We come to bury ChatGPT, not to praise it.

by Dan McQuillan

Mon 06 February 2023

Large language models (LLMs) like the GPT family learn the statistical structure of language by optimising their ability to predict missing words in sentences (as in 'The cat sat on the [BLANK]'). Despite the impressive technical ju-jitsu of transformer models and the billions of parameters they learn, it's still a computational guessing game. ChatGPT is, in technical terms, a 'bullshit generator'. If a generated sentence makes sense to you, the reader, it means the mathematical model has made sufficiently good guess to pass your sense-making filter. The language model has no idea what it's talking about because it has no idea about anything at all. It's more of a bullshitter than the most egregious egoist you'll ever meet, producing baseless assertions with unfailing confidence because that's what it's designed to do. It's a bonus for the parent corporation when journalists and academics respond by generating acres of breathless coverage, which works as PR even when expressing concerns about the end of human creativity.

Unsuspecting users who've been conditioned on Siri and Alexa assume that the smooth talking ChatGPT is somehow tapping into reliable sources of knowledge, but it can only draw on the (admittedly vast) proportion of the internet it ingested at training time. Try asking Google's BERT model about Covid or ChatGPT about the latest Russian attacks on Ukraine. Ironically, these models are unable to cite their own sources, even in instances where it's obvious

they're plagiarising their training data. The nature of ChatGPT as a bullshit generator makes it harmful, and it becomes more harmful the more optimised it becomes. If it produces plausible articles or computer code it means the inevitable hallucinations are becoming harder to spot. If a language model suckers us into trusting it then it has succeeded in becoming the industry's holy grail of 'trustworthy AI'; the problem is, trusting any form of machine learning is what leads to a single mother having their front door kicked open by social security officials because a predictive algorithm has fingered them as a probable fraudster, alongside many other instances of algorithmic violence.

Of course, the makers of GPT learned by experience that an untended LLM will tend to spew Islamophobia or other hatespeech in addition to talking nonsense. The technical addition in ChatGPT is known as Reinforcement Learning from Human Feedback (RHLF). While the whole point of an LLM is that the training data set is too huge for human labelling, a small subset of curated data is used to build a monitoring system which attempts to constrain output against criteria for relevance and non-toxicity. It can't change the fact that the underlying language patterns were learned from the raw internet, including all the ravings and conspiracy theories. While RLHF makes for a better brand of bullshit, it doesn't take too much ingenuity in user prompting to reveal the bile that can lie beneath. The more plausible ChatGPT becomes, the more it recapitulates the pseudo-authoritative rationalisations of race science. It also shows that despite the boast that LLMs are largely self-training, any real world system will require precaritised 'ghost work' to maintain its plausibility. It turns out that AI is not sci-fi but a techologised intensification of existing relations of labour and power. The $2/hour paid to outsourced workers in Kenya so they could be "tortured" by having to tag obscene material for removal is figurative of the invisible and gendered labour of care that always already holds up our existing systems of business and government.

As with the rest of AI, the dangers of ChatGPT go far deeper than bias and discrimination. Despite evidence that the model's powers of 'reasoning' are shallow heuristics based on the frequency of associations in the training data (meaning, as an illustrative example, that it's good at answering 'What is 24 x 18?' and poor at answering

'What is 23 x 18?') there are many in the AI community who insist on imputing emergent properties of reasoning and insight to ChatGPT. Its parent company, OpenAI, was set up "to ensure that artificial general intelligence benefits all of humanity", where 'artificial general intelligence' (AGI) is the insider term used for human-like intelligence that goes beyond narrow AI like facial recognition or self-driving cars. However, as I spell out in my book, the concept of AGI is inseparable from the kind of hierarchy of intelligence that has underpinned ideas of innate supremacy since the days of empire and colonialism. Hardly surprising, then, that the same Silicon Valley cultures that incubate enthusiasm for ChatGPT as emergent AGI also show allegiance to associated world views like Long Termism, where the immediate vulnerability of millions of ordinary people counts as nothing in relation to the prospects of a future space-faring super race.

In the mean time, OpenAI is acquiring billions of dollars of investment on the back of the ChatGPT hype. The point here is not only the pocketing of a pyramid-scale payoff but the reasons why institutions and governments are prepared to invest so much in these technologies. For these players, the seductive vision isn't real AI (whatever that is) but technologies that are good enough to replace human workers or, more importantly, to precaritise them and undermine them. ChatGPT isn't really new but simply an iteration of the class war that's been waged since the start of the industrial revolution. That allegedly well-informed commentators can infer that ChatGPT will be used for "cutting staff workloads" rather than for further staff cuts illustrates a general failure to understand AI as a political project. Contemporary AI, as I argue in my book, is an assemblage for automatising administrative violence and amplifying austerity. ChatGPT is a part of a reality distortion field that obscures the underlying extractivism and diverts us into asking the wrong questions and worrying about the wrong things. Instead of expressing wonder, we should be asking whether it's justifiable to burn energy at "eye watering" rates to power the world's largest bullshit machine.

Commentary that claims 'ChatGPT is here to stay and we just need to learn to live with it' are embracing the hopelessness of what I call 'AI Realism'. The compulsion to show 'balance' by always referring to AI's alleged potential for good should be dropped by

acknowledging that the social benefits are still speculative while the harms have been empirically demonstrated. Saying, as the OpenAI CEO does, that we are all 'stochastic parrots' like large language models, statistical generators of learned patterns that express nothing deeper, is a form of nihilism. Of course, the elites don't apply that to themselves, just to the rest of us. The structural injustices and supremacist perspectives layered into AI put it firmly on the path of eugenicist solutions to social problems.

Instead of reactionary solutionism, let us ask where the technologies are that people really need. Let us reclaim the idea of socially useful production, of technological developments that start from community needs. The post-Covid 'new normal' has turned out to involve both the normalisation of neural networks and a rise in necropolitics. Transformer models and diffusion models are not creative but carceral - they and other forms of AI imprison our ability to imagine real alternatives. It's not so long ago that we all woke up to the identity of truly essential workers; the people carrying out the precaritised roles of nursing, teaching, caring, delivering and cleaning, the very professions who are being forced to reinvent the idea of the general strike simply to regain the conditions for survival. Instead of being complicit with expensive toys running in carbon emitting data centres, we can focus instead on centring activities of care. As discussed in more detail in *'Resisting AI'*, a refusal of algorithmic immiseration goes along with a positive search for alternatives, and I lay out a programme of people's councils and commons-based solidarity to do just that. It's not time to chat with AI, but to resist it.
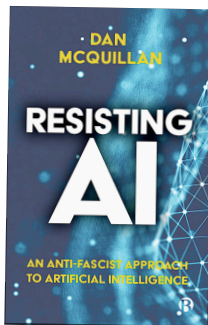
319 N. 11th St. #2I, PHL
iffybooks.net

Read the original blog post, with hyperlinked references:
https://www.danmcquillan.org/chatgpt.html