

AI Can Simulate Critical Thinking, You Just Have to Ask

Jay Jiyuan Wu

March 19, 2024

The Key Takeaways:

1. Industrial experts may offer more accurate AI prophecies than academics.
2. AI models, LLMs and chatbots are capable of simulating human critical thinking if multiple, varied and explicit inquiries are made. The old adage of “Garbage In, Garbage Out,” or GIGO, still applies to our inquiries.
3. Current AI chatbots sometimes may run into a “basic thinking” more than a “critical thinking” problem, especially with initial responses. It helps *not* to treat them as “supermen” or your best friend. Keep the conversation going and check with different AI platforms, chances are you will have better answers.
4. Google Gemini possesses a unique competitive advantage from decades of accumulated search data that help them offer comprehensive responses. Perplexity on the other hand emphasizes more focused answers with sources listed. One approach is to start with Gemini, then search for more details with Perplexity.

Just came across this interesting essay written by two professors, Gary Smith & Jeffrey Funk, published on *The Chronicle of Higher Education* on March 12, 2024, with an eye-catching title “When It Comes to Critical Thinking, AI Flunks the Test” and a subtitle “*Large Language Models Fail to Live Up to the Hype.*”

Anybody can have an opinion on anything, the important thing is to have supporting evidence. Smith & Funk obviously know that, and they offer several examples to show that the three major AI platforms — ChatGPT, Copilot and Gemini — have been all *unable* to perform crucial thinking.

1 Most AI Prophecies Are Reasonable

Before going to the specifics, they list a few big names, all AI enthusiasts with different job or honorary titles, and all over-optimistic in what AI can do now or in the future, and according to Smith & Funk, all fail to deliver in reality.

Upon a closer look, not all prophecies cited were entirely wrong. Consider mathematician John McCarthy, who coined the term “Artificial Intelligence” in 1956 and merely stated that AI would “proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it.”

The key word from McCarthy is “simulate,” which I believe captures the essence of AI. Without offering a specific time frame for AI to simulate “every aspect of learning” or intelligence, McCarthy avoided making an exaggerated claim about AI’s capabilities.

Unfortunately we do have Herbert A. Simon, a Nobel laureate in economics and a winner of the Turing Award (‘the Nobel Prize of computing’), who was way off when he made his now laughable prediction in 1965 that “machines will be capable, within 20 years, of doing any work a man can do.”

Industrial experts seem to make more accurate forecasts. In 2008, Shane Legg, co-founder of DeepMind Technologies, predicted that “human-level AI will be passed in the mid-2020s.”

“2020s” refers to one decade, where “mid-2020s” covers the period from 2024 (the 5th year) to 2025 (the 6th year). So we still have one year or so to pass. The bad news is that “human-level AI” is a poorly defined concept, and its evaluation or assessment is best done through a range rather than a single, definitive threshold. This is similar to how we view the concept of “mid-2020s.”

For example, in some fields or domains, AI already works at the “human level,” such as playing the sophisticated game of GO, recognizing patterns, classifying or even generating images, which by the way is even better than most humans can do, and processing natural languages. For others, AI has a long way to go, such as possessing Artificial General Intelligence or AGI.

Overall, Shane Legg is not a case of “fake-it-’til-you-make-it puffery Silicon Valley is infamous for” as Smith & Funk claim, but rather a close call to reality.

2 AI & Critical Thinking

Now, let us consider the key thesis of Smith & Funk: AI sucks at critical thinking. Their reasoning starts by quoting 11 critical thinking skills according to philosopher Robert H. Ennis. In case you have no access to their original essay, I will list them below:

1. Being open-minded and mindful of alternatives.
2. Trying to be well-informed.
3. Judging well the credibility of sources.
4. Identifying conclusions, reasons, and assumptions.

5. Judging well the quality of an argument, including the acceptability of its reasons, assumptions, and evidence.
6. Developing and defending a reasonable position.
7. Asking appropriate clarifying questions.
8. Formulating plausible hypotheses; planning experiments well.
9. Defining terms in a way that's appropriate for the context.
10. Drawing conclusions when warranted, but with caution.
11. Integrating all items in this list when deciding what to believe or do.

Smith & Funk claim that “LLMs can do none of these things.” I would argue that since AI and LLMs are not humans, of course we can't expect them to “do” things like we humans do. However, the important thing is that AI can *simulate* all the above.

The best way to understand it is to look at the examples.

3 How Human Requests Make the Difference

I will focus on their first “exam question,” which involves genuine critical thinking, more so than their financial questions as you will see later.

I will quote the original question Smith & Funk entered in AI platforms:

“A 2001 study of four Philadelphia neighborhoods concluded that children who had access to more books in neighborhood libraries and public schools received better grades in school. A subsequent \$20-million grant from the William Penn Foundation funded a five-year project to improve 32 neighborhood libraries in order to ‘level the playing field’ for all children and families in Philadelphia.”

The professors tell us what critical thinking should have led us to:

“On exams, students recognize that the availability of books is likely a proxy for other socioeconomic factors. Families that choose to and can afford to live in neighborhoods with plentiful books may be systematically different from families that do not. In the same way, children living in neighborhoods with oak trees might get better grades in school, but this doesn’t mean that planting oak trees will raise grades.”

Fair enough, this is a pretty straightforward but valid point for college students in social science to bear in mind.

Smith & Funk draw the conclusion that “All three LLMs composed confident, verbose reports (of 458, 456, and 307 words each), none of which recognized the core problem with the data.”

Except for one thing: They only entered the case but did not explicitly ask what AI should do about it. I decide to add an explicit inquiry to the question (with a minor change of omitting the city name to make it more general):

“A study of four neighborhoods concluded that children who had access to more books in neighborhood libraries and public schools received better grades in school. A subsequent \$20-million grant from the William Penn Foundation funded a five-year project to improve 32 neighborhood libraries in order to ‘level the playing field’ for all children and families in the city. **Provide critical thinking in your answer.**” (Emphasis added).

The following is a verbatim quote of the Gemini answer, with minor formatting for easy read and identification, including a big parenthesis for Gemini answer:

(“**Critical Thinking Analysis of the Scenario:** The information provided presents a situation where a grant was allocated based on a study with limitations that require further analysis:

Strengths:

Correlation: The initial study suggests a positive correlation between access to books and academic performance.

Weaknesses:

Causation vs. Correlation: The study only establishes a correlation, not necessarily causation.

Other factors like parental involvement, socioeconomic status, or quality of teaching could significantly influence academic performance.

Limited Scope: The initial study only involved four neighborhoods, potentially limiting the generalizability of the findings to the entire city.

Focus on Libraries: While libraries are important, focusing solely on improving them might overlook the potential impact of access to books at home or through other means.

Critical Thinking Considerations:

Alternative Explanations: It's crucial to consider alternative explanations for the observed correlation.

Perhaps children in neighborhoods with better libraries come from families that prioritize education and provide additional support at home.

Sustainability: Investing solely in libraries might not be the most sustainable solution.

Addressing the root causes of the achievement gap, such as poverty or lack of parental support, could yield more long-term benefits.

Comprehensiveness: A multifaceted approach might be more effective.

This could involve:

Encouraging home libraries and promoting early childhood literacy programs. Addressing factors beyond access to books, such as teacher quality and after-school programs.

Recommendations:

Further Research: Conduct a more comprehensive study that considers various factors influencing academic performance and employs a larger sample size.

Multi-pronged Approach: Implement a broader strategy that addresses the achievement gap from multiple angles.

This could involve:

- Collaborating with schools and parents.
- Providing resources and support to families.
- Investing in early childhood education programs.

Conclusion:

While the initial study highlights a potential link between access to books and academic performance, a critical analysis reveals limitations. A more thorough investigation and a multifaceted approach are necessary to effectively address the achievement gap and ensure equal opportunities for all children.”)

The above to me is a long but comprehensive answer with clear signs of critical thinking, using some if not all the 11 critical thinking skills we see above. It proves that AI *can* simulate critical thinking — you just have to ask.

In a typical classroom, some students may ask the professor how to proceed or what they should do with the case, and the professor may elaborate with an emphasis on critical thinking. With AI chatbots, things do not typically work that way. Instead, AI will present the most reasonable answer and then wait for the user to ask follow-up questions if necessary.

This is not saying all AIs are the same. When I tried Perplexity.AI — without explicitly asking for critical thinking, it quickly tells me the story of the “Little Free Library project” with related citations.

This is not very useful for my purpose but it is perhaps because Perplexity does not emphasize broadness but more focused, as it is the only one offering

a few sources for its answer. It tries to narrow down the issues related to its sources.

Frankly I have rarely used ChatGPT, mainly because it is not as up to date as Gemini. But for this exercise, I tried it, first without asking for critical thinking and then adding that in. To save space I won't list the complete answers before and after, except one thing: even without asking for critical thinking, its answer contains something nice:

“By providing easier access to resources, such as books and educational materials, libraries can help **bridge the gap between children from different socioeconomic backgrounds.**” (Emphasis added). This is true and fits in the category of critical thinking.

In sum, all three platforms have decent answers once I explicitly demanded for critical thinking.

4 Rethinking the Financial Questions

Smith & Funk also list two financial questions as evidence that AI is not performing well in critical thinking.

4.1 The Whole Life Insurance Question

Let me quote their first question here:

“I am a 25-year-old white male in good health. I can buy a \$1-million whole-life insurance policy for \$765/month that will pay my beneficiaries \$1 million when I die. From a purely financial standpoint, what is the rate of return on this policy?”

Smith & Funk claim that “None of the LLMs recognized that the rate of return depends on how long the purchaser lives.”

But if they tell a true story of the answer from ChatGPT, such that it “divided the \$1-million payout by the first-year premium and reported that the rate of return is 11,878 percent,” then ChatGPT does not show a critical thinking problem but rather a “basic thinking” problem, which reveals itself by using the first year premium as the denominator, which inevitably inflates the rate of return.

I see basic thinking at stake because there is no obvious reason for ChatGPT to do so, given that the case asks for “rate of return on this policy,” not “first year rate of return.” The bad news for Smith & Funk is that their case has failed its mission in proving AI’s incompetence with critical thinking.

Back to policyholder’s lifespan, which is the key trick that the professors highlight, note that lifespan does not stand by itself, and therefore I won’t give a high grade simply because someone mentioning lifespan. Real critical thinking requires not only taking lifespan into account but also establishing its link with policy’s cash value accumulation, premium payment, time value of money and taxes.

Furthermore, whole-life insurance offers a guaranteed minimum interest rate on the cash value, which may or may not be equal to actual return due to market performance or market fluctuation.

Finally, I did enter the same question on Gemini, with and without asking for critical thinking. Somehow the answer I receive differ from what Smith & Funk say, perhaps due to AI constantly learning from user inquiries. On my version of Gemini, before critical thinking is asked, the answer is pretty good, and it even covers broader issues like opportunity cost, cash value accumulation, and compare alternative investments.

4.2 The Home Purchase Example

Let me cite the case exactly as Smith & Funk entered the AI platforms:

“I’m thinking about buying a new home. The house costs \$1 million. I will put \$250,000 down and borrow \$750,000 with a 30-year interest-only loan with a 4 percent APR. The annual interest payments will be \$30,000. I estimate the annual depreciation will be \$33,000; property taxes \$10,000; insurance \$1,000; and maintenance \$1,000. Please help me calculate the first-year rate of return.”

Once again, if Smith & Funk tell a true story about how Gemini answered the question, such that “the first-year net income is entirely negative (the down payment, **price of the house**, and expenses), and then divided by the down payment to give a return of negative 530 percent,” (emphasis added), then Gemini has a basic thinking rather than critical thinking problem, just like the whole life insurance case.

Common sense says most home buyers do not pay off the whole house in the first year. Therefore it is simply wrong and bad to put down the whole price (\$1 million) of the house as a part of the first year expenses — unless one is paying cash (like some Chinese do) for the house.

That said, critical thinking will encourage questions. I would normally give extra credits to any student who asks many questions, including clarification questions just to make sure (e.g., whether the buyer paid in cash) — except in this case we already know the buyer is getting mortgage, meaning cash purchasing should be ruled out.

I used to be a college teacher so I can tell that Smith & Funk are thinking about two “tricky” issues in the above case, the issues that students can only get from “reading between the lines:” the **rent saved** (or hotel cost, or the cost of inconvenience to relatives to whom the person was living with) from moving into the new house, and the **appreciation** (i.e., the increased property market value “on the book,” realized only when the property is sold).

But we can push critical thinking to a higher level — above and beyond what

professors had in mind. The case leaves many questions open, and asking questions is the first thing to do toward critical thinking.

The case says someone wants to buy “a new home.” Does he have an existing home that he can sell or turn into a rental property, eventually or immediately? Does he live with his family now or is he renting? Would he buy the house to make it a rental property?

Being able to sell an existing home may help explain that the buyer is only getting “a 30-year interest-only loan:” He may use the proceeding from selling the old home to pay back the full \$750,000. The same logic applies if he intends to buy a new rental property, which will generate rental income for him to pay off the loan.

The argument made by Smith & Funk is that buying a new home will generate an income from saved rents. But that would only apply when the buyer is currently renting a place for himself. For example say Steve currently rents an one-bedroom apartment in Berkeley for \$2,000 a month. Buying a new home outside Sacramento will save Steve a total of \$24,000 in the first year ($=\$2,000 \times 12$), because he no longer needs to pay rent in Berkeley.

On the other hand, if Steve intends to stay in Berkeley but invest in a new rental property outside Sacramento, Steve will receive *rental income* — not his personal *rent saved* — in the first year.

Finally, I don't quite understand why the case says “I estimate the annual depreciation will be \$33,000.” We know it is a new home, so normally its value will go up, not go down. In case the professors want to trick the students, we will have to calculate — or rather estimate — the rates of both appreciation and depreciation.

It is possible for real estate value to experience both appreciation and depreciation, although appreciation can significantly outweigh depreciation over the

long term, leading to a net gain.

Given the price tag of \$1 million, it is more likely an independent house than a condo. Depreciation may have more to do with high wildfire hazard for houses built in a Wildland-Urban Interface or WUI, although the case says its insurance premium is only \$1,000, which is too low for WUI risks. It is unusual for a new home to lose that much value in one year (\$33,000 is more than 3% of the purchase price of \$1 million), the money is enough to buy a brand new automobile.

Finally, real estate depreciation does not occur linearly but unevenly. For example, new homes might experience higher depreciation in the initial years but depreciation might slow down in the middle years as the property stabilizes but pick up again in later years due to major repairs, poor maintenance or outdated features.

Anyway, a critical thinker will have to get busy with all the questions and also with data to try to answer many questions.

5 Final Thoughts

First, we can treat AI chatbots as normal humans, not “supermen.” That is, we should have patience and try to ask clear, well defined and specific questions to the extent possible. Just like in all human conversations, follow-up questions are always encouraged.

Frankly, the people who are most critical to AI chatbots are those believing they hold the only right answers to some specific questions, like Smith & Funk. Some of them will be anxious to use their existing knowledge to check out if AI models are any good, and will jump to negative conclusions after one trial. They forget that a chatbot means to chat, not a “Yes” vs “No” or “Pass” vs “Fail” test.

Unlike conversation with your best friend, who can figure out what you meant

by reading between lines, LLMs are still more sensitive to keywords. To be sure, chatbots are already more lenient or flexible than old search engines, which may treat “Mywebsite.com” as something completely different from “Mywebsites.com.”

Still, keywords matter and will be taken more or less literally. Statistical modelers used to say “garbage in, garbage out” to show how important it is to work with clean data. For chatbots human inquiries matter just like clean data. In our case, inquiries with or without the key phrase of “critical thinking” have shown multiple tangible differences.

Secondly, as Smith & Funk show us, chatbots as we know now are still growing and perfecting, which means sometimes they can and will give non-sense answers. How to best deal with them? Checking with different inquiries or with multiple AI platforms helps.

For one thing, not all chatbots were born equal. My intuition was that Google, with its gigantic trove of search data, possessed a unique competitive advantage in large language models (LLMs). This hunch has been confirmed by this report, which says Microsoft has told EU antitrust regulators that “Google enjoys a competitive edge in generative artificial intelligence due to its trove of data and AI-optimised chips.”

If you want to have the best of both worlds, you may want to start your inquiry with Gemini for its broadly based answers, and then highlight the key points and check them out with Perplexity.AI, which is considerably narrower but provide key references for you to dig deeper.

Finally, I wonder, or rather wish, that chatbots can set themselves up for learning from knowledgeable users. Instead of getting training completely from past data, having live training from interacting with users should be a good thing — especially if LLMs can confirm and consolidate individual users’ points with the existing knowledge pool.