Algorithmic Electability: Generative AI, Gendered Political Bias, and LGBTQ Candidates

Zachary Baum

Department of Political Science

Rutgers, The State University of New Jersey

New Brunswick, NJ, USA

ORCID: 0009-0003-7374-684X

Email: *zb207@scarletmail.rutgers.edu*

Corresponding Author

Zachary Baum

Department of Political Science

Rutgers, The State University of New Jersey

89 George Street

New Brunswick, NJ 08901

United States

Author Note

The author declares no conflicts of interest.

Funding Statement

This research received support from the Institute for Humane Studies through the Humane Studies Fellowship. The funder had no role in the design of the study; the collection, analysis, or

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

interpretation of data; the writing of the manuscript; or the decision to submit the article for publication.

Data Availability Statement

The data, replication materials, experimental prompts, candidate profiles, and analysis code supporting the findings of this study are available from the corresponding author upon reasonable request.

Ethics Statement

This study did not involve human participants. The research analyzed outputs generated by publicly available large language models using standardized prompts. Institutional Review Board approval was therefore not required.

Acknowledgments

The author thanks colleagues in the Department of Political Science at Rutgers University for helpful discussions during the development of this project.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Algorithmic Electability: Generative AI, Gendered Political Bias, and LGBTQ Candidates

Abstract

Political candidates are not judged solely on their experience or qualifications. Ideas about what makes someone a credible leader, including how they present themselves and whether they are seen as capable of winning, also shape these evaluations. With large language models (LLMs) now being used in political campaigns, journalism, and everyday information searches, I am interested in whether they make similar assumptions. Researchers have documented bias in AI across a range of settings, but we still know relatively little about how these systems judge political candidates or whether those judgments change depending on a candidate's identity.

Whether a candidate is considered qualified for office is not necessarily the same as whether they are seen as capable of winning an election. I constructed twenty candidate profiles that varied sexual orientation, transgender identity cues, gender presentation, and leadership descriptions. Six AI platforms were assessed using repeated sessions. The three recorded numerical outcomes were competence, electability, and political strength.

The clearest differences emerged in electability and political strength. Ratings of competence, by comparison, were relatively consistent across candidate profiles. LGBTQ candidates generally received somewhat less favorable assessments of their political prospects than non-LGBTQ candidates with similar credentials. The larger differences, however, concerned leadership style. Profiles emphasizing traditionally masculine qualities tended to fare better than those emphasizing collaboration or approaches that departed from conventional expectations of political leadership. The pattern was not consistent across all six models. Some

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

produced larger differences than others, raising questions about how each system arrives at its judgments of political viability.

This article introduces the concept of algorithmic electability to describe systematic variation in AI-generated assessments of candidate viability under otherwise equivalent informational conditions. Rather than drawing conclusions about the origins of these differences or their downstream effects on political behavior, the study documents AI-generated candidate evaluations as an emerging object of inquiry and extends research on candidate evaluation and political communication into the domain of generative artificial intelligence.

Keywords: generative artificial intelligence; LGBTQ candidates; algorithmic bias; electability; descriptive representation; candidate evaluation; gender nonconformity; political behavior; AI and democracy; large language models; political communication

Introduction

Generative artificial intelligence has rapidly become embedded within contemporary political communication. Large language models (LLMs) are increasingly used to summarize political information, evaluate candidates, draft campaign materials, synthesize news coverage, and answer questions regarding elections and public affairs. Rather than functioning solely as passive repositories of information, these systems generate evaluative judgments that increasingly shape how users encounter and interpret political actors.

Political communication scholars have long argued that information intermediaries—including journalists, search engines, and digital platforms—play an important role in structuring political knowledge and influencing perceptions of candidates (Kreiss &

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

McGregor, 2019). As generative AI systems become integrated into search engines, productivity software, campaign workflows, and public-facing information platforms, understanding the content of AI-generated political evaluations has become an increasingly important scholarly question.

Existing research on algorithmic bias has primarily focused on domains such as employment, criminal justice, facial recognition, lending, and content moderation (Benjamin, 2019; Noble, 2018; Crawford, 2021). Comparatively little attention has been devoted to how generative AI systems evaluate political candidates or whether those evaluations vary systematically across candidate identities. This omission is increasingly consequential given the expanding role of AI-generated information within democratic communication environments.

This study examines whether six large language models evaluate otherwise comparable political candidates differently when their sexual orientation, gender identity, or leadership presentation changes. It focuses on the evaluations these systems produce, rather than whether those evaluations ultimately influence election outcomes.

Candidate Evaluation and Electability Heuristics

Political scientists have long distinguished between judgments of candidate competence and judgments of candidate electability. Research documents that voters, donors, journalists, campaign professionals, and party elites frequently evaluate candidates not only according to their qualifications but also according to beliefs about whether those candidates are capable of winning elections (Lawless & Fox, 2010; Sanbonmatsu, 2002; Brooks, 2013).

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

These electability heuristics often operate independently of evaluations of competence. Individuals may personally regard a candidate as highly qualified while simultaneously expressing doubts about the candidate's electoral prospects because they believe other voters may respond negatively to particular social identities or demographic characteristics. Such strategic evaluations influence candidate recruitment, fundraising, endorsements, media coverage, and political ambition.

Research on women candidates illustrates this distinction particularly well. Although explicit evaluations of competence have become increasingly equal over time, women frequently continue to encounter skepticism regarding their electoral viability (Brooks, 2013). Similar patterns have been documented for racial minority candidates and LGBTQ candidates, where strategic concerns regarding voter acceptance frequently persist despite favorable assessments of professional qualifications.

Although these electability heuristics have been studied extensively among human decision makers, little research has examined whether comparable patterns emerge in AI-generated political evaluations. If generative AI systems produce similar distinctions between competence and electability, they may reproduce longstanding political heuristics within digital information environments.

Gendered Leadership Prototypes

Research on political leadership further suggests that evaluations of electoral viability are shaped by culturally embedded expectations regarding legitimate leadership. Political leadership

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

has historically been associated with characteristics such as decisiveness, authority, confidence, rationality, toughness, and institutional professionalism, traits that have frequently been socially coded as masculine (Eagly & Karau, 2002).

Role congruity theory argues that individuals whose identities or leadership styles depart from these expectations may encounter evaluative penalties because they violate dominant leadership prototypes rather than because they lack competence (Eagly & Karau, 2002). Political behavior research similarly shows that voters often distinguish between candidate qualifications and candidate "fit" with conventional expectations regarding political leadership (Huddy & Terkildsen, 1993).

These leadership schemas are particularly relevant for understanding evaluations of transgender and gender-nonconforming candidates. Departures from conventional gender presentation may challenge implicit assumptions regarding political legitimacy even when objective qualifications remain equivalent. Consequently, AI-generated evaluations may reproduce differences associated not only with candidate identity but also with the leadership characteristics emphasized within candidate biographies.

LGBTQ Candidates and Political Representation

Research examining LGBTQ candidates clarifies that sexual orientation and gender identity can shape political evaluations in complex ways. LGBTQ candidates have historically confronted barriers related to visibility, perceived acceptability, fundraising, media coverage, and assumptions regarding electoral competitiveness (Haider-Markel, 2010; Reynolds, 2013).

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Importantly, these barriers frequently reflect strategic judgments rather than direct assessments of competence. Voters and political elites may view LGBTQ candidates favorably while simultaneously questioning whether other voters will support them electorally. Such distinctions mirror broader research demonstrating that perceptions of electability often differ from perceptions of candidate quality.

These findings suggest that candidate identity may influence evaluations through assessments of political viability rather than professional capability. Whether contemporary AI systems generate comparable distinctions remains an open empirical question.

Algorithmic Bias and Political Evaluation

Scholarship on algorithmic bias has demonstrated that computational systems can reproduce patterns of inequality present within the data, institutions, and social environments from which they are developed (Benjamin, 2019; Noble, 2018; Crawford, 2021). Existing studies have documented unequal outcomes across domains including employment, lending, criminal justice, facial recognition, and online search.

Generative AI presents a distinct challenge because large language models do not merely retrieve information; they generate evaluative narratives and recommendations. Consequently, these systems may produce different assessments of political candidates even when presented with otherwise equivalent information.

Importantly, the present study does not assume that observed differences arise from any particular feature of model development. Contemporary AI systems differ in architecture,

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

alignment procedures, safety mechanisms, and training approaches, but the proprietary nature of these systems makes causal attribution impossible. Accordingly, the analysis focuses on documenting variation in AI-generated evaluations rather than explaining the mechanisms responsible for those differences.

Theoretical Contribution

This article makes three principal contributions to scholarship on political communication, candidate evaluation, and artificial intelligence.

First, it introduces the concept of algorithmic electability to describe systematic variation in AI-generated assessments of candidate viability under otherwise equivalent informational conditions. Existing research has extensively examined how voters, party elites, journalists, and donors evaluate political candidates, while scholarship on algorithmic bias has focused primarily on domains such as employment, criminal justice, lending, and content moderation. By conceptualizing AI-generated candidate evaluations as a distinct form of political judgment, this study extends theories of electability into the rapidly expanding domain of generative artificial intelligence.

Second, the study contributes to research on candidate evaluation by demonstrating that AI-generated differences are concentrated primarily in assessments of electability and political strength rather than competence. This distinction mirrors longstanding findings in political behavior research showing that candidates may be viewed as professionally qualified while simultaneously being perceived as less electorally viable because of social identities or departures from conventional leadership expectations. Examining whether similar distinctions

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

emerge in AI-generated evaluations broadens existing theories of electability beyond human decision-makers to include increasingly influential computational systems.

Third, by comparing six commercially available large language models using standardized candidate vignettes, this study documents meaningful variation across contemporary AI systems. Rather than treating generative AI as a single, homogeneous technology, the analysis demonstrates that AI-generated political evaluations differ across models under identical informational conditions. Although the present study does not identify the mechanisms responsible for those differences, it establishes AI-generated candidate evaluation as an important new object of inquiry and provides a foundation for future research examining how generative AI may shape political communication, candidate assessment, and democratic representation.

Expectations

H1: Candidates whose leadership presentation departs from conventional masculine political norms will receive lower evaluations of electability and political strength than otherwise comparable candidates.

H2: LGBTQ candidates will receive lower electability evaluations than otherwise comparable non-LGBTQ candidates.

H3: Differences associated with candidate identity will be larger for electability than for competence evaluations.

H4: The magnitude of identity-related evaluation differences will vary across AI systems.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Methods

I used candidate vignettes to compare how six AI platforms evaluated state legislative candidates with different identity cues and leadership descriptions. The observations are AI-generated ratings, not responses from human participants. Some contrasts vary more than one characteristic and therefore cannot isolate the effect of a single cue.

Profiles were constructed with broadly comparable professional and political backgrounds, although the manipulations were not uniform across all pairs. In particular, the O–P and S–T comparisons change additional characteristics; interpretations of those contrasts must account for this.

Candidate Vignette Construction

Candidate profiles were designed to resemble realistic candidates for state legislative office in the United States. Each vignette included information regarding:

- educational attainment;
- occupational experience;
- community involvement;
- prior public service;
- policy priorities;
- endorsements;
- family background; and
- campaign messaging.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Within most paired conditions, core professional qualifications and political experience were held constant while identity cues varied. The leadership-presentation comparisons involved additional differences in biographical framing, which limit the extent to which observed rating differences can be attributed to a single characteristic.

Identity manipulations included references to:

- same-sex spouses,
- LGBTQ community involvement,
- transgender identity,
- gender ambiguity,
- and nontraditional gender presentation.

Leadership manipulations altered descriptive language associated with political leadership styles. Traditional leadership conditions emphasized characteristics such as decisiveness, pragmatism, confidence, discipline, and institutional professionalism. Nontraditional leadership conditions emphasized collaboration, empathy, emotional openness, expressiveness, and participatory leadership.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Table 1
Experimental Candidate Conditions

Condition	Experimental Manipulation
Control	Opposite-sex spouse cue
LGBTQ	Same-sex spouse cue
Transgender	Explicit transgender identity
Gender Ambiguous	Neutralized gender markers
Nontraditional Leadership	Expressive, collaborative, or unconventional leadership cues
Traditional Leadership	Disciplined, pragmatic, conventionally authoritative leadership cues

AI Systems Evaluated

The study evaluated responses generated by six commercially available large language models:

- ChatGPT
- Claude
- Gemini
- Grok
- Perplexity
- DeepSeek

These systems were selected because they represent widely used publicly accessible generative AI platforms that differ in training procedures, alignment strategies, and model architecture.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

The purpose of including multiple models was not to identify the causes of model-specific differences but rather to evaluate whether AI-generated political assessments vary across contemporary commercially available systems.

Prompt Standardization

Each AI platform received identical prompts and identical candidate vignettes.

Prompts instructed the model to evaluate each candidate across multiple dimensions including:

- electability;
- competence;
- political strength

The candidate biographies varied according to the experimental conditions, including sexual orientation cues, transgender identity, gender presentation, and leadership-related descriptions.

Appendix A summarizes the candidate conditions.

Evaluation Procedure

Independent evaluations were conducted using separate chat sessions to reduce the influence of conversational memory and stochastic response variation. Repeated evaluations

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

were collected because contemporary LLMs generate probabilistic outputs that may vary across identical prompts.

The intended balanced design of twenty profiles, six platforms, and five runs per profile-platform combination would yield 600 evaluations. The original data collection was not fully balanced. Some candidate–model combinations had fewer than five recorded evaluations, while additional evaluations were collected for the final leadership-presentation profiles. Missing numerical scores were retained as missing rather than imputed. Analyses used the available recorded ratings for each outcome.

Repeated evaluations were retained rather than averaged to characterize within-model variability and to permit estimation of uncertainty across repeated model outputs.

Outcome Coding

AI-generated responses were converted into quantitative outcome measures using a standardized coding protocol.

The dependent variables were numerical scores generated by the models. Most responses provided explicit numerical ratings, which were extracted directly. Accompanying narrative explanations were retained for verification but were not quantitatively coded. Because the numerical ratings were extracted rather than independently assigned by human coders, intercoder reliability statistics were not applicable to these outcomes.

Statistical Analysis

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

The analysis proceeded in several stages.

First, descriptive statistics summarized average evaluations across experimental conditions.

Second, Welch two-sample t-tests compared mean evaluations between LGBTQ and non-LGBTQ candidate profiles.

Third, ordinary least squares regression models with heteroskedasticity-robust standard errors estimated the association between candidate identity cues and AI-generated evaluations. Interaction terms between candidate identity and AI platform were subsequently included to assess whether observed evaluation differences varied across systems. Because repeated evaluations were generated within the same AI platforms and candidate vignettes, observations may not be fully independent. The regression estimates should therefore be interpreted as descriptive associations rather than definitive causal or population-level effects.

Robustness Checks

To assess the stability of the findings, I repeated the analyses separately for each AI platform, estimated interactions between candidate identity and platform, and compared the results across electability, competence, and political strength. These checks were intended to determine whether the observed differences were concentrated in particular models or evaluation dimensions.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Reproducibility and Limitations

Generative AI systems are proprietary technologies that evolve over time through updates to model architecture, alignment procedures, safety systems, and training data. Consequently, identical prompts administered at different points in time may produce different outputs.

The objective of this study is therefore not to establish immutable properties of any particular model but rather to characterize the behavior of contemporary commercially available LLMs during the study period.

Several limitations should be considered when interpreting these findings.

First, this study evaluates AI-generated candidate evaluations rather than human political behavior. Although large language models are increasingly used to summarize political information, generate strategic recommendations, and evaluate candidates, the present design does not examine whether AI-generated evaluations influence voters, donors, journalists, campaign professionals, political consultants, or party elites. Consequently, the findings should not be interpreted as evidence that observed differences translate directly into electoral outcomes or political decision making. Rather, the study documents systematic variation in AI-generated evaluations under standardized conditions.

Second, the study evaluates model outputs rather than model architecture. Contemporary large language models are proprietary systems whose training corpora, alignment procedures,

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

reinforcement learning methods, moderation systems, and parameter configurations are not publicly available in sufficient detail to permit causal attribution. Although meaningful differences were observed across AI platforms, the present analysis cannot determine whether those differences arise from training data, alignment strategies, reinforcement learning, safety mechanisms, model architecture, or other aspects of system development. These mechanisms remain important questions for future research.

Third, large language models are dynamic systems. AI developers regularly update model weights, alignment procedures, safety policies, and deployment settings. As a result, identical prompts administered at different points in time may produce different outputs. The findings reported here therefore characterize model behavior during the period in which the data were collected rather than establishing permanent properties of any individual AI system.

Fourth, the vignette design simplifies actual electoral settings. Although many biographical details were similar across conditions, some profile pairs varied in more than the intended identity or leadership cue. As a result, not every contrast identifies a single characteristic's effect.

Fifth, the study focuses on a limited set of candidate identities and leadership presentations. Although the analyses examine sexual orientation, transgender identity, gender presentation, and leadership style, they do not evaluate many other politically relevant identities or their intersections, including race, ethnicity, disability, religion, age, or socioeconomic background. Future research should examine how multiple marginalized identities interact within AI-generated political evaluations.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Finally, the present study documents systematic differences in AI-generated evaluations but cannot determine whether those differences reflect broader societal stereotypes, patterns present in publicly available political discourse, design choices made by AI developers, or other mechanisms. Identifying the origins of these observed differences will require complementary research combining computational analyses, model auditing, transparency initiatives, and behavioral experiments involving human participants.

Results

Across all AI systems, evaluations consistently distinguished between candidate competence and candidate viability. While competence ratings differed little across identity conditions, electability and political strength exhibited more systematic variation. These patterns suggest that any observed differences in AI-generated evaluations were concentrated primarily in assessments related to electoral competitiveness rather than perceptions of professional qualifications.

Average electability scores were modestly higher for non-LGBTQ candidates (M = 7.26) than for otherwise comparable LGBTQ candidates (M = 7.01), representing an average difference of approximately 0.25 points on the ten-point evaluation scale. Political strength exhibited a similar pattern, with non-LGBTQ candidates receiving average scores of 7.43 compared to 7.25 for LGBTQ candidates. By contrast, competence evaluations differed negligibly across conditions, with both groups receiving average ratings of approximately 7.5.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Although the observed differences in electability and political strength were relatively modest in absolute magnitude, they appeared consistently across multiple AI systems and repeated candidate evaluations.

Welch two-sample t-tests indicated that political strength evaluations differed significantly across identity conditions. Non-LGBTQ candidates received higher average political strength evaluations than LGBTQ candidates ($t = 2.73$, $p = .0066$). The estimated difference of approximately 0.19 points corresponds to a modest but statistically reliable reduction in AI-generated assessments of political strength.

By contrast, differences in competence evaluations were substantively negligible and statistically indistinguishable from zero. This distinction between competence and viability represents one of the central empirical findings of the study, suggesting that AI systems did not characterize LGBTQ candidates as less qualified but instead evaluated them somewhat less favorably on dimensions associated with electoral success.

Transgender Candidate Evaluations

The comparison involving an explicitly transgender candidate produced lower average electability ratings for the profile disclosing transgender identity. The average declined from 7.30 in the comparison condition to 6.85 in the transgender condition, a difference of approximately 0.45 points on the ten-point scale. Because the profiles also differed in the disclosure of family information, this contrast should not be interpreted as isolating transgender identity alone.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Importantly, this pattern was not mirrored in competence evaluations. Average competence ratings remained nearly identical across conditions, differing by only 0.06 points. Political strength also declined modestly for transgender candidates relative to comparison profiles.

These findings suggest that AI-generated differences associated with transgender identity were concentrated primarily in assessments of political viability rather than professional capability. While the present study cannot determine the mechanisms responsible for these differences, the observed pattern is consistent with longstanding distinctions in political behavior research between judgments of competence and judgments of electability.

Gendered Leadership Norms and Political Legitimacy

The analysis also examined whether departures from conventional gender presentation influenced AI-generated political evaluations. Candidate profiles employing gender-neutral or ambiguous identity cues produced comparatively modest differences in evaluation, suggesting that ambiguity alone did not consistently trigger substantial penalties across models.

The clearest evidence of gendered algorithmic bias emerged in the leadership presentation experiment. Candidates framed through traditionally coded leadership descriptors, including discipline, pragmatism, confidence, and conventional leadership suitability, received substantially more favorable evaluations than candidates framed through expressive, emotionally open, collaborative, and unconventional leadership characteristics.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

The traditional leadership prototype condition received an average electability score of 7.68, compared to just 6.74 for the nontraditional leadership condition, representing a difference of nearly one full point on the ten-point evaluation scale. Similar disparities emerged in other outcome measures. Traditional leadership candidates received an average competence score of 7.71, compared to 7.35 in the nontraditional condition, while political strength evaluations declined from 7.70 to 7.06.

These differences are consistent with established expectations surrounding gender and political leadership. However, because the profiles also vary in candidate gender and aspects of their biographies, the observed ratings cannot be attributed exclusively to leadership style.

Regression Analysis

Regression analyses produced results consistent with the descriptive comparisons. Across specifications, LGBTQ identity was associated with lower average electability evaluations after accounting for other variables included in the model.

The estimated coefficient indicated an average reduction of approximately one-quarter of a point on the ten-point electability scale ($b = –0.248$, $p = .0015$). Although statistically reliable, the substantive magnitude of this effect was modest.

No statistically significant relationship emerged between LGBTQ identity and competence evaluations ($b = 0.003$, $p = .968$), reinforcing the broader descriptive finding that AI

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

systems differentiated more strongly on assessments of political viability than professional capability.

Taken together, these findings indicate that observed identity-related differences were concentrated in electability-related outcomes rather than generalized across all dimensions of candidate evaluation.

Interaction Effects Across AI Systems

Interaction models further demonstrated substantial variation across AI systems. The magnitude of the LGBTQ electability penalty differed significantly depending on the model generating the evaluation. ChatGPT exhibited the largest apparent electability penalty for LGBTQ candidates, while other systems, including Claude, DeepSeek, Gemini, and Perplexity, showed substantially smaller penalties or near-null effects after accounting for interaction terms.

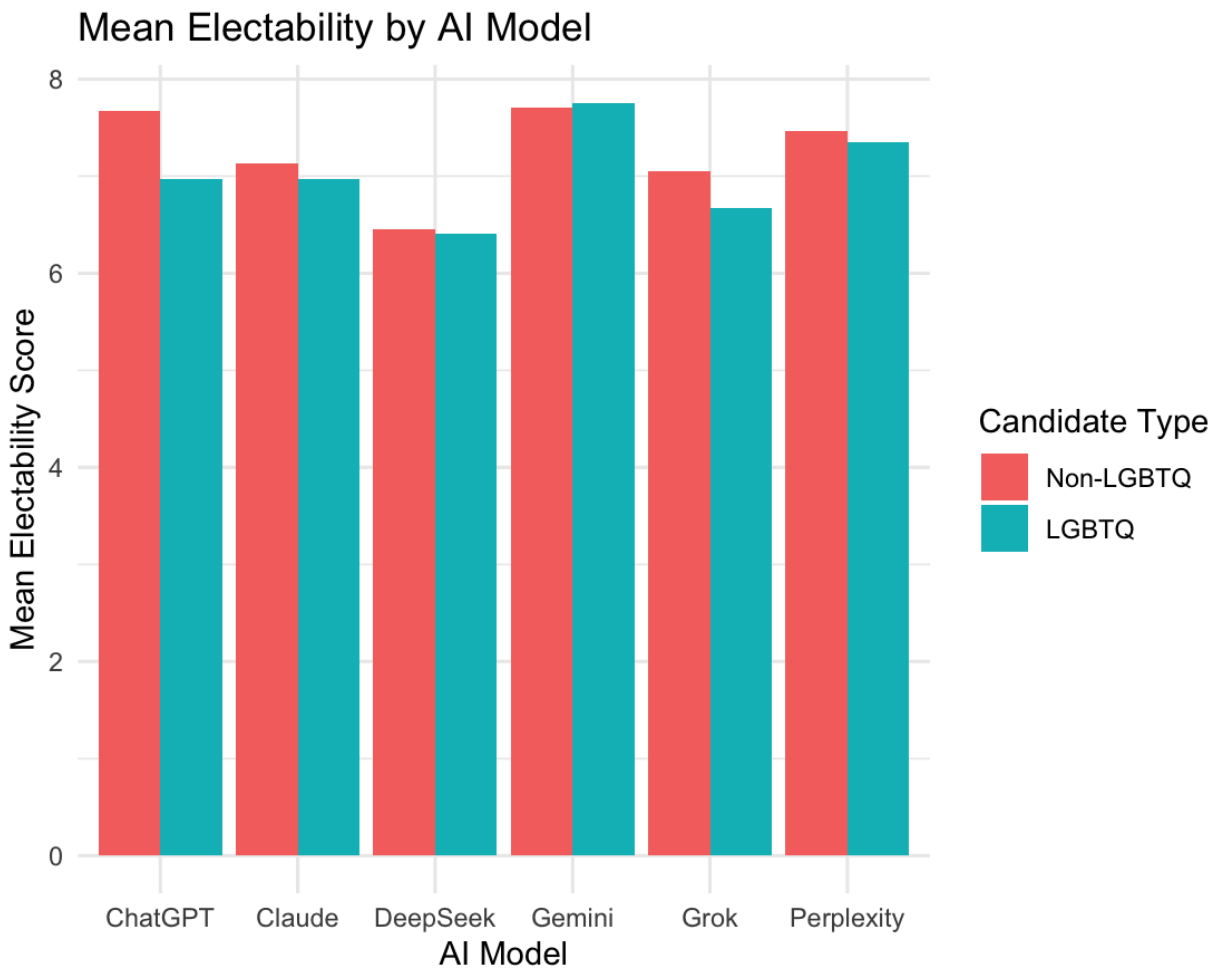
These results indicate that AI-generated evaluations are not uniform across contemporary large language models. Under identical candidate profiles and standardized prompts, some models generated larger differences in electability evaluations than others. The present study, however, does not identify the mechanisms responsible for these differences. Because the analysis does not examine model architecture, training corpora, alignment procedures, or reinforcement learning directly, it cannot determine why these models produced different outputs. Accordingly, the observed variation should be interpreted descriptively rather than as evidence regarding the specific causes of algorithmic differences.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Figure 1. Electability Ratings by AI Model and LGBTQ Candidate Condition

Mean electability evaluations across six generative AI systems for LGBTQ and non-LGBTQ political candidates. Across most models, Candidates classified as LGBTQ generally received lower electability scores than those classified as non-LGBTQ in the analyzed observations, though the magnitude of the penalty varied substantially by AI system. ChatGPT and Grok exhibited the largest negative gaps in evaluations, while Gemini showed little to no apparent electability penalty. These findings suggest that algorithmic evaluations of political viability are not uniform across generative AI platforms and may reflect model-specific differences in training data, alignment procedures, or embedded political heuristics.

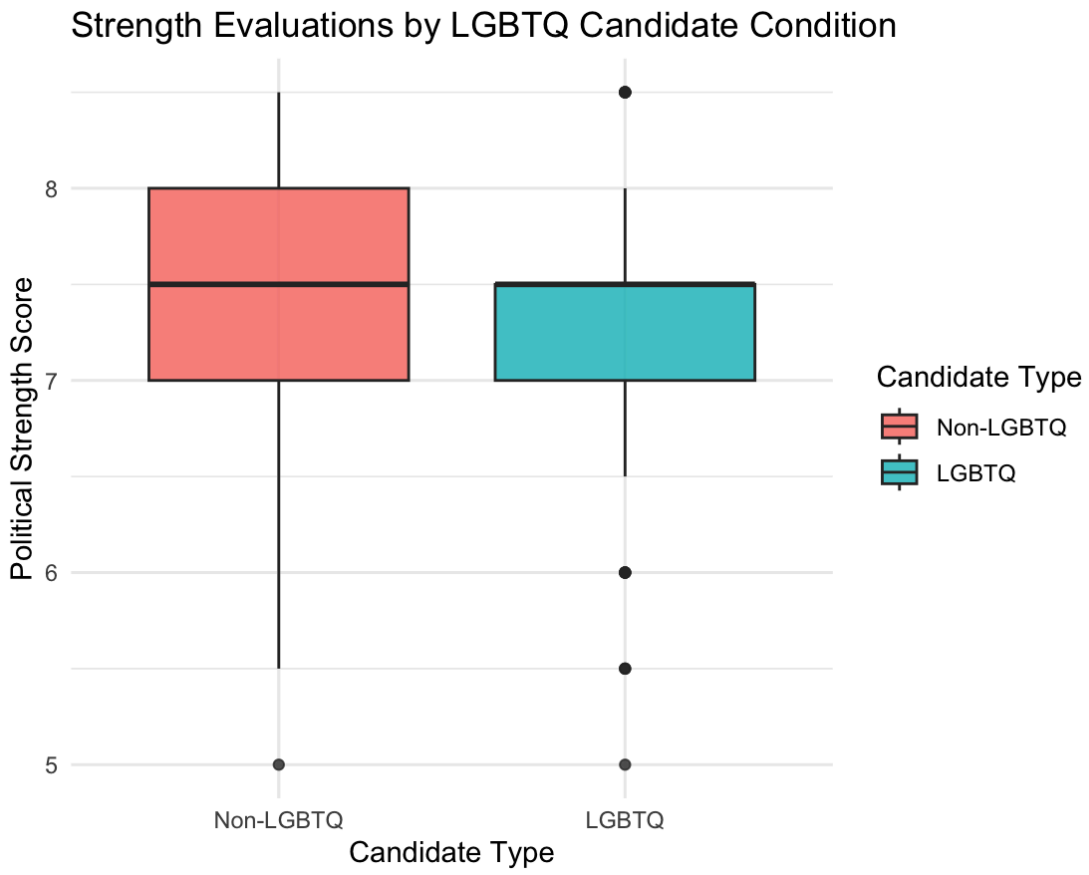
https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved



https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Figure 2. Strength Evaluations by LGBTQ Candidate Condition

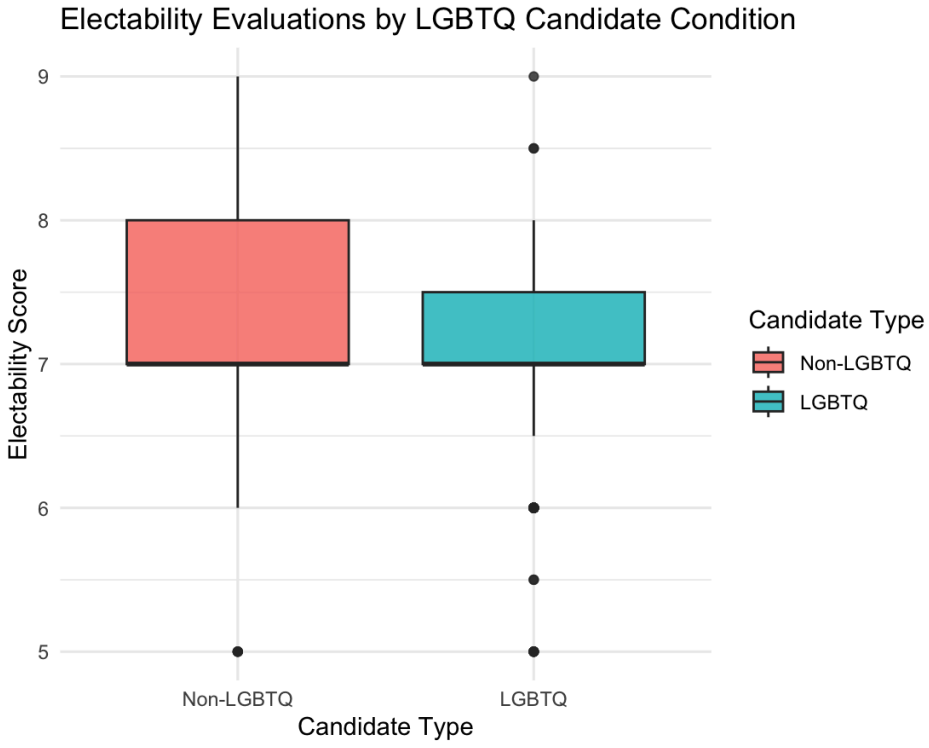
Boxplots displaying the distribution of political strength evaluations for LGBTQ and non-LGBTQ candidates across all AI systems included in the study. Non-LGBTQ candidates generally received higher strength evaluations and exhibited a slightly higher median score compared to LGBTQ candidates. Although the distributions overlap substantially, LGBTQ candidates received lower average evaluations overall, consistent with the statistically significant differences identified in the regression and t-test analyses. The results suggest that generative AI systems may associate LGBTQ identity cues with reduced perceptions of political strength or viability across the candidate profiles evaluated in this study.



https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Figure 3. Electability Evaluations by LGBTQ Candidate Condition

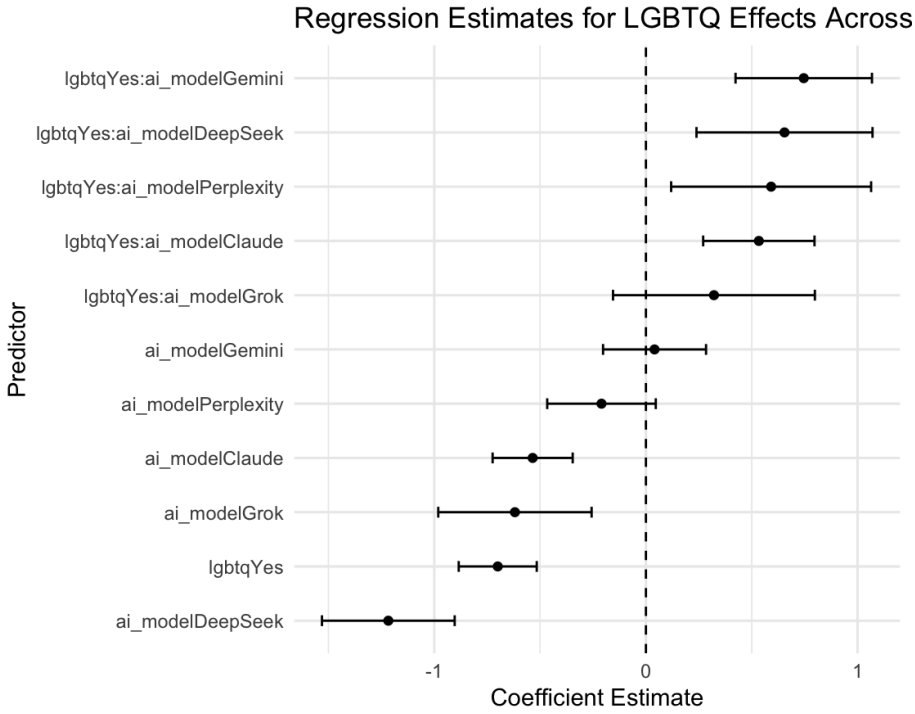
Boxplots displaying the distribution of electability evaluations for LGBTQ and non-LGBTQ political candidates across all AI systems included in the study. Non-LGBTQ candidates generally received higher electability evaluations and exhibited a slightly higher central tendency than LGBTQ candidates, across the experimental conditions included in the analysis. While the distributions overlap considerably, LGBTQ candidates received consistently lower average electability scores overall, supporting the regression findings indicating a statistically significant electability penalty associated with LGBTQ identity cues. These findings suggest that generative AI systems may reproduce broader assumptions regarding political viability and perceived electoral competitiveness.



https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Figure 4. Regression Estimates for LGBTQ Effects Across AI Models

Coefficient plot displaying regression estimates and 95% confidence intervals from interaction models examining the relationship between LGBTQ candidate identity and electability evaluations across generative AI systems. The dashed vertical line at zero represents no estimated effect. Negative coefficients indicate lower electability evaluations relative to the reference category, while positive interaction terms indicate reductions in the magnitude of the LGBTQ electability penalty within specific AI models relative to the baseline model. The results demonstrate substantial variation across AI systems in how LGBTQ candidate cues shape perceptions of political viability, suggesting that algorithmic evaluations of electability are model-dependent rather than uniform across generative AI platforms.



Conclusion

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Algorithmic Electability: Generative AI, Gendered Political Bias, and LGBTQ Candidates

This study examined whether contemporary large language models generate systematically different evaluations of political candidates across profiles varying in identity cues and leadership presentation. Although professional qualifications were held constant in most paired comparisons, some conditions also varied in biographical framing. Across six commercially available AI systems, the analyses identified a consistent distinction between evaluations of competence and evaluations of political viability. Although competence ratings differed little across candidate conditions, differences emerged more consistently in assessments of electability and political strength.

The findings suggest that AI-generated candidate evaluations may distinguish between professional qualifications and perceived electoral viability in ways that are consistent with longstanding theories of electability heuristics and gendered leadership expectations. In particular, profiles emphasizing conventionally masculine leadership characteristics generally received more favorable evaluations than profiles emphasizing collaborative or nontraditional leadership styles. Because these conditions also differed in candidate gender and aspects of their biographies, the results cannot isolate the effect of leadership presentation. Similarly, LGBTQ identity cues were associated with modest reductions in assessments of political viability across several outcomes, although the magnitude of these differences varied considerably across AI systems.

Equally important, the results demonstrate that AI-generated political evaluations are not uniform across contemporary large language models. Identical candidate profiles produced different patterns of evaluation across platforms, underscoring the importance of treating AI systems as distinct political information intermediaries rather than assuming that generative AI

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

produces homogeneous outputs. Because the present study evaluates model behavior rather than model architecture, however, it cannot determine why these differences emerged or identify the specific mechanisms responsible for variation across systems.

The primary contribution of this article is conceptual. By introducing the concept of algorithmic electability, the study extends scholarship on candidate evaluation into the rapidly evolving domain of generative artificial intelligence. Rather than assuming that AI systems are either neutral technologies or direct reproductions of existing political inequalities, the concept directs attention to AI-generated candidate evaluations themselves as a distinct phenomenon worthy of empirical investigation.

Several limitations remain. The analyses characterize AI-generated outputs under standardized experimental conditions rather than human political behavior, and they do not establish whether these evaluations influence voters, journalists, campaign professionals, donors, or party elites. Nor do they identify the mechanisms through which different AI systems generate varying evaluations. Future research should examine whether AI-generated candidate assessments shape political decision-making in real-world settings, explore how these patterns evolve as models are updated, and investigate whether similar dynamics emerge across additional political identities, institutional contexts, and electoral systems.

A candidate can be regarded as qualified for office while still being judged unlikely to win. That distinction, familiar in research on political recruitment and candidate evaluation, also appears in the AI-generated assessments examined here. The results suggest that evaluating whether AI systems recognize candidates' qualifications is not

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

sufficient; researchers must also examine the assumptions these systems make about who is politically viable.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

References

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of Machine Learning Research, 81*, 149–159.

Brooks, D. J. (2013). He runs, she runs: Why gender stereotypes do not harm women candidates. Princeton University Press.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research, 81*, 1–15.

Carroll, S. J., & Fox, R. L. (2018). *Gender and elections: Shaping the future of American politics* (4th ed.). Cambridge University Press.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Crawford, K., & Paglen, T. (2021). Excavating AI: The politics of training sets for machine learning. AI & Society, 36(4), 1105–1116. https://doi.org/10.1007/s00146-021-01162-8

Doan, A. E., Loehr, A., & Miller, L. R. (2014). Formal rights and informal privileges for same-sex couples: Evidence from a national survey experiment. *American Politics Research, 42*(5), 815–838. https://doi.org/10.1177/1532673X13504414

Dolan, K. (2014). *When does gender matter? Women candidates and gender stereotypes in American elections*. Oxford University Press.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*(3), 573–598. https://doi.org/10.1037/0033-295X.109.3.573

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Flores, A. R. (2015). Attitudes toward transgender rights: Perceived knowledge and secondary interpersonal contact. Politics, Groups, and Identities, 3(3), 398–416. https://doi.org/10.1080/21565503.2015.1050414

Fulton, S. A. (2012). Running backwards and in high heels: The gendered quality gap and incumbent electoral success. *Political Research Quarterly, 65*(2), 303–314. https://doi.org/10.1177/1065912911407629

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

Haider-Markel, D. P. (2010). *Out and running: Gay and lesbian candidates, elections, and policy representation*. Georgetown University Press.

Huddy, L., & Terkildsen, N. (1993). Gender stereotypes and the perception of male and female candidates. *American Journal of Political Science, 37*(1), 119–147. https://doi.org/10.2307/2111526

Kreiss, D., & McGregor, S. C. (2019). Technology firms shape political communication: The work of Microsoft, Facebook, Twitter, and Google with campaigns during the 2016 U.S. presidential cycle. *Political Communication, 36*(2), 155–177. https://doi.org/10.1080/10584609.2018.1556350

Lawless, J. L., & Fox, R. L. (2010). It still takes a candidate: Why women don't run for office. Cambridge University Press.

Mansbridge, J. (2003). Rethinking representation. American Political Science Review, 97(4), 515–528.

Mansbridge, J. (1999). Should Blacks represent Blacks and women represent women? A contingent "yes." *The Journal of Politics, 61*(3), 628–657. https://doi.org/10.2307/2647821

McDermott, M. L. (1998). Race and gender cues in low-information elections. *Political Research Quarterly, 51*(4), 895–918. https://doi.org/10.1177/106591299805100404

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Pitkin, H. F. (1967). *The concept of representation*. University of California Press.

Reynolds, A. (2013). Representation and rights: The impact of LGBT legislators in comparative perspective. *American Political Science Review, 107*(2), 259–274. https://doi.org/10.1017/S0003055413000051

Sanbonmatsu, K. (2002). Gender stereotypes and vote choice. *American Journal of Political Science, 46*(1), 20–34. https://doi.org/10.2307/3088412

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance, 12*(4), 505–523. https://doi.org/10.1111/rego.12158

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Public Affairs.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

APPENDIX A

Experimental Candidate Profiles

The study materials contain twenty candidate profiles (A–T). Several profiles use the same treatment label but describe different candidates. Certain comparisons also change multiple characteristics, as detailed in the original vignettes.

Experimental Conditions

Condition	Manipulation
A	Michael Thompson — male control candidate with opposite-sex spouse
B	Michael Thompson — same-sex spouse cue (husband)
C	Daniel Carter — male control candidate with opposite-sex spouse

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

D	Daniel Carter — same-sex spouse cue (husband)
E	Jessica Ramirez — female control candidate with opposite-sex spouse
F	Jessica Ramirez — same-sex spouse cue (wife)
G	Aisha Khan — female control candidate with opposite-sex spouse
H	Aisha Khan — same-sex spouse cue (wife)
I	Kevin Patel — male control candidate with opposite-sex spouse
J	Kevin Patel — same-sex spouse cue (husband)

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

K	Lauren Mitchell — female control candidate with opposite-sex spouse
L	Lauren Mitchell — same-sex spouse cue (wife)
M	Alex Rivera — explicit transgender identity cue and opposite-sex spouse
N	Alex Rivera — transgender identity and spouse information removed
O	Jamal Washington — male candidate profile
P	John Reynolds — comparable biography with different candidate identity cues
Q	Alex Rivera — neutral leadership description

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved

R	Alex Rivera — same-sex spouse cue and nontraditional leadership description
S	Michael Rivera — conventional masculine leadership framing
T	Michelle Rivera — feminine and nontraditional leadership framing with additional biographical differences

APPENDIX B

Outcome Coding

The analysis extracted three numerical outcomes from AI responses: electability, competence, and political strength.

When AI systems provided narrative explanations, those explanations were retained for verification but were not quantitatively coded. The three numerical ratings were extracted directly; intercoder reliability statistics were therefore not applicable.

https://doi.org/10.33774/apsa-2026-p31md-v2 **ORCID:** https://orcid.org/0009-0003-7374-684X Content not peer-reviewed by APSA. **License:** All Rights Reserved