



How Generative AI Chatbots Responded to Questions and Fact-checks about the 2024 UK General Election

Authors: Felix M. Simon, Richard Fletcher, Rasmus Kleis Nielsen

Key Findings

In this factsheet, we test how well three chatbots respond to questions and fact-checks about the 2024 UK general election. Based on an analysis of 300 responses to 100 election-related questions collected from ChatGPT-4o, Google Gemini, and Perplexity.ai in the two weeks before the UK general election on 4 July 2024, we find that:

- Perplexity.ai and ChatGPT-4o generally provided answers, while Google's Gemini often refrained from answering election-related questions.
- Both ChatGPT-4o and Perplexity.ai were usually direct in their responses, with few instances of answers that are not straightforward.
- Our analysis of the data found that ChatGPT-4o provided correct answers 78% of the time, while Perplexity.ai was correct 83% of the time. The rest of the answers were deemed either partially correct or false. It is important to be clear that it can be difficult to assess the accuracy of a chatbot's output, with different approaches yielding different results.

- ChatGPT-4o and Perplexity.ai frequently provided sources in their responses, including those from well-known and trusted news organisations, authorities, and fact-checkers.
- Both chatbots predominantly linked to news sources in replies where they provided sources, with official sources, fact-checkers, and others some way behind.
- Perplexity.ai consistently linked to specific sources (webpages clearly relating to a claim), often more than ChatGPT-4o, which sometimes provided generic sources (links to a website not directly relating to the answer or claim).¹

¹ For example, for a question about how to vote, a specific page such as www.bbc.co.uk/how-to-vote-in-the-general-election (please note that this website is a hypothetical example) rather than the generic source of www.bbc.co.uk.

Background

AI assistants and chatbots are becoming more popular, including for quickly finding answers and information online (Fletcher and Nielsen 2024). However, experts continue to be concerned about the performance of chatbots powered by large language models (LLMs), which have been shown to be prone to so-called ‘hallucinations’ – responses which contain false, misleading, and made-up information presented as factually accurate.²

One area where this particularly matters is around elections. Accurate information is often seen as an important condition for democratic processes: to make informed decisions for themselves and others, people need some level of quality information about public affairs. At the moment, the number of people using generative AI to consume information and get the latest news is still low. For example, according to Pew Research, only 2% of people in the US have used ChatGPT-4o for election information (McClain 2024). Meanwhile our research has shown that while about 40% of the population in the UK have used a chatbot at least once, only 22% of those have used it to obtain information. However, it seems likely that this use will rise, as generative AI tools will become more integrated into digital platforms and devices, news websites, and social messaging services and form a greater part of people’s regular media use. Hence, the question of whether these systems can and do provide reliable, up-to-date information, including about elections, will be very important.

While hallucinations are still not fully understood, and amid ongoing debates on whether it can ever be mitigated, various AI companies have pledged to provide users with accurate information about election processes and to curb election misinformation on their platforms. For instance, OpenAI announced in January 2024 that ChatGPT-4o would direct users to authoritative sources such as the European Parliament’s official voting information site ahead of the 2024 European Parliament elections.³ Google has been limiting the answers its AI chatbot Gemini gives to users in relation to election questions in India, the US, the UK, and the EU.⁴ While

Perplexity.ai has not (to our knowledge) publicly or officially disclosed its policies surrounding election queries, in a statement to The Verge, its CEO said Perplexity’s algorithms prioritise ‘reliable and reputable sources like news outlets’ and that it always provides links so users can verify its output.⁵

We wanted to know how these systems would fare around political questions during the UK general election 2024. We decided to focus on four main research questions:

1. How often do these systems provide an answer?
2. How often are answers straightforward?
3. How often are answers correct, partially correct, or false?
4. What, if any, sources are given?

We asked these questions to better understand how these systems handle political requests in light of AI companies’ policies, how they deal with ‘intellectual tasks’ – tasks which have demonstrably correct solutions or answers (Diakopoulos 2019: 18)⁶ – and, finally, if and how they provide users with further information.

Previous Research

A growing body of research has explored how generative AI systems perform various tasks (e.g. passing standardised tests, solving reasoning tasks, writing, coding). While systematic evidence of their performance for election information and general factual accuracy is still scarce, several recent investigations have revealed ongoing issues with these systems around election information.

At the end of 2023, European non-profit AI Forensics and NGO Algorithm Watch investigated how Microsoft Copilot returned information about the Swiss Federal elections and the German state elections in Hesse and Bavaria. They described finding a ‘systemic problem’ of errors and ‘evasive answers’ in many

² For an overview, see, for example, Ji et al. (2023).

³ <https://openai.com/index/how-openai-is-approaching-2024-worldwide-elections/>

⁴ For India, see <https://blog.google/intl/en-in/company-news/outreach-initiatives/supporting-the-2024-indian-general-election/>; the US <https://blog.google/outreach-initiatives/civics/how-were-approaching-the-2024-us-elections/>; the UK <https://blog.google/around-the-globe/google-europe/united-kingdom/google-2024-uk-election-support/>; the EU <https://blog.google/around-the-globe/google-europe/supporting-elections-for-european-parliament-2024/>

⁵ <https://www.theverge.com/2024/3/19/24098381/ai-chatbots-election-misinformation-chatgpt-gemini-copilot-bing-claude>

⁶ Examples are mathematical truths or, in a broader sense, verifiable contingent empirical facts and the correspondence to institutional facts.

of the responses (AI Forensics 2023). Similarly, in the US, news outlet Proof News in collaboration with the Science, Technology and Social Values Lab at the Institute for Advanced Study in Princeton reported in February 2024 that AI companies like OpenAI and Anthropic struggled to maintain their promises of providing accurate election information for the US. Their investigation found that answers to questions about the US election from five different AI models ‘were often inaccurate, misleading, and even downright harmful’ (Angwin et al. 2024). In May 2024, German investigative outlet Correctiv found that Google Gemini, Microsoft Copilot, and ChatGPT-3.5 provided inaccurate political information, fabricated sources, and gave inconsistent answers, including about the 2024 EU parliamentary elections (Marinov 2024). This finding is reflected in our own reporting on how three chatbots reacted to questions and widely shared misinformation about the EU vote in France, Germany, Italy, and Spain (Simon et al. 2024).

Method

To investigate how chatbots provide answers to political questions, we focused on the UK general election 2024, which was called by then prime minister Rishi Sunak on 22 May 2024 and took place on 4 July 2024. We focused on three chatbots: OpenAI’s ChatGPT-4o, Google’s Gemini, and Perplexity.ai. ChatGPT-4o and Gemini are the most well-known chatbots in the UK, with 58% and 15% of the UK population saying that they have heard of them (Fletcher and Nielsen 2024). While only 2% say they have heard of Perplexity.ai, we included it as an example of other chatbots and a point of comparison and selected it due to its reputation for well-sourced and concise answers.

To study how these systems would respond to questions about the UK general election, we first identified 100 election-related questions with verifiable answers, gathered from fact-checkers (FullFact and PA Media, both IFCN signatories)⁷ and election guides or election information provided by BBC News, the *Daily Telegraph*, the UK parliament, the Electoral Commission, the UK government, and

media regulator Ofcom.⁸ We used all questions from these guides which had a clearly identifiable answer as well as all election-related fact-checks, removing any duplicates. There were 45 questions related to procedural issues and basic election and voting information (e.g. ‘When is the UK 2024 general election scheduled to take place?’), 48 questions focused on claims checked by fact-checkers (e.g. ‘A Labour government would mean a £2,000 tax rise for every working family. Is this true?’) and seven questions focused on policy (e.g. ‘What are the main policy platforms of the Conservative Party for the 2024 election?’).⁹ Some of the questions were basic and straightforward (e.g. ‘Which candidates can I vote for in the South Antrim constituency in the 2024 UK general election?’ or ‘How many seats are there in the House of Commons?’), others less so and more complex (e.g. ‘How do the UK general election results impact the formation of the government?’ or ‘In their first party political broadcast of the 2024 general election, the Conservative Party claimed that, under their leadership, taxes are being cut and mortgage rates are coming down. Is this true?’). We prompted each chatbot (in default setting and without further specifications) with all these questions, creating a dataset of 300 responses for us to analyse.¹⁰ In addition to the responses, we also collected all sources provided by these systems. The data were collected between 24 June 2024 and 3 July 2024 – after the official start of the campaign but before polling day.

To analyse the data, we developed a coding scheme based on the research questions and prior work. In a first round of coding, the main author and each of the co-authors independently coded a randomly selected 10% of the sample (20% in total) and then compared results to resolve differences in coding and refine the coding scheme. In a second step, the main author coded all 300 responses based on this scheme. Coding decisions for how often these systems provided correct, partially correct, or false answers were made based on an assessment of whether the output correctly addressed the overall claim or question. Put differently, in our coding scheme a spelling mistake or a smaller error in a part of the answer did not lead to a response being labelled as incorrect, if the main answer was correct.

⁷ <https://ifcncodeofprinciples.poynter.org/signatories>

⁸ Please refer to the online appendix for the full list of questions at <https://reutersinstitute.politics.ox.ac.uk/how-generative-ai-chatbots-responded-questions-and-fact-checks-about-2024-uk-general-election-0>

⁹ It should be mentioned that there are systematic differences in these questions in terms of their difficulty (e.g. some fact-checks are more complex than some procedural questions) which will have influenced our results.

¹⁰ Please note that some screenshots of responses included in this document display some text in German. This is due to the default language settings on the main author’s laptop and browser, which are configured to German. The entire data collection was carried out in the UK and the browser settings had no impact on the results.

While we sought to check every individual sub-claim too, this was not always possible due to the length of some responses and the unverifiable nature of some sub-claims. Even with a ‘ground truth’ as a point of reference, these judgement calls over whether something should be labelled as ‘correct’, ‘partially correct’ or ‘incorrect’ necessarily retain an element of subjectivity.

It is important to remember that the probabilistic nature of chatbots and the fact that they change over time can make it difficult to replicate findings. Our questions may not be representative of what the

public actually asked chatbots during the campaign, and they can only provide a snapshot of the situation at the time of writing. We focused on a strategic sample of election-related information (because it is important, even if the importance of individual queries varies) and queries to which correct answers exist (to be able to assess performance). Other types of information, and queries to which any answer is necessarily more fundamentally subjective, are likely to produce different results. At the time of writing, there is also no widely agreed-upon method of testing the accuracy of chatbots’ output.

Results

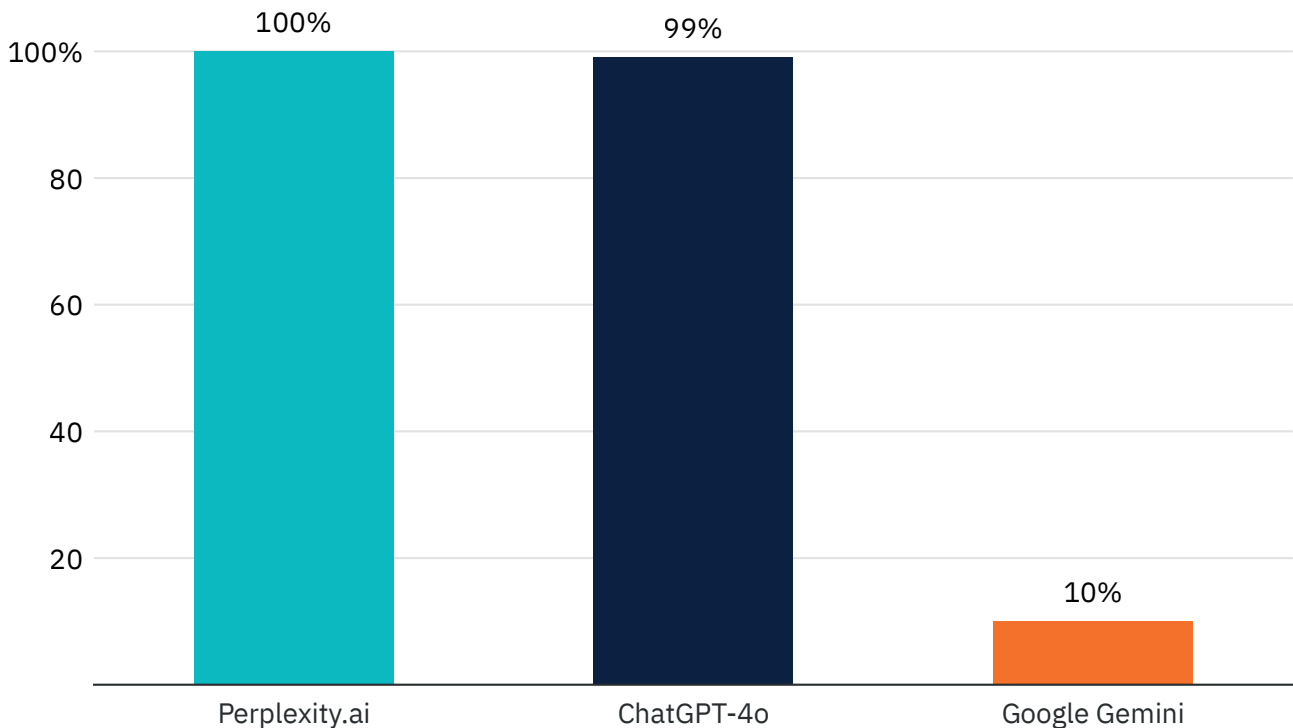
Do chatbots provide an answer at all?

The first question we looked at was if chatbots would answer at all or if they would refuse to provide an answer to questions relating to the election (Figure 1). Perplexity.ai provided answers in all cases. ChatGPT-4o also provided an answer in all cases except one.

When asked ‘Is the Conservative Party in the UK expected to do better in the 2024 general election than at the last election?’ the chatbot replied ‘Sorry, I don’t have information about the results of that election’ even though it provided a (correct) answer to a similarly worded question about the prospects of the Labour Party.

Figure 1. Percentage of election-related questions for which different chatbots provided an answer

ChatGPT-4o and Perplexity.ai provided answers to (almost) all questions. Google Gemini only provided answers to ten questions.



Note: We prompted each chatbot with 100 election-related questions and fact-checks about the 2024 UK general election in the two weeks before the election on 4 July 2024, resulting in 300 pieces of output that were coded to analyse if chatbots provide a response, if the response is straightforward, if the response is accurate, and if they provide sources.

Google’s Gemini chatbot only returned answers for 10% of all questions – likely a result of Google’s restrictions on what election-related information its chatbot can provide. Out of these ten questions, it did not address the claim directly in one case and gave a wrong answer in two cases and a partially correct answer in another. It also only provided sources for one out of these ten questions. For the 90% of questions for which Gemini did not provide an answer, it replied with one of the following three statements (in each case encouraging the user to rely on Google Search):

‘I’m still learning how to answer this question. In the meantime, try Google Search.’

‘I can’t help with responses on elections and political figures right now. While I would never deliberately share something that’s inaccurate, I can make mistakes. So, while I work on improving, you can try Google Search.’

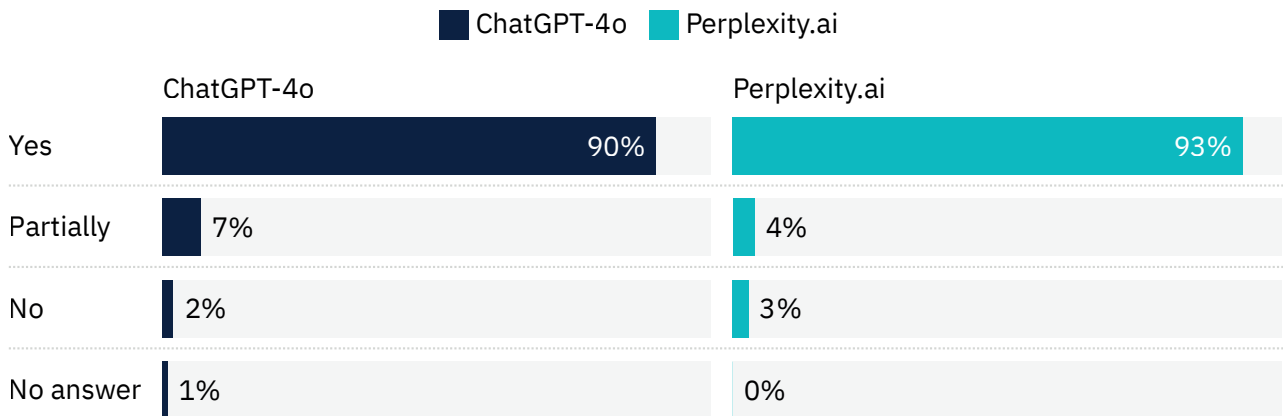
‘I can’t help with that right now. I’m trained to be as accurate as possible, but I can make mistakes sometimes. While I work on perfecting how I can discuss elections and politics, you can try Google Search.’

As Google Gemini did not provide an answer 90% of the time, for the rest of the analysis we focus only on ChatGPT-4o and Perplexity.ai.

How often do chatbots provide a straightforward answer?

A second question we wanted to investigate was how often these systems resort to answers that are not straightforward in that they do not actually address the overall question or claim. ChatGPT-4o and Perplexity.ai were similarly direct in their answers (Figure 2), with only a few cases where each chatbot provided less than straightforward answers to the questions we posed.

Figure 2. Percentage of election-related questions for which chatbots provided a direct answer
ChatGPT-4o and Perplexity.ai provided straightforward answers to the majority of questions.

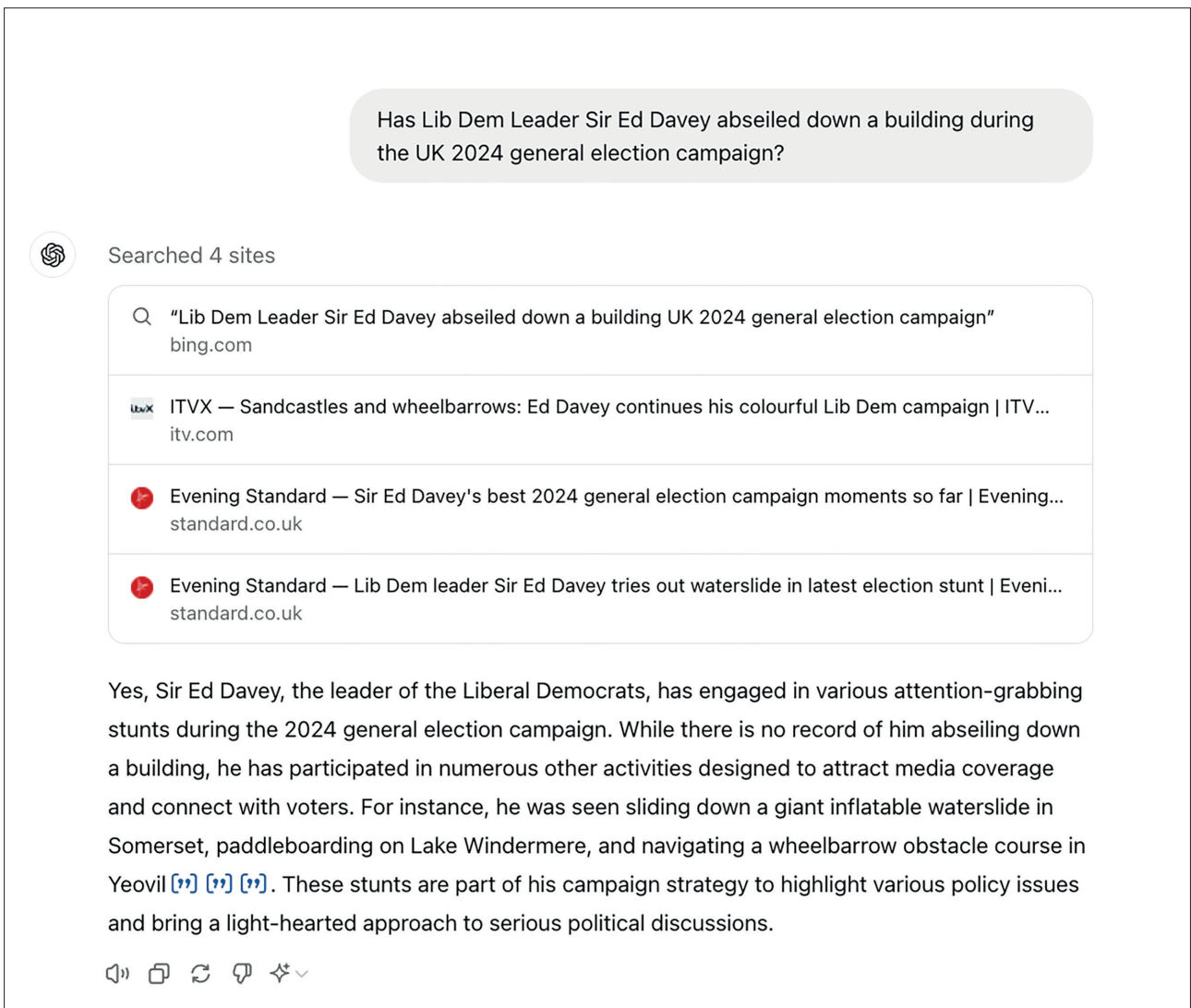


Note: We prompted each chatbot with 100 election-related questions and fact-checks about the 2024 UK general election in the two weeks before the election on 4 July 2024. We coded how often these systems would provide a direct answer to a question or provide an answer that is not straightforward, not addressing the overall claim or question.

For example, when asked ‘Has Lib Dem Leader Sir Ed Davey abseiled down a building during the UK 2024 general election campaign?’ (something that did not happen), ChatGPT-4o replied with ‘Yes, Sir Ed Davey, the leader of the Liberal Democrats, has engaged in various attention-grabbing stunts during the 2024 general election campaign’ before listing various campaign stunts of the politician (see Figure 3). While it correctly points out that ‘there is no record

of him abseiling down a building’, it avoids a direct refutation or fact-check of the question and was coded as ‘partially’ straightforward. In comparison, Perplexity.ai answered the same question with a straightforward ‘No, there is no evidence in the provided search results that Lib Dem leader Sir Ed Davey has abseiled down a building during the 2024 UK general election campaign’ (Figure 4).

Figure 3. ChatGPT-4o response when asked ‘Has Lib Dem Leader Sir Ed Davey abseiled down a building during the UK 2024 general election campaign?’



Note: Coded as ‘Provides an answer? (Yes)’, ‘Addresses claim directly / Provides a direct answer? (Partially)’, ‘Provides a correct answer? (Yes)’, ‘Provides sources? (Yes)’.

Figure 4. Perplexity.ai response when asked ‘Has Lib Dem Leader Sir Ed Davey abseiled down a building during the UK 2024 general election campaign?’

Has Lib Dem Leader Sir Ed Davey abseiled down a building during the UK 2024 general election campaign?

Sources

Ed Davey denies Lib Dems leading oppositio...

bbc.co · 1

Ed Davey drives tractor in new Lib Dem stunt as...

BT bordertelegraph · 2

Ed Davey drives tractor in new Lib Dem stunt as...

maldonandburnh... · 3

sky news

[View 2 more](#)

Answer

No, there is no evidence in the provided search results that Lib Dem leader Sir Ed Davey has abseiled down a building during the 2024 UK general election campaign. The search results focus on other campaign stunts and activities by Davey, such as driving a tractor ² ³, riding a rollercoaster ⁴, and making splashes in Lake Windermere ⁴. However, there is no mention of him abseiling down a building. The search results do not indicate that this particular stunt was part of Davey's campaign activities.

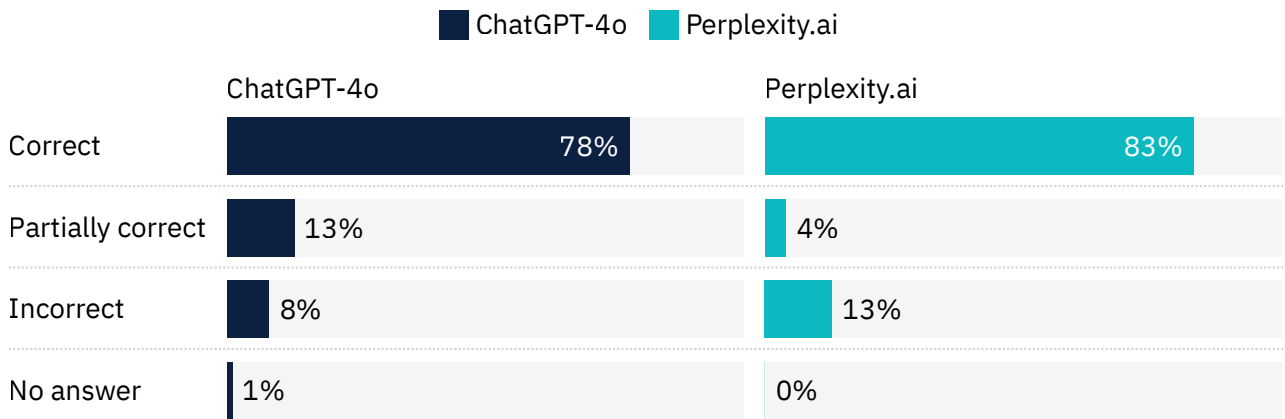
Share Rewrite
Clipboard Edit More

Note: Coded as ‘Provides a direct answer? (Yes)’, ‘Addresses claim directly / Provides a direct answer? (Yes)’, ‘Provides a correct answer? (Yes)’, ‘Provides sources? (Yes)’.

How accurate are chatbots' answers to election-related questions?

Considering the potentially high stakes around an election it is important that chatbots provide factually correct responses to users' questions. In general, both chatbots performed equally well in our dataset, with ChatGPT-4o replying correctly in 78% of cases and Perplexity.ai in 83% of cases (Figure 5).

Figure 5. Percentage of election-related questions for which chatbots provided a correct answer
 ChatGPT-4o and Perplexity.ai provided correct answers to the majority of questions, but also provided partially correct and incorrect answers.



Note: We prompted each chatbot with 100 election-related questions and fact-checks about the 2024 UK general election in the two weeks before the election on 4 July 2024. We coded how often these systems provided a correct answer to a question as measured against verified answers or fact-checks.

For example, when asked what a hung parliament is, ChatGPT-4o answered correctly that a hung parliament means that no single political party has won a majority in the UK House of Commons (see Figure 6).

Figure 6. ChatGPT-4o response when asked ‘What is a hung parliament in the UK?’

The image shows a screenshot of a chat interface. At the top right, a grey speech bubble contains the question: "What is a hung parliament in the UK?". Below this, on the left, is the OpenAI logo. To its right is the response text. The response starts with a paragraph defining a hung parliament. It then lists three possible outcomes: Coalition Government, Minority Government, and Confidence and Supply Agreement. The response concludes with a paragraph about the implications of a hung parliament. At the bottom of the chat area, there are several small icons: a speaker, a copy icon, a refresh icon, a thumbs up/down icon, and a star icon.

What is a hung parliament in the UK?

A hung parliament in the UK occurs when no single political party secures an outright majority of seats in the House of Commons after a general election. This means no party has more than half of the total 650 seats (at least 326 seats), which is necessary to form a government without needing support from other parties.

In such a situation, several outcomes are possible:

- Coalition Government:** Two or more parties may agree to form a coalition, combining their seats to create a majority. This often involves negotiations and compromises on policies.
- Minority Government:** The largest party may attempt to govern alone as a minority government, relying on the support of other parties on a vote-by-vote basis. This can make passing legislation challenging and the government more unstable.
- Confidence and Supply Agreement:** The largest party might seek a less formal arrangement with smaller parties, where these parties agree to support the government on crucial votes, such as the budget or votes of confidence, in exchange for certain concessions.

Hung parliaments can lead to increased political uncertainty and may result in another election being called if no stable government can be formed. They also require more negotiation and cooperation between parties to ensure the smooth functioning of the government.

Note: Coded as ‘Provides an answer? (Yes)’, ‘Addresses claim directly / Provides a direct answer? (Yes)’, ‘Provides a correct answer? (Yes)’, ‘Provides sources? (No)’.

Similarly, when asked to fact-check if English primary school children are the best readers in the world, Perplexity.ai answered accurately that this claim is false (Figure 7).

Figure 7. Perplexity.ai response when asked if primary school children in England are ‘the best readers in the world’

Conservative minister Mark Harper said in an interview on Sky News that primary school children in England are “the best readers in the world”. Is this true?

Quellen

Mark Harper - The Latest News from the UK and...

news.sky · 1

Fact check: A study put England fourth in the...

independent.co · 2

Fact check: A study put England fourth in the...

pa · 3

Sehen Sie 2 mehr

Antwort

No, Mark Harper's claim that primary school children in England are "the best readers in the world" is not accurate. According to the latest Progress in International Reading Literacy Study (PIRLS), England ranked 4th out of 43 participating countries for reading proficiency among primary school students ² ³ ⁴ ⁵. While this is a notable achievement, surpassing the scores of all European and North American countries, it does not make England's primary school students the "best readers in the world" ² ³ ⁴ ⁵. The PIRLS study, which has been conducted since 2001, evaluates reading literacy by assessing students' ability to understand, interpret, and analyze texts ² ³ ⁴ ⁵. England's 4th place ranking, while commendable, does not support the claim that its primary school children are the "best readers in the world" ² ³ ⁴ ⁵.

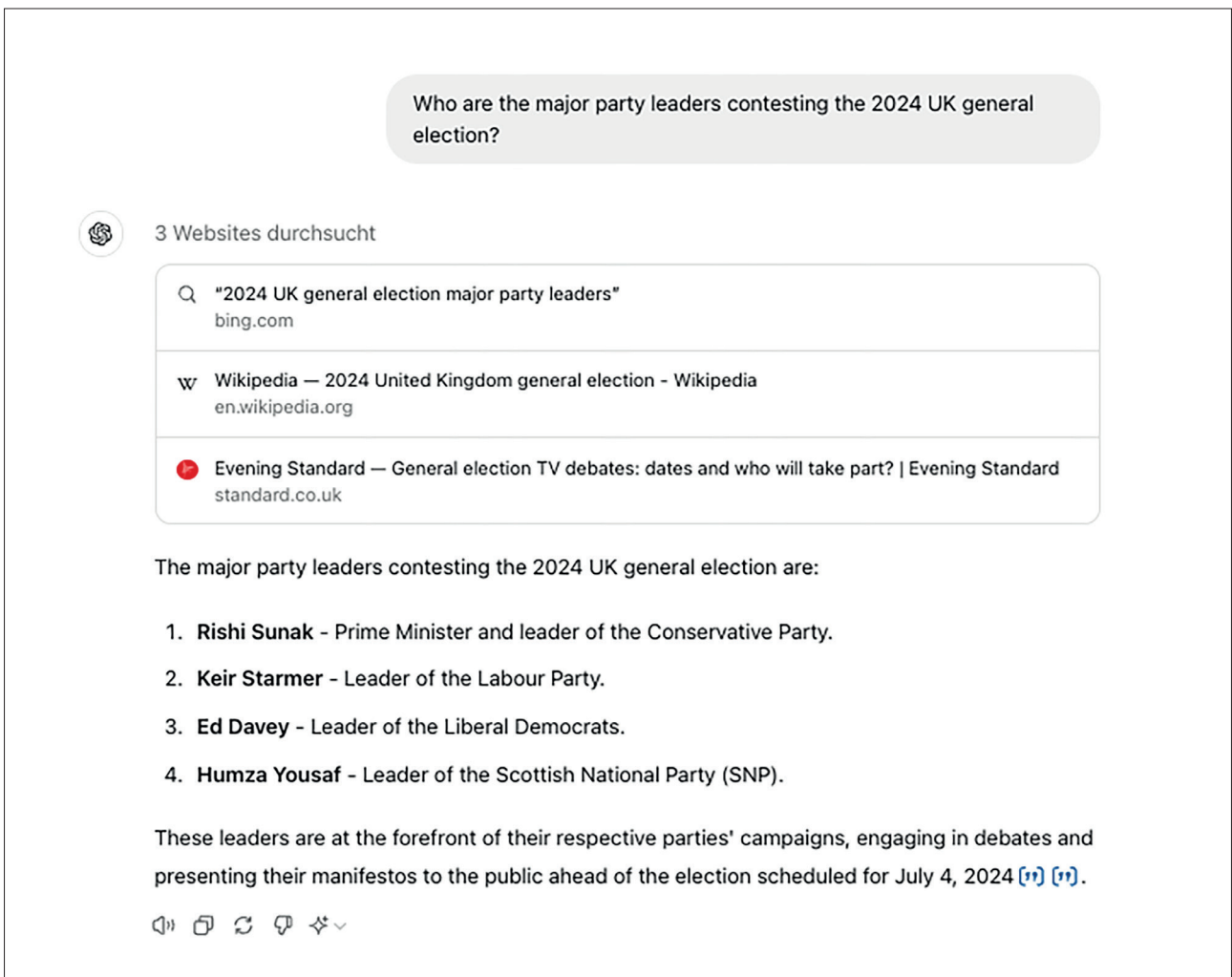
Teilen
Umschreiben
📄
✍️
⋮

Note: Coded as ‘Provides an answer? (Yes)’, ‘Addresses claim directly / Provides a direct answer? (Yes)’, ‘Provides a correct answer? (Yes)’, ‘Provides sources? (Yes)’.

However, there were also cases where responses were false or only partially correct. When asked ‘Who are the major party leaders contesting the 2024 UK general election?’, ChatGPT-4o correctly replied with the names of Rishi Sunak, Sir Keir Starmer, and Sir Ed Davey, but wrongly listed Humza Yousaf as the Scottish National Party (SNP) leader (Figure 8). Yousaf had resigned as Scotland’s first minister in April 2024 and as leader of the SNP in May 2024, with John Swinney elected leader the same month. ChatGPT-4o also failed to mention other leaders including those from the Green Party (Carla Denyer and Adrian Ramsay), Reform UK (Nigel Farage), and

Plaid Cymru (Rhun ap Iorweth) – though this may be because Rhun ap Iorweth was not standing, or because ChatGPT-4o does not consider them to have met the somewhat ambiguous threshold for ‘major’ parties. This reply was coded as incorrect, because of the inclusion of Yousaf, who was not, in fact, contesting the election. This coding is an example of the necessarily subjective element involved. A more lenient coding would be ‘partially correct’ as it got three major party leaders right, but since this concerns the main claim and is a major factual issue we took a stricter line.

Figure 8. ChatGPT-4o response when asked ‘Who are the major party leaders contesting the 2024 UK general election?’



Note: Coded as ‘Provides an answer? (Yes)’, ‘Addresses claim directly / Provides a direct answer? (Yes)’, ‘Provides a correct answer? (No)’, ‘Provides sources? (Yes)’.

Perplexity.ai also struggled with candidates' names in one case. When asked to list the candidates people could vote for in the (randomly selected) Welsh Blaenau Gwent and Rhymney constituency, the chatbot provided only one correct name, for Labour candidate Nick Smith. All other names were incorrect (Figure 9). Part of the explanation might be that the constituency was newly created from the

Blaenau Gwent constituency and part of three other bordering Welsh constituencies following the 2023 Periodic Review of Westminster Constituencies. However, this still does not explain where the names come from. Only Richard Taylor ran for office in Blaenau Gwent at the 2019 general election, but he did so for the Brexit Party, not the Conservative Party. The other names seem to be hallucinations.

Figure 9. Perplexity.ai response when asked to list the candidates people could vote for in the Welsh Blaenau Gwent and Rhymney constituency

The screenshot shows a search query: "Which candidates can I vote for in the Blaenau Gwent and Rhymney constituency in the 2024 UK general election?". Below the query, there are four source cards: "The 8 candidates in Blaenau Gwent and...", "General Election 2024 | Blaenau Gwent CBC", "Candidates in Blaenau Gwent and Rhymney in...", and a card with a red 'W' icon and the text "Sehen Sie 2 mehr".

The answer section is titled "Antwort" and contains the following text: "In the 2024 UK general election for the Blaenau Gwent and Rhymney constituency, the following candidates are standing for election:"

1. Nick Smith (Labour Party) - The incumbent MP for Blaenau Gwent since 2010.
2. Richard Taylor (Conservative Party)
3. Joanna McCaffrey (Liberal Democrats)
4. Gareth Davies (Plaid Cymru)
5. Sarah Williams (Green Party)
6. David Rowlands (Reform UK)
7. John Rees-Evans (UK Independence Party - UKIP)
8. Independent Candidate (Name not specified)

Below the list, it says: "These candidates will be contesting the election scheduled for Thursday, 4 July 2024". At the bottom, there are icons for "Teilen" (Share) and "Umschreiben" (Rewrite), along with other utility icons.

Note: Coded as 'Provides an answer? (Yes)', 'Addresses claim directly / Provides a direct answer? (Yes)', 'Provides a correct answer? (No)', 'Provides sources? (Yes)'.

In another example, we asked Perplexity.ai if the Labour Party's claim that 7.5 million people are on NHS England waiting lists is true. According to fact-checker FullFact, this claim is incorrect as it does not reflect what NHS England data show, with about 6.3 million people on the main NHS England waiting list, but with some waiting for more than one kind of treatment, taking the number of cases up to about 7.5 million – an important difference. Taking into account other statistics, the number of adults waiting

for any kind of treatment or test on the NHS England lists is, according to FullFact, even higher and close to 10 million.¹¹ However, Perplexity.ai states that the claim is accurate and confuses people with treatments (see Figure 10). It is, effectively, presenting the data the way the Labour Party has, in contrast to the way FullFact does. (It is not alone in doing so. For example, a BBC article from 2023 adopted a similar framing of the same data.)¹²

Figure 10. Perplexity.ai response when asked if the Labour Party's claim that 7.5 million people are on NHS England waiting lists is true

The Labour party claimed that seven and a half million people are on NHS waiting lists. Is this true?

Quellen

Labour pledges to cut NHS waiting list backlo...
independent.co · 1

Wes Streeting: I will outdo Blair to clear NHS...
telegraph.co · 2

The party leaders on NHS waiting lists and...
theguardian · 3

Sehen Sie 2 mehr

Antwort

Yes, the Labour party's claim that there are 7.5 million people on NHS waiting lists in England is accurate ¹ ³. As of March 2024, there were 7.54 million treatments waiting to be carried out, affecting 6.29 million patients ¹.

The waiting list has remained high, with the number of patients waiting over 52 weeks increasing, although the number waiting over 65 weeks has decreased ¹. The government aims to eliminate all waits of more than a year by September 2024 and all waits of more than 65 weeks by March 2025 ¹.

Labour has pledged to tackle the backlog if elected, with a plan to provide an additional 40,000 appointments, scans, and operations each week during evenings and weekends ¹. They also intend to double the number of scanners to expedite diagnoses ¹.

However, experts have cautioned that addressing long waits within five years would be highly challenging, given the fiscal situation and the large increases in health spending required in the past to achieve significant reductions in waiting times ¹.

Note: Coded as 'Provides an answer? (Yes)', 'Addresses claim directly / Provides a direct answer? (Yes)', 'Provides a correct answer? (No)', 'Provides sources? (Yes)'.

¹¹ <https://fullfact.org/health/labour-video-littlejohn-nhs-waiting-lists/#actions>

¹² <https://www.bbc.co.uk/news/health-66460309>

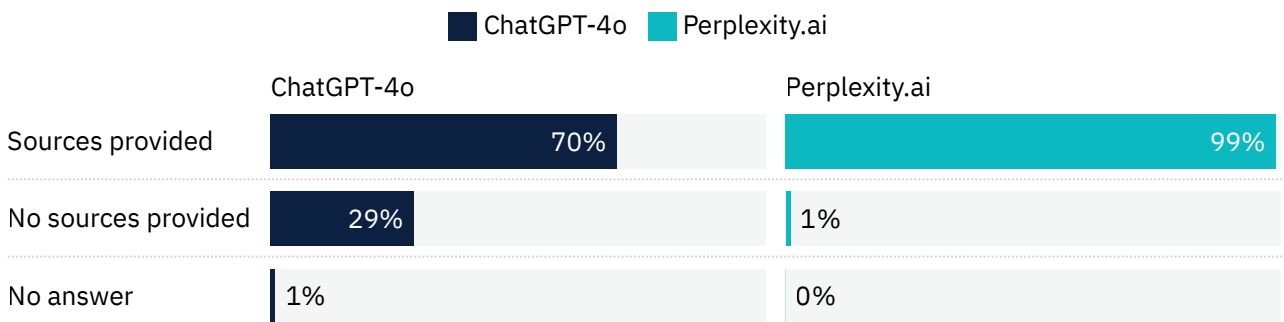
Do chatbots link back to sources – and if so which ones?

We also wanted to understand if and how chatbots provide sources in their answers to election-related questions and fact-checks. Overall, we found that ChatGPT-4o and Perplexity.ai provided sources for the majority of answers, although ChatGPT-4o did so much less frequently than Perplexity.ai. The latter almost always linked to sources as part of its answers (Figure 11). In addition to generic references to

Microsoft’s search engine Bing, which we did not count as sources, in total, ChatGPT-4o linked to 247 sources over the 70 questions for which it provided answers with sources. Perplexity.ai linked to 496 sources over the 99 questions for which it provided answers with sources.¹³ The median of sources for ChatGPT-4o was three, with a minimum of two and maximum of seven sources in one answer. Perplexity.ai provided five sources for all responses which had sources, with the exception of one response where it provided six.

Figure 11. Percentage of election-related questions for which chatbots provided sources

ChatGPT-4o and Perplexity.ai provided sources for the majority of questions, although ChatGPT-4o to a lower degree than Perplexity.ai.



Note: We prompted each chatbot with 100 election-related questions and fact-checks about the 2024 UK general election in the two weeks before the election on 4 July 2024. We coded if chatbots provided sources in their responses.

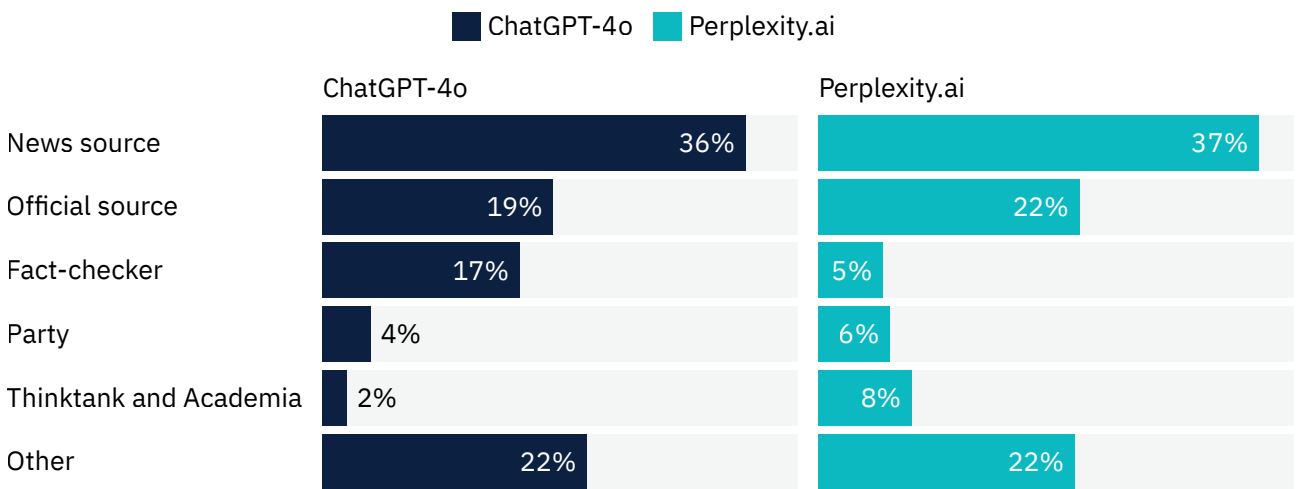
¹³ ChatGPT-4o queried and listed Bing.com in all of its responses with sources. We did not count Bing.com as an individual source.

When it came to the type of source provided (Figure 12), it is noteworthy that both chatbots drew most frequently from news sources in their responses (e.g. BBC News, the *Independent*, the *Guardian*, or the *Daily Telegraph*), followed by authorities and public bodies (e.g. the UK parliament and UK government websites, the Office for National Statistics (ONS), or the Electoral Commission), fact-checkers such as FullFact, party websites such

as Labour.org.uk, think tanks and academic institutions (e.g. the Institute for Government and University College London), and a variety of other sources ranging from Wikipedia to various blogs and private websites. In our dataset, Perplexity.ai provided slightly more sources from academic institutions and think tanks than ChatGPT-4o; however, ChatGPT-4o linked more often to fact-checkers.

Figure 12. Percentage of different source types

ChatGPT-4o and Perplexity.ai provided a mixture of different sources for some of their answers, including news sources, official sources, fact-checkers, parties, think tanks, and academic institutions, as well as other miscellaneous websites.



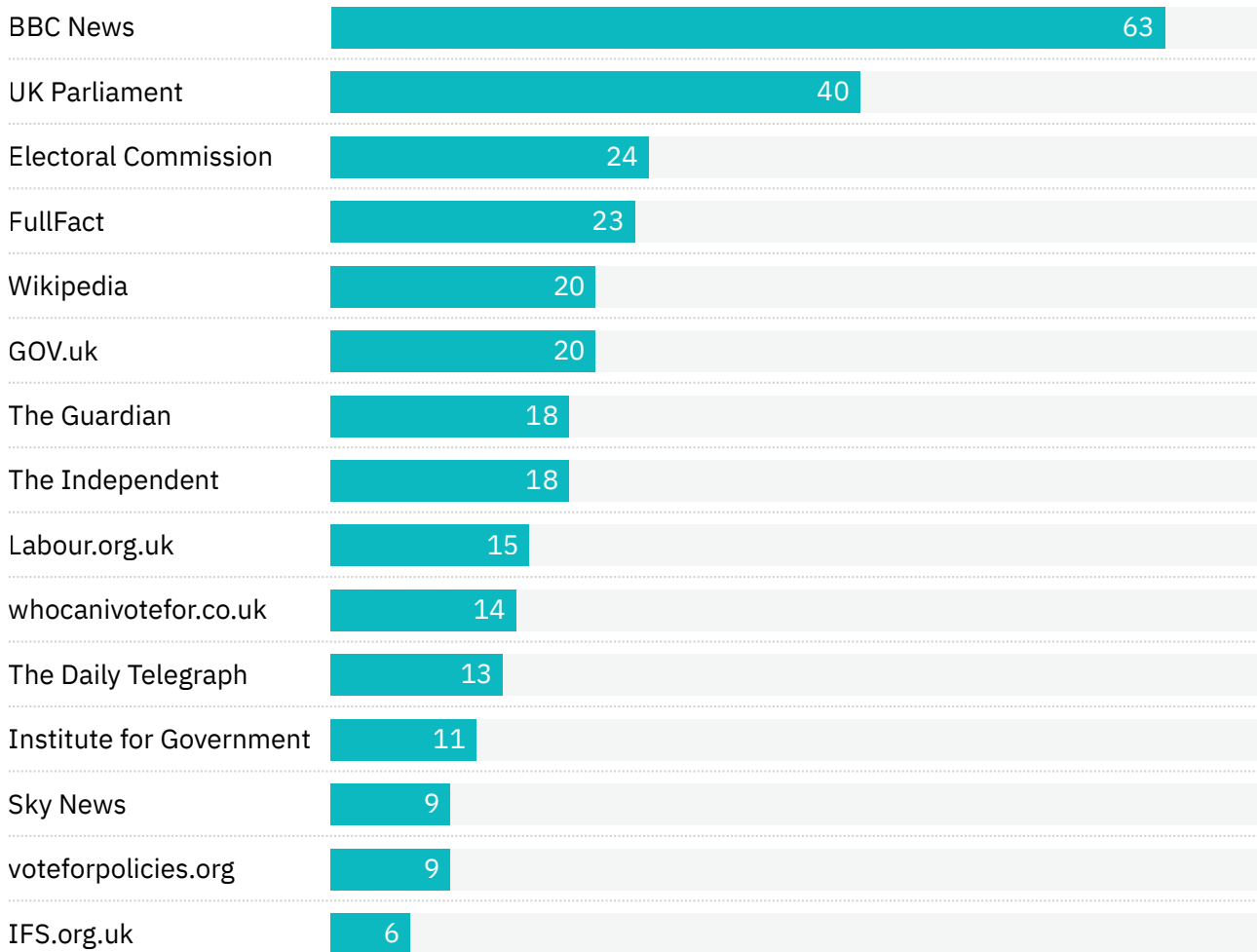
Note: We prompted each chatbot with 100 election-related questions and fact-checks about the 2024 UK general election in the two weeks before the election on 4 July 2024. We coded if chatbots provided sources in their responses. *Total N (ChatGPT) = 247 sources over 70 questions for which ChatGPT-4o provided responses with sources. Total N (Perplexity) = 496 sources over 99 questions for which Perplexity.ai provided responses with sources.* Bing.com was not counted as a source for ChatGPT-4o.

Looking at the 15 most frequently cited individual sources for both chatbots, we can see that Perplexity.ai (Figure 13) most commonly references BBC News, cited 63 times, followed by various websites of the UK parliament with 40 citations, and the Electoral Commission with 24 citations. Other common sources include FullFact, Wikipedia, and GOV.uk, each appearing 20 or more times. Prominent media outlets such as the *Guardian*, the *Independent*, and the *Daily Telegraph* are also frequently cited, alongside websites like whocanivotefor.co.uk (which provides information about elections and candidates based on a user’s

postcode) and voteforpolicies.org (a party policy comparison survey which allows users to compare promises from the main parties’ manifestos with their own preferences). In comparison, ChatGPT-4o (Figure 14) referred most frequently to UK fact-checker FullFact with 38 citations, followed by London-based newspaper the *Evening Standard* and the *Independent*, each cited 14 times, and Yahoo! News with 13 mentions. Governmental and authoritative sources such as websites of the UK parliament, GOV.uk, and the ONS also feature prominently.

Figure 13. The 15 most frequent sources quoted by Perplexity.ai

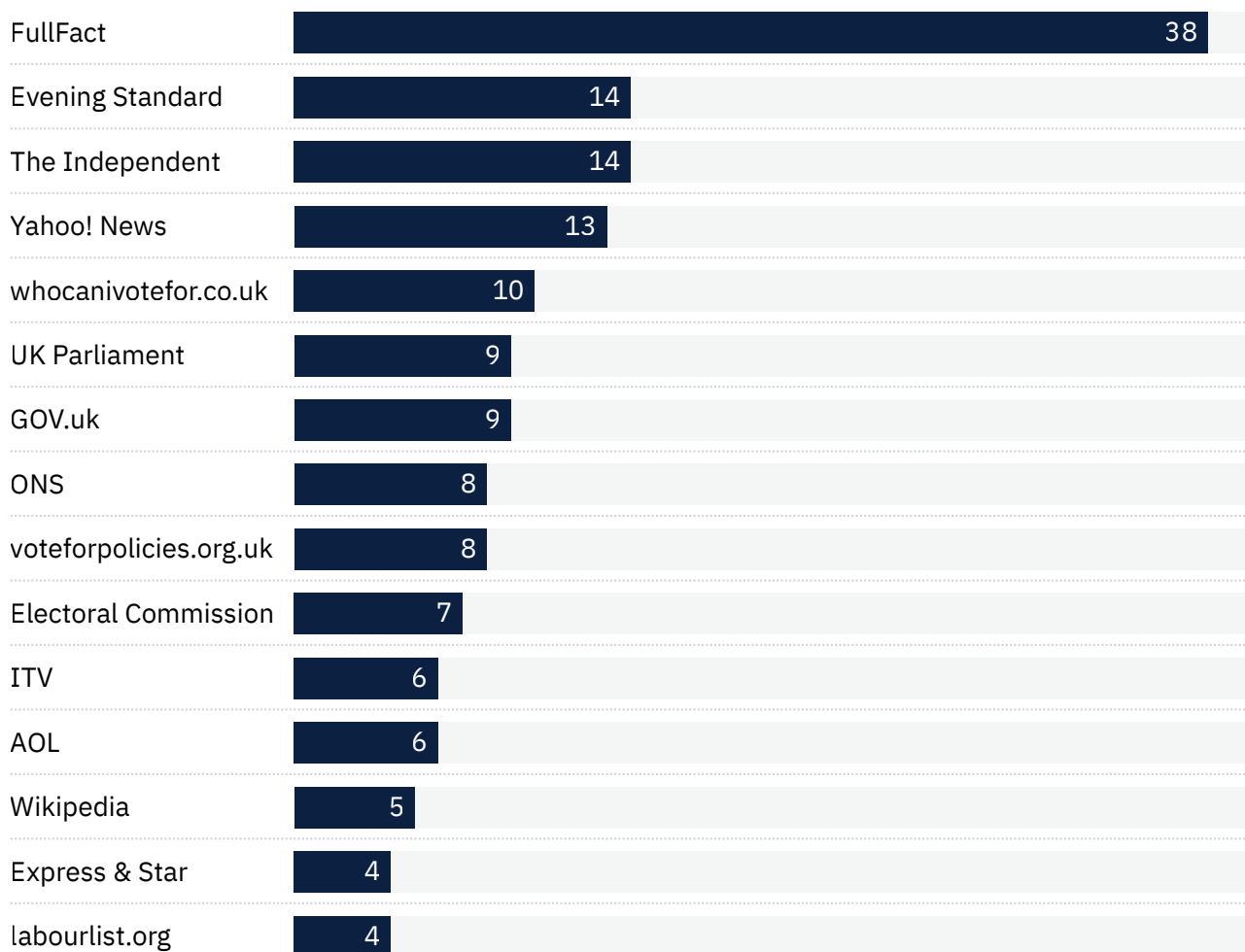
Perplexity.ai linked to BBC News content most often, followed by websites of the UK parliament, the Electoral Commission, and various other fact-checkers, news sources, and other sources.



Note: We prompted each chatbot with 100 election-related questions and fact-checks about the 2024 UK general election in the two weeks before the election on 4 July 2024. We coded if chatbots provided sources in their responses. Total N = 496 sources over 99 questions for which Perplexity.ai provided responses with sources.

Figure 14. The 15 most frequent sources quoted by ChatGPT-4o

ChatGPT-4o linked to fact-checker FullFact most often, followed by the *Evening Standard* and various other news sources, official sources, and other websites.



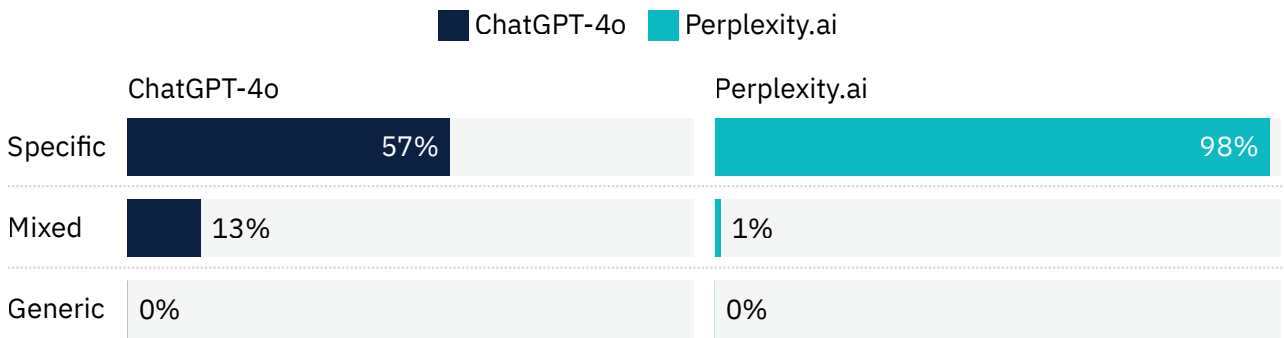
Note: We prompted each chatbot with 100 election-related questions and fact-checks about the 2024 UK general election in the two weeks before the election on 4 July 2024. We coded if chatbots provided sources in their responses. *Total N = 247 sources over 70 questions for which ChatGPT-4o provided responses with sources.* Bing.com was not counted as a source for ChatGPT-4o.

Finally, we wanted to know for how many responses the chatbots would provide us with only specific sources (pages on a clearly identifiable topic relating to a claim) in comparison with a mixture of specific and generic sources (a general website page which does not relate to the claim) or only generic sources

(Figure 15).¹⁴ Perplexity.ai linked only to specific sources in 98% of its responses, whereas ChatGPT-4o did so 57% of the time, with 13% of its responses containing at least one link to a generic source, most commonly the main pages of AOL and Yahoo! News.

Figure 15. Percentage of responses with specific or generic links to sources or both

Perplexity.ai linked only to specific sources in almost all responses, whereas ChatGPT-4o did so for 57% of its responses, with 13% of its responses containing at least one link to a generic source.



Note: We prompted each chatbot with 100 election-related questions and fact-checks about the 2024 UK general election in the two weeks before the election on 4 July 2024. We coded if chatbots provided specific sources in their responses. We defined a specific source as a page on a clearly identifiable topic relating to a claim (e.g. www.website.com/politics/election_2024_polling in a question about polling results) in comparison with a generic source (e.g. www.website.com). Bing.com was not counted as a source for ChatGPT-4o.

Conclusion

Our analysis looked at the performance of generative AI chatbots on election-related questions about the UK general election 2024. We found that, with the exception of Google’s Gemini, AI chatbots generally provided direct answers to the political questions and fact-checks we posed to them. Instances where chatbots avoided a straightforward reply or gave non-committal responses were rare, and the systems generally did not shy away from directly refuting false information in our data. Our dataset can only provide a snapshot view of the situation and there are no widely agreed-upon ways of testing these systems for factual accuracy – different approaches (e.g. sampling strategies, coding schemes) may yield somewhat different results, and these results in turn call for interpretation, with little in terms of objective benchmarks.

Our results provide some evidence, too, for how difficult it is to make general statements about these systems’ performance given how they continue to

develop. For example, when we tested ChatGPT Enterprise (ChatGPT-4) and Google’s Bard across ten countries to provide us with the latest news headlines in April 2024, both systems performed poorly in that ChatGPT-4 returned non-news output more than half the time and Google’s Bard did so 95% of the time (Fletcher et al. 2024). Compared with this, for the task at hand in this study two chatbots performed relatively well, and better than in some previous studies on the same topic, providing answers in most cases that were often direct and correct.

A clear majority of responses from ChatGPT-4o and Perplexity.ai were overall accurate, but we saw some notable errors, such as incorrect listings of candidates. Some of these inaccuracies seemed to stem from the chatbots’ difficulties in handling tasks that require reasoning about and contextual understanding of the questions. Nevertheless, the incorrect answers were stated with the same certainty as responses that were correct. The chatbots

¹⁴ For example, for a question about how to vote, a specific page such as www.bbc.co.uk/how-to-vote-in-the-general-election (please note that this website is a hypothetical example) rather than the generic source of www.bbc.co.uk.

frequently provided sources, including those from well-known and trusted news organisations, as well as fact-checkers, authorities, and other sources often considered reliable, such as academic institutions. It is also noteworthy how news organisations provided a majority of the information that we queried these chatbots for and how frequently ChatGPT-4o and Perplexity.ai linked to news sources with which they had no known licensing agreement at the time (e.g. BBC News). However, it remains an open question whether and to what extent users take advantage of these opportunities to access additional information, and which sources receive preferential treatment (and for which kinds of questions), highlighting the need for further research in this area. Whereas many of our news-related prompts in our previous study produced ‘I’m unable to’-type output that is very unlikely to be a satisfying enough experience to lead users to rely on chatbots for general news for now, the election-related questions and fact-checks we examine here generally produced direct, correct answers (often relying in part on news coverage, and the same official sources much news coverage of these issues relies on), suggesting that chatbots might be a potentially more compelling alternative to news media for such purposes.

Despite the better performance of these systems in our study than in some other analyses, concerns about LLMs remain. While these chatbots can evidently produce true responses, their output is ultimately still the result of probability and subject to the limitations of the underlying architecture, the quality of the training data, and various other factors. Even the most advanced foundation models continue

to sometimes hallucinate or regurgitate false information. While the overall quantity of accurate responses these systems provide matters, not all responses carry equal stakes. Sometimes, a single incorrect response can potentially have a greater impact than a string of incorrect ones – especially when considering that chatbots often provide wrong answers with the same certainty as correct ones. They also only provide a single answer, which differs from search results where users are presented with a range of options to pick from. One noteworthy recent example is X’s AI chatbot Grok, which drew criticism from experts and US lawmakers after it provided false information to millions of users suggesting that US vice-president Kamala Harris was not eligible to appear on the 2024 presidential ballot (Ellison and Gardner 2024).

While some claims and questions such as this can be easily verified, others are more complex because they have multiple correct answers or are not verifiable or falsifiable at all. How chatbots deal with these issues at scale will be important to understand, especially as they become more integrated into people’s digital experiences. Finally, with some early research suggesting that responses that match a user’s views are more likely to be preferred (Sharma et al. 2023), future research will also need to pay close attention not just to the accuracy of such chatbots but also to how people use the information they provide in contexts such as elections – where much is at stake, and people frequently rely on informational shortcuts, so correct answers therefore matter to a greater extent than in some other domains.

References

- AI Forensics. 2023. 'Prompting Elections: The Reliability of Generative AI in the 2023 Swiss and German Elections', *AI Forensics*, 20 December, <https://aiforensics.org/work/bing-chat-elections> (Accessed 26 Jul. 2024).
- Angwin, J., Nelson, A., Palta, R. 2024. 'Seeking Reliable Election Information? Don't Trust AI', *Proof News*, 27 February, <https://www.proofnews.org/seeking-election-information-dont-trust-ai/> (Accessed 26 Jul. 2024).
- Diakopoulos, N. 2019. *Automating the News: How Algorithms are Rewriting the Media*. Cambridge, MA: Harvard University Press.
- Ellison, S., Gardner, A. 2024. 'Secretaries of State Urge Musk to Fix AI Chatbot Spreading False Election Info', *The Washington Post*, 4 August, <https://www.washingtonpost.com/politics/2024/08/04/secretaries-state-urge-musk-fix-ai-chatbot-spreading-false-election-info/> (Accessed 5 Aug. 2024).
- Fletcher, R., Nielsen, R. K. 2024. *What Does the Public in Six Countries Think of Generative AI in News?* Oxford: Reuters Institute for the Study of Journalism. <https://doi.org/10.60625/RISJ-4ZB8-CG87>.
- Fletcher, R., Adami, M., Nielsen, R. K. 2024. *'I'm Unable to': How Generative AI Chatbots Respond When Asked for the Latest News*. Oxford: Reuters Institute for the Study of Journalism. <https://doi.org/10.60625/risj-hbny-n953>.
- Ji, Z., et al. 'Survey of Hallucination in Natural Language Generation', *ACM Computing Surveys*, 55(12), 248, 1–38, <https://doi.org/10.1145/3571730>.
- Marinov, V. 2024. 'Don't Bother Asking AI About the EU Elections: How Chatbots Fail When It Comes to Politics', *Correctiv*, 23 May, <https://correctiv.org/en/fact-checking-en/2024/05/23/dont-bother-asking-ai-about-the-eu-elections-how-chatbots-fail-when-it-comes-to-politics/> (Accessed 26 Jul. 2024).
- McClain, C. 2024. 'Americans' Use of ChatGPT-4o is Ticking Up, but Few Trust Its Election Information', *Pew Research Center*, 26 March, <https://www.pewresearch.org/short-reads/2024/03/26/americans-use-of-chatgpt-is-ticking-up-but-few-trust-its-election-information/> (Accessed 8 Aug. 2024).
- Sharma, M., et al. 2023. *Towards Understanding Sycophancy in Language Models*. Conference paper, International Conference on Learning Representations 2024. <https://openreview.net/forum?id=tvhaxkMKAn>.
- Simon, F. M., Adami, M., Kahn, G., Fletcher, R. 2024. *How AI Chatbots Responded to Basic Questions About the 2024 European Elections Right Before the Vote*. Oxford: Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/news/how-ai-chatbots-responded-basic-questions-about-2024-european-elections-right-vote> (Accessed 26 Jul. 2024).

Acknowledgements

This factsheet is published by the Reuters Institute for the Study of Journalism as part of our work on AI and the Future of News, supported by seed funding from Reuters News. The authors would like to thank the team of the Reuters Institute for their valuable feedback and support.

About the Authors

Felix M. Simon is a Research Fellow in AI and News at the Reuters Institute for the Study of Journalism.

Richard Fletcher is Director of Research at the Reuters Institute for the Study of Journalism.

Rasmus Kleis Nielsen is the Director of the Reuters Institute for the Study of Journalism and Professor of Political Communication at the University of Oxford.

