines, and inter-rater reliability statistics are available in the supplementary materials. To ensure coding reliability, the first 50 projects were coded by three validators to establish inter-rater agreement and refine coding guidelines. Discrepancies were resolved through discussion to develop consistent interpretation criteria for each variable.

Declarations

Data and Code Availability

The datasets generated and analyzed in the current study are available in the Harvard Dataverse repository: <https://doi.org/10.7910/DVN/1J3WW1>. The code is available at <https://github.com/mit-renewable-energy/re-opp-llm-analysis>.

Competing Interests

All authors declare no financial or non-financial competing interests.

Funding

Declaration

This study received no external funding.

Author Contribution

A.C. conceived and designed the study, developed the methodology, conducted all data collection and computational analysis, performed validation, and wrote the manuscript. J.C. and L.S. provided supervision and methodological guidance throughout the research process. All authors reviewed and approved the final manuscript.

Acknowledgement

We thank Ella Wang for initial dataset exploration and Sauhaarda (Raunak) Chowdhuri for support with developing the codebase and thread parallelization script for the LLM scoring pipeline. We are grateful to the MIT Energy Initiative, MIT Climate and Sustainability Consortium, and MIT Undergraduate Research Opportunities Program Office for supporting the lead author's work on this project. We also thank those who participated in exploratory conversations on renewable energy siting that informed our keyword selection methodology. This study received no external funding.

Data Availability

The datasets generated and analyzed in the current study are available in the Harvard Dataverse repository:

<https://doi.org/10.7910/DVN/1J3WW1> . The code is available at <https://github.com/mit-renewable-energy/re-opp-llm-analysis> .

References

1. Eisenson, M. et al. Opposition to Renewable Energy Facilities in the United States: June 2025 Edition. *Sabin Cent. Clim. Change Law* https://scholarship.law.columbia.edu/sabin_climate_change/251 (2025).
2. World Energy Outlook 2023. *IEA* <https://www.iea.org/reports/world-energy-outlook-2023> (2023).
3. Wüstenhagen, R., Wolsink, M. & Bürer, M. J. Social acceptance of renewable energy innovation: An introduction to the concept. *Energy Policy* 35, 2683–2691 (2007).
4. Baxter, J. Energy justice: Participation promotes acceptance. *Nat. Energy* 2, 17128 (2017).
5. Walker, C., Stephenson, L. & Baxter, J. "His main platform is 'stop the turbines' ": Political discourse, partisanship and local responses to wind energy in Canada. *Energy Policy* 123, 670–681 (2018).
6. Bidwell, D. The role of values in public beliefs and attitudes towards commercial wind energy. *Energy Policy* 58, 189–199 (2013).
7. Carley, S., Konisky, D. M., Atiq, Z. & Land, N. Energy infrastructure, NIMBYism, and public opinion: a systematic literature review of three decades of empirical survey literature. *Environ. Res. Lett.* 15, 093007 (2020).
8. Bidwell, D. The Effects of Information on Public Attitudes Toward Renewable Energy. *Environ. Behav.* 48, 743–768 (2016).
9. O'Shaughnessy, E., Wiser, R., Hoen, B., Rand, J. & Elmallah, S. Drivers and energy justice implications of renewable energy project siting in the United States. *J. Environ. Policy Plan.* 0, 1–15 (2022).
10. Fergen, J. & B. Jacquet, J. Beauty in motion: Expectations, attitudes, and values of wind energy development in the rural U.S. *Energy Res. Soc. Sci.* 11, 133–141 (2016).
11. Firestone, J. et al. Reconsidering barriers to wind power projects: community engagement, developer transparency and place. *J. Environ. Policy Plan.* 20, 370–386 (2018).

12. Bates, A. & Firestone, J. A comparative assessment of proposed offshore wind power demonstration projects in the United States. *Energy Res. Soc. Sci.* 10, 192–205 (2015).
13. Swofford, J. & Slattery, M. Public attitudes of wind energy in Texas: Local communities in close proximity to wind farms and their effect on decision-making. *Energy Policy* 38, 2508–2519 (2010).
14. Bessette, D. L. & Mills, S. B. Farmers vs. lakers: Agriculture, amenity, and community in predicting opposition to United States wind energy development. *Energy Res. Soc. Sci.* 72, 101873 (2021).
15. Groth, T. M. & Vogt, C. Residents' perceptions of wind turbines: An analysis of two townships in Michigan. *Energy Policy* 65, 251–260 (2014).
16. Bailey, I. & Darkal, H. (Not) talking about justice: justice self-recognition and the integration of energy and environmental-social justice into renewable energy siting. *Local Environ.* 23, 335–351 (2018).
17. Mills, S. B. Preserving Agriculture through Wind Energy Development: A Study of the Social, Economic, and Land Use Effects of Windfarms on Rural Landowners and Their Communities. in (2015).
18. Mills, S. B., Bessette, D. & Smith, H. Exploring landowners' post-construction changes in perceptions of wind energy in Michigan. *Land Use Policy* 82, 754–762 (2019).
19. Klain, S. C., Satterfield, T., MacDonald, S., Battista, N. & Chan, K. M. A. Will communities "open-up" to offshore wind? Lessons learned from New England islands in the United States. *Energy Res. Soc. Sci.* 34, 13–26 (2017).
20. Petrova, M. A. From NIMBY to acceptance: Toward a novel framework – VESPA – For organizing and interpreting community concerns. *Renew. Energy* 86, 1280–1294 (2016).
21. Cass, N., Walker, G. & Devine-Wright, P. Good Neighbours, Public Relations and Bribes: The Politics and Perceptions of Community Benefit Provision in Renewable Energy Development in the UK. *J. Environ. Policy Plan.* 12, 255–275 (2010).
22. Susskind, L. et al. Sources of opposition to renewable energy projects in the United States. *Energy Policy* 165, 112922 (2022).
23. Elmallah, S. & Rand, J. "After the leases are signed, it's a done deal": Exploring procedural injustices for utility-scale wind energy planning in the United States. *Energy Res. Soc. Sci.* 89, 102549 (2022).
24. Ottinger, G. The Winds of Change: Environmental Justice in Energy Transitions. *Sci. Cult.* 22, 222–229 (2013).
25. Ottinger, G., Hargrave, T. J. & Hopson, E. Procedural justice in wind facility siting: Recommendations for state-led siting processes. *Energy Policy* 65, 662–669 (2014).
26. Liu, L., Bouman, T., Perlaviciute, G. & Steg, L. Public participation in decision making, perceived procedural fairness and public acceptability of renewable energy projects. *Energy Clim. Change* 1, 100013 (2020).
27. Baxter, J., Morzaria, R. & Hirsch, R. A case-control study of support/opposition to wind turbines: Perceptions of health risk, economic benefits, and community conflict. *Energy Policy* 61, 931–943 (2013).
28. Jami, A. A. & Walsh, P. R. From consultation to collaboration: A participatory framework for positive community engagement with wind energy projects in Ontario, Canada. *Energy Res. Soc. Sci.* 27, 14–24 (2017).
29. Banerjee, A., Prehoda, E., Sidortsov, R. & Schelly, C. Renewable, ethical? Assessing the energy justice potential of renewable electricity. *AIMS Energy* 5, 768–797 (2017).
30. Fast, S. Assessing public participation tools during wind energy siting. *J. Environ. Stud. Sci.* 7, 386–393 (2017).
31. Stokes, L. C., Franzblau, E., Lovering, J. R. & Miljanich, C. Prevalence and predictors of wind energy opposition in North America. *Proc. Natl. Acad. Sci.* 120, e2302313120 (2023).
32. Rand, J. et al. *Perceptions of Large-Scale Solar Project Neighbors: Results From a National Survey*. 79 <https://emp.lbl.gov/publications/perceptions-large-scale-solar-project> (2024).
33. Creutzig, F. et al. Digitalization and the Anthropocene. *Annu. Rev. Environ. Resour.* 47, 479–509 (2022).
34. Debnath, R., Creutzig, F., Sovacool, B. K. & Shuckburgh, E. Harnessing human and machine intelligence for planetary-level climate action. *Npj Clim. Action* 2, 20 (2023).
35. Ziems, C. et al. Can Large Language Models Transform Computational Social Science? *Comput. Linguist.* 50, 237–291 (2024).
36. Buster, G. et al. Supporting energy policy research with large language models: A case study in wind energy siting ordinances. *Energy AI* 18, 100431 (2024).
37. Antonesi, G. et al. A systematic review of transformers and large language models in the energy sector: towards agentic digital twins. *Appl. Energy* 401, 126670 (2025).
38. Burns, C. et al. Weak-to-Strong Generalization: Eliciting Strong Capabilities With Weak Supervision. Preprint at <https://doi.org/10.48550/arXiv.2312.09390> (2023).

39. Phan, H. et al. RAG vs. Long Context: Examining Frontier Large Language Models for Environmental Review Document Comprehension. Preprint at <https://doi.org/10.48550/arXiv.2407.07321> (2024).
40. Coeckelbergh, M. AI for climate: freedom, justice, and other ethical and political challenges. *AI Ethics* 1, 67–72 (2021).
41. van Wynsberghe, A. Sustainable AI: AI for sustainability and the sustainability of AI. *AI Ethics* 1, 213–218 (2021).
42. Strubell, E., Ganesh, A. & McCallum, A. Energy and Policy Considerations for Deep Learning in NLP. in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (eds Korhonen, A., Traum, D. & Màrquez, L.) 3645–3650 (Association for Computational Linguistics, Florence, Italy, 2019). doi:10.18653/v1/P19-1355.
43. Lanelongue, L. et al. GREENER principles for environmentally sustainable computational science. *Nat. Comput. Sci.* 3, 514–521 (2023).
44. Vera, L. A. et al. When data justice and environmental justice meet: formulating a response to extractive logic through environmental data justice. *Inf. Commun. Soc.* 22, 1012–1028 (2019).
45. Mah, A. Environmental justice in the age of big data: challenging toxic blind spots of voice, speed, and expertise. *Environ. Sociol.* <https://www.tandfonline.com/doi/abs/10.1080/23251042.2016.1220849> (2017).
46. Marjanovic, O., Cecez-Kecmanovic, D. & Vidgen, R. Theorising Algorithmic Justice. *Eur. J. Inf. Syst.* <https://www.tandfonline.com/doi/abs/10.1080/0960085X.2021.1934130> (2022).
47. Klein, L. & D'Ignazio, C. Data Feminism for AI. in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* 100–112 (Association for Computing Machinery, New York, NY, USA, 2024). doi:10.1145/3630106.3658543.
48. Osipov, V. S. & Skryl, T. V. AI's contribution to combating climate change and achieving environmental justice in the global economy. *Front. Environ. Sci.* 10, (2022).
49. Form EIA-860. U.S. Energy Information Administration (2023).
50. Recchia, G., Mangat, C. S., Li, I. & Krishnakumar, G. FindTheFlaws: Annotated Errors for Detecting Flawed Reasoning and Scalable Oversight Research.
51. Brinkley, C. & Wagner, J. Who Is Planning for Environmental Justice—and How? *J. Am. Plann. Assoc.* <https://www.tandfonline.com/doi/abs/10.1080/01944363.2022.2118155> (2024).
52. Bai, Y. et al. Constitutional AI: Harmlessness from AI Feedback. Preprint at <https://doi.org/10.48550/arXiv.2212.08073> (2022).
53. SERP API Start Guide. *Bright Data* <https://brightdata.com/products/serp-api/start-guide> (2024).
54. Jina Reader API. *Jina AI* <https://jina.ai/reader> (2024).
55. Unstructured 0.13.0 documentation. *unstructured.io* <https://unstructured-io.github.io/unstructured/core.html> (2024).
56. Claude 3 Haiku: our fastest model yet. *Anthropic* <https://www.anthropic.com/news/claude-3-haiku> (2024).
57. Introducing the next generation of Claude. *Anthropic* <https://www.anthropic.com/news/claude-3-family> (2024).

Figures

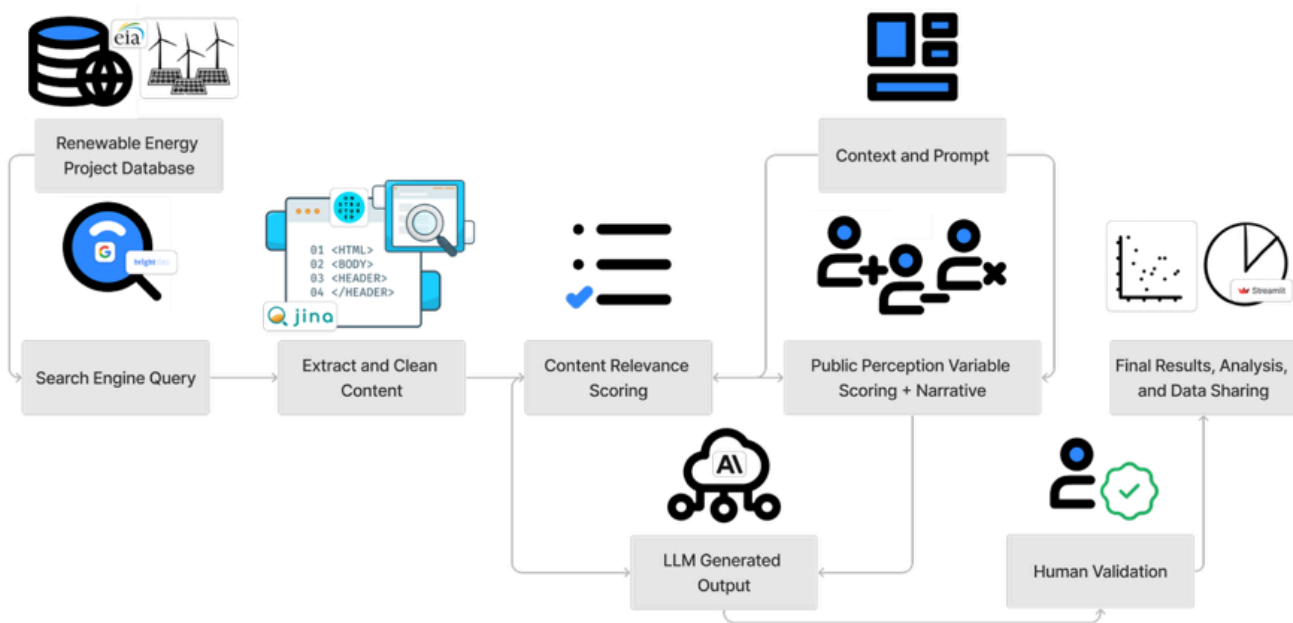


Figure 1

Pipeline overview showing the stages of data collection, relevance scoring, opposition/support classification, and human validation.

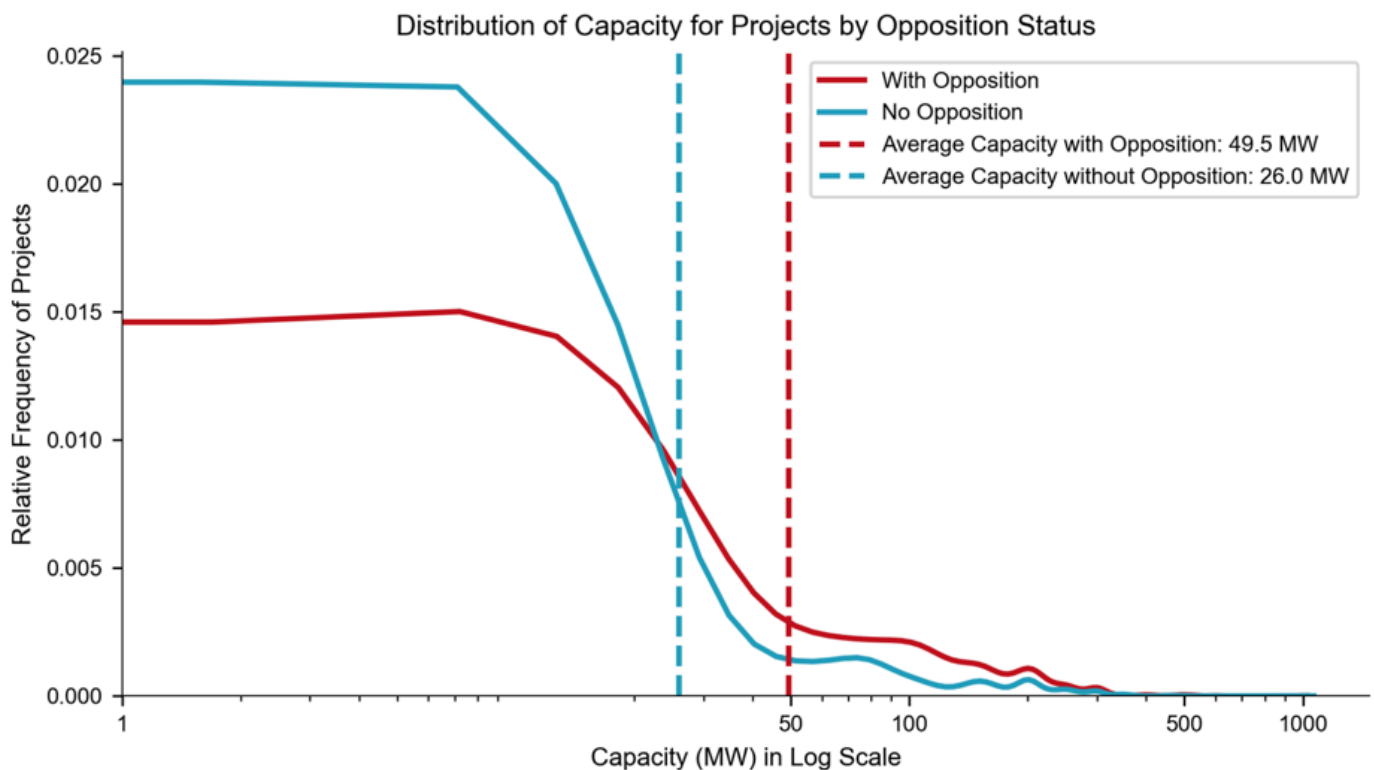


Figure 2

Distribution of project capacity by opposition status. Below approximately 26 MW, projects without documented opposition (blue) appear at higher relative frequencies than those with opposition (red). The distributions cross near 26 MW; above this threshold, projects with documented opposition consistently outnumber those without, suggesting a capacity threshold effect for community resistance.

Opposition and Support for Projects Over Time

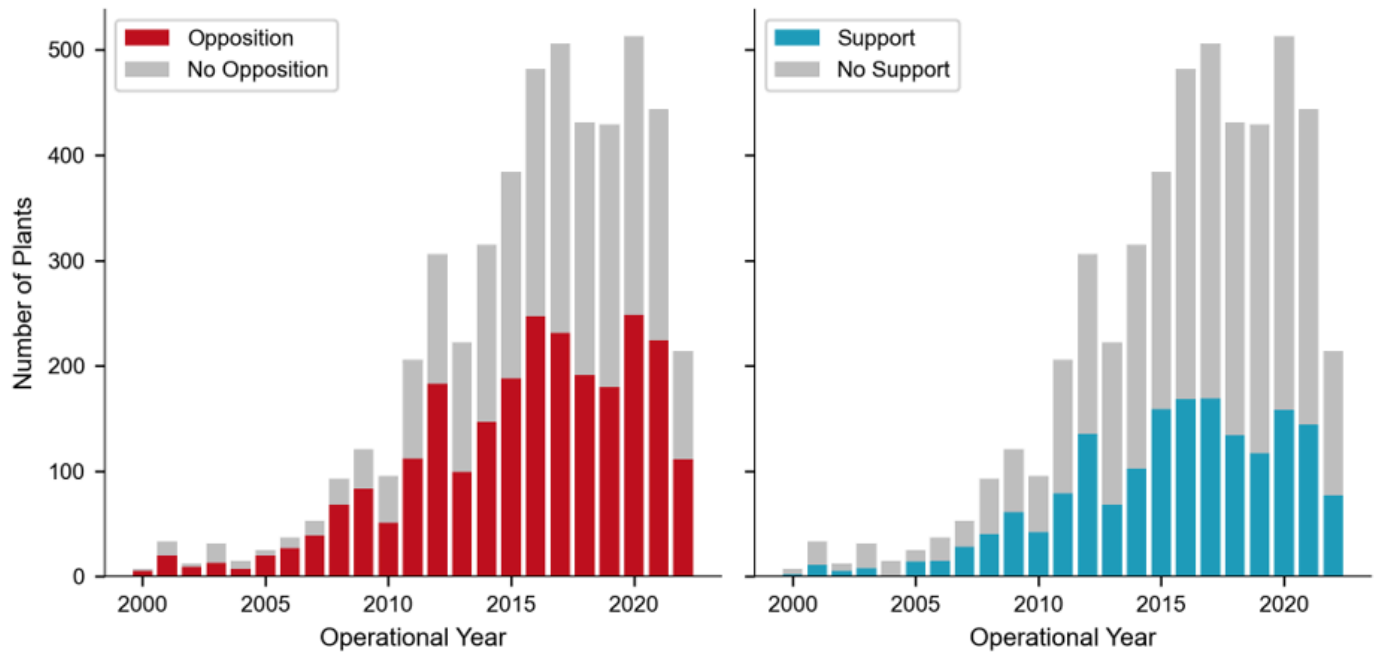


Figure 3

Temporal trends in mentions of opposition vs. support as identified by the LLM over time (2000-2022).

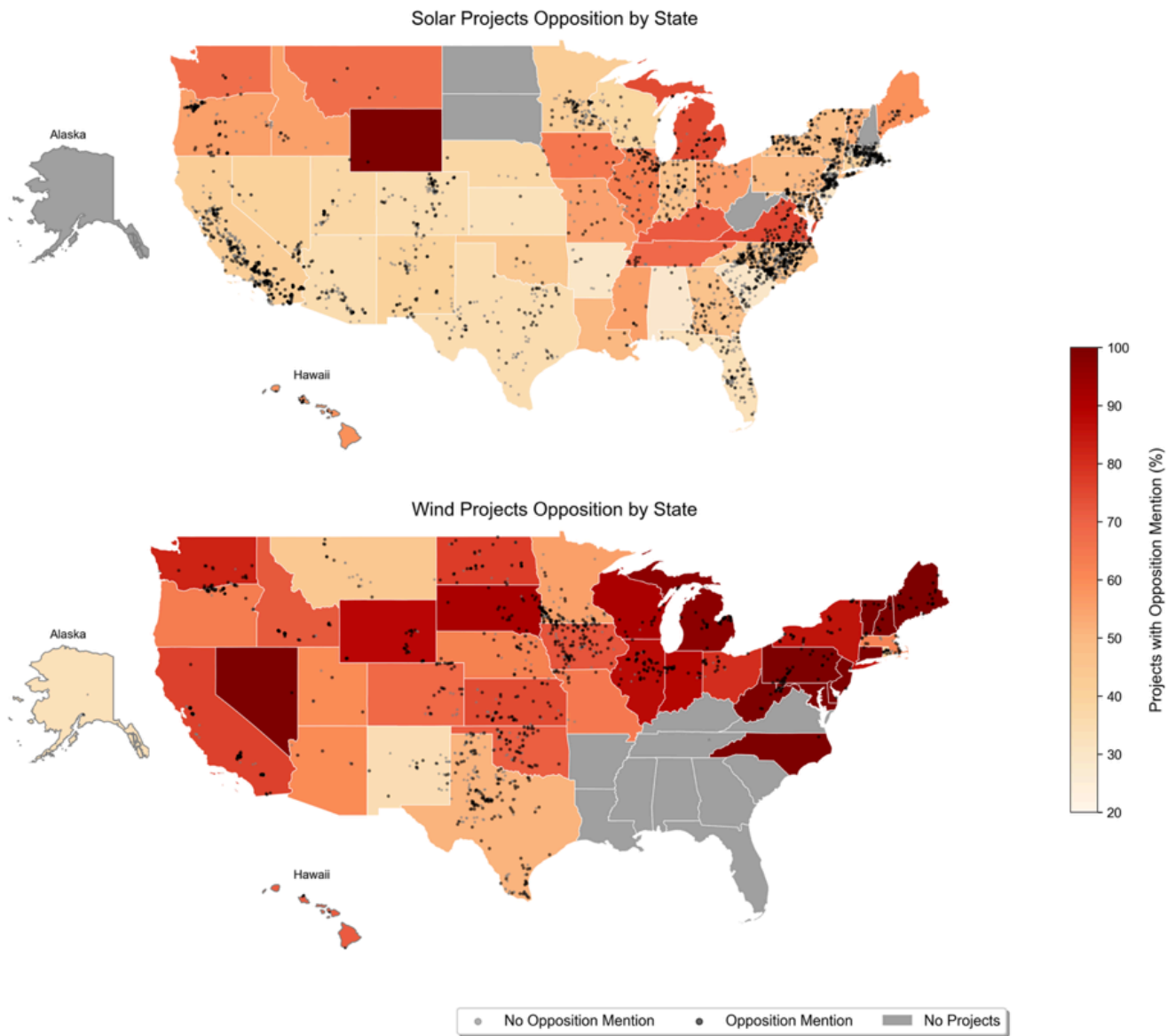


Figure 4

Spatial distribution of the percentage of projects with evidence of opposition by state.

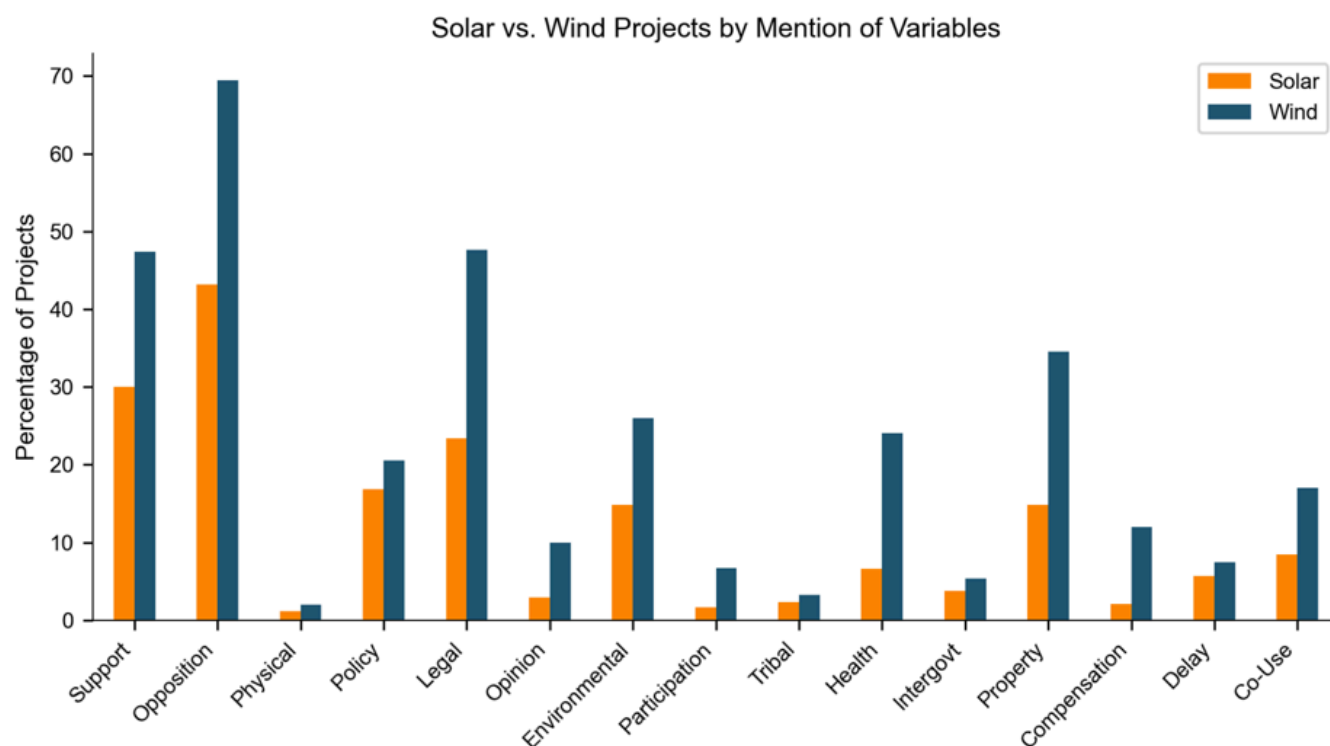


Figure 5

Percentage of solar and wind projects with mentions of each public perception variable.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalInformation.docx](#)