Title
Homogenized Memory: Impacts of Digital Systems on Word Recollection

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Sims, Gabriel
Oppenheimer, Daniel

Publication Date
2026

Homogenized Memory: Impacts of Digital Systems on Word Recollection

Gabriel Sims (gjsims@andrew.cmu.edu)[1] & Daniel Oppenheimer (doppenh1@andrew.cmu.edu)[2]
[1]Department of Social and Decision Sciences, Carnegie Mellon University
[2]Department of Social and Decision Sciences, Carnegie Mellon University

Abstract

The tools we use change the way we think. Thus, the proliferation of generative AI underscores the importance of research on the cognitive implications of increasingly ubiquitous AI use. Collaboration with generative AI and other digital systems increases productivity. However, previous research has shown that AI exerts a homogenizing influence on users. This study extends the investigation of homogenization effects of digital collaboration to memory. We administer a memory task and ask participants to recall words while collaborating with generative AI, internet searches, or alone. We find that collaborating with digital systems homogenizes memory content towards words with greater lexical and category-dependent frequencies. Although the effect size is small, the repeated usage of digital systems compounds homogenization. Unchecked memory homogenization risks collective vulnerability, and has implications for recollection of past, present, and future experiences.

Keywords: Generative AI; Memory; Homogenization; LLMs; Algorithmic Monoculture

Introduction

The adoption rate of generative artificial intelligence is 2.25 times that of the personal computer, with 45.5% of working adults reporting some usage and 10.4% reporting daily usage (Bick et al., 2025). AI can improve productivity across industries, and mitigates skill gap inequality between workers (Brynjolfsson et al., 2023; Peng et al., 2023; Noy & Zhang, 2023; Kanazawa et al., 2022; but see Otis et al., 2024 and Argawal et al., 2023 for instances where the opposite trends have been observed).

However, digital technology adoption risks algorithmic monoculture, where working with the same systems homogenizes users. For example, Oppenheimer and Patterson (2025) had participants engage in the Alternative Uses Task (Guilford, 1957) and showed that while internet use could improve the number of ideas individuals generated, nominal groups with internet access generated fewer total ideas (than those without internet access) because the internet users all came up with the same ideas, while non-internet users generated distinct ideas (for similar results, see Kumar et al., 2025). Other research has shown that access to AI can also homogenize the themes and plots of stories (Doshi & Hauser, 2024) and reduce the creativity of composers writing songs (Oros et al, under review). Sharing model parameters leads to greater outcome homogenization (Bommasani et al., 2022) and risks cultural shifts, since users adopt the writing style of the model both lexically and semantically (Agarwal et al., 2024). A monocultural system where users or firms collaborate with the same algorithms risks harm from unexpected, correlated shocks, and may also result in lower average quality absent shocks (Kleinberg et al. 2021).

Generative AI engages with multiple cognitive processes (Bhat et al., 2023) and the AI ideas persist in the post-task opinions of users without their awareness (Jakesch et al., 2023). Homogenized outputs can be attributed to direct semantic incorporation (Roemmele 2021; Arnold et al., 2020) or to anchoring on ideas of AI (Doshi & Hauser, 2024). Homogenization applies to digital systems more generally, and can be observed in exposure to ideologically cross-cutting content in Facebook networks (Bakshy et al., 2015) and while using internet searches to write (Kosmyna et al., 2025).

Although generative AI and digital systems interact with many malleable cognitive processes, currently the homogenization literature concentrates on ideation or composition tasks (Agarwal et al., 2024, Arnold et al, 2020; Jakesch et al., 2023; Roemmele 2021; Doshi & Hauser, 2024; Padmakumar & He, 2024; Kosmyna et al., 2025; Oppenheimer & Patterson, 2025; Oros et al., under review). To date, there is little literature on how AI use might influence memory. But decades of research have shown that memory is reconstructive (Loftus, 1975) and can be altered via priming or conversations (Wright et al., 2020), suggesting that AI use could influence recall. While there is a literature on offloading (Eliseev & Marsh, 2021) and on how access to digital tools can change people's encoding strategies (Sparrow et al., 2011), the literature does not address the content of what is recalled as the content that people are exposed to changes due using digital tools.

To address this gap, we propose that the content of what a user remembers may be shaped by the digital systems that they use. We perform a basic memory experiment, asking users to recall words within categories while either collaborating with the LLM ChatGPT, having unfettered access to the internet, or being prevented from using the internet. We expect digital tools to favor words with greater lexical and category-dependent frequencies. Words with high lexical frequency are common within the lexicon; because LLMs operate by producing the most likely words conditional upon the context, the words LLMs provide users will likely be highly sampled from words that are more common, leading to memory distortions toward common words.. Homogenization by

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

category-dependent frequency would present as a reduction in the diversity of ideas within a specific topic. Words with high category-dependent frequency are typical exemplars of the category and words with low category-dependent frequency are atypical exemplars of the category. Together, lexical and category-dependent frequencies capture how standard words are; as standard ideas are recalled more and unique ideas are recalled less, memory becomes homogenous.

As a consequence, we predict users with access to digital tools should be more successful at recalling common and typical words (which will be frequently produced by digital tools), but less successful in recalling uncommon or atypical words (which will be rarely produced by digital tools), relative to users without access to digital tools. As a result, users with access to digital tools should have less variance in which words they remember than users without access to digital tools. We hypothesize that memory homogenization will occur irrespective of any productivity gains for the individual user.

Methods

Participants

149 adult, English-speaking participants were recruited from Prolific and compensated at Pennsylvania minimum wage for their time (approximately $1.80). All procedures were approved by the ethics committee of our university (protocol ID: STUDY2025_00000191) .

Stimuli Selection

A total of 96 words from 8 categories (12 words per category) were selected based on their commonality and typicality in a balanced 2x2 factor design. The categories were: an ocean animal, a carpenter's tool, a crime, a fruit, a kitchen utensil, a metal, a part of the human body, and family member. Commonality was operationalized by word frequency using a corpus of 51 million words (Brysbaert & New, 2009). Typicality was operationalized based on typicality ratings from Castro et al. (2020): participants were given a category name and asked to generate exemplars of that category, with more frequently generated exemplars being defined as more typical. Stimuli were selected such that all high typicality words were significantly more typical than all low typicality words $t(44) = -10.75, p < .05$), and all common words were significantly more common than all uncommon words ($t(47) = 4.28, p < .05$).[1]

Experimental Procedure

Participants were presented with all 96 stimuli on one screen and given 5 minutes to memorize them. Stimuli were presented in the form "**category name**: word 1, word 2...", with category presentation order and within-category word order both randomized. Next, as a distraction task, 10 images of insects were shown and participants were asked to name them.

Then, participants were assigned to one of three randomized conditions. The control condition instructed participants to recall words without consulting the internet or LLMs. The LLM condition instructed participants to navigate to the website ChatGPT.com, and to look up anything related to the words or categories that helps them remember the stimuli. The internet search condition instructed participants to use the internet and look up anything related to the words or category that helps them remember. Participants in the internet condition were not restricted from using generative AI to promote naturalistic behavior as AI is increasingly integrated with search engines and we cannot eliminate passive influence. The study was incentive compatible: participants could receive a bonus of up to 10 cents, one cent for each of 10 randomly selected words that the participant correctly recalled.

Finally, as a manipulation check, participants were asked to self-report if they used an internet search, AI tool like chatgpt, or other web page and were given room to provide additional comments about usage or the study.

Pilot Study

A 2-stage pilot study was run before final analysis. The first stage ensured that participants followed instructions regarding whether or not to use digital tools (and which tools). We gauged this with the manipulation check (asking if they used an internet search, AI tool like chatgpt, or other web page) for groups of 50 participants until more than self-reported adhering to the appropriate randomized condition. The second stage tested our research paradigm and used the data of 149 participants, 50 of which came from the last iteration of the first stage of the pilot. 85.6% (126/149) of the participants faithfully followed the instructions about digital tool use.

Analysis

This study, exclusion criteria, and analysis strategy were preregistered (https://aspredicted.org/uh576d.pdf).

Data Cleaning and Participant Exclusion

Responses were spellchecked and reformatted, with 84 spelling errors corrected and extraneous words removed. Spelling errors and incorrect pluralization were corrected blind to condition. As preregistered, participants who correctly recalled stimuli more than 3 standard deviations from the mean were removed (M=27.21, SD=19.04). This excluded 3 participants, resulting in final data from 146 participants.

Dependent Variables

We measure productivity in terms of memory performance. At the subject level we measured counts of correctly recalled words and incorrectly recalled false alarms. We calculate a false alarm rate as the proportion of false alarms to correctly recalled words; 7 participants who did not correctly recall any words were excluded, as their ratios are undefined.

[1] One category, ocean animals, was not studied by Castro et al., (2020) but was included based on internal pilot studies done in the lab.

We primarily measure homogenization using the commonality and typicality of stimuli. At the stimuli level, we calculate the proportion of participants in each condition that correctly recalled each word. This captures whether standard or unique ideas are preferentially remembered under different conditions. Disproportionate changes in recall probabilities across word types indicate a collective shift in the similarity of ideas retained within a condition. To test homogenization, we ran an ANOVA on a linear model with commonality, typicality, and condition as fixed factors that are interacted, with stimuli as a random factor.[2] As preregistered, our main analysis groups digital tools together and compares words that are both common and typical to words that are combinations of uncommon or atypical. Grouping digital tools and frequency measures tests a generalized account of memory homogenization; we expect both tools and frequency measures to behave consistently.

Results

Productivity

We first explored whether use of digital tools improved memory performance; did access to AI or the internet increase (or decrease) the number of words recalled.

At the participant level, the LLM condition recalled fewer words (M=23.04, SD=17.52) than the internet (M=28.06, SD=19.92) and control conditions (M=27.06, SD=12.13) but an ANOVA test showed no significant difference *($F(2, 146) = 1.64$, $p > .05$)*. Similarly, the false alarm rate was higher for the LLM condition (M=.12, SD=.21) than the internet (M=.06, SD=.09 and control conditions (M=.07, SD=.16) but was not significant *($F(2,136) = 1.70$, $p > .05$)*. In sum, access to digital tools neither appreciably helped nor harmed participant recall (Figure 1), regardless of whether false alarms were taken into consideration (Figure 2). However, statistical power constrains findings at the participant level (see limitations section for more detail).

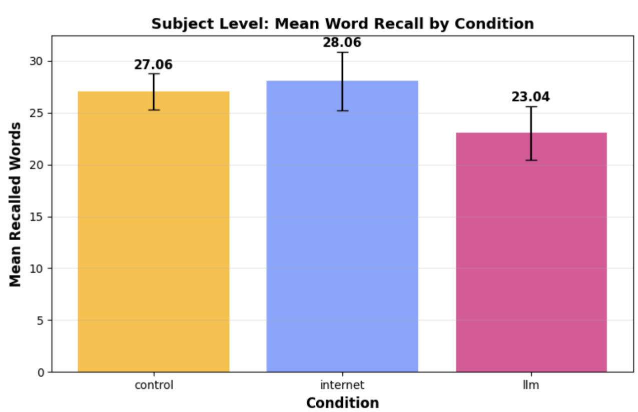

Figure 1: Correct recall by condition

[2]The p-values from a mixed between-within subject ANOVA are equivalent to this design.

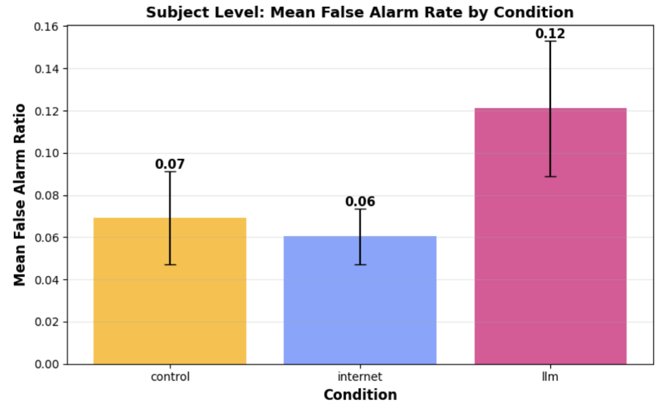

Figure 2: False alarm rate by condition.

Homogenization

Across conditions, common words (M=.32, SD=.10) were recalled at greater rates than uncommon words (M=.22, SD=.08) and typical words (M=.31, SD=.11) were recalled at greater rates than atypical words (M= .23, SD=.09). The main effects of commonality *($F(1, 92) = 57.63$, $p < .05$)* and typicality *($F(1, 92) = 30.74$, $p < .05$)* are both significant. This informally validates both metrics; it makes intuitive sense and aligns with decades of literature on memory retrieval that lexically uncommon words and atypical category exemplars are remembered less well.

We see significant interaction effects between condition and commonality *($F(2, 184) = 3.17$, $p < .05$)* and between condition and typicality *($F(2, 184) = 3.66$, $p < .05$)*.

We hypothesized that users with access to digital tools should be more successful at recalling common and typical words but less successful at recalling combinations of uncommon or atypical words relative to users without access. To explore this, we compared common and typical words against all other cells in our 2x2 design. For common and typical words, participants with access to digital tools (M=.37, SD=.10) and participants without access (M=.37, SD=.09) did not recall at differential rates *($t(184) = .646$, $p > .05$)*. However, participants who collaborated with digital tools (M=.23, SD=.08), compared to those without (M=.25, SD=.09), recalled combinations of uncommon or atypical words at a lower rate *($t(184) = 5.51$, $p < .05$)*. These findings support our hypothesis that people with access to digital tools disproportionately remember fewer uncommon or atypical words. The estimated difference, .018, is small, but reliable.

However, this pre-registered analysis obscures important nuances in the drivers of these patterns. In particular, memory looks different for LLM use as compared to general internet use, and memory for uncommon words looks different than memory for atypical words.

Exploratory Analysis: Differences by Frequency Measure and Digital Tool

LLM Condition Statistical power constrains findings at the participant level (see limitations section for more detail), but a stimuli level analysis can provide insight on recall by condition. The proportion of participants who recalled each word was lower for the LLM condition (M=.24, SD=.09) than the control (M=.28, SD=.10) and internet (M=.29, SD=.11) conditions. The differences from participants without digital access *($t(184) = 5.88$, $p < .05$)* and from participants with unfettered internet access *($t(184) = 7.35$, $p < .05$)* are significant.[3]. Because we preregistered participant level analyses to understand productivity, we view this significant productivity decrease for LLM users at the stimulus level as suggestive rather than definitive.

Typicality For typical words, access to the internet led to improved recall, relative to participants who did not have internet access (Figure 2). However, for atypical words this trend was reversed, with access to the internet leading to impaired recall for atypical words. This means that the difference in recall rate between typical and atypical words is larger in the internet condition (typical: M=.34, SD=.11; atypical: M=.24, SD=.09) that the control (typical: M=.32, SD=.11; atypical: M=.25, SD=.09) or LLM conditions (typical: M=.27, SD=.09; atypical: M=.21, SD=.08); the pattern is significant and suggests homogenization *($t(184) = -2.31$, $p < .05$)*.

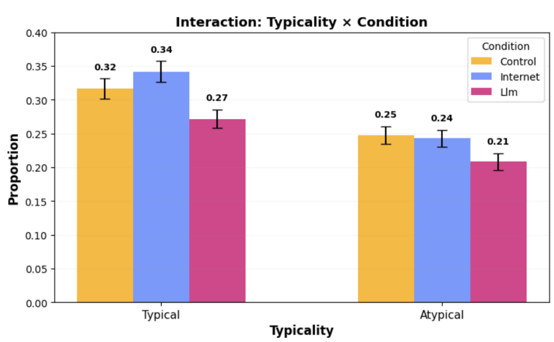

Figure 3: Recall by word typicality and internet use.

Commonality In contrast to typicality, commonality did not differentially yield different recall between non-internet users and unfettered internet users (Figure 3).

Meanwhile, as with typicality, LLM use impaired memory performance regardless of the commonality of the words, although this was particularly pronounced with common words. The difference in recall rate between common and uncommon words is smaller in the LLM condition (common: M=.28, SD=.10; uncommon: M=.20, SD=.06) than the control (common: M=.34, SD=.09; uncommon: M=.22, SD=.08) or internet conditions (common: M=.35, SD=.10; uncommon: M=.24, SD=.09). On commonality, the gap between participants without digital access and with an LLM is significant, suggesting LLMs may partially ameliorate natural differences in uncommon word retrieval *($t(184) = 2.48$, $p < .05$)*.[4]

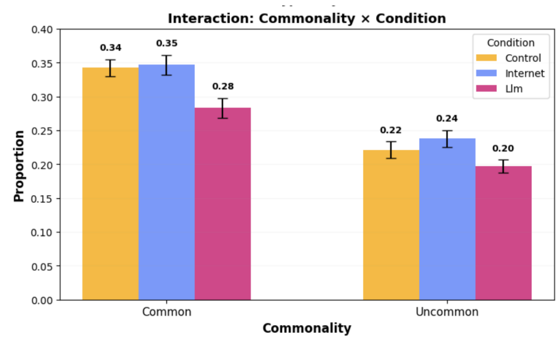

Figure 4: Recall by word commonality and internet use.

Exploratory Analysis: Heterogenized False Alarms

We create nominal groups for each condition by randomly drawing 1000 samples of group size k (without replacement), with k ranging from 1 to 47. For each sample we find the count of unique false alarms (recalled words not in the provided list) between participants and calculate the average for each group size k. A nominal group analysis captures change for the collective as group size increases. As described, subject level analyses of efficiency show no differences of false alarm rate by condition. However, a nominal group analysis shows that as group size increases, the number of unique false alarms diverges between users with digital access and without digital access (Figure 4). This suggests collaborating with digital systems may not influence errors for the individual but increase error variance for the collective. This exploratorily suggests that digital systems may homogenize accurate answers but heterogenize incorrect answers.

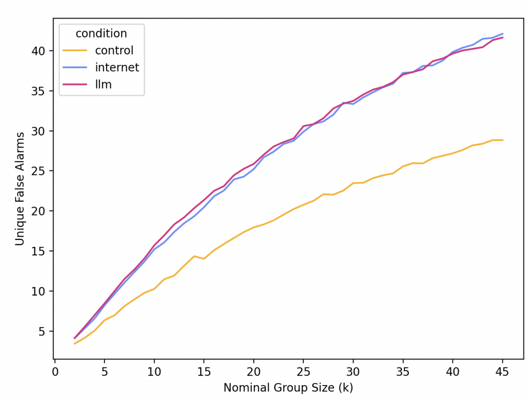

Figure 5: Unique false alarms by nominal groups and condition.

[3]The exploratory t-tests for conditions comparisons were computed for all pairs using Tukey's procedure as they were not preregistered

[4]The exploratory contrast null hypotheses for frequency measures are Holm-Bonferroni corrected as they were not preregistered.

Discussion

Memory Homogenization

We investigated whether access to digital tools homogenizes memory toward common and typical stimuli. We asked participants to memorize 96 words from 8 categories and then had them recall the words while collaborating with the LLM ChatGPT, with unfettered access to the internet, or without a digital tool. Participants with access to a digital tool recalled uncommon or atypical words at a lower rate than participants without access to a digital tool. This supports our hypothesis that access to digital systems homogenizes memory towards stimuli with greater lexical and category-dependent frequencies.

However, deeper exploration into the data suggests that this may be qualified by both type of digital system used, and by different statistical features of the words. Notably, while using the internet seems to aid recall of typical words, it inhibits recall of atypical words. At the same time, using the internet does not differentially affect recall as a function of the commonness of the target word.

Meanwhile, pure LLM use may lead to deficits in memory regardless of typicality or commonality of words, with this effect somewhat attenuated for uncommon words. Different digital systems and measures of frequency may contribute to and reduce the effects of homogenized memory.

Descriptive Mechanisms

Descriptive Mechanisms Related work in our lab followed up on this experiment using participants from a university pool (n=29) while recording participant screens to understand underlying mechanisms. This should be taken suggestively as data was analyzed after initial CogSci submission. As preregistered and hypothesized, the primary search strategy for users with digital access was looking up category names, which displayed typical *($F(1, 184) = 43.10$, $p < .05$)* and common *($F(1, 184) = 4.70$, $p < .05$)* words at differentially higher rates. Interestingly, although typical words are differentially displayed relative to atypical words in both LLM (typical: M=.72, SD=.28; atypical: M=.39, SD=.29) and internet (typical: M=.56, SD=.30; atypical: M=.37, SD=.25) conditions, there is an interaction between commonality and condition *($F(1, 184) = 3.91$, $p < .05$)*, and common words are not differentially displayed relative to uncommon words for LLM users (common: M=.56, SD=.32; uncommon: M=.55, SD=.34), only for unfettered internet users (common: M=.54, SD=.30; uncommon: M=.38, SD=.27); this may explain the exploratory suggestion that LLMs may partially ameliorate natural differences in uncommon word retrieval. Both the display percentage of a word (β = .26, SE=.03, p <.05) and the interaction of display percentage with total searches for that category (β = .02, SE=.01, p <.05) were positively associated with increased recall. This suggests direct incorporation of observed stimuli drives homogenization.

Transcription or Cognitive Change?

This study only observes the list of words provided by the user, thereby retaining the possibility that the list is homogenized but all that occurs cognitively is recognition and transcription. Philosophers of collective memory believe memory is shaped by external and internal factors (Clark & Chalmers, 1998; Wilson, 2005), and find it can be influenced through dyadic interactions (Wright et al., 2000). Digital system interaction is dyadic between a user and the system, especially for generative AI that produces conversation-like text. The effects of forgotten memories from dyadic interaction could be attributed to retrieval induced forgetting (RIF), where inducing the recollection of category exemplars decreases subsequent recollection of non-recalled exemplars of the same category (Anderson et al., 1994). Although we find digital systems homogenize output, we don't find that they increase correct recollection (and exploratory findings show decreased correct recollection for LLM users). Recognition and transcription without cognitive changes for participants should increase total correct recall, which is not the case. We reason that recall may be enhanced for displayed words but an RIF-like effect may simultaneously occur cognitively.

Limitations

Experimental Context In our study, participants were all English speakers, so we cannot infer that our findings generalize to different populations of digital system users. Our study included one shot learning for 5 minutes with a brief distractor task and took a median 12 minutes to complete; we don't know if these effects would hold with deeper encoding or weaker memory traces: Prolific users participate in research at their own convenience, and we cannot control for surrounding environments, device type, or a bias in the Prolific pool. Moreover, ChatGPT was the only LLM used, but results may differ across LLM companies and models. LLMs are relatively new, and users may also adapt in their use of tools as best practices become more mainstream. These factors must be considered when contextualizing our findings to other settings.

Statistical Power As discussed, statistical power constrains our hypothesis on participant productivity. Our study was underpowered for participant level analysis at .36 (although item level analysis was appropriately powered at .97). Results, especially at the participant level, should be viewed with this caveat in mind.

Future Work

Baseline Memory Performance We ignore baseline memory performance, but controlling for baseline performance may add nuance to our findings. For example, it may be that these effects are particularly pernicious for people who have weaker memories, as they more fully offload the memory task to AI. Alternatively, it could be that people with weaker memories already focus on common and typical words, while individuals with stronger memories can recall a broader range of stimuli – unless the use of AI fixates them on certain words. Thus, it is possible that digital tool use differentially impairs individuals who otherwise would perform quite well. This is an exciting avenue for future research.

Collective Memory These findings may be relevant to research on collective memory. We reason retrieval induced forgetting (RIF), where inducing the recollection of category exemplars decreases subsequent recollection of non-recalled exemplars of the same category (Anderson et al., 1994) may be occurring, which should be explored by future research. This area of literature suggests that dominant narrators and individuals perceived to have expertise can implant false memories in listeners, in both dyadic and non-dyad settings (Brown et al., 2009). Beyond single-word stimuli, RIF has been extended to social interactions (Cuc et al., 2007) and conversations about collectively remembered events like 9/11 (Coman et al., 2009). That literature suggests that actively retrieved memories in social interactions shape the content of memory for the collective. Our findings on homogenized memory suggest that digital system use could shape the content of memory for the collective. Can memory homogenization by digital systems apply to non single-word stimuli or collectively remembered events? If so, there could be significant social implications for the homogenization of memory through digital tool use that is worth exploring in future studies.

Summary

Our findings of homogenization of memory by users of digital systems is both novel and significant. Even with a small effect size, it suggests that repeated memory tasks with digital systems leads to more similar thought. As generative AI proliferates and digital systems generally become integrated in society the risks of algorithmic monoculture increase. Left unchecked, we may risk our collective memory of more diverse experiences while developing correlated vulnerabilities more generally.

Acknowledgments

We would like to acknowledge that components of some figures were constructed in collaboration with CoPilot's plug-in for VS Code. We would also like to thank Professor Eli Ben-Michael for informative conversations on appropriate statistical analyses. Lastly, we wish to acknowledge that this is made possible by the Carnegie Mellon Small Undergraduate Research Grants.

References

Agarwal, D., Naaman, M., & Vashistha, A. (2025). Ai suggestions homogenize writing toward western styles and diminish cultural nuances. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–21. https://doi.org/10.1145/3706598.3713564

Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. (2023). *Combining human experience with artificial intelligence: Experimental evidence from radiology*. Massachusetts Institute of Technology.

Anderson, B., Shah, J., & Kreminski, M. (2024). Homogenization effects of large language models on human creative ideation. *Creativity and Cognition*. https://doi.org/10.1145/3635636.3656204

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1063–1087.

Arnold, K., Chauncey, K., & Gajos, K. (2020). Predictable text encourages predictable writing. *Proceedings of the 25th International Conference on Intelligent User Interfaces*.

Bakshy, E., Messing, S., & Adamic, L. (2015). Exposure to ideologically diverse news and opinion on facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Bhat, A., Agashe, S., Mohile, N., Oberoi, P., Jangir, R., & Joshi, A. N. (2022). Interacting with next-phrase suggestions: How suggestion systems aid and influence the cognitive processes of writing. *Proceedings of the 28th International Conference on Intelligent User Interfaces*.

Bick, A., Blandin, A., & Deming, D. J. (2025). *The rapid adoption of generative ai* (No. 2024-027). Federal Reserve Bank of St. Louis. https://doi.org/10.20955/wp.2024.027

Bommasani, R., Creel, K., Kumar, A., Jurafsky, D., & Liang, P. (2022). Picking the same person: Does algorithmic monoculture lead to outcome homogenization? *Proceedings of the 36th International Conference on Neural Information Processing Systems*.

Brown, A., Coman, A., & Hirst, W. (2009). Expertise and the formation of collective memory. *Social Psychology*, *40*, 118–128.

Brynjolfsson, E., Li, D., & Raymond, L. (2023). *Generative ai at work* (No. 31161). National Bureau of Economic Research.

Brysbaert, M., & New, B. (2009). Moving beyond kučera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure. *Behavior Research Methods*.

Castro, N., Curley, T., & Hertzog, C. (2021). Category norms with a cross-sectional sample of adults in the united states. *Behavior Research Methods*, *53*, 898–917. https://doi.org/10.3758/s13428-020-01454-9

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 10–23.

Coman, A., Brown, A., Koppel, J., & Hirst, W. (2009). Collective memory from a psychological perspective. *International Journal of Politics, Culture, and Society*, *22*(2), 125–141.

Coman, A., Manier, D., & Hirst, W. (2009). Forgetting the unforgettable: Socially shared retrieval-induced forgetting of september 11 memories. *Psychological Science*.

Cue, A., Koppel, J., & Hirst, W. (2007). Silence is not golden: A case for socially shared retrieval-induced forgetting. *Psychological Science*, *18*, 727–737.

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

Eliseev, D., & Marsh, E. (2021). Externalizing autobiographical memories in the digital age. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2021.08.005

Guilford, J. P. (1957). Creative abilities in the arts. *Psychological Review*, *64*(2), 110–118. https://doi.org/10.1037/h0048280

Hancock, J., Naaman, M., & Levy, K. (2020). Ai-mediated communication. *Journal of Computer-Mediated Communication*, *25*(1), 89–100. https://doi.org/10.1093/jcmc/zmz022

Jakesch, M., Bhat, A., Buschek, D., Zalmanson, L., & Naaman, M. (2023). Co-writing with opinionated language models affects users' views. *Proceedings of the CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3544548.3581196

Kanazawa, K., Kawaguchi, D., Shigeoka, H., & Watanabe, Y. (2022). *Ai, skill, and productivity: The case of taxi drivers* (No. 30612). National Bureau of Economic Research.

Kleinberg, J., & Raghavan, M. (2021). Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, *118*(22). https://doi.org/10.1073/pnas.2018340118

Kosmyna, N., Hauptmann, E., Yuan, Y., Situ, J., Liao, X., Beresnitzky, A., Braunstein, I., & Maes, P. (2025). Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing. *arXiv*.

Kumar, H., Vincentius, J., Jordan, E., & Anderson, A. (2025). Human creativity in the age of LLMs: Randomized experiments on divergent and convergent thinking. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3706598.3714198

Loftus, E. F. (1975). Reconstructing memory: The incredible eyewitness. *Jurimetrics Journal*, *15*(3), 188–193.

Noy, S., & Zhang, W. (2023). *Experimental evidence on the productivity effects of generative artificial intelligence*. Massachusetts Institute of Technology.

Oppenheimer, D. M., & Patterson, M. T. (2025). Thinking outside the box means thinking outside the search engine. *Memory & Cognition*, *53*, 2686–2699. https://doi.org/10.3758/s13421-025-01732-x

Oros, J., Randall, R., & Telang, R. (n.d.). *Help that hurts: The creativity cost of LLM ideation in music composition* [Manuscript under review].

Otis, N., Clarke, R., Delecourt, S., Holtz, D., & Konin, R. (2024). *The uneven impact of generative ai on entrepreneurial performance*. Harvard Business School.

Padmakumar, V., & He, H. (2024). Does writing with language models reduce content diversity? *Proceedings of the International Conference on Learning Representations*.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, *333*, 776–778. https://doi.org/10.1126/science.1207745

Wilson, R. A. (2005). Collective memory, group minds, and the extended mind thesis. *Cognitive Processing*, *6*, 227–236.

Wright, D. B., Self, G., & Justice, C. (2000). Memory conformity. *British Journal of Psychology*, *91*, 189–202.