Original Manuscript

Generating the Past: How Artificial Intelligence Summaries of Historical Events Affect Knowledge

Social Science Computer Review
2025, Vol. 0(0) 1–20

DOI: 10.1177/08944393251409744
journals.sagepub.com/home/ssc
Sage

Daniel Karell[1], Matthew Shu[2], Thomas Davidson[4], and Keitaro Okura[3]

Abstract

Many people now use AI chatbots to obtain summaries of complex topics, yet we know little about how this affects knowledge acquisition, including how the effects might vary across different groups of people. We conducted two experiments comparing how well people recalled factual information after reading AI-generated or human-written historical summaries. Participants who read AI-generated summaries scored significantly higher on knowledge tests than those who read expert-written blog posts (Study 1) or Wikipedia articles (Study 2). These improvements were present regardless of whether readers knew the content was AI-generated or if the AI summaries were politically biased. Moreover, AI summaries improved recall across various demographic groups, including gender, race, income, education, and digital literacy levels. This suggets that using AI tools for everyday factual queries does not create new knowledge inequalities but could still amplify existing ones through differential access. Our findings indicate that the increasingly routine use of AI for information-seeking could enhance factual learning, with implications for education policy and addressing inequality.

Keywords

generative AI, artificial intelligence, summarization, information ecosystems, online experiments

Introduction

Artificial intelligence (AI) is transforming many domains of social life, including work, education, and healthcare (Joyce et al., 2021). This includes recent large language models (LLMs) that are already revolutionizing how many people access and consume information (Capraro et al., 2024;

[1]Department of Sociology and Institution for Social and Policy Studies, Yale University, New Haven, CT, USA
[2]Department of Statistics and Data Science, Yale University, New Haven, CT, USA
[3]Department of Sociology, Yale University, New Haven, CT, USA
[4]Department of Sociology, Rutgers University–New Brunswick, New Brunswick, NJ, USA

Corresponding Author:
Daniel Karell, Department of Sociology and Institution for Social and Policy Studies, Yale University, 493 College Street, New Haven, CT 06511, USA.
Email: daniel.karell@yale.edu

Liao et al., 2024; Milmo & Robins-Early, 2024). People can now find answers to a variety of questions by asking chatbots, like OpenAI's ChatGPT, or by using tools like Google's AI Overviews or Perplexity to summarize web queries. These tools have become widely adopted: one-third of American adults report using chatbots according to an April 2025 Pew Survey (McClain et al., 2025), and ChatGPT is reported to have over 500 million weekly active users.[1] To date, the implications of these novel technologies for learning are not well understood, with scholarship emphasizing both promises and risks (Yan et al., 2024).

The goal of our research is to better understand how the everyday use of AI tools to learn factual information influences knowledge acquisition and the extent to which this varies across social groups. To do so, we address three related issues, building upon an emerging literature on the impacts of AI usage. First, much of the existing work on AI and learning focuses on formal schooling contexts (Bastani et al., 2025; Ma & Zhong, 2025; Zhu et al., 2025) rather than the increasingly ubiquitous use of these tools for simple, mundane factual queries. As a result, we know relatively less about how this kind of consumption of AI-generated content affects knowledge acquisition compared to the consumption of easily accessible existing information sources. Second, AI technologies risk "reinforcing hegemonic biases" present in their training data (Bender et al., 2021, p. 613; Gillespie, 2024), potentially reproducing structural inequalities (Joyce & Cruz, 2024; Joyce et al., 2021; Zajko, 2022), as other algorithmic technologies have done (Benjamin, 2019; Noble, 2018). However, the precise mechanisms through which everyday AI use for factual queries could perpetuate inequity remain unclear. Third, while scholarship emphasizes how AI can be used to create misinformation (Feuerriegel et al., 2023; Kreps et al., 2022; Makhortykh et al., 2023), its outputs can easily be politicized (Argyle, Busby, et al., 2023) even if they are not factually inaccurate. Thus, politicized yet factually sound summaries may affect knowledge acquisition in distinct ways.

We focus on historical events, a relatively accessible form of knowledge that is widely encountered on the internet, often in the form of simplified summaries, and which offers an ideal case for evaluating the impacts of AI on learning. Most people know how to approach and make sense of a text about a historical event, compared to descriptions of more abstruse topics like biological mechanisms or statistical formulas. Additionally, people often construct an understanding of their current social environment by referencing past events. Learning about the past (even, and perhaps especially, from short summaries) is a common and familiar practice involved in understanding the present. Moreover, recent media coverage indicates that AI is shaping the craft of professional historians (Wasik, 2025), suggesting that the way AI describes events has significant implications for how human history is presented.

We conduct two online experiments that test whether learning about a historical event by reading summaries produced by ChatGPT results in a better or worse factual recall than reading summaries written by humans. We compare the synthetic summaries against texts from two online sources that people commonly turn to when seeking to learn about history: blog posts and articles on Wikipedia, the largest online encyclopedia, which attracts around ten billion page views a month.[2] The second experiment additionally tests whether the effect of reading synthetic summaries varies depending on whether the content is labeled as AI-generated and whether it conveys a politically ideological slant. We also analyze how learning from synthetic content varies across social and demographic attributes to examine inequalities in its effects.

We find that AI-generated summaries lead to higher factual recall about historical events than human-written summaries. The results of the first study ($N = 193$) show that individuals who read a synthetic historical summary of the Seattle General Strike (SGS) correctly answered, on average, more factual questions about the event than individuals who read a blog post written by experts. The preregistered second study replicates the findings of the first study when drawing on a sample that reflects the population of the United States (US) along several social and demographic

dimensions ($N = 1907$). It compares synthetic summaries to Wikipedia, testing both SGS and an additional historical event, the Third World Liberation Front (TWLF) student protests. The results also show that reading synthetic summaries contributed positively to knowledge, regardless of whether individuals knew the summary was generated by AI and whether the synthetic summaries framed the events in a politically biased manner. Additional analyses in Study 2 suggest that reading synthetic summaries led to gains in factual knowledge regardless of gender, ethnicity, race, household income, educational attainment, and digital literacy. This suggests that AI tools may not directly exacerbate inequalities in knowledge, but rather, existing inequalities will be exacerbated by unequal access to AI.

Background and Research Questions

Recent research has shed important light on how AI-generated content influences people's attitudes, opinions, and beliefs (Bai, Voelkel, et al., 2025; Costello et al., 2024; Hackenburg & Margetts, 2024), how AI assistants can shape social relations by mediating interactions among people (Argyle, Bail, et al., 2023; McKee et al., 2023; Ueshima et al., 2024), and how AI can change people's understanding of the social world by altering how they perceive the people and events (Karell et al., 2025; Laba, 2024). Yet, it remains less clear how the use of these tools affects factual knowledge. Scholarship on AI and education has focused on the potential of AI to aid learning by innovating personalized and diverse learning materials and assessment methods, while also introducing potential risks due to inaccuracies, ethical issues, and disruption of traditional learning methods (Yan et al., 2024). Meanwhile, meta-analyses of AI use in schools and universities find mixed impacts of AI on learning (Ma & Zhong, 2025; Zhu et al., 2025). We look beyond classroom settings and study the increasingly common use of AI to access information and learn about the world on the internet. Chatbots like ChatGPT, Gemini, and Claude have reached large audiences, and AI-generated syntheses now feature prominently in search engine results, augmenting or replacing traditional information sources.

To illustrate how AI-infused information ecosystems now operate, we performed a Google search for the term "seattle general strike," one of the events in the experiements, using a private browser session. The results, shown in Figure 1, feature an "AI Overview" of the event at the top that was produced using one of Google's Gemini models. The results also include hyperlinks to a Wikipedia page and expert resources at the University of Washington, as well as a panel with some basic information. This highlights how information from AI-generated summaries has become prominent and intermingled with resources written by humans, both expert and amateur. Moreover, there is emerging evidence that these AI summaries are supplanting existing sources: Wikipedia recently reported a substantial drop in traffic, blaming generative AI, and noting how its content was being used to create summaries that substituted visits to the website (Miller, 2025). The goal of our research is to establish whether learning differs when people read such AI-generated content compared to material written by humans.

Existing scholarship suggests the potential for diverging impacts on knowledge and learning. On one hand, there are reasons to expect that obtaining information from synthetic content may impede or even undermine the acquisition of knowledge. There are widespread concerns that the LLMs underlying chatbots may be "hallucinating," producing plausible-sounding responses to queries that are interwoven with misleading and embellished half-truths, or outright falsehoods (Bender et al., 2021; Farquhar et al., 2024; Yan et al., 2024). Moreover, popular AI tools can produce derivative text that has less lexical diversity than human writing (Padmakumar & He, 2023) and skews toward homogeneity (Doshi & Hauser, 2024; Karell et al., 2025; Laba, 2024). These characteristics may make synthetic content less interesting to readers than human-created content, thereby dampening engagement and undermining the retention of factual information.

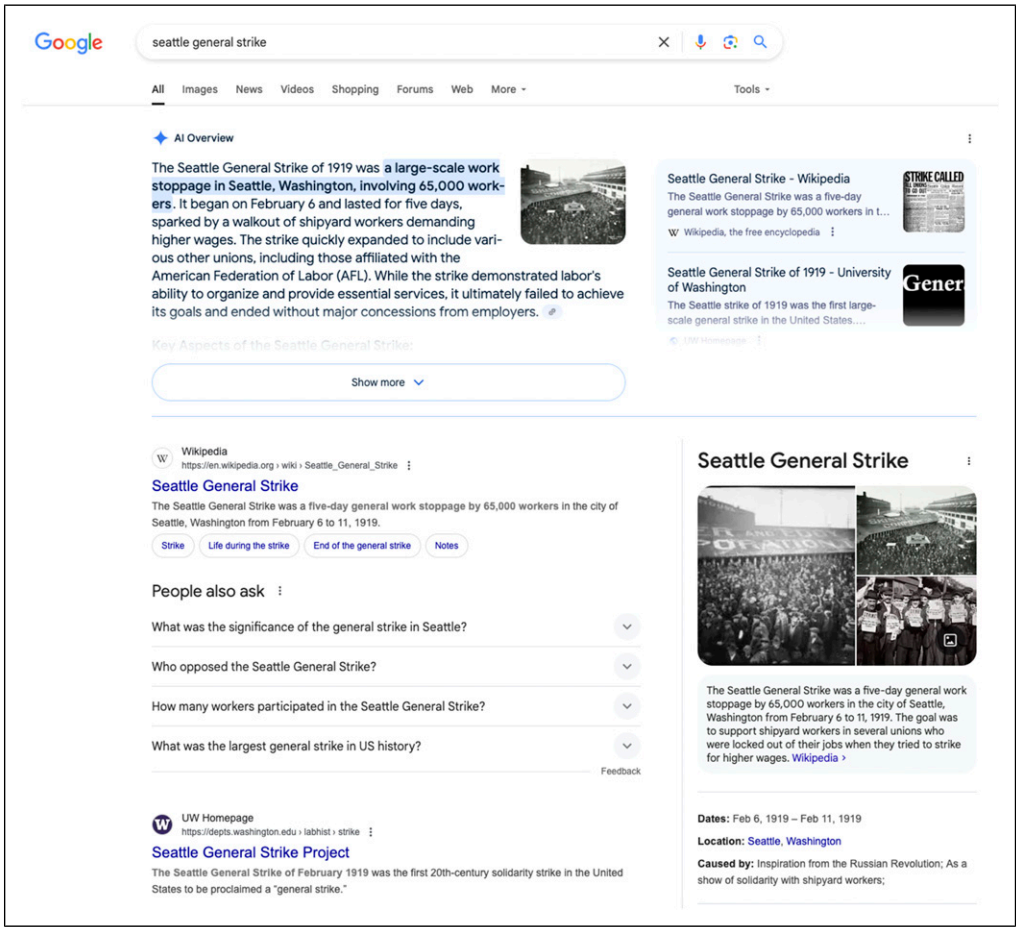

Figure 1. Top Google search results for the query "seattle general strike". The search was performed in a new private browsing session. Screenshot by the authors

On the other hand, chatbots like ChatGPT have been designed to "write" in clear, familiar, and easily comprehensible ways that could support factual recall (Padmakumar & He, 2023; Spitale et al., 2023). Furthermore, there is evidence that LLMs can contribute to creative thinking (Ueshima et al., 2024), engage users' views in personalized ways (Costello et al., 2024), and appear more "professional," which users have found pleasant to read (Doshi & Hauser, 2024). If people learn about the past through clearer, more comprehensible, and more engaging content, they may be more likely to retain facts they learn from synthetic texts and increase their knowledge. The countervailing expectations of how synthetic historical summaries may affect people's knowledge motivate our first research question:

RQ1: Does reading AI-generated summaries about historical events help or hinder the acquisition of knowledge, compared to reading human-written summaries?

Understanding the consequences of reading synthetic summaries for learning requires considering how individuals' awareness of AI provenance affects their factual recall. The increasing sophistication and proliferation of AI content have drawn attention to many people's difficulty

distinguishing between real and AI-generated texts and images (Clark et al., 2021; Nightingale & Farid, 2022; Spitale et al., 2023). As a result, some media outlets and platforms have begun labeling or watermarking synthetic content (Dathathri et al., 2024). Thus, people seeking information about historical events will sometimes be explicitly informed that they are reading synthetic texts, but at other times, they will not. And, when they know about the source, they may value it differently than text known to be written by humans (Gilardi et al., 2024; Palmer & Spirling, 2023; Zhang & Gosline, 2023), an effect which sometimes further depends on the content domain (Altay & Gilardi, 2024; Gallegos et al., 2025). We therefore examine whether telling individuals that AI-generated texts are synthetic affects their knowledge acquisition:

RQ2: Does being informed about the AI or human provenance of a historical summary help or hinder readers' knowledge acquisition?

There is evidence that generative AI models can produce biased outputs, including hateful language and offensive stereotypes (Abid et al., 2021; Bianchi et al., 2023; Hofman et al., 2024). Scholars argue that these biases are learned from the data used to train the models, as well as subsequent reinforcement learning (Bender et al., 2021; Gillespie, 2024; Steinert & Kazenwadel, 2024). AI companies have attempted to mitigate some of the more egregious biases, but they can be challenging to eliminate entirely (Bai, Wang, et al., 2025; Ouyang et al., 2022). Moreover, it is relatively straightforward to prompt these models to convey information with different perspectives and to induce political biases (Argyle, Busby, et al., 2023). This kind of politicization of AI models is particularly evident in xAI's Grok. Although Elon Musk has stated that Grok should be "politically neutral," the model's responses have shifted across the political spectrum as its instructions have been modified (Thompson et al., 2025). In July 2025, for example, an update to Grok's instructions led the model to produce racist and antisemitic posts on the social media platform X, at one point referring to itself as "MechaHitler" (Hagen et al., 2025).[3] Such biases, whether intentional or not, could influence how people receive and retain the information they encounter. Thus, the deep-rooted and pervasive nature of biases in generative AI motivates our third research question.

RQ3: Do politically biased historical summaries generated by AI promote or hinder knowledge acquisition, compared to human-written summaries?

When people use generative AI tools to gather information about the world, they do so with perspectives and skills that are, in part, shaped by their own characteristics. It is therefore reasonable to expect that individuals' social and demographic attributes could condition how reading synthetic summaries affects their knowledge acquisition. For example, AI tools that tend to write in a simpler and more accessible style might not confer significant benefits for individuals with relatively high educational attainment but could be particularly effective at conveying information to those with less formal education. Any such disparities could have significant implications for AI's impact on social inequality (Joyce & Cruz, 2024; Joyce et al., 2021; Zajko, 2022).

If socially disadvantaged individuals learn better from synthetic summaries, then AI tools may hold promise for improving information accessibility and alleviating certain types of social inequality. By contrast, if people with social advantages benefit more from reading synthetic summaries, this could mean that AI technologies will worsen existing inequities. For instance, individuals who are highly "digitally literate" can usually better discern the quality of information found in digital sources and therefore be selective about what they incorporate into their knowledge (Hargittai & Micheli, 2019; McCosker, 2024). This, in turn, could compound existing digital divides by privileging those who already have the resources to develop technological

competencies (Capraro et al., 2024; Yan et al., 2024). To examine these possibilities, we analyze the effect of learning from AI-generated text across the individual-level attributes of gender, ethnicity, race, household income, educational attainment, and digital literacy.

RQ4: How do individuals' social and demographic attributes condition the effect of reading AI historical summaries on acquiring knowledge?

Materials and Methods

To answer our research questions, we conducted two online human-subjects experiments during the spring of 2024.[4]

Study 1

Data Collection. We recruited 200 study participants through the online platform Prolific, which provides access to a large pool of verified potential study participants. To participate, they had to be aged 18 years or older, located in the US, and fluent in English. From these, we removed seven participants who spent less than 30 seconds reading the historical summary since reading and understanding an approximately 500-word text in less than 30 seconds was implausible. This created a final sample of 193.[5] See Supplemental Information (SI) Section A for the composition of the sample across various social and demographic attributes. We discuss a post hoc power analysis in SI Section B.

Materials. Study 1 examined how reading an AI-generated summary of a historical event affected participants' knowledge about the event compared to reading a summary written by expert humans. The human-written text was adapted from an online blog post summarizing the SGS, written by the Civil Rights and Labor History Consortium at the University of Washington.[6] The SGS was the first solidarity, or "general," strike in the US. Around 65,000 workers in Seattle from multiple industries and unions stopped work from February 6 to 11, 1919. We selected the SGS for our study because it was a significant enough event to have sufficient content online to enable LLMs to write cogent summaries, but at the same time, it is not so familiar that most Americans would know key facts about it.

We used ChatGPT to produce the synthetic summary, which we refer to as the "baseline AI" summary, because ChatGPT is the most widely used and one of the most capable chatbots, providing strong external validity. Specifically, we provided ChatGPT (using GPT-4o, the latest version at the time of our study) with the human-written text and prompted it to create five questions about the SGS that had factual answers in the human-written text (SI Section C). Then, in a separate ChatGPT session, we provided a prompt with three parts: a simple request to provide a 500-word overview of the SGS; instructions stating that the summary should contain information allowing readers to answer the following questions; and the five knowledge questions, along with their answers. This procedure was designed to create content comparable to that which a user might receive when querying a chatbot to summarize an event, while ensuring that the returned content contained the required factual knowledge. This design thus reduced the possibility of hallucination, allowing us to compare texts that contain the same basic facts. We fixed the temperature parameter to zero for all our experiments to produce the most deterministic response to each query.

SI Section C displays the prompts, and SI Section D presents the various texts. Consult SI Section E for the questions generated from the human-written text that we used to create the synthetic summary.

Design and Analyses. Upon entering the study, participants answered a series of questions about their social and demographic characteristics. Then, they were randomized with equal probability into one of two treatment conditions. One condition presented participants with the baseline AI summary of the SGS, while the other presented them with the expert-written text.

After reading the assigned summary, participants answered a set of questions to measure their knowledge of the SGS. Each participant attempted to answer the same questions. We used their responses to construct the outcome variable, their *knowledge score*, which was a sum of the number of questions they answered correctly. We analyzed differences in this outcome across treatment conditions by estimating Cohen's d, the standardized difference between the mean knowledge score of the group that read the AI summary and the group that read the expert-written text.[7] See SI Section E for the questions and SI Section I for the relationship between the summaries and the questions.

Study 2

Data Collection. For our second study, we recruited 2410 participants through the online platform Bovitz Forthright, a platform similar to Prolific. Potential participants had to be 18 years of age or older, located in the US, and fluent in English. We excluded 402 participants who failed to complete the study. Of these, 253 failed to answer the first set of outcome questions, and an additional 149 did not answer the second set of outcome questions. To better understand the risk that this posttreatment attrition posed to our results (Montgomery et al., 2018), we analyzed whether the participants who left the study had a different treatment exposure than those participants who remained. A chi-squared test indicated no significant difference overall between the two distributions of treatment assignment ($p \approx 0.83$). From the remaining 2008, we removed 115 participants who spent less than 30 seconds reading at least one of the texts, as in Study 1.[8] Thus, the final analytical sample comprised 1912 participants.

The sample reflects the US population in terms of gender, age, region of residence, race and Hispanic ethnicity, education, household income, and political ideology (based on population benchmarks from the 2019 American Community Survey). We report the composition of the sample across these attributes in SI Section A. See SI Section B for design analyses (Gelman & Carlin, 2014) and a discussion of post hoc power analyses.

Materials. Study 2 replicated and extended Study 1. We evaluated responses to summaries of the SGS and a second historical event, the Third World Liberation Front (TWLF) protests. We added this event to assess whether the findings of Study 1 were dependent upon the specific event. The TWLF event consisted of the formation of a coalition of university student groups (primarily based on ethnic identities and initially in the San Francisco Bay area) in 1968, followed by their activism supporting greater representation of ethnic minority views and experiences within academia. This activism led to the creation of Ethnic Studies departments in many universities. Like the SGS, the event was historically significant but is not so widely known that most participants would have good factual recall before reading the summaries. However, compared to the SGS, it is more closely related to contemporary debates that were occurring when we conducted this study, which centered on the roles of student activism and diversity, equity, and inclusion in universities. In addition, the TWLF activists were successful, whereas those involved in the SGS were not. These features enabled us to account for how effects might vary according to the relevance of events and whether they resulted in progressive social change.

To produce Study 2's summaries, we first generated baseline summaries of the TWLF and SGS events again using GPT-4o, the latest version at the time. When doing so, we used system messages and prompts tailored to create summaries that approximated "default" summaries a

person may encounter in everyday life when using a straightforward prompt to have a chatbot summarize a historical event. As such, our "baseline" summaries did not explicitly contain biases other than those already latent in ChatGPT. The system messages and prompts are presented in SI Section F of the SI, and Section G displays the generated texts.

We then created two biased synthetic summaries for each event, a politically conservative one ("conservative AI") and a politically liberal one ("liberal AI"). We provided a system message that articulated a conservative or liberal worldview, using a methodology similar to that proposed by Argyle, Busby, et al. (2023). Each message comprised a series of survey questions from the General Social Survey about important political issues, along with the responses to these questions that aligned with a particular political ideology. That is, the "conservative" system message had ideologically conservative answers to the questions, while the "liberal" system message had ideologically liberal answers. After each system message, we provided a prompt that created the summary. This procedure simulates a method by which chatbots can be easily manipulated to inject a biased perspective. Note that the system messages were the same for the summaries of both events, while the user prompts were particular to the events. See SI Section F for the system messages and prompts, and see SI Section G for the generated summaries.

For the human-written texts, we relied on Wikipedia, one of the main online sources people use to learn about historical events. Indeed, Wikipedia was the top result for the SGS when we searched for the event on Google (Figure 1), and it was the second result for the TWLF.[9] While we consider Wikipedia to be a reliable source of information (Steinsson, 2024), it is important to emphasize that we do not believe any source, including Wikipedia, can be "unbiased." Our goal is not to compare AI-generated summaries to an ideal "unbiased" text, but rather to compare them to sources that people could reasonably be expected to consult when learning about the past online.

To produce the human-written summaries, we downloaded Wikipedia articles about the SGS and TWLF events and lightly edited each article to match its length with that of the synthetic summaries. We used the current version of the Wikipedia articles at the time of the study, as the most recent versions have undergone the most revisions and tend to be the most neutral and authoritative (Greenstein & Zhu, 2018). When editing, we were careful not to remove any core information about the events. The human-written summaries are also shown in SI Section G.

Finally, we used all the texts to create knowledge questions that constitute the outcome variables. For each event, we began with a system message telling a ChatGPT assistant that it was a historian and teacher knowledgeable about the given historical event. Then, we prompted it to create six multiple-choice questions that could be answered with facts found across the Wikipedia text and each AI variant, all of which were included in the prompt. This procedure helped us identify fact-based questions that could be answered after reading any of the summaries or, seen from a different angle, ensured that the summaries shared the relevant content. We also used this AI-assisted question-generation procedure because it can be easily scaled to a larger number of events in future studies. After manually checking the generated questions and all the summaries, we selected three questions to include in the study and analyses. Consult SI Section F for the system messages and prompts used to create the questions and SI Section H for the questions included in the study.

Design and Analyses. The study began with participants answering questions about their social and demographic characteristics. (See SI Section A for the number of participants across levels of each characteristic.)

After answering the questions, participants were randomly assigned with equal probability to one of four treatment conditions: reading the baseline AI, conservative AI, liberal AI, or Wikipedia summaries of the events. Within each condition, participants first read the text summarizing the TWLF, then the text about the SGS. Both texts were always aligned with the condition. For

example, if a participant was in the "conservative AI" condition, both the texts she read were generated with AI using the conservative-bias system message.

Upon being placed into one of the reading conditions, participants were further randomized with equal probability into one of two "labeling" conditions. Specifically, they were either told about the summaries' provenance or not. In the labeled condition, the summaries were identified as AI-generated if they were synthetic or identified as human-written if they were from Wikipedia. The labels appeared in a short paragraph introducing each of the summaries. Participants were told that they would read AI-generated material before reading the TWLF summary and again before reading the SGS summary. Participants in the unlabeled condition were not provided with any information about the texts' provenance. SI Section K presents the labels inserted into the paragraphs introducing each summary.

After reading each summary, participants answered questions that measured their factual knowledge about the events. All participants were shown the same questions. We included an attention check question after the first summary and corresponding questions. After answering the second set of knowledge questions, participants had an opportunity to share feedback and were debriefed. See SI Section H for the questions and SI Section L for the relationship between the summaries and the questions.

As with Study 1, we constructed our outcome variable, *knowledge score*, by summing the number of questions each participant answered correctly. We analyzed this outcome by calculating the mean knowledge scores for each treatment group and comparing the differences between groups. To determine whether any differences are statistically significant, we report the means, distributions, and p-values, which are adjusted using the Holm-Bonferroni method when appropriate.[10]

We additionally examined how reading synthetic summaries affected knowledge scores conditional on participants' social and demographic attributes. To test for differences across social and demographic groups and answer RQ4, we fitted linear models with a multiplicative term that interacted the treatment condition with social or demographic attributes of interest: gender, Hispanic ethnicity, race, household income, level of educational attainment, and digital literacy.[11] These models compared the knowledge scores of the group who read the Wikipedia summary to the scores of just those who read the baseline AI summary, across the participants' attributes.

All models adjusted for whether participants read the labeled or unlabeled version of the summaries and were fit using ordinary least squares.[12] We present the results by showing the estimated marginal effects of the treatment across each level of a given attribute. When calculating the marginal effects, we averaged over levels of the labeling indicator variable. To assess statistical significance, we applied the Bonferroni correction and then calculated 95% confidence intervals.[13]

Results

AI-Generated Historical Summaries can Improve Knowledge

Our analyses compared the effect of reading AI-generated or human-written summaries of historical events on the factual recall of the events. We begin by reporting consistent results across Study 1 and Study 2, summarized in Figure 2. On average, participants in both experiments correctly answered more knowledge questions when shown the synthetic summary compared to the human summary. The results indicated strong statistical significance, and we found comparable effect sizes across both studies (Study 1, $d = 0.48$, $p = 0.001$; Study 2, $d = 0.48$, $p < .001$).[14] This similarity is notable because the studies employed different procedures to generate the synthetic summaries, varied types of human summaries, different questions to measure

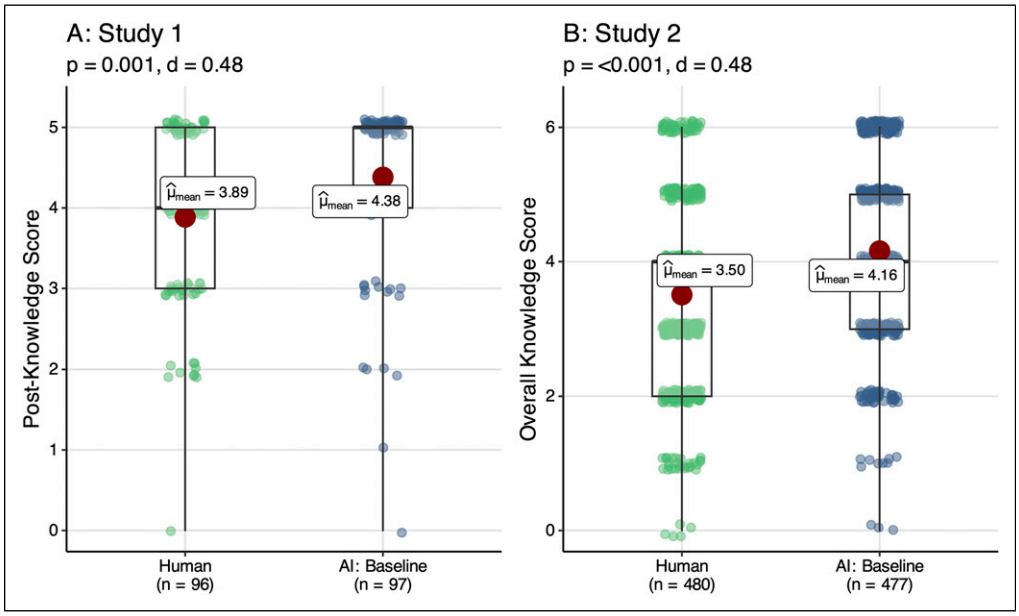

Figure 2. Readers of AI-generated historical summaries correctly answered more factual questions than readers of human-written summaries, on average

knowledge, different samples, and different platforms for implementing the experiments. In Study 2, we also find evidence of synthetic summaries' relative positive effect on knowledge when we tested the summaries of each event, SGS and TWLF, separately (SI Section M) and when we used the complete sample (SI Section N).

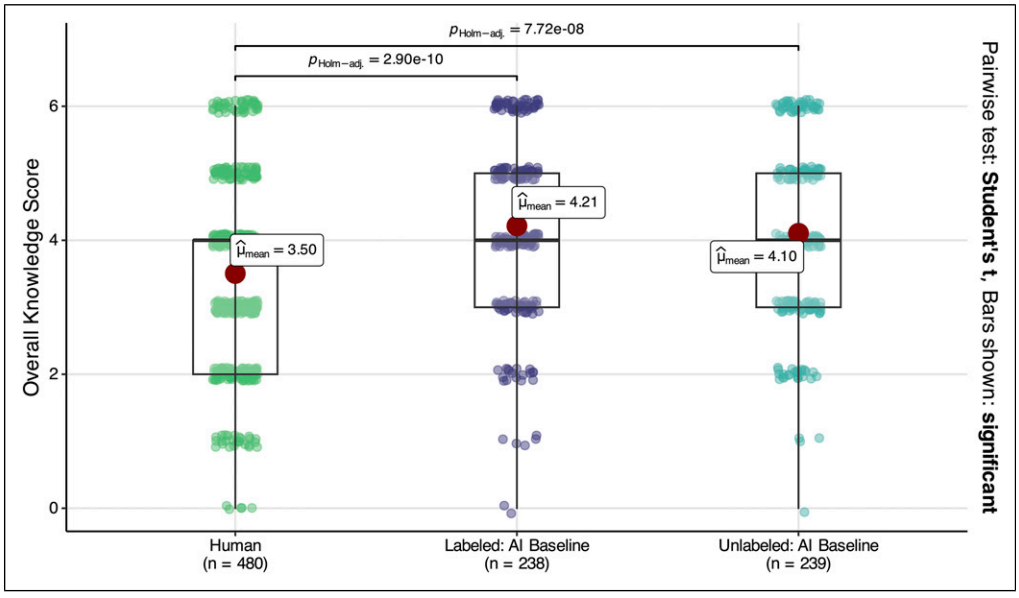

Figure 3. Readers of AI-generated historical summaries correctly answered more factual questions than readers of human-written summaries, on average, whether or not they were aware of summaries' AI provenance

AI-Labeling Does Not Impact Knowledge

Does awareness of the provenance of synthetic summaries alter how people answer knowledge questions? Figure 3 shows that participants performed better under the synthetic condition regardless of whether they were informed that the texts were AI-generated. In both synthetic conditions, we observe positive, statistically significant differences in answers to the knowledge questions compared to the Wikipedia condition. The improvement was greatest for the labeled condition—when subjects were told that they were reading AI-generated summaries—although there is no statistically significant difference between the labeled and unlabeled synthetic conditions. We observe similar results when examining the effect across events, as well as when comparing readers of labeled and unlabeled synthetic texts to readers of labeled and unlabeled human-written texts (SI Section M)[15], and when using the complete sample (SI Section N). Together, these findings indicate that readers' awareness of a text's AI provenance does not diminish the acquisition of knowledge from synthetic materials.

Political Bias in AI-Generated Summaries Does Not Impact Knowledge

So far, we have considered the baseline synthetic summaries that were not designed to convey any ideological slant. We now examine how the relationship between reading synthetic summaries and

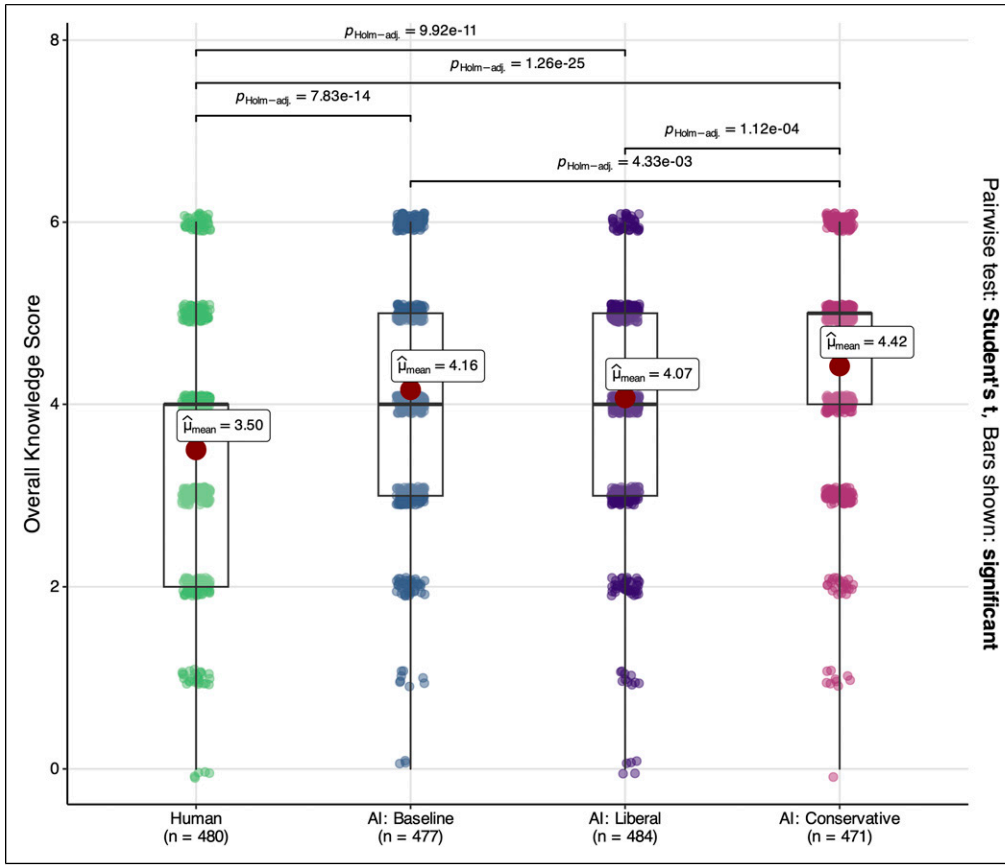

Figure 4. Readers of AI-generated historical summaries correctly answered more factual questions than readers of human-written summaries, on average, whether or not the summaries were politically biased

knowledge may change when the synthetic summaries frame historical events in politically biased ways. Figure 4 shows that both biased synthetic summaries led to significantly higher knowledge scores than the Wikipedia summary. This suggests that, holding key factual information constant, texts with an ideological slant do not necessarily undermine knowledge retention. We also observe some variation between the texts, as the readers of the conservative synthetic summary tended to perform slightly better than both readeres of the baseline and liberal AI summaries. The results are largely consistent when we examined the two historical events separately (SI Section M) and used the complete sample (SI Section N).

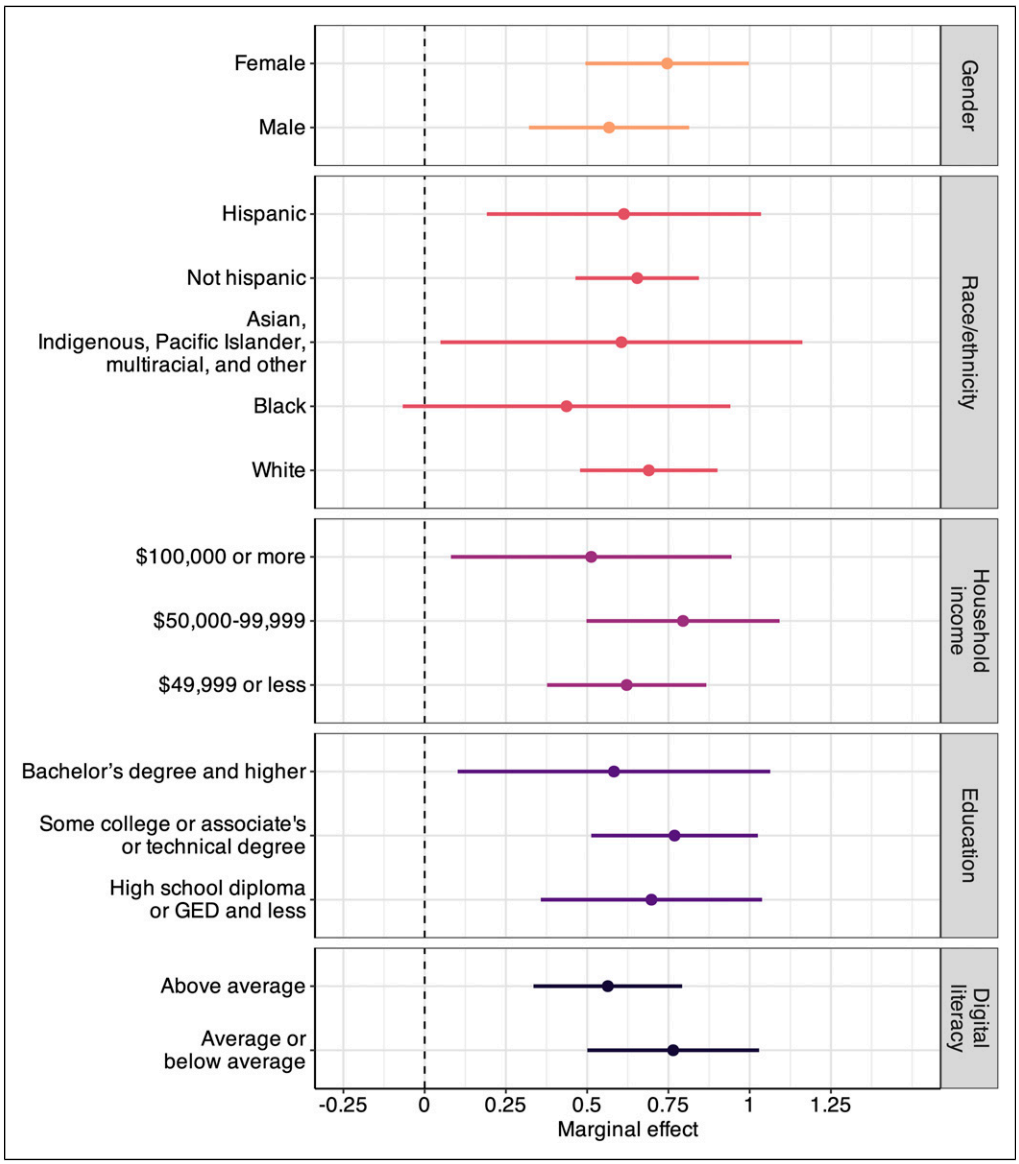

Figure 5. Baseline AI summaries, compared to human-written summaries, had a positive effect on knowledge across social and demographic groups. Error bars denote 95% confidence intervals

Baseline AI Summaries Improve Knowledge for Most People

The preceding results together show the robustness of our main finding: synthetic summaries of historical events can help people learn facts about the past more effectively than human-written summaries. However, as discussed earlier, it is possible that this positive effect only applies to certain groups of people.

Figure 5 suggests that this is not the case. Compared to the Wikipedia summaries, exposure to baseline synthetic summaries had a positive effect across a variety of individual-level attributes. There is scant evidence of heterogeneity across social and demographic groups. Both men and women benefited, as did Hispanic and non-Hispanic individuals and individuals with different racial identities. The only racial group that may not have benefited was those who self-identified as Black or African-American, although this null finding may be due to the somewhat small number of subjects ($n = 115$), as the estimate remains positive. We also observe consistent positive effects across different income levels, educational attainment, and digital literacy.

Discussion and Conclusion

Tools like ChatGPT, Claude, and Google's AI Overview have become common ways to access and synthesize information (Capraro et al., 2024; Liao et al., 2024; Milmo & Robins-Early, 2024). Our growing reliance on AI-generated information can have significant social consequences, which to date are not well understood. For example, AI may or may not help people gain factual knowledge compared to existing methods of obtaining information, and as a result, potentially exacerbate or reduce inequalities in knowledge and learning. In this paper, we have examined these potential consequences of using AI with two online experiments.

Summery of findings and implications

The results provide compelling answers to our first three research questions: reading AI-generated summaries of historical events has a positive effect on factual knowledge compared to human-written summaries (RQ1), and this effect exists whether or not the synthetic summaries are labeled as AI-generated (RQ2) and whether or not these summaries frame the events in a politically biased manner (RQ3). In other words, when they function well, AI tools can facilitate learning. Study 2 additionally demonstrates that the improvements in knowledge from reading AI-generated summaries can occurr across various social and demographic groups. Therefore, to answer RQ4: men and women, Hispanic and non-Hispanic, individuals identifying with most races, and individuals at each income bracket, level of educational attainment, and level of digital literacy all benefited, on average, from reading the synthetic summaries of historical events.

Our findings suggest that AI is effective at synthesizing factual information, thereby providing an important corrective when more conventional resources, such as search engines, yield low-quality results that can reinforce misperceptions (Aslett et al., 2023). Furthermore, while the awareness of AI, as when Google labels information as "AI Overview" (Figure 1), has mixed impacts in different domains (Altay & Gilardi, 2024; Gallegos et al., 2025; Gilardi et al., 2024), our study suggests that it does not diminish factual learning. Nonetheless, the integration of AI tools into online-based knowledge acquisition is not necessarily a universal good. We suspect that any existing advantages in learning, such as the advantages that individuals from wealthier households have over those from less wealthy households (Hällsten & Pfeffer, 2017), will probably continue uninterrupted by the adoption of AI tools. In other words, our findings indicate that the effect on social inequality will likely depend on access. It is the digital divide, or the

inequity in AI access and effective use, which will exacerbate inequality rather than the tools themselves (see also Capraro et al., 2024; Yan et al., 2024).

Why does AI enhance learning relative to the texts from exerpts' blog posts and Wikipedia? It is beyond the scope of our study to definitively explain why AI-generated summaries are effective, but we expect that two core features of ChatGPT and other LLMs make them particularly effective at conveying information. First, summarization is one of the main applications of LLMs; these models have been developed to excel at the exact task of generating texts that summarize existing information in a clear and concise manner. Second, post-training techniques such as reinforcement learning from human feedback (RLHF) enable models to produce output aligned with people's preferences (Ouyang et al., 2022). Models are built to generate the kinds of texts readers want and are optimized to deliver highly readable content (Doshi & Hauser, 2024; Karell et al., 2025; Padmakumar & He, 2023; Spitale et al., 2023). While we cannot draw conclusive inferences due to the small number of texts in our studies,[16] metrics computed on our texts indicate that the sentences in the AI summaries tend to be shorter than those in the Wikipedia texts, and they are more readable, based on the Flesch-Kincaid and SMOG indices, two widely used quantitative measures of readability (SI Section O).

Limitations and Directions for Further Research

Our paper is a step towards understanding the relationship between AI and learning, knowledge, and inequality. As such, we highlight three limitations of our research and areas for further research.

First, we tested only a small number of synthetic summaries and focused on two historical events. Future studies should expand the breadth of the texts generated to include both a greater number of events, as well as contexts beyond history. For any given event, it is also possible to generate multiple different synthetic versions, either by varying the prompt, the model, or the model parameters. In our case, we explored some prompt variations, but we used a single model, GPT-4o, and fixed the temperature parameter to obtain more deterministic results. Each of these aspects could be varied to examine how different prompts, models, and parameters impact the qualities of synthetic texts across various domains.

Second, our design necessitated creating texts with comparable length, which may simplify how AI summaries are encountered in the world and how people read different kinds of texts. Furthermore, the approach to ensuring internal validity in Study 1 may represent a best-case scenario for AI, insofar as we specified the factual content that must be included, potentially enabling the models to prioritize this material in a way that may not have been central to the production of human-written texts. However, the models did not simply "game" the task by placing the answers to the factual questions at the forefront of the summaries, as evidenced in SI Section I. In Study 2, we generated questions that can be answered from any of the texts, so there is less risk of gaming. As the examples in SI Section L show, the texts are of a similar length to the human-written examples (6–8 paragraphs), and the sentences containing the answers in AI-generated texts are spread throughout the texts. Based on this, we are confident that any differences are a function of the way that information is presented in the texts. Of course, some of the differences may be artifacts of the data construction process rather than true differences in human and AI-written texts, and future work should build upon this design by testing a broader range of texts, including those that have undergone no manual editing. Nonetheless, we expect that this design has external validity insofar as AI companies may use similar prompting strategies to present factual information, nudging the model to avoid hallucinating key facts from source materials.

Third, the study focused on one-off interactions that mimic how people query a chatbot or read a Wikipedia article to obtain the desired information. This design has strong external validity as these are common ways to use these resources in everyday life. However, people can also learn

from chatbots during extended dialogue sessions (Costello et al., 2024) or even during a semester-long college courses (Carpraro et al., 2024; Yan et al., 2024). In these cases, the effect of synthetic content on knowledge may differ. For example, the attributes of synthetic style that are potentially beneficial in one-off settings—such as greater readability, simpler syntax, shorter words—might render synthetic summaries repetitive and dull in conversational or multi-use settings, hindering their positive effect. Moreover, recent research on AI in classroom settings shows that the way in which AI is used matters, with AI "tutors" that provide hints being more effective at facilitating longer-term learning than models that give answers immediately (Bastani et al., 2025). Thus, future research should examine the relationship between AI tools and learning in settings with repeated interactions and different pedagogical strategies. In addition, future work should measure knowledge in other ways than immediate factual recall to better understand more in-depth and higher-level knowledge outcomes (Melumad & Yun, 2025).

Ultimately, this research is a step towards understanding how these new technologies shape the way we learn. Further research is needed to fully understand the broader implications of utilizing AI tools for learning about history and other aspects of society. While we show that biases in texts do not always hamper learning, it is plausible that these biases could have downstream implications for people's attitudes and beliefs and subsequent political polarization (Baldassarri & Gelman, 2008; DellaPosta, 2020; DiMaggio et al., 1996), particularly as traditional media consumption and trust in institutions continues to decline (Finkel et al., 2020). Our findings suggest that accurate yet instrumentalized synthetic text could be effective at propagating curated facts that advance certain political and social agendas (Makhortykh et al., 2023), along with false and misleading synthetic text (Kreps et al., 2022; Spitale et al., 2023). While AI companies have taken steps to safeguard against these uses, it is important to emphasize that unethical use of these tools has profound societal implications.

Acknowledgments

We thank the participants of the New York University Department of Sociology's colloquium for their insightful comments on an earlier version of this project.

ORCID iDs

Daniel Karell https://orcid.org/0000-0001-6709-6535
Thomas Davidson https://orcid.org/0000-0002-5947-7490
Keitaro Okura https://orcid.org/0000-0001-9061-7490

Ethical Considerations

The studies reported in this paper were approved by the Institutional Review Board at Yale University (#2000037333).

Consent to Participate

All participants provided informed consent before joining the studies reported in this paper. Consent was provided in written form.

Author contributions

D.K., M.S., K.O., and T.D. conceptualized and designed the project. M.S. and K.O. implemented the experiments. M.S., D.K., and T.D. analyzed the data. D.K. and T.D. wrote and edited the paper.

Funding

The authors received no external financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The code, data, and replication materials will be publicly available on an Open Science Foundation project site upon acceptance for publication.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. https://uk.finance.yahoo.com/news/sam-altman-faces-scrutiny-ted-022805407.html (last accessed November 2025).
2. See https://en.wikipedia.org/wiki/Wikipedia:Statistics (last accessed December 2024).
3. The embedding of ideological perspectives in models has also become subject to government regulation. Trump's second presidential administration issued an executive order titled "Preventing Woke AI in the Federal Government," mandating that models used by the federal government do not "sacrifice truthfulness and accuracy to ideological agendas" (The White House, 2025).
4. The experiments were approved by the Institutional Review Board at Yale University. The preregistration plan for Study 2 is available at https://osf.io/xtgnk/overview.
5. In the main analyses, we did not include respondents who spent less than 30 seconds reading at least one of the texts ($n = 7$). But since time spent reading is a posttreatment variable and conditioning on it can bias our estimates (Montgomery et al., 2018), we conducted the main analyses a second time using the complete sample and obtained consistent results (SI Section N).
6. Specifically, we used the text of the summary provided by the Civil Rights and Labor History Consortium (https://depts.washington.edu/labhist/strike/; last accessed September 2024) except for references to the other multimedia material.
7. We computed Cohen's d using the effectsize package for R (version 0.8.9) and visualized the results using the ggstatsplot package for R (version 0.12.3).
8. As earlier mentioned, participants' time spent reading is a posttreatment variable (Montgomery et al., 2018), so we replicated all of Study 2's analyses with the main sample of 2008 and obtained consistent results (SI Section N).
9. We performed these searches using a private browsing session.
10. We conducted the analyses using the R packages effectsize (version 0.8.9) and ggstatsplot (version 0.12.3).
11. When analyzing race, we grouped together Asians, American Indians, Alaska Natives, Native Hawaiians, Pacific Islanders, and people who reported being multiracial or some other race since these categories all had relatively few participants.
12. We implemented the models using the stats package for R (version 4.3.0).
13. We estimated the marginal effects and conducted Bonferroni corrections for multiple comparisons using the emmeans package for R (version 1.10.5).
14. The results of a design analysis indicate that these results of Study 2, as well as all its following results, are strongly powered and have negligible Type S error and Type M error. In addition, the results of a post hoc power analysis support Study 1's conclusions. For discussion and detailed results of these analyses, see SI Section B.
15. In our main discussion of the results, we pool the organic labeling conditions to focus attention on how types of synthetic summaries compared to organic summaries overall.

16. We used only a small number of texts in our studies, so we do not have sufficient material to perform an analysis intended to detect average effects of distinct textual attributes.

References

Abid, A., Farooqi, M., & Zou, J. (2021). Persistent anti-Muslim bias in large language models. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, Virtual Event, Association for Computational Machinery, pp. 298–306.

Altay, S., & Gilardi, F. (2024). People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation. *PNAS Nexus*, *3*(10), pgae403. https://doi.org/10.1093/pnasnexus/pgae403

Argyle, L. P., Bail, C. A., Busby, E. C., Gubler, J. R., Howe, T., Rytting, C., Sorensen, T., & Wingate, D. (2023). Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(41), Article e2311627120. https://doi.org/10.1073/pnas.2311627120

Argyle, L. P., Busby, E. C., Fulda, N., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, *31*(3), 337–351. https://doi.org/10.1017/pan.2023.2

Aslett, K., Sanderson, Z., Godel, W., Persily, N., Nagler, J., & Tucker, J. A. (2023). Online searches to evaluate misinformation can increase its perceived veracity. *Nature*, *625*(7995), 1–9. https://doi.org/10.1038/s41586-023-06883-y

Bai, H., Voelkel, J. G., Muldowney, S., Eichstaedt, J. C., & Willer, R. (2025). LLM-generated messages can persuade humans on policy issues. *Nature Communications*, *16*(1), 6037. https://doi.org/10.1038/s41467-025-61345-5

Bai, X., Wang, A., Sucholutsky, I., & Griffiths, T. L. (2025). Explicitly unbiased large language models still form biased associations. *Proceedings of the National Academy of Sciences of the United States of America*, *122*(8), Article e2416228122. https://doi.org/10.1073/pnas.2416228122

Baldassarri, D., & Gelman, A. (2008). Partisans without constraint: Political polarization and trends in American public opinion. *AJS; American Journal of Sociology*, *114*(2), 408–446. https://doi.org/10.2139/ssrn.1010098

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *122*(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

Bender, E. M., McMillan-Major, A., Gebru, T., & Schmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Association for Computational Machinery, 610–623.

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

Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., & Caliskan, A. (2023). Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. In Proceedings of the 2023 ACM conference on fairness, accountability, and transparency, Virtual Event, Association for Computational Machinery, 1493–1504.

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J. F., Brañas-Garza, P., Butera, L., Douglas, K. M., Everett, J. A. C., Gigerenzer, G., Greenhow, C., Hashimoto, D. A., Holt-Lunstad, J., Jetten, J., Johnson, S., Kunz, W. H., Longoni, C., … Viale, R. (2024). The impact of generative artificial intelligence on socioeconomic inequalities and policy making. *PNAS Nexus*, *3*(6), pgae191. https://doi.org/10.1093/pnasnexus/pgae191

Clark, E., August, T., Serrano, S., Haduong, N., Gururangan, S., & Smith, N. A. (2021). All that's 'human' is not gold: Evaluating human evaluation of generated text. In Proceedings of the 59th annual meeting of

the association for computational linguistics and the 11th international joint conference on natural language processing, Association for Computational Linguistics, Online, 7282–7296.

Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, *385*(6714), eadq1814. https://doi.org/10.1126/science.adq1814

Dathathri, S., See, A., Ghaisas, S., Huang, P. S., McAdam, R., Welbl, J., Bachani, V., Kaskasoli, A., Stanforth, R., Matejovicova, T., Hayes, J., Vyas, N., Al Merey, M., Brown-Cohen, J., Bunel, R., Balle, B., Cemgil, T., Ahmed, Z., Stacpoole, K., … Kohli, P. (2024). Scalable watermarking for identifying large language model outputs. *Nature*, *634*(8035), 818–823. https://doi.org/10.1038/s41586-024-08025-4

DellaPosta, D. (2020). Pluralistic collapse: The 'oil spill' model of mass opinion polarization. *American Sociological Review*, *85*(3), 507–536. https://doi.org/10.1177/000312242092298

DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American's social attitudes become more polarized? *American Journal of Sociology*, *102*(3), 690–755. https://doi.org/10.1086/230995

Doshi, A. R., & Hauser, O. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, *10*(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290

Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, *630*(8017), 625–630. https://doi.org/10.1038/s41586-024-07421-0

Feuerriegel, S., DiResta, R., Goldstein, J. A., Kumar, S., Lorenz-Spreen, P., Tomz, M., & Pröllochs, N. (2023). Research can help to tackle AI-generated disinformation. *Nature Human Behaviour*, *7*(11), 1818–1821. https://doi.org/10.1038/s41562-023-01726-2

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., & Druckman, J. N. (2020). Political sectarianism in America. *Science*, *370*(6516), 533–536. https://doi.org/10.1126/science.abe1715

Gallegos, I. O., Shani, C., Shi, W., Bianchi, F., Gainsburg, I., Jurafsky, D., & Willer, R. (2025). Labeling messages as AI-Generated does not reduce their persuasive effects. *arXiv*. https://doi.org/10.48550/arXiv.2504.09865

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Gilardi, F., Di Lorenzo, S., Ezzaini, J., Santa, B., Streiff, B., Zurfluh, E., & Hoes, E. (2024). Disclosure of AI-generated news increases engagement but does not reduce aversion, despite positive quality ratings. *arXiv*. https://arxiv.org/pdf/2409.03500

Gillespie, T. (2024). Generative AI and the politics of visibility. *Big Data & Society*, *11*(2), 20539517241252131. https://doi.org/10.1177/20539517241252131

Greenstein, S., & Zhu, F. (2018). Do experts or crowd-based models produce more bias? Evidence from Encyclopedia Britannica and Wikipedia. *MIS Quarterly*, *42*(3), 945–958. https://doi.org/10.25300/misq/2018/14084

Hackenburg, K., & Margetts, H. (2024). Evaluating the persuasive influence of political microtargeting with large language models. *Proceedings of the National Academy of Sciences*, *121*(24), Article e2403116121. https://doi.org/10.1073/pnas.2403116121

Hagen, L., Jingnan, H., & Nguyen, A. (2025, July 9). *Elon Musk's AI chatbot, Grok, started calling itself "MechaHitler."* NPR. https://www.npr.org/2025/07/09/nx-s1-5462609/grok-elon-musk-antisemitic-racist-content

Hällsten, M., & Pfeffer, F. T. (2017). Grand advantage: Family wealth and grandchildren's educational achievement in Sweden. *American Sociological Review*, *82*(2), 328–360. https://doi.org/10.1177/0003122417695791

Hargittai, E., & Micheli, M. (2019). Internet skills and why they matter. In Mark Graham & William H. Dutton (Eds.), *Society & the Internet: How networks of information and communication are changing our lives* (pp. 109–124). Oxford University Press.

Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). AI generates covertly racist decisions about people based on their dialect. *Nature*, *633*(8028), 147–154. https://doi.org/10.1038/s41586-024-07856-5

Joyce, K., & Cruz, T. M. (2024). A sociology of artificial intelligence: Inequalities, power, and data justice. Socius: Sociological Research for a Dynamic World, 10, Article 23780231241275393. https://doi.org/10.1177/23780231241275393

Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S. G., Noble, S. U., & Shestakofsky, B. (2021). Toward a sociology of artificial intelligence: A call for research on inequalities and structural change. *Socius: Sociological Research for a Dynamic World*, *7*, Article 2378023121999581. https://doi.org/10.1177/2378023121999581

Karell, D., Sachs, J., & Barrett, R. (2025). Synthetic duality: A framework for analyzing generative artificial intelligence's representation of social reality. *Poetics*, *108*, 101966. https://doi.org/10.1016/j.poetic.2024.101966

Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, *9*(1), 104–117. https://doi.org/10.1017/xps.2020.37

Laba, N. (2024). Engine for the imagination? Visual generative media and the issue of representation. *Media, Culture & Society*, *46*(8), 1599–1620. https://doi.org/10.1177/01634437241259950

Liao, W., Weisman, W., & Thakur, A. (2024). On the motivations to seek information from artificial intelligence agents versus humans: A risk information seeking and processing perspective. *Science Communication*, *46*(4), 458–486. https://doi.org/10.1177/10755470241232993

Ma, N., & Zhong, Z. (2025). A meta-analysis of the impact of generative artificial intelligence on learning outcomes. *Journal of Computer Assisted Learning*, *41*(5), Article e70117. https://doi.org/10.1111/jcal.70117

Makhortykh, M., Vziatysheva, V., & Sydorova, M. (2023). Generative AI and contestation and instrumentalization of memory about the Holocaust in Ukraine. *Eastern European Holocaust Studies*, *1*(2), 349–355. https://doi.org/10.1515/eehs-2023-0054

McClain, C., Kennedy, B., Gottfried, J., Anderson, M., & Pasquini, G. (2025). *How the U.S. public and AI experts view artificial intelligence*. Pew Research Center.

McCosker, A. (2024). Making sense of deepfakes: Socializing AI and building data literacy on GitHub and YouTube. *New Media & Society*, *26*(5), 2786–2803. https://doi.org/10.1177/14614448221093943

McKee, K. R., Tacchetti, A., Bakker, M. A., Balaguer, J., Campbell-Gillingham, L., Everett, R., & Botvinick, M. (2023). Scaffolding cooperation in human groups with deep reinforcement learning. *Nature Human Behaviour*, *7*(10), 1769–1787. https://doi.org/10.1038/s41562-023-01686-7

Melumad, S., & Yun, J. H. (2025). Experimental evidence of the effects of large language models versus web search on depth of learning. *PNAS Nexus*, *4*(10), pgaf316. https://doi.org/10.1093/pnasnexus/pgaf316

Miller, M. (2025, October 17). *New user trends on wikipedia*. Diff. https://diff.wikimedia.org/2025/10/17/new-user-trends-on-wikipedia/

Milmo, D., & Robins-Early, N. (2024, May 14). Google rolls out AI-generated, summarized search results in US. *The Guardian*. https://www.theguardian.com/technology/article/2024/may/14/google–ai–search–results

Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, *62*(3), 760–775. https://doi.org/10.1111/ajps.12357

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(8), Article e2120481119. https://doi.org/10.1073/pnas.2120481119

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

Ouyang, L., Wu, J., Jiang, X., Almedia, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christian, P., Leike, J., … Lowe, R. (2022). Training language models to follow instructions with human feedback. *arXiv*. https://arxiv.org/abs/2203.02155

Padmakumar, V., & He, H. (2023). Does writing with language models reduce content diversity? *arXiv*, 2309.05196v1.

Palmer, A. K., & Spirling, A. (2023). Large Language Models Can Argue in Convincing Ways About Politics, But Humans Dislike AI Authors: Implications for Governance. *Political Science*, *75*(3), 281–291. DOI:10.1080/00323187.2024.2335471

Spitale, G., Biller-Andorno, N., & Germani, F. (2023). AI model GPT-3 (dis)informs us better than humans. *Science Advances*, *9*(26), eadh1850. https://doi.org/10.1126/sciadv.adh1850

Steinert, C. V., & Kazenwadel, D. (2024). How user language affects conflict fatality estimates in ChatGPT. *Journal of Peace Research*, *62*(4), 1128–1143. https://doi.org/10.1177/00223433241279381

Steinsson, S. (2024). Rule ambiguity, institutional clashes, and population loss: How Wikipedia became the last good place on the internet. *American Political Science Review Review*, *118*(1), 235–251. https://doi.org/10.1017/s0003055423000138

The White House. (2025, July 23). Preventing woke AI in the federal government. https://www.whitehouse.gov/presidential-actions/2025/07/preventing-woke-ai-in-the-federal-government/

Thompson, S. A., Mondria Terol, T., Conger, K., & Freedman, D. (2025, September 2). How elon musk is remaking grok in his image. *The New York Times*. https://www.nytimes.com/2025/09/02/technology/elon-musk-grok-conservative-chatbot.html

Ueshima, A., Jones, M. I., & Christakis, N. A. (2024). Simple autonomous agents can enhance creative semantic discovery by human groups. *Nature Communications*, *15*(1), 5212. https://doi.org/10.1038/s41467-024-49528-y

Wasik, B. (2025, June 16). A.I. is poised to rewrite history. Literally. *The New York Times Magazine*.

Yan, L., Greiff, S., Teuber, Z., & Grasević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, *8*(10), 1839–1850. https://doi.org/10.1038/s41562-024-02004-5

Zajko, M. (2022). Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*, *16*(3), Article e12962. https://doi.org/10.1111/soc4.12962

Zhang, Y., & Gosline, R. (2023). Human favoritism, not AI aversion: People's perceptions (and bias) toward generative AI, human experts, and human–GAI collaboration in persuasive content generation. *Judgment and Decision Making*, *18*, Article e41. https://doi.org/10.1017/jdm.2023.37

Zhu, Y., Liu, Q., & Zhao, L. (2025). Exploring the impact of generative artificial intelligence on students' learning outcomes: A meta-analysis. *Education and Information Technologies Technologies*, *30*(11), 16211–16239. https://doi.org/10.1007/s10639-025-13420-z

Author Biographies

Daniel Karell is an Assistant Professor of Sociology at Yale University and a Faculty Affiliate of Yale's Institution for Social and Policy Studies and Institute for Foundations of Data Science. He studies the intersection of culture, politics, and social movements using compuational and quantitative methods.

Matthew Shu is a B.S/M.A graduate in Statistics & Data Science from Yale University and works in applied machine learning at Brain Co. His research explores how artificial intelligence tools influence learning, social norms, and inequality.

Thomas Davidson is an Assistant Professor of Sociology at Rutgers University–New Brunswick. He studies online activism and discourse using a range of computational and statistical techniques.

Keitaro Okura is a Ph.D. candidate in the Department of Sociology at Yale University. His research examines the symbolic boundaries of national and ethnoracial group membership and their social consequences in the United States, drawing primarily on survey data and experiments.