Man v machine: everything you need to know about AI

Publication Date: 2023-05-06

Author: Alex Hern

Section: Technology

Tags: Artificial intelligence (AI), The Observer, ChatGPT, Chatbots, Computing, Google, Microsoft, Bing, explainers

Article URL: https://www.theguardian.com/technology/2023/may/06/man-v-machine-everything-you-need-to-know-about-ai

Where might I start to encounter more chatbots or AI content? Almost anywhere you currently interact with other people is being eagerly assessed for AI-based disruption. Chatbots in customer service roles are nothing new but, as AI systems become more capable, expect to encounter more and more of them, handling increasingly complex tasks. Voice synthesis and recognition technology means they'll also answer the phone, and even call you. The systems will also power low-cost content generation across the web. Already, they're being used to fill "content farm" news sites, with one recent study finding almost 50 news websites that hosted some form of obviously AI-generated material, rarely labelled as such. And then there are the less obvious cases. The systems can be used to label and organise data, to help in the creation of simple programs, to summarise and generate work emails – anything where text is available, someone will try to hand it to a chatbot. What are the differences between ChatGTP, Bard and Bing AI? All three systems are built on the same foundation, a type of AI technology called a "large language model", or LLM, but with small differences in application that can lead to large variety in output. ChatGPT is based on OpenAI's GPT LLM, fine-tuned with a system called "reinforcement-learned human feedback" (RLHF). In giant "call centres", staffed by workers paid as little as $2 an hour, the company asked human trainers to hold, and rate, millions of chat-style conversations with GPT, teaching the AI what a good response is and what a bad response is. However, ChatGPT can't know the answer to any question after its training data was set, in around 2021. Microsoft has revealed little about how Bing chat works behind the scenes, but it seems to take a simpler approach, called "prompting". The bot, also built on top of OpenAI's GPT, is invisibly given the same text input before each conversation, telling it that, for instance, it is a helpful assistant, it is expected to be polite and friendly, and that it should not answer questions that might be dangerous. Bing also has an ace up its sleeve: a live connection to the web, which it can use to pull information in to supplement its answers. The approach is cheap and mostly effective, but opens the system up to "prompt injection" attacks, where users trick the AI into ignoring its own rules in favour of new ones instead. Sometimes, a prompt injection can also come from the web information Bing tries to read to answer queries. Google's Bard sits somewhere between the two. It is built on the company's own Palm system, again fine-tuned with the same RLHF system as ChatGPT. Like Bing, though, Bard can also look information up on the internet, bringing live data in to update its knowledge. Are they going to get more powerful and capable, and how quickly? We don't know, but probably. One of the key breakthroughs in recent years has been that quantity trumps quality: the more processing power and the more data an AI has, the better it is. Efforts to only give it good data are less important than simply giving it more and more. And on that metric, we are only getting started: AI systems have been fed a substantial amount of the public text on the internet, for instance, but nowhere near all the data a company like Google holds when private data is considered. And the cost of computing power for a system like GPT-4 was around $100m –

we don't know what will happen when they're handed billions. There could be a limit. If there are diminishing returns to more data, and we're running out of sources, it might get hard to improve systems much beyond where they are today. But there could also be a "flywheel effect", where AI systems can be used to make AI systems better. Some approaches, for instance, have tried training AI using data generated by other AI – and they seem to work. The uses sound quite benign. Why are experts linking AI to the end of humanity or society as we know it?! We don't know what happens if we build an AI system that is smarter than humans at everything it does. Perhaps a future version of ChatGPT, for instance, decides that the best way it can help people answer questions is by slowly manipulating people into putting it in charge. Or an authoritarian government hands too much autonomy to a battlefield robotics system, which decides the best way to achieve its task of winning a war is to first hold a coup in its own country. "You need to imagine something more intelligent than us by the same difference that we're more intelligent than a frog," says Geoffrey Hinton, one of the inventors of the neural network. Can I trust what a chatbot tells me? ChatGPT, Bing and Bard have produced factual errors, or hallucinations as they are known in the industry jargon. For instance, ChatGPT falsely accused an American law professor of sexual harassment and cited a non-existent Washington Post report, while a promotional video for Bard gave an inaccurate answer to a query about the James Webb Space Telescope. Chatbots are trained on astronomical amounts of data taken from the internet. Operating in a way akin to predictive text, they build a model to predict the likeliest word or sentence to come after the user's prompt. This can result in factual errors, but the plausible nature of the responses can trick users into thinking a response is 100% correct. There are also concerns that the technology behind chatbots could be used to produce disinformation at a significant scale. Last week, Microsoft's chief economist warned that AI could "do a lot of damage in the hands of spammers with elections and so on". How can I tell if my job is at risk from AI? Listen to the tech executives. Asked recently what jobs would be disrupted by AI, Sundar Pichai, the Google chief executive, said: "Knowledge workers." This means writers, accountants, architects, lawyers, software engineers – and more. OpenAI's CEO, Sam Altman, has identified customer service as a vulnerable category where he says there would be "just way fewer jobs relatively soon". The boss of technology group IBM, Arvind Krishna, has said he expects nearly 8,000 back-office jobs at the business, like human resources roles, to be replaced by AI over a five-year period. Last week, shares in education firms were hit after Chegg, a US provider of online help for students' writing and maths assignments, warned that ChatGPT was hitting customer growth. Also last week, the World Economic Forum published a survey of more than 800 companies with a total of 11.3 million employees. A quarter of the firms said they expected AI to create job losses, although 50% said they expected it to spur jobs growth. Who will make money from AI? The big tech companies at the forefront of AI development are San Francisco-based OpenAI, Google's parent Alphabet and Microsoft, which is also an investor in OpenAI. Prominent AI startups include British firm Stability AI – the company behind image generator Stable Diffusion – and Anthropic. For now, the private sector is leading the development race and is in a leading position to gain financially. According to the annual AI Index Report, the tech industry produced 32 significant machine-learning models last year, compared with three produced by academia. In terms of companies that will make money from applying generative AI (the term for chatbots, sound and image generators that produce plausible text, images and sound in response to human prompts), then the development of open-source AI models might throw open the potential gains to the wider economy. Some of it sounds dangerous; why is it being released to the public without regulation? The recent history of tech regulation is that governments and regulators scramble into action once the technology has already been unleashed. For instance, nearly two decades after the launch of Facebook, the UK government is only just on the verge of implementing the online safety bill, which seeks to limit the harms caused by social media. The same is happening with AI. Last week, the White House announced measures to address concerns about unchecked AI development, but they will not halt the AI arms race on their own . The UK competition watchdog, the Competition and Markets Authority, launched a review into the sector, but that will not report initial findings until September. The EU parliament will hold a vote on the AI act, although negotiations on shaping the legislation will continue after that. In the meantime, the most overt calls from restraint are coming from AI professionals. In March, Elon Musk – a co-founder of OpenAI – was among the signatories to a letter calling for a pause in major AI projects. Pichai, who has said he loses sleep over the pace of AI development, has called for nuclear arms-style global regulation of the technology. So far there is no sign of the development race slowing, or the emergence of a global framework to moderate it.