

© 2024. Reuse of manuscript is restricted to non-commercial, and no derivative uses.

For more information please visit:

<https://uk.sagepub.com/en-gb/eur/journal-author-archiving-policies-and-re-use>

Accepted on 15-July-2024 for publication in Political Studies Review.

<https://us.sagepub.com/en-us/nam/political-studies-review/journal202480>

This is the author accepted manuscript.

Early Results: Assessing ChatGPT as a tool for research on US state and territory politics

Introduction

Artificial intelligence (AI) is driving change across academic disciplines and societal spheres. As is often the case with new technologies, some observers are wary of its impacts (Ferrer et al., 2021) and others are cautious about government responses (Schopmans, 2022). Early analyses identify issues with the accuracy of the outputs AI can produce (Burger et al., 2023), with others, whilst not denying potential downsides, seeing the possibility for positive societal impacts depending on how AI is utilised (Fry, 2019; König and Wenzelburger, 2020). Another set of analyses notes the potential for bad actors to use AI to generate and spread mis- and dis-information (Satter, 2023; Milmo and Hern, 2023); in the context of politics and governance, many experts are thus wary of the potential of AI to influence public opinion and political outcomes (Schick, 2020) and are concerned about the uses of AI in political decision making (Saetra, 2021). As such, it is important that, as in other fields, politically-focused AI-generated content is accurate.

ChatGPT, a product created by the AI research and development company OpenAI, is probably the best-known of the generative AI platforms developed within the last few years. These platforms use “large language models” to generate content based on user-initiated prompts (Sundar and Mok, 2023). Generative AI platforms are typically “trained” on huge amounts of previously-published materials; when asked to provide information by a user, the algorithms guiding the technology generate a response based on similar previous questions.¹

The 50 States or Bust! project, within which this early results paper sits, aims to contribute to a systematic understanding of the accuracy and value of politically-focused AI-generated content. This occurs by the analysis of politically-focused profiles generated via standardised ChatGPT4 prompts for all US states and territories, and engagement with experts on the politics of all US states and territories. One of the goals of the project is to compile an interview with an expert in each state or territory that can be shared publicly in order to compile a repository of high-quality information about US states and territories.

After introducing the 50 States or Bust! project research question and design, this paper explains the project purposes, highlights the variety among the profiles generated within ChatGPT, and provides details of the work completed to date. The results section documents problems with references provided by ChatGPT, including the vague nature of some sources provided by ChatGPT, and, more concerningly, details the inaccuracy or even wholesale fabrication of other sources. It documents the mixed response to the profiles by experts interviewed thus far, and records their observations about the uses and limits of ChatGPT for the study of subnational US politics. Interestingly, expert scores given for different aspects of the profiles during this pilot suggest ChatGPT is better suited to accurately capturing the history of states as opposed to providing information about politics or the insights available from academic literature. The final section reflects on the project and provides initial thoughts on the future of this strand of research.

Research Question and Method

¹ When asked, for the purpose of this paper, how to explain generative AI to someone unfamiliar with it, ChatGPT responded: “Generative AI is like a creative robot brain that learns from lots of examples to create new things on its own.”

The research question motivating the 50 States or Bust! project is:

To what extent is ChatGPT useful as a research tool for exploring state and territory level US politics?

As the question suggests, the level of analysis of the project is state and territory US politics. This question is answered via research team engagement with profiles on all states and territories generated in ChatGPT4 via standardised prompts, and by drawing on experts on states and territory level politics. Ethical approval for interviews of such experts, defined as research participants, was obtained from [University X] in June 2023.

To begin to answer the above research question, in Spring 2023 a standard list of ChatGPT prompts that could be tweaked for all US states and territories was developed (50 States or Bust, 2023a). These prompts were designed to provide insight into how ChatGPT would respond to inquiries about the history and politics of US states and territories, and which sources of information would be provided when requested by the prompts. In all cases, the prompts asked ChatGPT to cite relevant sources. Between 22 June and 27 June 2023, these prompts were put into ChatGPT4 for all states and territories. This generated 56 profiles² (50 States or Bust, 2023b).³

The average profile length is 464 words, though there is variety on either side of the average. The shortest, New Hampshire, is just 123 words, and the term “profile” is generous, as no details are provided. Instead, ChatGPT explains that ‘as an AI model developed by OpenAI with [a] [...] knowledge cutoff in September 2021, I do not have access to the internet to provide real-time data or sources post 2021, nor do I have the ability to cite specific academic work’ (50 States or Bust, 2023b). The brevity and lack of content of this response is particularly surprising given New Hampshire’s significant role in US electoral politics.

At the other end of the spectrum, ChatGPT generated 705 words in response to the standard prompts for the state of Alabama. A 157-word answer was generated, for instance, in response to a prompt to provide details about state history. Among other things this response touched briefly on various topics including Alabama’s admittance to the union, pre-colonial and colonial history, and civil rights. In most cases, separate text was generated for all prompts, though this was not always the case. For Virginia, for instance, questions were explicitly answered collectively rather than individually, while for Missouri, sequential numbering was not provided (50 States or Bust, 2023b).

Concurrently with the creation of the profiles, the authors crafted standardized approach materials, including an email template, list of interview questions, and informed consent resources to solicit participation from experts in state politics in all US states and territories. The term “expert” was conceived broadly and includes academics, practitioners and activists, and journalists; however, to date each of the experts interviewed is affiliated with an academic institution. To identify experts to approach in each of the 50 states and territories, searches of newspaper articles about state politics, reviews of the state politics literature in journals such as *State Politics and Policy Quarterly*, and standard Google searches were used. Once an expert was identified through one or more of these methods, they were sent the approach materials via email. Sometimes, this led to the recommendation of a different expert through a version of snowball sampling, in which the expert initially contacted suggested that someone else they knew was more of an expert. For example, one scholar who was approached declined and offered: ‘I pay what is an average “good citizen” amount of attention ...but definitely not

² There are two profiles for Washington DC with tweaked prompts. This was to see if there was a difference in outputs when referring to Washington DC as a territory *or* a federal district.

³ Links to all project profiles files are available on this [project page].

“political scientist” amount of attention. I could recommend some other [state university] folks that might be a better fit if that would be helpful.’ In many cases, initial outreach yielded no response from the identified expert, requiring the research team to begin the process anew. When an expert agreed to participate, interviews with a member of the research team were scheduled at mutually convenient times. With the experts’ permission, these interviews were recorded for analysis and for the creation of podcasts about each state or territory.

At the time of this writing, a total of 17 experts have been interviewed. While this number represents approximately one-third of US states, and therefore does not yet reach the goal of a complete set of state and territory interviews, it is large enough that it has generated several important insights, which are discussed below. That just 17 interviews have been conducted over a ten-month period reflects the challenges with recruiting participants and scheduling interviews. The interviews are scheduled as the participants are available, and participants have tended to be most available during traditional university recesses and holidays. Nevertheless, because the unit of analysis is each individual state, the order of the interviews should not affect the outcomes of the study, although as additional interviews are obtained, a fuller picture of the value of generative AI for addressing political questions will emerge.

To test the methodology and ensure that the questions about the ChatGPT-generated profiles would yield substantive responses from expert respondents, a pilot study consisting of three interviews focused on the politics of Virginia, North Carolina, and California was carried out in late August and early September 2023. Professor Rich Meagher was interviewed about Virginia on 16 August, with Professor Christopher A. Cooper interviewed on 30 August about North Carolina. Finally, on 6 September, Professor Jeffrey Cummins was interviewed about California. These states were selected to be the pilot set because these experts were the first to agree to participate in the project; each expert agreed to participate in an extended interview that included questions about the study methodology itself.

The pilot interviews had three parts. Part 1 interviewed participants about the politics of the respective state, using the prompts fed into ChatGPT to guide but not limit the semi-structured interviews. Neither ChatGPT in general, nor the state profiles, were discussed in Part 1. The aim for Part 1 is twofold; first, to have an output that could act as a general introduction to the politics of each state for the non-expert and second, in terms of the project specifically, to have a point of comparison for the state profiles generated by AI. In Part 2, participants were asked to reflect on the ChatGPT state profiles with regard to details provided on history and politics, and the profiles’ engagement with academic work. In Part 3, participants were asked to reflect on their experience as research participants and suggest methodological developments for the main study. In general, feedback for Part 3 was very positive, with reflections and suggestions for development also arising. Expert suggestions included tweaking the wording and location of one of the questions in Part 2; this suggestion was implemented. One of the initial participants noted the potential priming effect of sharing the ChatGPT profiles prior to interviews. The potential for this is accepted, although it was clear that for other research participants, having the profiles prior to the interview had helped them in their preparation and led to more nuanced and reflective insights about the profiles. As such, it was decided to continue sharing profiles prior to interviews for the main study. Following the analysis of the pilot interviews, the project proceeded with the adjustment to question wording described above.

The process of carrying out the remaining interviews is ongoing. Since early September 2023, 14 participants beyond the initial three in the pilot have been asked the same battery of questions in Parts 1 and 2 of the interview protocol. In order to compare the experts’ responses to the profiles, in Part 2, all interview subjects were asked to provide scores between 1 (low) and 10 (high) for three dimensions of the ChatGPT profiles for their states: how well ChatGPT

captured the history of the state; how well ChatGPT captured the politics of the state, and; how well ChatGPT incorporated extant academic literature about the state. Table 1, below, presents the scores on these question for all 17 responses collected thus far. As Table 1 demonstrates, the overall average rating on the 10-point scale is 5.6, with ChatGPT doing somewhat better in the experts' view in summarizing state histories than in summarizing state politics or incorporating insights from the academic literature. These results are described in greater detail below; it must be noted, however, that because the experts were not provided with specific criteria by which to judge the profiles, their numerical responses can be compared only insofar as they offer a general sense of how the experts individually and collectively responded to the ChatGPT profiles. Therefore, in addition to the numerical scores, the results section below also offers representative commentary from across the interviews, to contextualize the experts' ratings.

Results

This section discusses the project's early results on two themes: first, the sources cited in the ChatGPT profiles and next, the experts' responses to the profiles in Part 2 of the interviews, which asked the experts to assess how well the profiles for their respective states captured the realities of that state's history and politics. These themes are the focus of the analysis because answering the question of how useful ChatGPT is for understanding state- and territory-level US politics requires interrogating the algorithm's ability to identify and synthesize legitimate information sources.

Source Usage

ChatGPT's difficulty with identifying and properly citing sources has been noted previously. When asked to cite its sources, ChatGPT frequently provides incorrectly-used, wrongly-listed, jumbled-up, or—as per Burger et al. (2023), in a term adopted to capture instances when AI appears to make things up—“hallucinated” references. Existing studies of AI hallucination rates are not granular enough to differentiate rates of made-up sources from rates of made-up facts, but recent research suggests that ChatGPT's overall hallucination rate is around just 3% (Metz, 2023; Tyson 2023). The analysis in this paper suggests otherwise.

TABLE 1 HERE

As noted above, all ChatGPT prompts requested that sources be given for the information provided. In its responses, the sources provided were largely online sources and often included only vague information, such as “Britannica,” “*New York Times*,” or a state government source with .gov after the state name. Notably, Nevada's profile is largely based on Wikipedia (50 States or Bust, 2023b).

Chat GPT's sourcing can, at times, be even vaguer, with a lack of any verifiable information provided. For example, in the Ohio profile, the reference given for a prompt asking for ChatGPT to cite ‘academic work’ was: ‘(Sources: Academic articles on Ohio politics, demographic and election data analyses).’ In many cases this meant identifying and interrogating the sources cited was impossible. Moreover, where source information was provided, it was often outdated, with many references from the 1990s and 2000s. There is nothing wrong with such sources *per se*, but they are obviously problematic in terms of providing meaningful contemporary summaries.

In addition to the vague citations ChatGPT provided, some sources it references are simply made up. For example, the Delaware profile cites ‘*The Delaware Citizen*’ by Roger A.

Martin and Carol E. Hoffecker (50 States or Bust, 2023b). However, *The Delaware Citizen* is a book published in 1952 and 1967 written by Cy Liberman and James M. Rosbrow (Liberman and Rosbrow, 1952), while Martin and Hoffecker both also wrote other books on Delaware (Hoffecker, 2004; Martin, 1991). As such, the detail provided in the profile is an interesting mix of information but is fundamentally incorrect. Similar issues are present in, among others, profiles for Virginia, Illinois, Louisiana, Missouri, and Hawai'i. The Missouri profile cites 'The Show-Me Paradox: Resilient Republicanism in "Bellwether" Missouri' by Marvin Overby, but a Google search for the title turns up nothing and no such work appears on Professor Overby's *curriculum vitae* (Overby, 2023). Vermont's profile is largely sourced by incorrectly cited popular press articles (50 States or Bust, 2023b). In all, to date at least one-third of all state profiles ChatGPT generated have suffered from hallucinated citations; additional others included hallucinated facts.

Expert Assessments

To some extent reflecting the issues with sources identified above, the experts offered a mixed response when asked to evaluate the state profiles. Many of the research participants were ambivalent about the information ChatGPT provided, with profiles being characterised as 'accurate' but also 'incomplete' (New York) (50 States or Bust, 2024); 'fine' but 'anodyne' (Virginia) (50 States or Bust, 2023c); 'a thumbnail sketch' (North Carolina) (50 States or Bust, 2023d); a 'superficial view of things' (Nevada) (50 States or Bust, 2023f) and 'lack[ing] specifics and details' (California) (50 States or Bust, 2023f). Some of the experts could not find anything explicitly inaccurate in how the profile captured the history or politics of their state; for example, reflecting on how ChatGPT covered the history of Virginia, Professor Rich Meagher said 'it's not bad' and that '[t]here's nothing [...] inaccurate' (50 States or Bust, 2023c). But, echoing SUNY Binghamton Professor Wendy Martinek's assessment of the New York profile, Randolph-Macon College's Meagher added, 'it's just very much incomplete, even for a brief overview' (50 States or Bust, 2024). While most experts echoed these criticisms about brevity and lack of completeness, other experts found that the profile for their state included incorrect information. For example, Miami University of Ohio Professor John Forren noted that the Ohio profile incorrectly claimed that Ohio's congressional delegation numbered 16, when in fact since 2022, the number has been 15 (50 States or Bust, 2023i). Forren noted that the incorrect information likely resulted from ChatGPT's programming cutoff in 2021. But the 2021 cutoff date cannot explain the incorrect information that Iowa State University Professor Karen Kedrowski identified in Iowa's profile: 'It says that [Iowa's] current boundaries were established in 1851, when it was admitted to the union in 1846. That's not just counterintuitive, that's wrong' (50 States or Bust, 2023j).

Finally, as noted above, most experts found that ChatGPT did not adequately incorporate existing insights from extant academic work on US state and territorial politics. Despite ChatGPT being given the prompt 'Citing academic work, in 150 words explain the politics of the US state of X' (with X swapped out for the relevant state), the experts' responses paint a relatively bleak picture across the board. In the case of North Carolina, the profile does not contain a single reference for any questions, and Professor Chris Cooper noted that 'It doesn't even really imply any academic work' even though 'there are books. There are articles that you know, using North Carolina as an example [...]. And they're just not. They're just not mentioned' (50 States or Bust, 2023d). Likewise, Ohio's expert, John Forren, said: '[I]t does not use the academic literature much at all. It tends to use sort of A[ssociated] P[ress] coverage or websites that themselves are compilers, secondary compilers of information' (50 States or Bust, 2023i).

However, while expressing disappointment in ChatGPT's use of academic literature, some of the experts hastened to add that while ChatGPT was not great at incorporating academic insights, some of the problem is the lack of sufficient scholarship on subnational US politics, such that there is a very limited set of academic materials for ChatGPT to incorporate. Utah's Leah Murray, a professor at Weber State University, noted: 'The problem is ... political science doesn't really cover States like Utah all that much. ... I think sometimes with us fly-over States, we don't get a lot of love. And the other thing I would say to you, which I think is interesting: parts of it [are] very rural and I don't think rural politics get up to a lot of love in political science. ... I don't think it's just about Utah. I think political science likes urban, coastal [states]' (50 States or Bust, 2023g). Professor Jaclyn Kettler at Boise State University in Idaho concurred: 'Some states get a lot of attention, and for good reason. They're either big, or there's a lot of scholars there, a lot of universities...But then other states get way less attention. But there's often still really interesting dynamics happening that can help us better understand American politics in general' (50 States or Bust, 2023h).

These narratives are reflected in the data in *Table 1* above. In all cases scores for the capturing of the history of states are the highest, leading to an average score of 6.9 out of 10. In the case of capturing the reality of state level politics, the average score was 6.1. By far the lowest scores were volunteered in response to the prompt asking how well the profiles reflected the insights from academic literature about the states. The average score across all 17 profiles on this dimension was 3.7, although this is skewed somewhat high (by comparison, the median value is 3) because two of the profiles—Vermont and South Dakota—were rated 9 and 10, respectively, for their incorporation of insights from the academic literature. Together these led to an average overall rating of 5.6 for the 17 profiles included to date.

Discussion and Next Steps

The initial work of the 50 States or Bust! project has demonstrated issues with the content and sources cited by ChatGPT, both in terms of their often vague, and thus hard to verify, nature and in terms of hallucinated references. While teaching faculty might take heart that these hallucinated sources, at least at present, permit identification of AI-generated content, they also point to a broader concern about the use of generative AI as a tool for political analysis, at least in the context of subnational US politics. These results support evidence from other work that has demonstrated the existence of hallucinations of false information, sometimes as composites of correct information reformulated into false sources. The results also suggest that ChatGPT may be less accurate at providing information about subnational US politics than it is at providing other types of information (Metz, 2023; Tyson 2023). The authors and experts have thus far identified factual inaccuracies or hallucinated references in more than 40 percent of the 17 profiles analysed. This certainly limits its utility as a resource for understanding subnational US politics.

Moreover, although this study is focused on subnational US politics, it identifies potential pitfalls for the use of AI in the study of politics, more generally, including at the national and international levels. For example, when ChatGPT was given the prompt: 'Citing academic sources, explain the politics of Luxembourg,' a country selected at random, it produced a 440-word essay, complete with in-text citations and a list of five works cited. However, not even one of the citations exists. To offer one exemplar, ChatGPT returned the following: 'Engel, C. (2015). "Luxembourg's Role in the European Union." *European Review*,

23(3), 416-435.’ But *European Review*’s third issue in volume 23 does not include any article on any subject with an author whose surname is Engel.⁴

Though more work is needed, interviews with experts from one-third of US states suggest that while ChatGPT can capture important elements of the *history* of US states, it is less well suited to do so with relation to state *politics*. This is almost certainly because a state’s history is far more static than a state’s politics, particularly in states with competitive elections and frequent changes in party control. ChatGPT4’s training cutoff in 2021—the same year in which US census-driven legislative redistricting occurred—means that it has a tendency to make mistakes about how many legislative seats a state has; the fact that every state held statewide elections either in 2022 or 2023 also makes it far less useful for understanding any aspect of a state’s politics that might have been changed by the outcomes of those elections.

As further interviews with experts on states and territories are carried out, avenues for additional research on US subnational politics, and likely other subfields, will emerge. These could, for instance, include research into profiles on specific regions (e.g., New England) or on the profiles focused on territories. There will also be the possibility to analyse themes within the interviews with experts (e.g., how well academic work covers states and territories, along the lines noted above, as well as if/how the trend toward the nationalisation of US politics is reflected in the profiles). Finally, there will be opportunities to use the project materials, both the profiles and interview materials, to support educational resources exploring the value of AI as a tool for research into the political realm in general, and US state and territory level politics in particular. While the work on this project is still ongoing, the results obtained from the project thus far offer scholars from a variety of disciplines and subfields one potential approach for evaluating the utility of ChatGPT in their own work. Furthermore, the evidence presented here of the propensity of ChatGPT to generate hallucinated sources suggests a mechanism for scholars, practitioners, and educators to verify the accuracy of AI-generated information in their own work.

References

Project Materials Cited

50 States or Bust (2023a) *50 States! June '23 Prompts Document*. Available:
<https://ukpsaapg.co.uk/wp-content/uploads/2023/12/50-States-June-23-Prompts.pdf>

50 States or Bust (2023b) *50 States or Bust! Project Page*. Available:
<https://ukpsaapg.co.uk/50-states-or-bust/>

50 States or Bust (2023c) *Project interview with Rich Meagher*, August 16, 2023. Interview by X.

⁴ In addition, pages 416-435 span three different articles: ‘Literary and Cultural Circulation: Machado de Assis and Théodule-Armand Ribot’ (pp. 406-420); ‘Japan, the First Quest of Modernization in East Asia’ (pp. 421-425); and ‘An Unknown Masterpiece: On Pak Kyongni’s *Land* and World Literature’ (pp. 426-438).

- 50 States or Bust (2023d) *Project interview with Christopher A. Cooper*, August 30, 2023. Interview by X.
- 50 States or Bust (2023e) *Project interview with Jeffrey Cummins*, September 6, 2023. Interview by X.
- 50 States or Bust (2023f) *Project interview with David Damore*, November 20, 2023. Interview by X.
- 50 States or Bust (2023g) *Project interview with Leah Murray*, November 28, 2023. Interview by X.
- 50 States or Bust (2023h) *Project interview with Jaclyn Kettler*, December 11, 2023. Interview by X.
- 50 States or Bust (2023i) *Project interview with John Forren*, December 12, 2023. Interview by X.
- 50 States or Bust (2023j) *Project interview with Karen Kedrowski*, December 13, 2023. Interview by X.
- 50 States or Bust (2024) *Project interview with Wendy Martinek*, January 2, 2024. Interview by X.

References Cited

- Burger B, Kanback, DK, Kraus S, Breier M, and Corvello V. 2023. On the Use of AI-Based Tools Like ChatGPT to Support Management Research, *European Journal of Innovation Management* 26: 233-241.
- Ferrer X, van Nuenen T, Such JM, Coté M, and Criado N (2021) Bias and Discrimination in AI: A Cross-Disciplinary Perspective, *Technology and Society Magazine*. 72-80.
- Fry A (2019) *Hello World: How to be Human in the Age of the Machine*. London: Penguin.
- Hoffecker C (2004) *Democracy in Delaware: The Story of the First State's General Assembly*. Wilmington, DE: Cedar Tree Books.
- König P and Wenzelburger G (2020) Opportunity for renewal or disruptive force? How artificial intelligence alters democratic politics. *Government Information Quarterly* 37. DOI: <https://doi.org/10.1016/j.giq.2020.101489>.
- Lieberman C and Rosbrow JM (1952) *The Delaware Citizen: The Guide to Active Citizenship in the First State*. Houston, TX: Elsevier Press.
- Martin R (1991) *Tales of Delaware*. No publisher or place of publication information available.
- Metz, C. (2023) Chatbots May ‘Hallucinate’ More Often Than Many Realize. *The New York Times*. 6 November 2023.
- Milmo, D. and Hern, A. (2023). Elections in UK and US at risk from AI-driven disinformation, say experts. *The Guardian*. Retrieved May 18, 2023 from <https://www.theguardian.com/technology/2023/may/20/elections-in-uk-and-us-at-risk-from-ai-driven-disinformation-say-experts>.

- Overby, M. (2023) Curriculum Vitae. Retrieved June 6, 2024 from <https://harrisburg.psu.edu/files/pdf/21226/2020/01/15/m-overby-cv-2023.pdf>.
- Sætra, H.S. (2021) A Typology of AI Applications in Politics. In: Visvizi, A., Bodziany, M. (eds) *Artificial Intelligence and Its Contexts*. Advanced Sciences and Technologies for Security Applications. Springer, Cham. https://doi.org/10.1007/978-3-030-88972-2_3.
- Satter, R. (2023) Exclusive: AI being used for hacking and misinformation, top Canadian cyber official says. Reuters.com. Retrieved May 18, 2023 from <https://www.reuters.com/technology/ai-being-used-hacking-misinfo-top-canadian-cyber-official-says-2023-07-20/>.
- Schick, N (2020) *Deep Fakes: The Coming Infopocalypse*. New York: Twelve/Hachette Books.
- Schopmans H (2022) From Coded Bias to Existential Threat: Expert Frames and the Epistemic Politics of AI Governance. In *Proceedings of 2022 AAAI/ACM Conference on AI, Ethics, and Society*, August 1-3, Oxford, United Kingdom. DOI: <https://doi.org/10.1145/3514094.3534161>.
- Sundar, S and Mok, A. (2023) What is ChatGPT? Here's everything you need to know about ChatGPT, the chatbot everyone's still talking about. *Business Insider*. Retrieved January 17, 2024 from <https://www.businessinsider.com/everything-you-need-to-know-about-chat-gpt-2023-1?r=US&IR=T>.
- Tyson, M (2023) AI Models Ranked By Hallucinations: ChatGPT is Best, Palm-Chat Needs to Sober Up. *Tom's Hardware*. Retrieved May 18, 2024 from <https://www.tomshardware.com/news/ai-hallucinations-ranked-chatgpt-best>.

Table 1: Scores Assigned to Aspects of 50 States or Bust! ChatGPT Profiles by State Experts

State	Captured realities of history the state (1-10)	Captured realities of state politics (1-10)	Captured insights from academic literature (1-10)	State average
Virginia [†]	8	4	4	5.3
North Carolina [†]	6	4	0	3.3
California [†]	9	8	2 ⁵	6.3
New Hampshire	0	0	0	0.0
Iowa	5	5	3	4.3
Illinois	5	6	2	4.3
Idaho	10	7	4	7.0
Utah	5	7	7	6.3
Ohio	9	6	0	5.0
New York	9	9	0	6.0
Vermont	8	7	9	8.0
Minnesota	8	6	8 ⁶	7.3
South Dakota	9	9	10	9.0
Florida	4	4	2	3.3
Massachusetts	9	6	3	6.0
Missouri	6.5	7	4	5.8
Nevada	8 ⁷	7	7	7.3
Overall average	6.9	6.1	3.7	5.6

[†] Pilot states.

⁵ X did not ask for this score in the initial interview, but in a later email. Email chain on file with X.

⁶ Extrapolated from the interviewee's verbal response. If this value is omitted, the average score is reduced to 7.0, which is not substantively different from the current value.

⁷ Extrapolated from the interviewee's verbal response. If this value is omitted, the average score increases to 7.5, which is not substantively different from the current value.