W3: Prediction Machines: Chatbots & Generative AI

> Imagine if a pharmaceutical company released a new drug with no clinical trials and said it was testing the medication on the wider public. Or a food company released an experimental preservative with little scrutiny. That was how large tech firms were about to start deploying large language models to the public, because in their race to profit from such powerful tools, there were zero regulatory standards to follow. It was up to the safety and ethics researchers to study all the risks from inside these firms, but they were hardly a force to be reckoned with. At Google, their leaders had been fired. At DeepMind, they represented a tiny proportion of the research team. A signal was emerging more clearly each day. Get on board with the mission to build something bigger, or leave.

Reading Parmy Olson's account of Microsoft's AI coding assistant, GitHub Copilot, at the start of chapter 13 of her book about ChatGPT, I couldn't help but smile. I'm writing this post in Visual Studio, the IDE (Integrated Development Environment) that as Olson mentions, was the app used by coders in which GitHub Copilot was first implemented. Not being a programmer, I only started using Copilot this past summer, but I haven't looked back. Not only does the current version predictively suggest code, as Olson explains, it also suggests textual content as you're actually writing. It also isn't limited to Visual Studio—I also use it as a plugin for RStudio, the IDE that I use for web authoring. After enabling it last summer, when I started authoring posts I was shocked to discover it literally trying to finish my sentences—all I needed to do was hit TAB to accept the suggestion and a whole paragraph describing, for example, a videogame, would appear before my eyes. My initial instinct was to disable it, but I quickly realized that the rather generic overview that I'd been starting to write, and that the AI had auto-suggested to me, wasn't all that bad, with some editing. I could certainly use this sentence, and maybe this one…

I even used GitHub Copilot as an assistant while authoring the syllabus for this course—not to suggest reading assignments, but simply to complete citations for sources I was planning to use (in the bibliography, etc.) without having to type them out in full. What I also quickly noticed, however, was that the citations that were magically appearing before my eyes were frequently **wrong**, in terms of the information about sources that they were providing, even though they were impeccably formatted. This actually created more work as I realized that I had to fact-check every citation that Copilot was generating, apparently out of thin air.

So this is the dilemma that we all now increasingly find ourselves in. The fox is definitively in the henhouse (or whatever the metaphor is–Trojan Horse? Nah, it's already been used), and there seems no likelihood of ejecting it anytime soon. In many ways, indeed, the fox's presence not only seems not all that bad but has the potential to save us countless hours that had previously been spent on repetitive, time-consuming tasks. Aware of it or not, we are already using AI every day, whether in textual authoring (Grammerly), image processing (everything Adobe), or music DAWs (Ableton Live). AI is already on our laptops, our tablets, and our mobile devices.

Since our course is focused specifically on **mobile**, handheld technologies and their relation to social media, while reading Parmy Olson's book I've been thinking specifically about current implementations of AI on mobile devices. The recent launch of Apple's own AI funcionality, Apple Intelligence—tagged "AI for the rest of us"—is a case in point, although since it isn't supported either on my relatively ancient 2016 Macbook or my iPhone 13 Mini, I'm not going to be able to try it out until I upgrade to the more recent models that support it.

But by now, Apple Intelligence itself is rather late to the party: type "ChatGPT" as a search term into the App Store and you will be presented with literally dozens of chatbot apps, all promising AI-enhanced search. I encourage you to do this. Look more closely, though, and you'll see that a large proportion of these apps announce themselves as **powered by** a relatively small number of the LLMs (Large Language Models) discussed in Olson's book: OpenAI's ChatGPT, Anthropic's Claude, Google's Gemini, DALL-E 2, Stable Diffusion, or Mid-journey. In other words, the apps are basically just front-end interfaces to the models, act-ing as middelmen between us and them via the magic of APIs - essentially the code that runs much of the internet these days, that enables apps to share data between themselves.

So my first question this week would be about your own use—if any— of AI tools on your cellphones or tablets: do you now routinely use generative AI and/or any of this new gen-eration of research tools in your everyday life? If so, which ones? How have you been using them, and what has been your experience so far? Have you encountered any of the prob-lems that I mentioned earlier, about generative AI auto-suggesting incorrect information, or ChatGPT (or some other tool) returning random answers that have become known as **hallucinations**?

Parallel to the generative AI revolution itself, over the past five years there has been an explosion in both academic and popular books about the revolution itself and its wider im-plications across all dimensions of 21C society, whether ethical, political, economic, social, cultural, or aesthetic. Obviously we don't have time to explore all of these dimensions here, but personally I have been especially interested in exploring the cultural and aesthetic im-plications of generative AI. In this course, I wanted to assign to you some of the most in-teresting texts I've come across in this area in recent years, as well as some of the most recent and up-to-date ones: when the field itself seems to be transforming from one day to the next, before our very eyes, it's particularly important to keep up with current devel-opments, whether in publications or online platforms such as Medium or Substack. So part of what I wanted to do this week is simply to share with you some of the resources that I've found most informative and though-provoking in recent years on the much-debated topic of generative AI.

Chapter 12 of Parmy Olson's book is useful in laying out some of the background of the current debates around generative AI. In particular, I wanted you to read about Emily Ben-der's (et al.) influential Stochastic Parrots co-written article, which I'm linking to here. The article itself, although quite short, is an academic paper and therefore quite technical. It's important to know about and worth taking a look at, but a more user-friendly introduction is this interview with Emily Bender herself in **New York** magazine.

As you'll have seen, the title of Olson's chapter is "Mythbusters." A lot of the myth-busting around generative AI has been coming from linguists like Emily Bender, who have sought to debunk many of the wild and frequently dystopian claims about AI surpassing human-level intelligence, or even achieving consciousness (aka "sentience") in the case of ex-Google employee Blake Lemoine. As Gary Marcus and Ernest Jones put it in a chapter of their co-authored book **Rebooting AI** "If Computers Are So Smart, Why Can't They Read?" (also linked here). As the authors demonstrate in the chapter, even the most state-of-the-art AI today is not even capable of actually **understanding** a simple children's story in any meaningful human sense; it is only capable of extrapolating explicit information by pat-tern matching strings of words. Depending on the context, this may be a very useful af-fordance, but it is not even close to the richness of human understanding, which depends on something called **inferential** meanings. Marcus and Jones explain all this brilliantly in the chapter and I strongly encourage you to read it. The rest of their book is devoted to explaining what it would take for us to be able to design a genuine form of artificial intel-ligence, but as they argue in the book, this is somthing that is unlikely to happen in the near or even more distant future. Instead, we will have more convincing versions of what

Gary Marcus & Ernest Davis, **Rebo AI: Building Artificial Intelligen Can Trust** (New York: Pantheon B 2019).

we have now: simulations of intelligence, and chatbots that can pass as human under very restricted conditions.

Another source of critique of generative AI centers on **identity** and inclusivity, involving "bias" and assymetries in the representation particularly of women and minority identities in generative AI outputs–itself the result of insufficient inclusivity in the datasets on which LLMs are modeled. This in itself is a large area of debate. If you're interested in exploring this aspect of generative AI further, a useful documentary to watch is **Coded Bias** (Shalini Kantayya, 2020), available on Netflix.

Since this post is already too long, I'll finish by just including some links to useful (human-authored) sources for further exploration of the subject of generative AI in relation to social media and mobile technologies. Feel free to message me if you have any questions, if you'd like further recommendations, or you just want to chat about the subject!

Looking forward to this week's discussion on Discord!

Sources (referenced in chs. 12-13 of Parmy Olson's book **Supremacy**)

Jennifer Nicholson, "The Gender Bias Inside GPT-3" (**Medium**, 8 March 2022).

Parmy Olson, "My Girlfriend is a Chatbot" (**Wall Street Journal**, 10 April 2020).

Tom Simonite, "What Really Happened When Google Ousted Timnit Gebru" (**WIRED**, 8 June 2021).