AI as Artificial Ignorance

Working paper

By

Bent Flyvbjerg[*]

Oxford University

IT University of Copenhagen

Abstract

AI and bullshit (in the strong philosophical sense of Harry Frankfurt) are similar in the sense that both prioritize rhetoric over truth. They mix true, false, and ambiguous statements in ways that make it difficult to distinguish which is which. AI sounds convincing even when it's wrong. As such, current AI is more about persuasion than about truth. This is a problem because it means AI produces faulty and ignorant results. For now, we need to be highly skeptical of AI.

Keywords: AI, artificial intelligence, artificial ignorance, large language models (LLM), general artificial intelligence (GAI), ChatGPT, Perplexity, bullshit

Full reference: Flyvbjerg, Bent, 2025, "AI as Artificial Ignorance," working paper, University of Oxford, IT University of Copenhagen.

[*] Professor Emeritus, University of Oxford; Villum Kann Rasmussen Professor and Chair, IT University of Copenhagen.

Introduction

In the Prologue to his book *The Coming Wave*, Mustafa Suleyman, CEO of Microsoft AI and cofounder of DeepMind, says, "Never before have we witnessed technologies with such transformative potential ... With AI, we could unlock the secrets of the universe, cure diseases that have long eluded us, and create new forms of art and culture that stretch the bounds of imagination."[1]

There is a catch, however. At the end of the quoted Prologue, Suleyman reveals that he did not actually write it. An AI did.

But Suleyman's own musings about AI are just as boosterish as the AI writing about itself. "I am convinced we're on the cusp of the most important transformation of our lifetimes," he explains.[2] To make this transformation come about we need to "distill the essence of what makes us humans so productive and capable into software, into an algorithm." The goal is to "replicate the very thing that makes us unique as a species, our intelligence." Doing this would help us tackle "awesome challenges" facing humankind, like climate change, sustainable food, and aging populations.[3]

Today, AI is already producing impressive results in focused areas like speech transcription, language translation, and face recognition, but it looks set "to reach human-level performance across a very wide range of tasks within the next three years," Suleyman predicted in 2023.[4] The future may prove him right. But we're not there yet, not even close. Instead, we may be witnessing yet another hypefest for AI.

[1] Suleyman, Mustafa with Michael Bhaskar, 2023, *The Coming Wave: AI, Power, and the Twenty-First Century's Greatest Dilemma*, The Bodley Head, p. 3.

[2] Suleyman, 2023, *The Coming Wave*, p. 16.

[3] Suleyman, 2023, *The Coming Wave*, pp. 7-8.

[4] Suleyman, 2023, *The Coming Wave*, p. 9.

A Simple Test of AI

To test the hype, simply ask a leading AI a question about something you are knowledgeable about. For instance, I asked ChatGPT to "give me a list of ten megaprojects with cost overruns above 200%, baselined at FID [the final investment decision], with the overrun for each." ChatGPT immediately produced what looked like a convincing list, with a well-formulated introduction and ending. Except for one thing.

On its list, ChatGPT included two projects under different names that are actually the same project, namely (a) "The Big Dig in Boston, USA (2007) - cost overrun of 210%" and (b) "The Boston Central Artery/Tunnel Project (2007) - cost overrun of 190%."[5] The "Big Dig" is the nickname of the "Central Artery/Tunnel project," which is the official name of the project. Any Bostonian knows this, as do many Americans and most megaproject experts. But ChatGPT did not know, which is surprising, especially as the Big Dig was the costliest US public works project in history when it was planned and built. The project was in the news constantly for almost two decades and books were written about it. By mistaking the two different names for two different projects and listing different cost overruns for each, ChatGPT did not produce artificial intelligence but, instead, *artificial ignorance*, even confusion.

You can easily generate examples like this on your own. In fact, you cannot avoid it if you use current versions of AI. To check ChatGPT's findings against another AI, I asked Perplexity, "What was the percentage cost overrun on the Big Dig?" Perplexity answered that the Big Dig "experienced a significant cost overrun ... of approximately 478%."[6] This is more than twice the overrun found by ChatGPT above. The difference could be due to Perplexity including inflation when calculating

[5] https://chat.openai.com/c/7e54e282-9068-4e00-9e8d-86073f260788, retrieved February 17, 2024.

[6] https://www.perplexity.ai/search/Please-make-me-s7MSotZVTkmrtBTrrY4O5Q, retrieved February 17, 2024. Unlike ChatGPT, Perplexity explicitly identified the Big Dig and the Central Artery/Tunnel Project as one and the same project.

overrun, although this would be non-standard. So I checked for this by directly asking Perplexity, which assured me, "The cost overrun for the Big Dig ... is given in real terms," that is, excluding inflation. [7]

On that background, it is clear that either ChatGPT or Perplexity or both got the cost overrun of the Big Dig wrong. In fact, both got it wrong, with Perplexity being significantly more off than ChatGPT. The correct number for the Big Dig cost overrun is 220 percent, which is a number published in widely cited, peer reviewed research, and which it is therefore surprising that neither ChatGPT nor Perplexity knows, as would an intelligent person researching the issue. But we don't even need to know the accurate number to see that the AI cannot be trusted regarding project names and project cost overrun, the blatant inconsistencies suffice. Again, the 478 percent cost overrun found by Perplexity is not artificial intelligence but artificial ignorance, wrong by a large margin.

More Tests, and a Verdict

In an earlier experience with AI, a few years ago, my front door needed repainting. It opens on a hallway with other doors of the same color – a rare dark blue, almost black. So I had to get the color right, or I'd be in trouble with my neighbors and the owners' association. I explained this to my painter. "No problem," he said and whipped out an app on his phone with an AI that would determine the exact color of my door by simply taking a photo of it with the phone's camera. The app would also produce a code for the paint shop, so they could get the mix of the paint exactly right. "Great," I said, "let's do it!" The photo was taken, the code sent to the paint shop, the paint was mixed, picked up, applied to my door, and allowed to dry. The result was not so great, however. The door now had an obviously different color from before and from the other doors in the hallway. The app had failed. My painter acknowledged this, and reverted to old-fashioned trial and error, mixing samples and applying them to

[7] I also checked that Perplexity used the same baseline as ChatGPT (the final investment decision) for calculating cost overrun, which was the case.

my door until we agreed, by visual inspection, that we had the same color as the old paint. Then the door was painted again, with the desired outcome but taking twice as long and costing twice as much as it would have if we had ignored the AI in the first place. Artificial intelligence turned out to be a real waste of time and money in this case. This is some years ago, so the app may have improved in the meantime, or been replaced by a better one. Nevertheless, current AI has similar flaws, as illustrated by the examples above, with errors that are more difficult to detect than the wrong color of a door.

Shortly after I ran my experiments with ChatGPT and Perplexity, Nassim Nicholas Taleb did something similar, specifically for ChatGPT, and published his conclusions on X (formerly Twitter): "**VERDICT ON ChatGPT**: It is ONLY useable if you know the subject very, very well. It makes embarrassing mistakes that only a connoisseur can detect ... So if you **must** know the subject, why use ChatGPT?" (bold and caps in the original).[8] Taleb's conclusion fits my own in emphasizing the irony that in order to make sense of the results from ChatGPT you need to know the subject at a level where you don't need ChatGPT.

Explaining the Difference between Hype and Reality

It is easy to explain the stark contrast between the faulty "intelligence" of ChatGPT and the happy hype of Mustafa Suleyman above. Suleyman is talking about artificial *general* intelligence (AGI) whereas ChatGPT is a generative artificial intelligence that is *limited* to what works on the basis of large language models (LLM). AGI does not exist today, so it cannot be tested. This allows wide scope for speculation, postulates, and hype à la Suleyman. LLMs are used for text generation using large volumes of already existing text as input and then repeatedly predicting the next word in a manner that seems right (but often is not), when compared with the existing text. LLMs are much

[8] https://twitter.com/nntaleb/status/1759234709949710753, February 18, 2024.

more limited than AGI and cannot be said to be truly intelligent, as illustrated by the examples above. LLMs have no logic or facts by which truth may be determined. LLMs simply generate text that *sounds* right when compared with existing text without knowing whether the generated text is *actually* right.

The limited intelligence of ChatGPT does not mean that there are not specific areas where it may be useful. Taleb mentions generating code, writing condolence letters, and fabricating quotations. It seems clear, however, that at its present level of development the real risk in using ChatGPT and similar AI is *not* that the AI will prove better than human intelligence and make humans redundant. The real risk is that humans begin to trust an AI that is in fact ignorant and faulty, which could prove disastrous. Current AIs are well-formulated and persuasive, even when they are wrong, because they were designed that way. That makes it all-too-easy to trust an AI, especially in areas where as user you do not know the subject well. Our biggest risk is, as usual, ourselves.

An Industry Perspective

The car industry seems to acknowledge this, at least for now. Based on a trial of AI in cars commenced in late 2023, *Car Magazine*, a leading outlet for the industry, questioned the AI's claim to cutting-edge relevance. The magazine found that the knowledge base for the AI was not up to date, which seems to be a general problem. An intelligence that is not up to date can hardly be said to be intelligent. Worse still, the AI made things up in the typical fashion of large language models. Mercedes' chief technology officer, Markus Schäfer, commented, "If you sit in a car and ChatGPT tells you something that's absolute nonsense, you might be exposed to product liability cases."[9] For industry, ignorance and liability come hand in hand, including for artificial ignorance, which is what

[9] Groves, Jake and Tom Webster, 2024, "Will AI Make Your New Car Better?" *Car Magazine*, Issue 740, March, p. 25.

the AI systematically produced, according to the trial. Schäfer therefore warned cautiousness for AI in cars.

ChatGPT as a Bullshit Generator

Bullshit and generative AI are not the same. They are similar, however, in the sense that both mix true, false, and ambiguous statements in ways that make it difficult or impossible to distinguish which is which. ChatGPT has been designed to sound convincing, whether right or wrong. As such, current AI is more about rhetoric and persuasiveness than about truth.[10] Current AI is therefore closer to bullshit than it is to truth. This is a problem because it means that AI will produce faulty and ignorant results, even if unintentionally.

It is therefore interesting to note that Professor Alan Blackwell of Cambridge University's Department of Computer Science and Technology does not hesitate to call ChatGPT "a bullshit generator."[11] This may sound like a flippant remark, but in fact Blackwell chooses his words carefully by not using the term bullshit in its everyday sense. He uses Harry Frankfurt's "scientific" (Blackwell's word) definition, like we did above. Based on this, Blackwell concludes, "there is no algorithm in ChatGPT to check which parts are true. The output is literally bullshit, exactly as defined by philosopher Harry Frankfurt."[12] Further referencing University of Toronto professor Geoffrey Hinton – widely known as the "Godfather of AI" – Blackwell goes on to stress that, "one of the greatest risks [of AI] is not that

[10] AI implies a pragmatic theory of truth, that is, statements that work are considered true. But pragmatic theories of truth are unviable, as argued below.

[11] Blackwell, Alan, 2023, "Oops! We Automated Bullshit," Alan Blackwell's blog, November 9, https://www.cst.cam.ac.uk/blog/afb21/oops-we-automated-bullshit, retrieved, April 3, 2024. See also Blackwell, Alan, 2024, *Moral Codes: Designing Alternatives to AI*, MIT Press.

[12] Blackwell, Alan, 2023, "Oops! We Automated Bullshit," Alan Blackwell's blog, November 9, https://www.cst.cam.ac.uk/blog/afb21/oops-we-automated-bullshit, retrieved, April 3, 2024. See also Blackwell, Alan, 2024, *Moral Codes: Designing Alternatives to AI*, MIT Press.

chatbots will become super-intelligent, but that they will generate text that is super-persuasive without being intelligent, in the manner of Donald Trump or Boris Johnson. In a world where evidence and logic are not respected in public debate ... systems operating without evidence or logic could become our overlords by becoming superhumanly persuasive, imitating and supplanting the worst kinds of political leader."[13]

Conclusion

Judging by the available evidence, current AI – which is generative AI based on large language models – entails artificial ignorance more than artificial intelligence. That needs to change for AI to become a trusted and effective tool in science, technology, policy, and management. AI needs criteria for what truth is and what gets to count as truth. It is not enough to *sound* right, like current AI does. You need to *be* right. And to be right, you need to know the truth about things, like AI does not. This is a core problem with today's AI: it is surprisingly bad at distinguishing between truth and untruth – exactly like bullshit – producing artificial ignorance as much as artificial intelligence with little ability to discriminate between the two.

Nevertheless, the perhaps most fundamental question we can ask of AI is that *if* it succeeds in getting better than humans, as already happens in some areas, like playing AlphaZero, would that represent the advancement of knowledge, even when humans do not understand how the AI works, which is typical? Or would it represent knowledge receding from humans? If the latter, is that desirable and can we afford it?[14]

[13] Blackwell, Alan, 2023, "Oops! We Automated Bullshit," Alan Blackwell's blog, November 9, https://www.cst.cam.ac.uk/blog/afb21/oops-we-automated-bullshit, retrieved, April 3, 2024.

[14] Kissinger, Henry, Eric Schmidt, and Daniel Huttenlocher, 2021, *The Age of AI and Our Human Future*, Little, Brown and Company.