OPINION PAPER

The consequences of AI hype

Kevin LaGrandeur[1,2]

Abstract

AI promises to be a potentially beneficial innovation if it can be wisely built and adopted. One of the impediments to its wise use that is not discussed enough is the tendency toward exaggerating the capabilities of AI, or "hyping" it. AI hype increases the possibility of bad consequences for society, including compromised public safety, and even faulty social, business, and educational practices. Three case studies given here regarding AI used for self-driving automobile functions, law practice, and worker displacement illustrate these dangers.

Keywords AI · Hype · Hype examples · Hype consequences

1 Introduction

The possibilities for abuse of AI, especially recent types such as generative AI, are currently the main focus of worry about that technology by governments, news companies, and even business leaders. These worries about unregulated, rapid growth of this new and powerful technology is justified. However, there are other concerns about AI that remain important, even if they are currently overshadowed by anxieties regarding generative AI. One of these overshadowed issues is exaggeration, or hype, of AI's capabilities in general, because hype that distorts expectations can be dangerous to society as well.

The rapid advances in generative AI, especially the sudden, splashy arrival of ChatGPT last year, have caused companies to panic that they will be left behind and to consequently trip over each other in their race to develop this kind of AI for themselves. Examples of this phenomenon include companies such as GM, Coca-Cola, Salesforce [1], and Accenture, which has sunk $3 billion into AI development to add to its consulting services [2]. As Accenture's case implies, the frenzy to develop AI applications also means a sudden rush by venture capitalists and others to inject lots of money into any project with the word AI in it. In fact, it is estimated that just by the end of 2023 alone, investment in generative AI will reach $42.6 billion [2]. This is despite the fact that nobody has any idea yet what the long-term prospects of AI are for business—much less the short-term prospects. Even Sam Altman, the CEO of OpenAI, which invented ChatGPT and the GPT-3 application on which it runs, has said that AI is "wildly overhyped in the short term" [3].

To make matters worse, the gold rush caused by the recent chaos surrounding generative AI has caused businesses to misleadingly label any software they produce as AI—an example of marketing hype. For instance, there are tools that can clean up the voices of singers in old recordings; these have been used recently and most famously by Paul McCartney to clean up the voice of John Lennon in an old recording. But, as Scott Rosenberg says in his article in *Axios*, this is misleading hype. The tool used in this case is substantially similar to older "audiovisual pattern recognition programs that have been in use for decades, [and which operate] like the 'magic wand' in Adobe Photoshop that isolates a foreground image from a background." It did not actually bring Lennon's voice "back from the dead" [4]. It just cleaned it up.

Hype regarding AI's capacities is detrimental because it leads to rushed and irresponsible development of innovations by companies that are afraid of being left behind, and also to misunderstandings by society at large about exactly what AI's real competencies and dangers might be. Consequently, AI hype increases the possibility of bad consequences for society, including compromised public safety,

✉ Kevin LaGrandeur
  kevin.lagrandeur@gmail.com

1 Global AI Ethics Institute, Paris, France

2 Santa Fe, NM, USA

and even faulty social, business, and educational practices. Let us look at some cases that demonstrate these dangers.

2 Case 1: a safety problem: exaggeration of Tesla's "autopilot" capabilities causes death and injuries

Elon Musk's and Tesla's exaggeration of Tesla's Autopilot capabilities in Full Self-Driving (FSD) mode has proven dangerous to human lives. Musk said in 2019, "My guess as to when we would think it is safe for somebody to essentially fall asleep and wake up at their destination: probably toward the end of next year. I would say I am certain of that. That is not a question mark" [5]. Meanwhile, a number of people have died or been injured because of FSD malfunctions. These include a man changing a tire on the side of the road in New York, a person whose autopilot ran him into the back of a truck in Florida, a pedestrian who was killed by a Tesla that ran through an intersection in Florida, a person whose autopilot drove without warning into a highway barrier, and others.

According to a recent article in *New York Times Magazine*, the various claims in the lawsuits against Tesla boil down to this one issue: that "Tesla consistently inflated consumer expectations and played down the dangers involved" with its AI-powered autopilot. In fact, claims the article's author,

> Ever since Autopilot was released in October 2015, Musk has encouraged drivers to think of it as more advanced than it was, stating in January 2016 that it was 'probably better' than a human driver. That November, the company released a video of a Tesla navigating the roads of the Bay Area with the disclaimer: 'The person in the driver's seat is only there for legal reasons. He is not doing anything. The car is driving itself.' Musk also rejected the name 'Copilot' in favor of 'Autopilot' [5].

In fact, despite warnings in the small print of the user's manual that drivers needed to keep their eyes on the road and monitor the FSD, Musk himself encouraged viral messages to the contrary, because he wanted users to think that the autopilot was better than it is. He said, for instance, in 2019, that, "If you have a system that's at or below human-level reliability, then driver monitoring makes sense. But if your system is dramatically better, more reliable than a human, then monitoring does not help much" [5].

As a result of these kinds of statements, which were given much more attention than the various cautions to drivers contained in the users' manual, "A large number of drivers seemed genuinely confused about Autopilot's capabilities. (Tesla also declined to disclose that the car in the 2016 video [promoting FSD] crashed in the company's parking lot.)" [5].

As a result of all this, a number of lawsuits have been filed against Tesla. Recently, as of 15 February, 2023, the federal government of the US has added weight to these lawsuits against Musk's company by issuing a recall of all 360,000 Teslas that have FSD [6].

The exaggerated claims of Musk and his representatives are clearly the main cause of all of the problematic expectations of Tesla's consumers. But some of this is also the US government's fault because of lax regulation. Although it demands that most modes of transportation get pre-approval to use innovations to their technology, it does not demand this of automobiles (Europe does, and that is why this Tesla issue is not a problem in Europe.) [7].

3 Case 2: hype about generative AI encourages bad legal practice

Perhaps ChatGPT's most salient characteristics are that it writes pretty smooth prose, can carry on fairly fluent conversations with a human, and can provide information very rapidly in a conversational format; it is certainly easier to use than older computer search applications. But some users have allowed these characteristics to convince themselves that this AI is more competent than it is. In fact, too many in the general public seem to have succumbed to hype about its capabilities. In the most extreme case, one of Google's engineers went on record claiming that ChatGPT was sentient and had intelligence parallel to a 7- or 8-year-old human (even ChatGPT says that neither it nor any other AI is sentient, by the way) [8].

Less extreme but maybe more serious in terms of social consequences is the fact that people have started using generative AI in reckless ways at work. A good example of this is a case where a lawyer had ChatGPT write a legal brief for him and then presented it in court on behalf of his client, without first checking its content. It turned out that all six of the previous cases the AI cited as precedents for the reasoning contained in the brief were complete fabrications (or hallucinations, in the nomenclature of AI engineering). The lawyer who used ChatGPT to make his brief had been hasty and simply bought into the hype about the AI; he clearly knew it wrote well, but he claimed that he was not aware that it made up facts [9].

In a similar recent incident, Joshua Browder, the CEO of an AI startup called DoNotPay announced that he was going to have an AI-powered "robot lawyer" represent a client in a court case involving a traffic ticket [10]. The defendant would wear a pair of smart glasses that would record the proceedings and transmit the recording to a combination of generative AIs, including ChatGPT and DaVinci. The AIs would

transmit back into an earpiece connected to the glasses a script for the defendant to repeat in his defense. Before any of this could take place, however, Browder received letters from the California bar association threatening him with jail time for unauthorized practice of the law. Considering the reliability of ChatGPT regarding its handling of facts, demonstrated in the previous incident above, Browder and his client may have been lucky that the bar association stepped in when it did.

The wisdom of those trying to use generative AI in court is certainly questionable, but both the lawyer who tried to use AI to write his brief and the CEO of the robot lawyer company can themselves also be seen as victims of hype—in the latter case, maybe the CIO's own hype of his product—and they were clearly over-confident of the capabilities of this new technology. The consequent ethical problems are clear, especially for the lawyer. He failed all the following ethical criteria of his profession [9]:

- Duty of Competence: lawyers are obligated to show competence in their practice, including not just knowing the law, but also knowing about the technology they use, and its reliability. This lawyer relied blindly on technology that is relatively new.
- Duty of Confidentiality: this lawyer risked revealing his client's data to the company that makes ChatGPT (OpenAI) because they routinely use user data to train the AI. The lawyer should have known this.
- Responsibilities Regarding Nonlawyer Assistance: As of 2012, the American Bar Association dictated that lawyers must supervise not only any person who provides assistance to them on a case, but also any machine providing help. "That means lawyers must supervise the work of AI programs and understand the technology well enough to make sure it meets the ethics standards that attorneys must uphold [9]."

4 Case 3: hype about capabilities of AI causes unnecessary anxiety about worker displacement

AI, and ChatGPT in particular, has people worried that it may potentially replace human jobs because it appears upon casual use to be able to do amazingly human-level thinking and communication. For instance, it has people who write for a living worried about being replaced at their jobs, and teachers worried about being replaced by AI chatbots that can teach classes online by themselves. But, although AI and automation will certainly replace some jobs, there are various reasons that the anxiety about this is overblown. First, as John Naughton says in his *The Guardian* article, we humans "generally overestimate the short-term impact of new […] technologies, while grossly underestimating their long-term implications" [11].

In other words, in the short term, AI does have promising aptitudes, but mostly those aptitudes are still limited. Automation still does not have human-like intelligence (it is not artificial *general* intelligence), and so it still can only do certain, circumscribed things well. As the author of an article in the *Harvard Review* puts it, "On a technical level, it doesn't work differently than previous AI systems, it's just better at what it does." Yet many users overestimate its capabilities because it can do some new and flashy things. "Since its release," for instance, "Twitter has been flooded with examples of people using it to strange and absurd ends: writing weight-loss plans and children's books" [12].

Because of its limitations, smart technology does not generally replace whole jobs, but rather certain activities within jobs. Typically, these activities are ones that are dangerous, dreary, or dirty; so human workers benefit by offloading those activities. This also frees them up to do other, higher level activities. For instance, my accountant uses software programs to do all of the number crunching for my taxes. This allows him to serve more clients because he can get done faster; it also frees up his time to do more complicated duties of his job, like the communicative "handholding" about tax laws that makes him especially helpful to clients—advising them about tax codes, what the tax authorities expect, and so forth.

As for the "underestimating" of "long-term implications," those implications are usually positive ones. If we look at history, we can see that throughout the previous three industrial revolutions, new technologies have caused job dislocations in the short run, but in the long run have spurred many more jobs than they destroyed. This historical trend is clear enough that the famous economist Joseph Schumpeter came up with the well-known principle of "creative destruction" to describe exactly this phenomenon.

Even ChatGPT, as good as it is at writing prose and computer code, is not as smart as most casual users might assume, based on its fluency. For one thing, it makes too many content errors—often making up its own facts whole cloth (hallucinating). This was the cause of the unfortunate demise of the legal case I discussed above. One can see this for oneself by trying an easy experiment. Try doing a search on yourself using ChatGPT. When I did this, some of the material in the biography that the AI wrote about me was correct, but key things were complete fiction. For example, it said that I worked at the United Nations and that I had written a book on Alan Turing titled, "The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus The Secrets of Enigma." That book sounds great. I wish I had written it. By the way, the title is far wordier than I would have liked.

These kinds of limitations are obvious to experts in the fields in which ChatGPT operates. Most of those who are proficient at jobs that ChatGPT can do, such as writing and

programming, say that it can do the initial work required for tasks like writing and coding, but that it takes an expert in those things to do the final work. For example, one software engineer who used ChatGPT to help him code a project said, "It's incredibly good at helping you get started in a new project. It takes all of the research and thinking and looking things up and eliminates it…. In 5 min you can have the stub of something working that previously would've taken a few hours to get up and running" [11]. But to make use of that stub, and to complete the project, a person has to know about programming and how it works. In addition, as Naughton concludes in his article, "That seems to me to be the beginning of wisdom about ChatGPT: at best, it's an assistant, a tool that augments human capabilities….In that sense, it reminds me, oddly enough, of spreadsheet software," and nothing more ominous [11].

5 Conclusion

There is a lot of promise in the growth of AI. It can help us detect and diagnose diseases faster and more accurately, improving healthcare outcomes. It can also help us develop new materials and technologies that are more sustainable and environmentally friendly, addressing climate change. AI can also help us improve education and lifelong learning by providing personalized and adaptive learning experiences.

However, we have to be careful not to let hype distort our expectations and reactions to it. As I hope my foregoing discussion shows, such distortion can lead us to depend too much or inappropriately on smart technology, and this will be to our own detriment; it may even threaten our lives, as the example of consumers' dependence on Tesla's autopilot demonstrates.

Unfortunately, technology hype is so common that Gartner, the technology research and consulting firm, has been able to define a common pattern that tech hype usually follows, called the Gartner Hype Cycle. It has five stages: technology trigger, peak of inflated expectations, trough of disillusionment, slope of enlightenment, and finally the plateau of productivity [13]. As a society, we seem to be at the "peak of inflated expectations" phase, or close to it. Let us hope that we can see our way safely through this phase all the way to the relative safety of a plateau of productivity.

Declarations

Conflict of interest On behalf of all the authors, the corresponding author states that there is no conflict of interest.

References

1. Rotman, D.: Chat GPT is about to revolutionize the economy. We need to decide what that looks like. MIT Technology Review https://www.technologyreview.com/2023/03/25/1070275/chatgpt-revolutionize-economy-decide-what-looks-like (2023). Published 25 Mar 2023
2. Sevilla, G.: Accenture's $3B investment in AI signals a shift in big business. Insider Intelligence https://www.insiderintelligence.com/content/accenture-s-3b-investment-ai-signals-shift-big-business (2023). Published 14 Jun 2023
3. Prakash, P.: Even OpenAI CEO Sam Altman thinks people are going a little too crazy over A.I.: 'It's wildly overhyped in the short term'. Fortune https://fortune.com/2023/06/08/openai-ceo-sam-altman-a-i-wildly-overhyped (2023). Published 8 Jun 2023
4. Rosenberg, S.: In tech, everything is labeled 'AI' now. Axios https://www.axios.com/2023/06/20/ai-everything-definition-label-hype (2023). Published 20 Jun 2023
5. Cox, C.: Elon Musk's appetite for destruction. New York Times Magazine. https://www.nytimes.com/2023/01/17/magazine/tesla-autopilot-self-driving-elon-musk.html (2023). Published 17 Jan 2023
6. National Highway Transportation Safety Administration: [Recall] 2021 Tesla Model 3. NHTSA https://www.nhtsa.gov/vehicle/2021/TESLA/MODEL%2525203/4%252520DR/RWD#recalls (2023). Published 15 Feb 2023
7. Zipper, D.: The massive Tesla recall isn't just Elon Musk's fault. Slate https://slate.com/technology/2023/02/tesla-recall-full-self-driving-nhtsa-musk-regulation.html (2023). Published 16 Feb 2023
8. O'Connell, B,: Former Google engineer issues grave warning about sentient AI. TheStreet https://www.thestreet.com/breaking-news/this-former-google-engineer-calls-ai-bots-sentient (2023). Published 1 Mar 2023
9. Sloan, K.: A lawyer used ChatGPT to cite bogus cases. What are the ethics? Reuters https://www.reuters.com/legal/transactional/lawyer-used-chatgpt-cite-bogus-cases-what-are-ethics-2023-05-30/ (2023). Published 30 May 2023
10. Allyn, B.: A robot was scheduled to argue in court, then came the jail threats. NPR https://www.npr.org/2023/01/25/1151435033/a-robot-was-scheduled-to-argue-in-court-then-came-the-jail-threats (2023). Published 25 Jan 2023
11. Naughton, J.: The ChatGPT bot is causing panic now – but it'll soon be as mundane a tool as Excel. The Guardian (2023) https://www.theguardian.com/commentisfree/2023/jan/07/chatgpt-bot-excel-ai-chatbot-tech. Published 7 Jan 2023
12. Mollick, E.: ChatGPT is a tipping point for AI. Harvard Review https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai (2023). Published 14 December 2022
13. Gartner glossary: Hype cycle. Gartner https://www.gartner.com/en/information-technology/glossary/hype-cycle#:~:text=Gartner%27s%20Hype%20Cycle%20is%20a,technology%20maturity%20and%20future%20potential (2023). No publication date