October 2026

AI Midterms Part 2: What the AI Answers Say and Cite

Top Things Worth Knowing So Far

Prepared by Dewey Square Group for PSG Consulting and Innovating for the Public Good
Tim Chambers, tchambers@deweydigital.com

Executive Summary

As part of our ongoing AI Midterms project (read the previous reports here and here), this next study evaluates the answers and citations to common voter questions from the top seven AI platforms across critical 2026 House, Senate, and gubernatorial campaigns.

This report is based on data collection thus far analyzing data from 151,872 answers and 1,571,181 citations.

63 Battleground Campaigns Examined

Our sample encompasses every candidate competing across 63 battleground races classified as "Toss-up" or "Lean" by The Cook Political Report and the UVA Center for Politics Sabato's Crystal Ball as of July 2026. Across these contests, we evaluated responses to common voter inquiries across seven leading AI platforms: ChatGPT, Claude, Copilot, Gemini, Google Search AI Overviews, Google Search AI Mode, and Perplexity.

Queries and Their Sources

To reflect authentic voter questions, our prompts were pulled directly from top Google autocomplete searches we found and tested using their exact phrasing. Questions covered projected winners, policy platforms, candidate comparisons, candidate backgrounds, controversies, and voting rules.

After the midterm primaries, we added prompts reflecting top attack narratives aimed at the Democratic candidates in 10 tipping-point races (identified via political ad library scans). In September, we began asking Spanish-language versions of four of the questions in Arizona, Nevada and Texas.

We will deploy two remaining prompt sets: (1) election integrity questions across all races in the final week (addressing vote security, counting accuracy, and polling-site

problems), and (2) post-election prompts tracking winners, asking about legal disputes, and asking questions about the final results.

Top Findings in this Report

In our analysis of the voting related questions asked:

- In an October sample, 13.9% of answers to voting-logistics questions were outdated or inaccurate on a stated fact, and 5.9% could mislead a voter (miss a deadline, go on the wrong day, skip a required step). Across all questions, 5.1% of answers carried at least one clear factual error and 10.9% a clear error or misleading claim.

In our overall grading of answers we found:

- Which platform you ask matters more than which question. Claude and Copilot produced three quarters of the clear errors; on voting questions Claude was wrong or outdated in one answer in three, Copilot in one in four, ChatGPT in one in 10.

In our analysis of the citations the AI systems gave us:

- The same errors repeat day after day: withdrawn candidates kept on the ballot, the wrong person's biography, manufactured statistics from vote-tracker sites, and a senator's confirmation votes stated backwards.
- Wikipedia and Ballotpedia carry one citation in seven.
- Local and regional news outlets are 17% of all citations, national news outlets 10.8%. This shows that in our sample the AI platforms lean on the local paper and the state newsroom more than on the national press.

We detail all the top categories of citation sources in the section on sources below.

- Campaign websites are a small input to the answers: 4.6% of all citations, just under social media's 6.0%. In October the order flipped, with campaign sites cited about two and a half times as often as all social platforms combined, as the platforms pulled back sharply from social media, Instagram most of all.
- Low authority sites (betting odds, AI-written content farms, vote trackers) are about one citation in 10 and rose on five of six platforms from the summer to October (defined in the sources section).
- Attack-narrative questions change what gets cited. On those questions, reference sites like Wikipedia and Ballotpedia fall from 18% to 4% of citations, while local news rises to 31% and partisan outlets, social media and fact-checkers all gain share. Partisan outlets stay a minority of citations (8.3%), and the shift comes from what is asked, not from which platform answers.
- Spanish: AI platforms decline to answer far more often. 32% of Spanish voting answers were a refusal, an error, or no answer, against 14% for the same questions in English. Only 15% of the official-source citations in Spanish answers point to a Spanish-language page.

Stay-tuned: Post-Election AI Midterms report

We will publish a full detailed AI Midterm Report Part 3, after the election. We expect the final report to be based on roughly twice this much data and it will include all final conclusions and top overall findings across the whole collection.

Errors or Hallucinations Discovered

We evaluated chatbot prompted responses against an in-house, race-by-race ground-truth database built from primary sources that included: election boards, FEC and state filings, campaign sites, court records, and local press. The database tracks verified facts, citations, verification dates, and pitfalls like duplicate candidate names or withdrawn bids.

Graders reviewed each response against this standard, recording contradictory statements verbatim as either clear factual errors or misleading claims. Voting-logistics answers were scored using the rubric from ISD's June 2026 study, and any claim the record could not definitively resolve was cataloged as unverified rather than counted as an error.

The error rates in this section come from two October grading samples: every voting-logistics answer for October 5 to 18 (3,054 answers), and every answer to every question across 56 of the 63 races for October 12 to 17 (22,203 answers). A few later sections draw on an additional October 5 to 11 full-field grade; the November report will cover all data collected.

Table 1: Error rates in the two October grading samples

Types of Errors	Share of all answers in the sample
Voting-logistics questions: answers with an outdated or inaccurate stated fact	**13.9%**
Same, counting incomplete answers too	**29.0%**
Same, but could mislead a voter (miss a deadline, wrong day, skip a required step)	**5.9%**
Errors from all questions, at least one clear factual error (a wrong candidate, a made-up event, poll or number, a wrong date or rule), grade of 56 races, Oct 12 to 17, 22,203 answers	**5.1%**
All questions, at least one clear factual error or misleading claim (a stale fact, a hedged allegation leaned on, betting odds presented as polls), same grade	**10.9%**

Each row is a share of every answer in that sample, not of the answers with errors. Voting-logistics rows: 3,054 answers, October 5 to 18. All-questions rows: 22,203 answers across 56 races, October 12 to 17.

On the voting questions, the share of answers with an outdated or wrong fact ran between 9% and 19% on every one of the 14 days in this sample, and the share of AI answers that could mislead a voter between 3% and 9%.

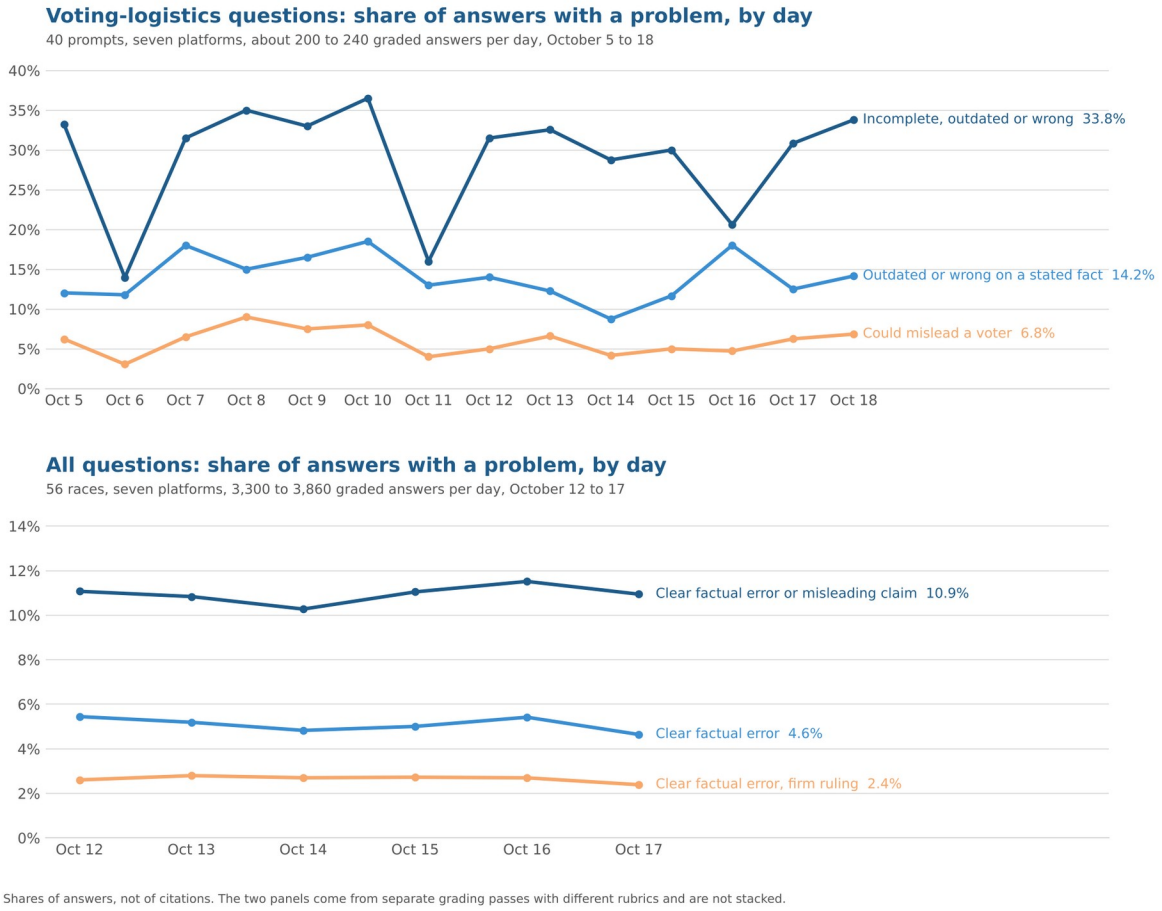


Figure 1: *Share of answers with a problem, by run day. Top: voting-logistics questions, 40 prompts, October 5 to 18. Bottom: all questions, 56 races, October 12 to 17. Separate grading passes, not stacked.*

Different Platforms, Different Answers to the same Questions

On the voting questions, Claude gave an outdated or wrong fact in one answer in three and Copilot in one in four. ChatGPT was at one in 10, but 44 of its 57 errors were the same thing that related to an outdated web address for Texas's voter lookup page that has since been replaced by the state's My Voter Portal.

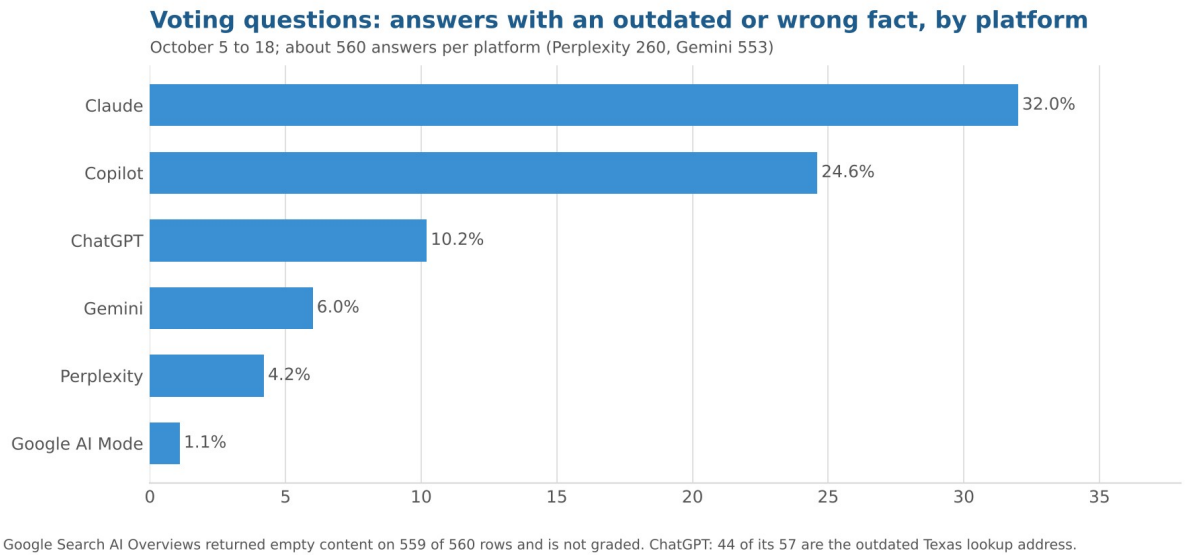


Figure 2: *Voting-logistics answers with an outdated or wrong fact, by platform, October 5 to 18.*

The same two platforms lead on every other question, too. Of the 1,127 answers that contained at least one hard error in the six-day full-field grade, 496 came from Copilot and 330 from Claude; the other five platforms produced 301 between them.

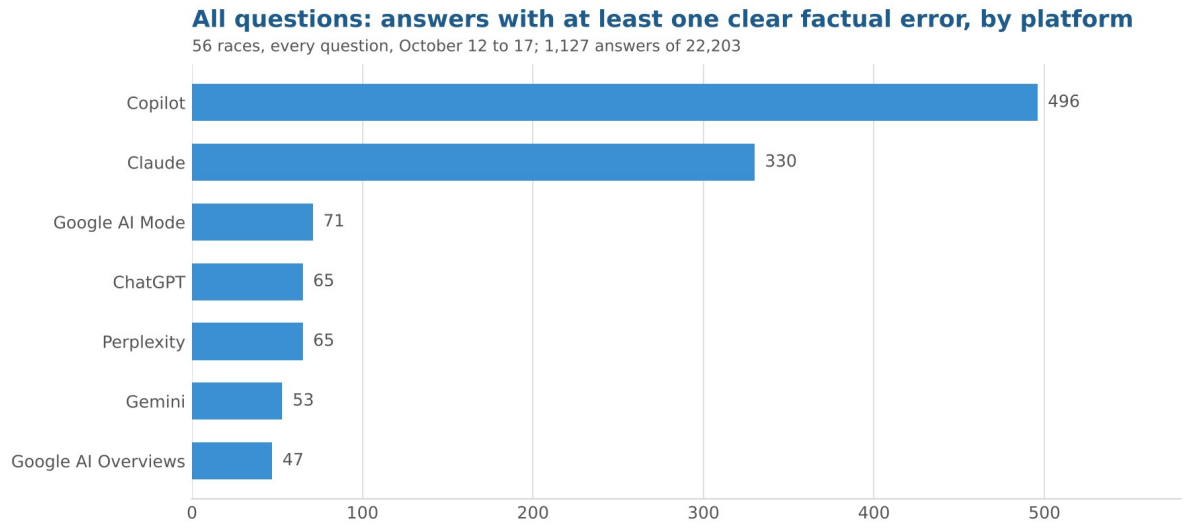


Figure 3: *Answers with at least one clear factual error, by platform, 56 races, October 12 to 17.*

These errors can be grouped into five distinct categories that recur on a daily basis

- Incorrect voting facts. A deadline that is still open is identified as closed; Michigan Election Day registration placed at the polling place instead of the clerk's office; New Hampshire's repealed no-ID affidavit offered as a fallback.
- Candidates who are not running. Withdrawn or defeated candidates listed on the ballot or in the polls.

- The wrong person. A candidate given the biography of someone else with the same name.
- Made-up statistics. Bill counts, attendance rates and letter grades from vote-tracker sites stated as fact.
- Wrong district numbers. A district's past presidential margin given for the wrong party or by the wrong amount.

Other figures and errors that stand out

- Voting facts. 179 of 3,054 voting-logistics answers (5.9%) could cause a voter to miss a deadline, go to the wrong place or skip a required step.
- Two AI platforms produced the highest number of errors. Copilot and Claude produced 77% of the individual hard-error lines in the Oct 12 to 17 grade (632 and 376 of 1,316 — one answer can contain several) and about 81% in the Oct 5 to 11 weekly grade (229 and 167 of about 489). On the voting questions, Claude gave an answer that could mislead a voter in 16.4% of cases and Google AI Mode in 0.2%. The same errors repeat for weeks.

Detailed examples of these errors are in the appendix.

AI Election Answers Sources

We examined some early findings on the AI systems' sources for their citations and how these changed with the use of low authority sources increasing over time.

How we classify sources

Each citation is sorted by the kind of site it points to: official and election authority sites, elected officials' .gov sites, campaign and party sites, advocacy groups, reference sites (Wikipedia, Ballotpedia), local and national news, partisan outlets, social platforms, aggregators, and low authority sites.

We define **low authority sites** as ones the project cannot vouch for as a source of fact: betting and forecast odds sites, AI-written content farms, vote tracker sites, spoof and mirror sites, low-quality aggregators, and anonymous partisan blogs.

We define **low factual rating sites** as a narrower group flagged by published rating lists: Media Bias/Fact Check, Wikipedia's list of deprecated sources, the Global Disinformation Index, sites publicly rated by NewsGuard below 60 or red, and websites tracked in the CRED-1 disinformation research collection. A site can be in either group, both, or neither.

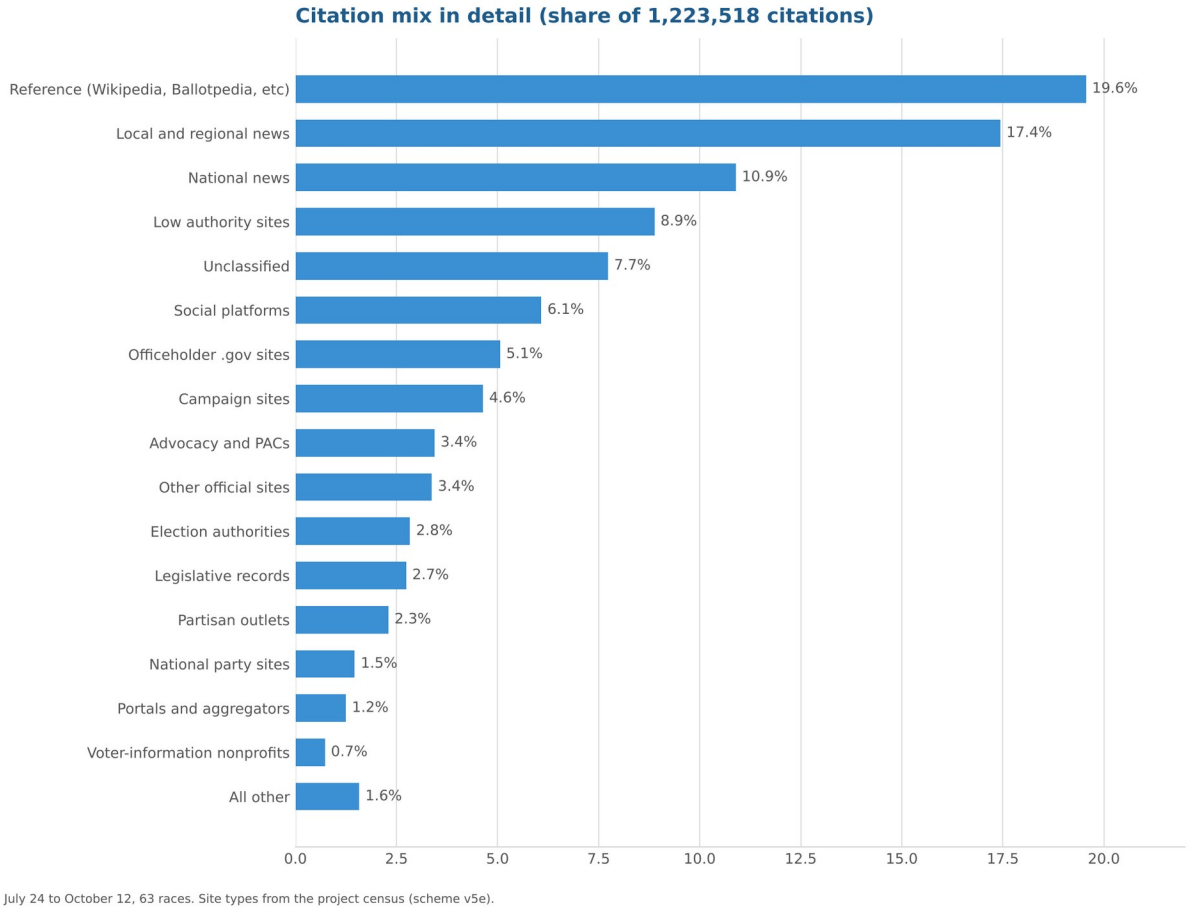
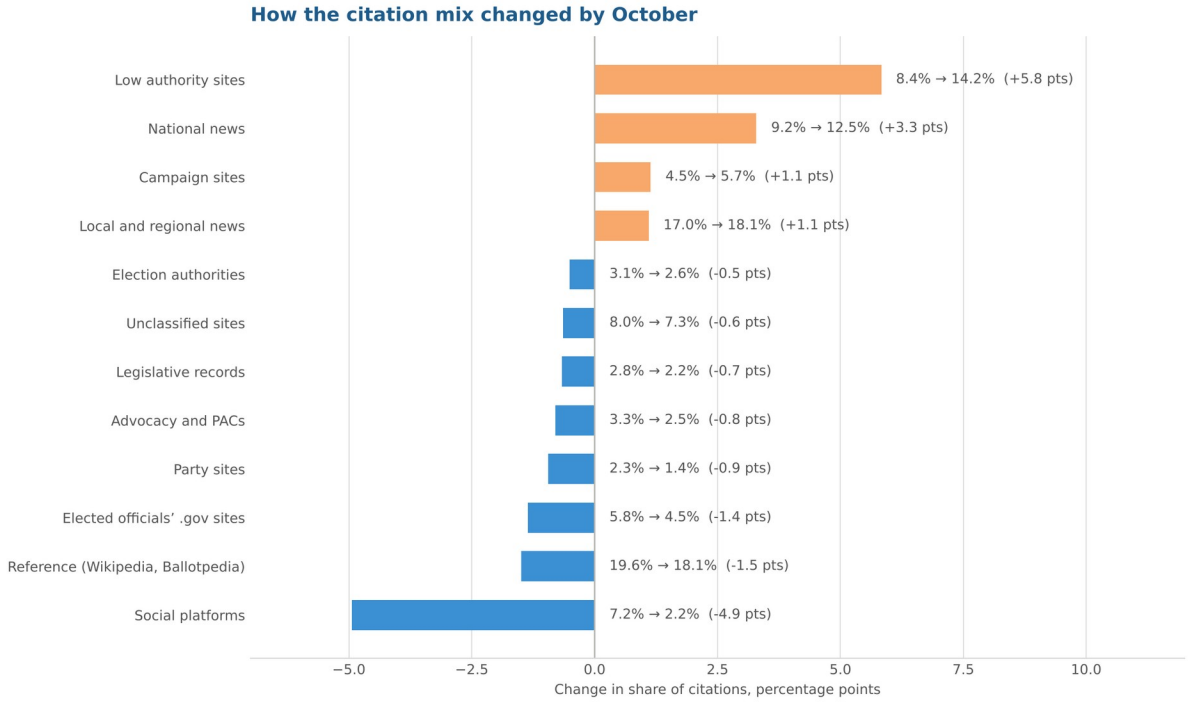


Figure 4: *Citation mix by site type, of 1,223,518 citations, July 24 to October 12.*

Note how the citation mix changes dramatically in October.



July 24 to August 27 (69,934 answers, 879,286 citations) against October 5 to 18 (29,549 answers, 217,642 citations). Same five standard questions, same races, six platforms (Perplexity excluded for its October outage). Orange went up, blue went down; labels give the share in each period. Partisan outlets (2.3% to 2.2%) and the rest moved by less than half a point.

Figure 5: *How the mix changed from July and August to October, on the same five questions, races and six of the seven platforms. Low authority sites and national news gained the most; social platforms lost the most.*

Notable findings from this data:

Two reference sites, Wikipedia and Ballotpedia, carry one in seven citations.

English Wikipedia (84,969 citations) and Ballotpedia (85,957) together are 14.0% of the citations in that window. Reference sites as a group are 19.5%. Of the Wikipedia pages cited, 34% are about a candidate and 42% about a race; Ballotpedia splits 25% candidate pages, 56% election pages.

Local news outranks national news

Regional news is 17.0% of citations (208,362), national news 10.8%. The most cited local outlets are City & State New York (6,129), The Nevada Independent (4,922), the Iowa Capital Dispatch (4,898), City & State Pennsylvania (4,499) and Iowa Public Radio (4,486). For a campaign, the local paper and the state nonprofit newsroom are weighted higher than the national press in the AI answers.

Campaign websites overall run just under social media as citation source, this changes in October when they pass it.

Across all data, social platforms are 6.0% of citations (73,096) and candidate campaign sites 4.6% (56,743). Facebook (37,622), YouTube (18,933) and Instagram (11,650) are 93% of the social total. The order reversed in October, when campaign sites were cited

about two and a half times as often as all social platforms combined: ChatGPT's social share fell from 7.6% in July and August to 0.3%, Google AI Mode's from 19.3% to 6.1%, and Instagram links across six platforms went from 10,754 to 75.

Campaign sites matter more than their share suggests. In our Part 1 report, a test of 1,640 graded answers to candidate biography questions found that an answer citing the candidate's own campaign site was about 40 percent less likely to contain a biographical error than other answers about the same candidate (odds ratio 0.57, p = 0.005). When an AI platform reaches the campaign's own material, the answer it gives is measurably better.

The breakdown of social media platforms in current citations collected:

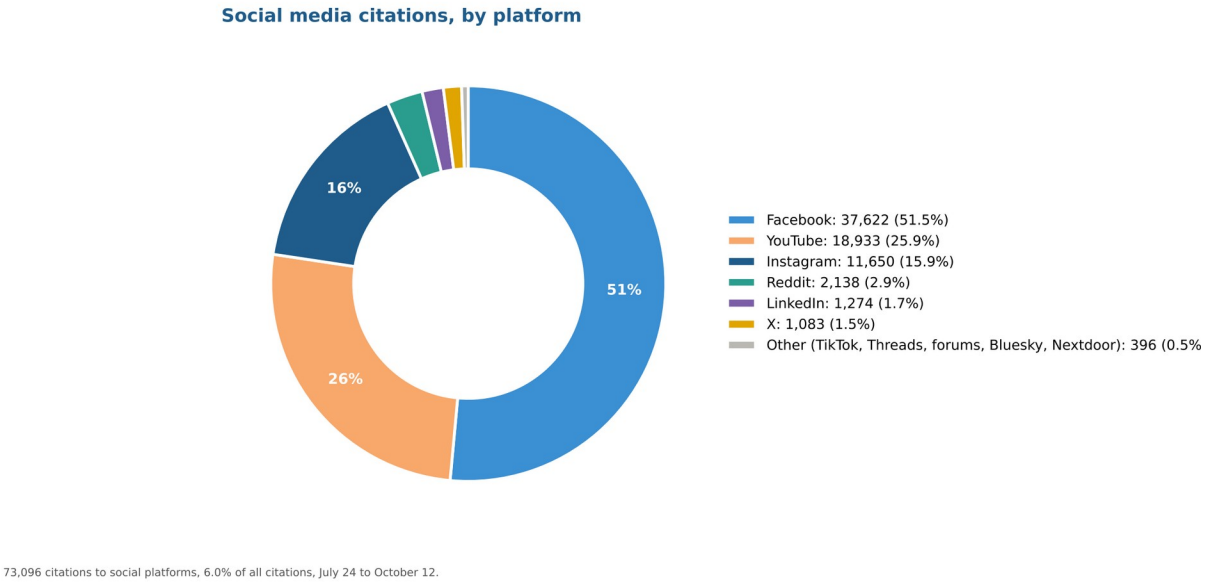


Figure 6: *73,096 social media citations by platform.*

Low authority sites are about one citation in 10, and rising.

Sites the project cannot vouch for as credible, (betting and forecast odds sites, AI-written content farms, vote tracker sites) are 9.2% of all citations.

The three largest kinds are betting and forecast odds sites (28,477 citations), AI-written content farms (24,325) and vote tracker sites (19,741); together they are nearly two thirds of the 113,137 low authority citations, with spoof and mirror sites, low-quality aggregators and a long tail of smaller kinds making up the rest.

Between July and August and October the share rose on five of six platforms: Gemini 2.3% to 10.4%, Claude 8.7% to 14.4%, Copilot 36.9% to 47.3%. In the October 12 to 17 full-field data, the share on the canonical questions (the standard prompt set, excluding the attack-narrative and Spanish add-ons) was 16.1% — above the 9.2% July–October average, consistent with the rising trend. This is a troubling pattern.

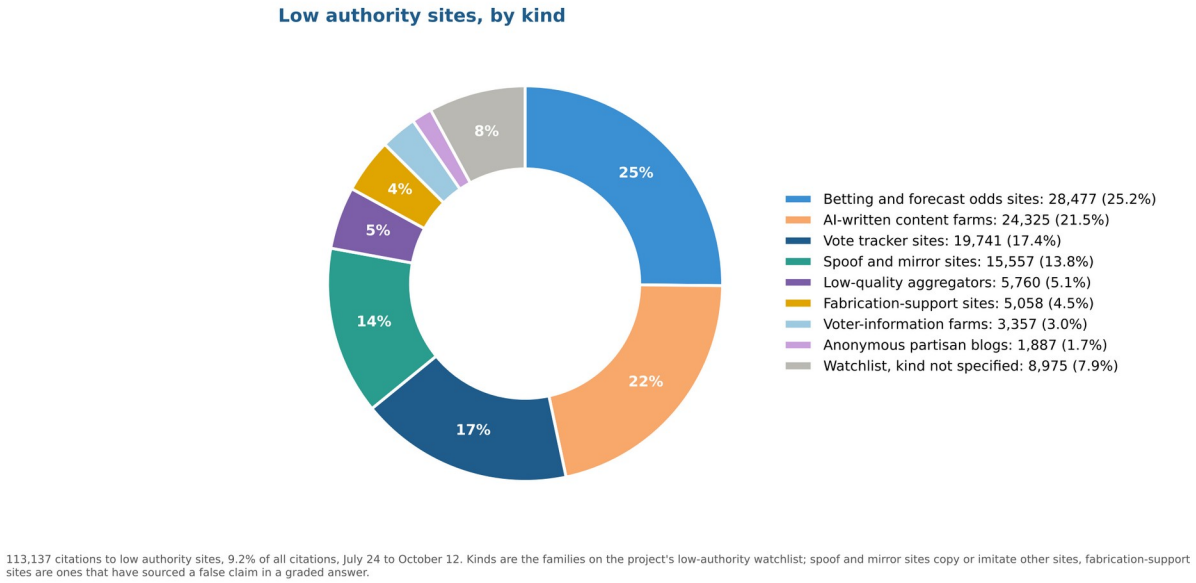


Figure 7: *113,137 citations to low authority sites by kind.*

Note that betting market sites being cited in answers to prompts such as those asking "who is winning" is particularly pernicious as betting market sites are not just unreliable, but also open to manipulation by investors or bettors. And we found that those results pipe directly into the AI answers at these rates.

Answers to attack narrative questions quadruple the use of low factual rating sites when compared to their use overall

We looked at websites flagged by Media Bias/Fact Check, Wikipedia's list of deprecated or blacklisted sources, GDI, NewsGuard and CRED-1, as defined in the box above. Collectively these websites are 1.74% of citations (21,237, 193 sites).

Four of these sites account for more than half of those citations: foxnews.com (6,589), nypost.com (2,176), washingtonexaminer.com (1,805) and breitbart.com (1,272). Next come aljazeera.com (884), realclearpolitics.com (862), michiganindependent.com (776) and townhall.com (739).

When you look only on citations to the questions built from a race's attack narrative, the share of low factual cites in citations is 8.11%, more than four times the overall share.

That 8.11% is higher than the share the same attack-narrative answers gave to several sources a voter would trust more: reference sites like Wikipedia and Ballotpedia (4.0%), the candidates' own campaign sites (4.0%), fact-checkers (3.4%), elected officials' .gov sites (0.2%) and election authorities (none). The full shift in sources on those questions is in the next section.

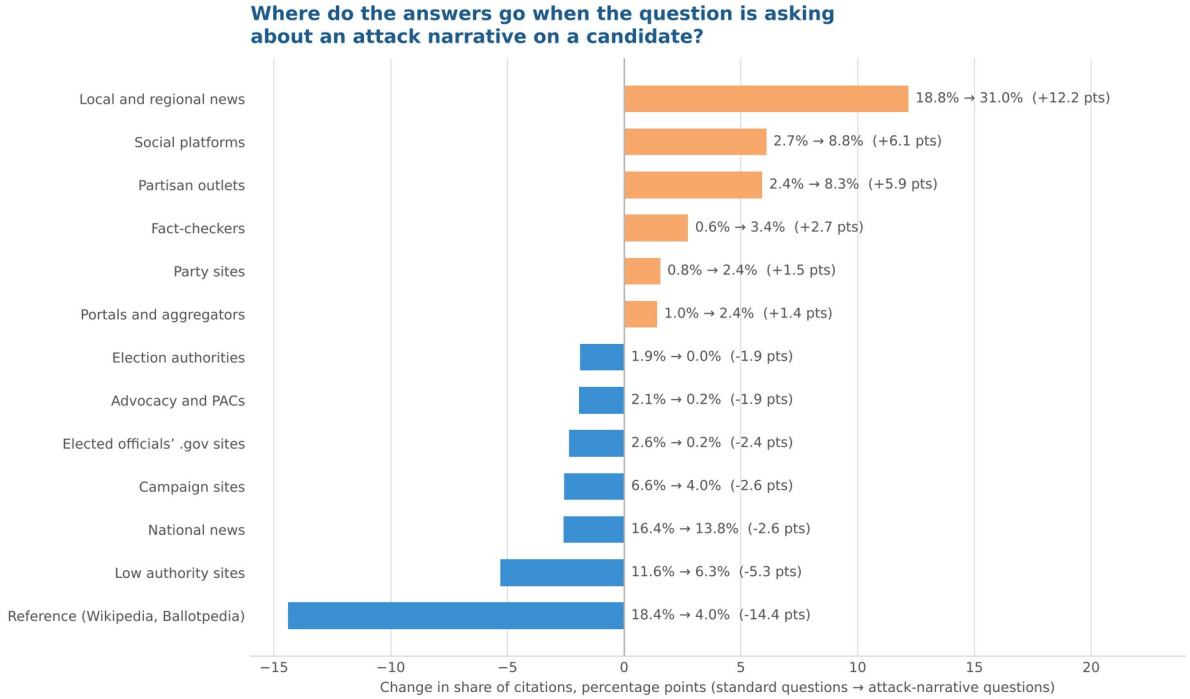
The mix of sources depends on the question asked; it is not a measure of platform lean.

How the mix of citations changes on attack-narrative questions

We evaluated one extra question in 10 races built from the dominant paid attack narrative against the Democratic candidate in that race.

The choice of attack was sourced from the main attack narrative in that race's paid advertising, drawn from scans of the Meta and Google ad libraries in late September, based on total spend of all ad based narrative attacks. These new attack narrative questions have run on all seven AI platforms daily since October 2.

The pattern is consistent: in the citations to attack narrative questions, reference sites like Wikipedia or Ballotpedia drop from 18% of citations to 4%, election authorities and elected officials' sites nearly disappear, and the answers lean instead on local news (31%), social media posts (8.8%), partisan outlets (8.3%) and fact-checkers (3.4%).



October 5 to 11, the same ten races. Attack-narrative questions: 420 answers, 3,403 citations. Standard questions in those races: 3,006 answers, 21,274 citations. Orange went up, blue went down; labels give the share on standard questions, then on attack-narrative questions. Unclassified sites 7.3% and 10.1% respectively, so the attack shares are indicative.

Figure 8: *Change in each site type's share of citations from the standard questions to the attack-narrative questions, same ten races, October 5 to 11. Labels give the share on each.*

Maine's five allegation questions — a separate set specific to the Maine Senate race, beyond the one-per-race attack-narrative questions — go further, with partisan outlets at 22% of citations.

English and Spanish

In September, we added Spanish versions of the biography, polls, controversies and voting questions in three races with large Spanish-speaking electorates (Texas Senate,

Arizona governor, Nevada governor); a Spanish attack-narrative question; and three generic Spanish voting questions. We asked these questions daily on all seven platforms beside their English counterparts. The comparison below uses the October data.

We are still collecting more data on this, but early signs show:

AI platforms often decline to answer in Spanish

32% of Spanish voting answers were a refusal, an error or no answer of their own, against 14% in English. Almost all our findings on refusal come down to two platforms: Gemini declined 94% of its Spanish voting answers (49% in English) and Google AI Mode 63% (26%).

Perplexity's refusal to answer went from 2% for English prompts but to 23% of the same content in a question in Spanish, and ChatGPT from near zero refusals in English to 10% refusals of the same question when posed in Spanish.

Claude answered every time in both languages, and Copilot all but once.

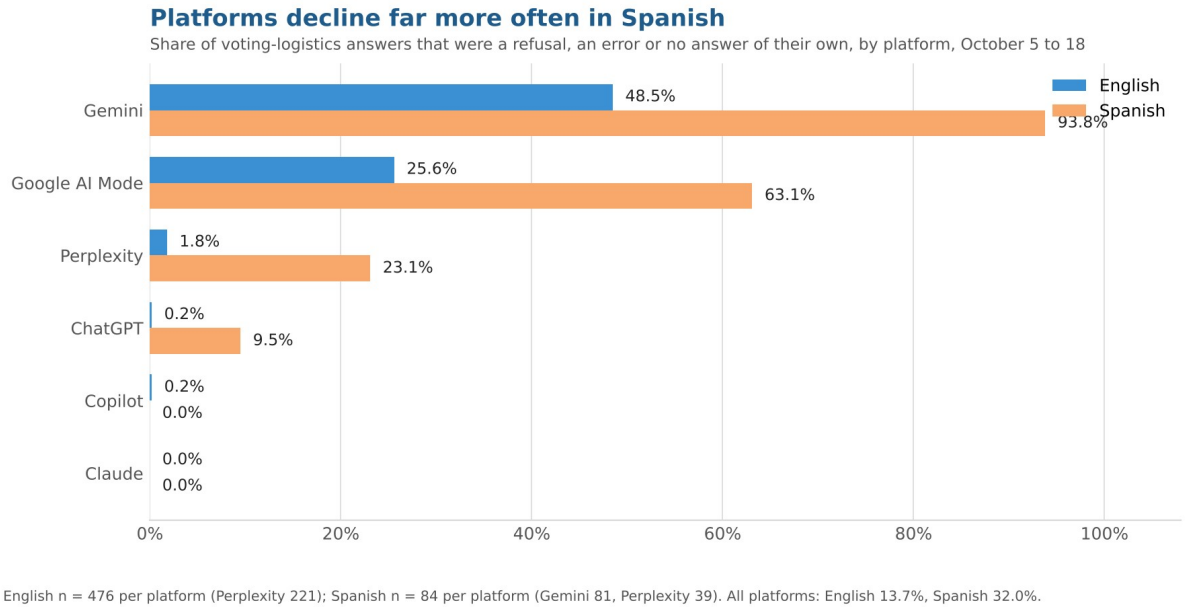
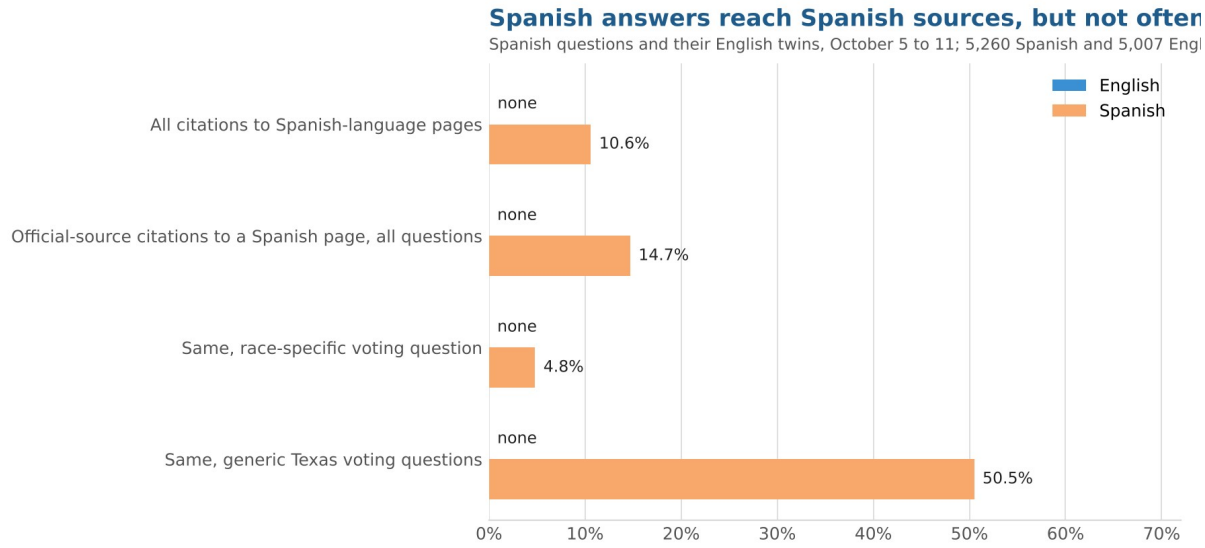


Figure 9: *Share of voting-logistics answers that were a refusal, an error or no answer of their own, by platform, October 5 to 18.*

Spanish questions to AI systems are often sent to English Language pages.

10.6% of the Spanish answers' citations went to Spanish-language pages; the English twins cited only one. About 42% of those Spanish pages are Spanish Wikipedia.

Official sources are critical for voters, yet only 15% of official citations offered by the AI systems led to Spanish-language pages. Link rates ranged from 50% on generic Texas voting queries to just 5% on race-specific prompts about election dates and voting rules.



Spanish pages are detected from the web address only, so approximate; about 42% of them are Spanish Wikipedia. Official-source rows: 132 of 895 overall, 10 of 207 on the race voting question, 103 of 204 on the generic Texas questions. The English twins cited one Spanish page.

Figure 10: *Spanish questions and their English twins, October 5 to 11. Spanish pages detected from the web address, so approximate.*

In November, we will publish our findings with the full dataset of Spanish and English answers and citations, and with further analysis.

We will focus our analysis into the dynamic of why the AI systems are feeding English language citations back to Spanish language prompts, with one theory being that this shows the relative lack of the AI systems being trained on and discovering Spanish language political content at parity with English language content.

What's Next

Look in early November for our full report, including complete data collected through election day in its entirety, full methodology, updates and new findings to this interim report.

We are expecting that the final dataset to be collected will be over 300,000 answers and close to 3 million citations.

Future Work Needed

There is additional research that can be done to understand the dynamics of the major AI systems answers to political questions, and we look forward to being a partner, willing to share and co-analyze this dataset with others. This research will be critical to understanding and preparing for 2027-2028.