Will Gemini Make Us Stupid?

Benjamin Lira[a,1]

[a]University of Pennsylvania

Generative AI (GenAI) applications—especially those based around large language models like Gemini and ChatGPT—are now ubiquitous, and their adoption rate is increasing. They are potentially the most powerful piece of technology ever developed, matching or surpassing human performance in many complex tasks. As AI becomes embedded in our email clients, word processors, and smartphones, it becomes critically important to understand its effects on how we think. I plan to delineate how AI influences our cognitive processing, specifically by encouraging System 1 thinking, which is fast, associative, automatic, and fluent, vs. System 2 thinking, which is slower, resource-intensive, and deliberative. I propose a series of experiments in which participants are randomly assigned to engage with an AI-chatbot or traditional web search in a research task, and evaluate the resulting changes in cognition during the task. This research will (1) establish whether the ease of processing AI responses prompts System 1 thinking, or if, alternatively, these interactions free up cognitive resources, allowing for increased System 2 processing; (2) identify the key mediators that drive this effect; and (3) develop manipulations, targeting user interaction and/or interface design, that encourage more vs. less System 2 thinking when engaging with AI. This research elucidate how AI interactions shape our thinking, which will enable better alignment between AI systems and users.

Fifteen years ago, The Atlantic published an essay that went viral. It was titled "Is Google making us stupid?", and it argued that the convenience of internet search was reducing people's capacity for deep, critical thinking. It is clear that search engines did not spell the end of complex human thought, but rather made information more easily accessible. The rise of generative artificial intelligence raises the question: "Is *generative AI* making us stupid?"

Without doubt, interacting with digital technologies can shape how we think. Students randomly assigned to take notes on laptops performed worse on conceptual questions than students who took notes longhand [1]. Similarly, the use of search engines has been shown to influence memory processes, for example, even when directed to remember facts, we are more likely to forget them if we know they will be accessible in a computer; and may better remember where to access information rather than the information itself [2].With long term use, technological tools can shape our brains. For example, taxi drivers in London, who must complete an exam without the aid of GPS, had significant posterior hippocampi relative to control subjects [3,4].

Generative AI tools—programs capable of generating seemingly new, meaningful content such as text, images, or audio [5]—have been shown to increase worker productivity, [6–8] but research remains silent on the psychological mechanisms behind these changes in performance. Studies have sampled different kinds of work tasks, such as writing, [6] customer support, [7] and consulting-related tasks like idea generation and product innovation [8]. In studies where tasks were simpler (e.g., customer support calls [7]) using generative AI improves performance—especially for users at the bottom of the performance distribution. In more complex tasks, however AI use led to higher error rates [8]. Similarly, an analysis of 150 million lines of code, shows that since the inception of AI coding assistants, written code is more likely to be repeated, violating principles of code maintainability [9].

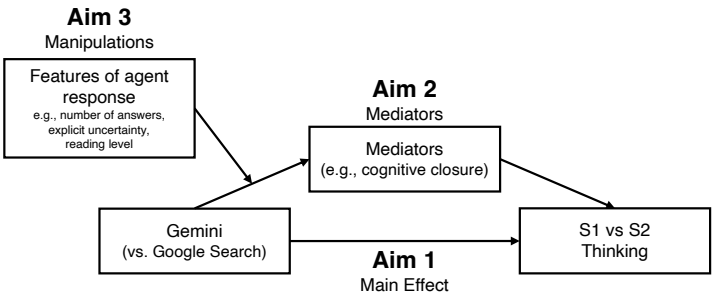

Fig. 1. Specific aims

Dual-process theories of reasoning [10–12] provide a useful framework for analyzing the effects of generative AI on human cognition. These models posit that the mind has two ways of processing information. System 1, characterized by associative pattern matching, is fast, effortless, and intuitive, underpinning our unconscious, automatic processes. In contrast, System 2 is deliberative, embodying our conscious, slow, and effortful cognitive activities that are more critical and reflective in nature. While there is debate regarding the details and the validity of this approach, [13–16] it is undeniable that when encountering problems we sometimes think deeply, and sometimes do not.

What determines whether System two will become engaged? One contemporary model [17] posits that when encountering a problem, System 1 will produce intuitions and monitor the amount of uncertainty, given the (potentially contradictory) intuitions. System 2 becomes activated when uncertainty exceeds a certain threshold. It will then engage in deliberative processing. The results of these processing will feed back into System 1, making some intuitions more active and others less so, which will change the level of uncertainty in turn. If uncertainty never passes the threshold, an intuitive response will be effected, if System 2 processing reduces uncertainty, then a more deliberative response will be effected.

While more deliberation may seem like desired outcome, there are many times when System 1 yields correct answers more efficiently than System 2. Take for instance chess masters, who are able to intuit the right move, even though they have no better general working memory and search for plays no more extensively than novices [18]. Despite this, there are plenty of times when over-relying on intuitions lead us astray, and more thinking would produce better results. For instance, if a decision is important, there is time available for deliberation, and we are would not be better off outsourcing the task to another person, then more deliberation is likely going to produce better results.

How might interactions with AI influence the way we think? On one hand, interactions with AI may allow us to engage more in System 2 deliberative thinking. By reducing cognitive load, and thereby freeing up cognitive resources, users can then allocate mental energy to more critical and reflective System 2 processing. There is evidence that people are able to more effectively engage in System 2 thinking when cognitive load is reduced [19]. Likewise, individuals with higher cognitive ability, are also more effective in using System 2 [20].

On the other hand, AI produces responses that are complete and easy to process. This fluency may lead to a greater subjective feelings of certainty (i.e, higher *feelings of rightness* [21,22]), which

[1]To whom correspondence should be addressed. E-mail: blira@sas.upenn.edu

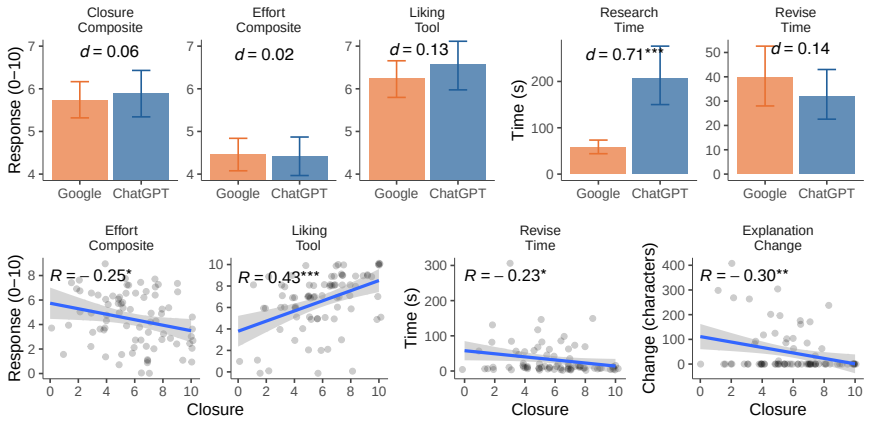

Fig. 2. Differences in user perceptions and engagement time of Google Search vs. Chatbot users in a judgment task. Correlations between closure and liking and effort metrics.

function as a metacognitive signal that no more thinking is necessary, and thus might reduce the amount of System 2 thinking. Past research suggests that the more some content is more easily processed, we are more likely to engage in shallower processing and to unquestioningly accept that content as true—a phenomenon known as fluency bias[23,24]. Thus, I predict that AI interactions will prompt more System 1 processing. The apparent contradiction in the role of cognitive load is resolved by the fact that reducing cognitive load is only helpful when what is being eliminated is not directly related with the task at hand (i.e., extraneous cognitive load[25]). When cognitive load is related to the main task (rather than a distractor task), higher amounts of cognitive load (i.e., deeper processing) often leads to increased performance,[23] even if we misperceive this effort as poor learning[26].

I propose a series of experiments to clarify how using generative AI prompts System 1 vs. System 2 thinking, relative to other kinds of interactions, such as traditional web search. I plan to shed light on the mechanisms by which generative AI changes the way we think, and the downstream consequences of these shifts on users' persuasiveness. I will then design interventions to help users engage with GenAI in ways that lead to faster System 1 thinking, or more careful System 2 thinking when appropriate. This research will uncover the cognitive consequences of our evolving AI landscape, and inform efforts at alignment of AI technologies with users most valued goals.

1. Specific Aims

As shown in **Figure 1,** This proposal has three main objectives.

Aim 1: How does GenAI change thinking? (Main effect). I will conduct a series of experiments to evaluate the degree to which engaging with GenAI (as opposed to traditional web search) for a research task results in increased automatic thinking.

Aim 2: What the mechanisms by which this happens? (Mediators). Once the main effect is established, a second set of experiments will tease out the characteristics of the AI interaction which drive the effect observed in Aim 1.

Aim 3: What manipulations can promote deliberation? (Manipulations). Finally, I will test whether manipulating the identified mediators with manipulations targeting both user behavior and interface design can result in more deliberative thinking.

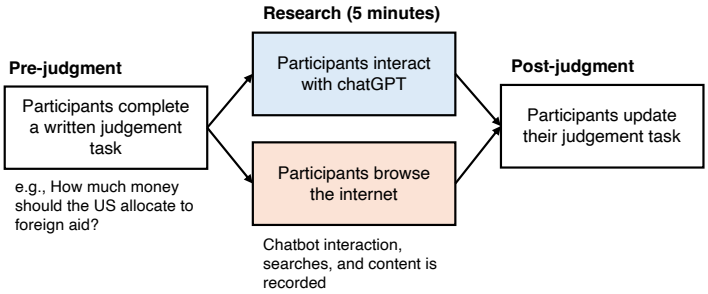

Fig. 3. Experimental task

2. Research Strategy

Pilot Data. As shown in **Figure 2,** preliminary pilot data supports the hypothesis that interacting with AI prompts participants for automatic thinking, consistent with fluency bias[23,24]. In a pilot study of 178 people, I found that participants interacting with AI-chatbot based on GPT-4 as opposed to traditional web search, spent 10 fewer seconds writing their revised response ($d = 0.14$), and reported exerting about the same effort ($d = 0.02$), despite having had done research for about four times as long ($d = 0.71^{***}$). Participants liked using the chatbot for research more than they did Google Search ($d = 0.13$), and they liked chatGPT more when they felt it produced more closure ($r = .43^{***}$). In ChatGPT users, closure, in turn, predicted less self-reported effort ($r = -.25^{*}$), less revision time ($r = -.23^{*}$), and smaller changes to their answers ($r = -.30^{**}$).

Aim 1: Main effect. To establish whether interacting with GenAI prompts automatic (System 1) or deliberative (System 2) thinking, I will conduct an experiment where participants will be assigned to complete a task using either traditional web search (e.g., Google), or a GenAI chatbot.

Our basic task procedure is shown in **Figure 3.** Participants will be asked to write a paragraph on a topic for which they might have strong priors (e.g., how much money should the U.S. allocate for foreign aid) but could benefit from learning additional information. They will then be randomized to learn more about the topic, either by using Google, or by interacting with an AI-chatbot. I will fully record their interactions with Google and the AI-chatbot. After up to 15 minutes, participants will be asked to re-write their views in the topic.

I will assess the extent of System 1 thinking by analyzing participants' perceived effort and confidence in their response, as well as

the time spent researching and writing, and the amount of work they produced. I will also use natural language processing and human raters to evaluate the quality, complexity, independence, and persuasiveness of the writing. I also introduce an incentivized behavioral outcome where naive participants produce allocations which they can later adjust based on the user's response. Additionally, I will assess the degree of effortful cognitive processing via pupillometry at the Wharton Behavioral Lab. I intend to use stimulus sampling[27] to make sure that results are not unique to a particular judgment task. See **Table 1** for a potential list of tasks, spanning judgment, decision, and forecasts.

Task Type	Example Tasks
Judgment	How much money do you think the US should allocate for aid?
Forecast	How likely is it that the Russia-Ukraine conflict ends within the next year?
Decision	If you were buying a car, would you buy a Toyota Corolla or a Ford Focus?

Table 1. Proposed judgement, forecasting, and decision tasks

To adequately power studies, I am working under the assumption that differences would be in the range of $.10 < d < .20$, which implies a sample size of between 395 and 1570 participants in each condition to achieve power of 0.80.

Aim 2: Mediators. The goal of Aim 2 is to pinpoint the specific characteristics of AI interactions that mediate the cognitive shift towards the hypothesized increase in System 1 thinking, as observed in Aim 1. I predict that processing AI responses will be less cognitively demanding than integrating information obtained through google search. This will trigger feelings of fluency and cognitive closure,[21] which will then explain the observed changes in cognitive style.

I will extend the methodology outlined in Aim 1, to measure these mediators by analyzing the content users engaged with in two ways.

1. **User Perceptions.** Users will report how they perceived their interaction with the AI and the search websites.

2. **Natural Language Processing.** I will use natural language processing to identify the stylistic features of AI responses that explain reduced deliberation.

The findings from this aim will provide crucial insights into how AI can be designed and used in a manner that promotes more deliberate cognitive processing. This understanding will be instrumental in developing manipulations (as outlined in Aim 3) to promote System 2 thinking when engaging with AI.

Aim 3: Manipulations. In the third phase of my research, I plan to develop dual-faceted manipulations aimed at increasing deliberative engagement with AI by targeting the mediators identified during Aim 2. These interventions will target interface design, as well as user interaction.

1. **Interface design.** I plan to modifying the AI-chatbot interfaces to foster more reflective and analytical engagement. By manipulating prompts and features of the user interface, I plan to reduce feelings of fluency and confidence.

2. **User interaction.** I plan to provide users with strategies to encourage more System 2 thinking while interacting with AI. These interventions will put users in a more deliberative mindset that will counteract automatic thinking.

A 2×2 factorial design will allow me to compare the effectiveness of strategies aimed at users as well as the interface design. I predict that interventions aimed at the interface will be more effective, and will be perceived as less intrusive by users. This line of work will be useful for designers building AI applications for situations where more deliberation is effective.

3. Conclusion

The proposed research will shed light on how the rapidly changing capabilities of generative artificial intelligence will shape our minds. From a **theoretical perspective,** it will elucidate some mechanisms that—with or without AI— prompt more automatic vs. more deliberative thinking. Additionally, it will inform on the cognitive mechanisms behind the mixed effects of AI-augmentation on workplace performance[6–8]. In more **practical terms,** my findings will inform the design of AI systems that better aligns to our goals and values.

If this proposal is successful, it would open new avenues for research. Specifically:

Short-term spillover effects. First, future research should address whether cognitive changes persist on an unrelated task after the AI interaction is over.

Long-term effects on cognitive style. Second, future research should address whether continued interaction with AI produces more enduring changes in cognitive style and learning motivation.

Generalizability. Third, this proposal focuses on a small number of tasks and compares chatbots to a single other information-seeking behavior. Future research should compare AI to other ways in which people get information, such as social media, or face-to-face human interaction. Likewise, future research should explore whether effects generalize over a wider range of tasks. Finally, while I focus on chatbots as the interface, future research should explore other alternatives such as autocomplete, email, and word processors.

Developmental impacts. I focus on the effects on adults who are transitioning to an AI world. Future research should address the effects of AI interactions in development, in particular as it relates to the development of cognitive skills that might be replaced by large language models.

Artificial intelligence will no doubt have a significant impact on human thinking. The goal of my research is to begin to understand this unexplored phenomenon. Specifically, this research will illuminate the nuances of how engagement with AI can steer our thinking towards more automatic or deliberative processes. As AI continues to permeate various facets of human life, understanding its cognitive implications is paramount. My hope is that findings from this work will help align the evolution of artificial intelligence with the enhancement of the cognitive abilities that make us human, rather than inadvertently reducing them.

4. References

1. PA Mueller, DM Oppenheimer, The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychol. Sci.* (2014).
2. B Sparrow, J Liu, DM Wegner, Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. *Science* **333**, 776–778 (2011).
3. EA Maguire, et al., Navigation-related structural change in the hippocampi of taxi drivers. *Proc. Natl. Acad. Sci.* **97**, 4398–4403 (2000).
4. EM Griesbauer, E Manley, JM Wiener, HJ Spiers, London taxi drivers: A review of neurocognitive studies and an exploration of how they build their cognitive map of london. *Hippocampus* **32**, 3–20 (2022).
5. S Feuerriegel, J Hartmann, C Janiesch, P Zschech, Generative ai. *Bus. & Inf. Syst. Eng.* **66**, 111–126 (2024).
6. S Noy, W Zhang, Experimental evidence on the productivity effects of generative artificial intelligence. (2023).
7. E Brynjolfsson, D Li, LR Raymond, Generative AI at work. *NBER Work. Pap. Ser.* **Working Paper 31161** (2023).
8. F Dell'Acqua, et al., Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electron. J.* (2023).
9. W Harding, M Kloster, Coding on Copilot: 2023 Data Shows Downward Pressure on Code Quality, Technical report (2024).
10. D Kahneman, *Thinking, Fast and Slow.* (Penguin Books), (2011).
11. P Wason, J.St.B.T Evans, Dual processes in reasoning? *Cognition* **3**, 141–154 (1974).
12. JSBT Evans, Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annu. Rev. Psychol.* **59**, 255–278 (2008).
13. W De Neys, On Dual- and Single-Process Models of Thinking. *Perspectives on Psychol. Sci.* **16**, 1412–1427 (2021).
14. DE Melnikoff, JA Bargh, The Mythical Number Two. *Trends Cogn. Sci.* **22**, 280–293 (2018).
15. G Keren, Y Schul, *Two Is Not Always Better than One: A Critical Evaluation of Two-System Theories.* Vol. 4, (2009).
16. G Keren, A Tale of Two Systems: A Scientific Advance or a Theoretical Stone Soup? Commentary on Evans & Stanovich (2013). *Perspectives on Psychol. Sci.* **8**, 257–262 (2013).
17. W De Neys, Advancing theorizing about fast-and-slow thinking. *Behav. Brain Sci.* **46**, e111 (2023).
18. WG Chase, HA Simon, Perception in chess. *Cogn. Psychol.* **4**, 55–81 (1973).
19. ED Johnson, E Tubau, W De Neys, The unbearable burden of executive load on cognitive reflection: A validation of dual process theory. (2014).
20. KE Stanovich, RF West, Advancing the rationality debate. *Behav. Brain Sci.* **23**, 701–717 (2000).
21. VA Thompson, JA Prowse Turner, G Pennycook, Intuition, reason, and metacognition. *Cogn. Psychol.* **63**, 107–140 (2011).
22. VA Thompson, SC Johnson, Conflict, metacognition, and analytic thinking. *Think. & Reason.* **20**, 215–244 (2014).
23. DM Oppenheimer, The secret life of fluency. *Trends Cogn. Sci.* **12**, 237–241 (2008).
24. AL Alter, DM Oppenheimer, Uniting the Tribes of Fluency to Form a Metacognitive Nation. *Pers. Soc. Psychol. Rev.* **13**, 219–235 (2009).
25. J Sweller, Cognitive Load Theory in *Psychology of Learning and Motivation.* (Elsevier) Vol. 55, pp. 37–76 (2011).
26. A Kirk-Johnson, BM Galla, SH Fraundorf, Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cogn. Psychol.* **115**, 101237 (2019/12/01/).
27. M Author, *Citation comming soon.* (Missing University Press), (9999).

5. Will Gemini Make Us Stupid? - 1000 Character Abstract

Generative AI (GenAI) applications based on large language models like Gemini are now ubiquitous. As AI becomes embedded in our email clients, word processors, and smartphones, it becomes crucial to understand how it influences how we think. I propose a series of experiments where participants are assigned to do research with an AI-Chatbot or traditional web search to explore how AI influences System 1 thinking, which is fast, associative, and automatic, vs. System 2 thinking, which is slower, resource-intensive, and deliberative. I plan to (1) establish whether the ease with which AI responses are processed prompts System 1 thinking, or whether the reduced cognitive load allows for increased System 2 processing; (2) identify the key mediators that drive this effect; and (3) develop interventions—targeting user interaction and interface design—that encourage more System 2 thinking. This research is pivotal for optimizing AI design to enhance user cognition and decision-making.

6. Student CV and Supervisor CV (1 page)

I asked Ashley for a one-page CV for Angela. I am updating my CV.

7. Transcripts

8. Describe the desired impact your research will make on the field and society, and why this is important to you. Include any personal, educational and/or professional experiences that have motivated your research interests.

I was already a grownup when smartphones and social media came along. I remember how exciting it felt to be in touch with friends and family members who lived many miles away. Today, this early excitement has given way to concern. I am constantly distracted by my phone. Social media drives political polarization around the world. Young people, especially girls, face mental health issues. A lively conversation after dinner is often replaced with sitting side-by-side staring into our phones and ignoring each other.

As we integrate ever more powerful generative AI (GenAI) into our daily lives, the lessons learned from past innovations weigh heavily on my mind. I often find myself chatting with GenAI models for leisure and work. Many of these interactions are fulfilling. I get to solve pressing problems much quicker and learn in the process much faster than I otherwise would. Sometimes, however, I find myself interacting with AI passively. Over relying on it to tell me what to think or write, and thoughtlessly accepting its first suggestion.

The way we think and behave is heavily dependent on the structure of our environments. This means that the way in which we design and engage with AI will determine whether it elevates the very capacities for thought and engagement that make us human, or whether it replaces them.

My desire is that my research will guide GenAI designers and users towards enriching human capabilities rather than inadvertently diminishing them. By understanding the mechanisms of how AI systems shape the way we think, we will be able to design interventions that help GenAI surface the best of us. My hope is that this knowledge will help steer our developing AI landscape closer to one where we can do more rather than less, think deep rather than shallow, and enhance the capabilities that make us human rather than mindlessly outsourcing them.

9. Describe an example of your leadership experience in which you have positively influenced others, helped resolve disputes or contributed to group efforts over time. (A leadership role can mean more than just a title. It can mean being a mentor to others, acting as the person in charge of a specific task, or taking the lead role in organizing an event or project. Think about what you accomplished and what you learned from the experience. What were your responsibilities? Did you lead a team? How did your experience change your perspective on leading others? Did you help to resolve an important dispute at your school, church, in your community or an organization? And your leadership role doesn't necessarily have to be limited to school activities. For example, do you help out or take care of your family?).

In many ways, I am now right where I was supposed to end up. And yet, without a fair bit of luck, I would not be. Very few graduates of Peruvian universities are admitted to PhD programs at world class institutions. Language and culture aside, Peruvian scholars who dream of becoming scientists at the cutting edge don't have access to meaningful research experiences, and local universities lack the Ivy League pedigree.

I now have mentorship, resources, and opportunities beyond the wildest dreams of my younger self. I feel beholden to help those who—in one way or another—look like an earlier version of myself. My hope is that they too will crack into an environment where their talents can be expressed. I have taken opportunities to advance this goal with undergraduate researchers and students navigating the graduate admission process, both from Peru and the US.

I have been lucky to work closely with undergraduate students exploring what psychological research looks like. At first, I was tempted to intervene too much, too soon. I quickly realized that my role was to make suggestions, provide feedback, and some support.

Likewise, I've been enriched by learning from PhD applicants. Providing feedback on their research statements and conducting mock interviews with them has allowed them to expand my own views. My questioning and probing have allowed them to refine their research programs.

It's easy to think that leadership is only about propelling a project forward by telling people what needs to be done. These experiences have taught me that leadership is often much more about listening, empathy, and connecting meaningfully with others. I am only powerful to help others when I listen to them, ask questions rather than give answers, and help others untangle their thoughts on their own.

I certainly put in a lot of effort and perseverance in getting to where I am. I was also extremely lucky. I hope I can contribute a little so that others from backgrounds like mine, might rely a little less on luck and more on a supportive community that I strive to nurture.