AI machines aren't 'hallucinating'. But their makers are

Naomi Klein

Tech CEOs want us to believe that generative AI will benefit humanity. They are kidding themselves

Inside the many debates swirling around the rapid rollout of so-called artificial intelligence, there is a relatively obscure skirmish focused on the choice of the word "hallucinate".

This is the term that architects and boosters of generative AI have settled on to characterize responses served up by chatbots that are wholly manufactured, or flat-out wrong. Like, for instance, when you ask a bot for a definition of something that doesn't exist and it, rather convincingly, gives you one, complete with made-up footnotes. "No one in the field has yet solved the hallucination problems," Sundar Pichai, the CEO of Google and Alphabet, told an interviewer recently.

That's true – but why call the errors "hallucinations" at all? Why not algorithmic junk? Or glitches? Well, hallucination refers to the mysterious capacity of the human brain to perceive phenomena that are not present, at least not in conventional, materialist terms. By appropriating a word commonly used in psychology, psychedelics and various forms of mysticism, AI's boosters, while acknowledging the fallibility of their machines, are simultaneously feeding the sector's most cherished mythology: that by building these large language models, and training them on everything that we humans have written, said and represented visually, they are in the process of birthing an animate intelligence on the cusp of sparking an evolutionary leap for our species. How else could bots like Bing and Bard be tripping out there in the ether?

Warped hallucinations are indeed afoot in the world of AI, however – but it's not the bots that are having them; it's the tech CEOs who unleashed them, along with a phalanx of their fans, who are in the grips of wild hallucinations, both individually and collectively. Here I am defining hallucination not in the mystical or psychedelic sense, mind-altered states that can indeed assist in accessing profound, previously unperceived truths. No. These folks are just tripping: seeing, or at least claiming to see, evidence that is not there at all, even conjuring entire worlds that will put their products to use for our universal elevation and education.

Generative AI will end poverty, they tell us. It will cure all disease. It will solve climate change. It will make our jobs more meaningful and exciting. It will unleash lives of leisure and contemplation, helping us reclaim the humanity we have lost to late capitalist mechanization. It will end loneliness. It will make our governments rational and responsive. These, I fear, are the real AI hallucinations and we have all been hearing them on a loop ever since Chat GPT launched at the end of last year.

There is a world in which generative AI, as a powerful predictive research tool and a performer of tedious tasks, could indeed be marshalled to benefit humanity, other species and our shared home. But for that to happen, these technologies would need to be deployed inside a vastly different economic and social order than our own, one that had as its purpose the meeting of human needs and the protection of the planetary systems that support all life.

And as those of us who are not currently tripping well understand, our current system is nothing like that. Rather, it is built to maximize the extraction of wealth and profit – from

both humans and the natural world – a reality that has brought us to what we might think of it as capitalism's techno-necro stage. In that reality of hyper-concentrated power and wealth, AI – far from living up to all those utopian hallucinations – is much more likely to become a fearsome tool of further dispossession and despoilation.

I'll dig into why that is so. But first, it's helpful to think about the *purpose* the utopian hallucinations about AI are serving. What work are these benevolent stories doing in the culture as we encounter these strange new tools? Here is one hypothesis: they are the powerful and enticing cover stories for what may turn out to be the largest and most consequential theft in human history. Because what we are witnessing is the wealthiest companies in history (Microsoft, Apple, Google, Meta, Amazon …) unilaterally seizing the sum total of human knowledge that exists in digital, scrapable form and walling it off inside proprietary products, many of which will take direct aim at the humans whose lifetime of labor trained the machines without giving permission or consent.

This should not be legal. In the case of copyrighted material that we now know trained the models (including this newspaper), various lawsuits have been filed that will argue this was clearly illegal. Why, for instance, should a for-profit company be permitted to feed the paintings, drawings and photographs of living artists into a program like Stable Diffusion or Dall-E 2 so it can then be used to generate doppelganger versions of those very artists' work, with the benefits flowing to everyone but the artists themselves?

The painter and illustrator Molly Crabapple is helping lead a movement of artists challenging this theft. "AI art generators are trained on enormous datasets, containing millions upon millions of copyrighted images, harvested without their creator's knowledge, let alone compensation or consent. This is effectively the greatest art heist in history. Perpetrated by respectable-seeming corporate entities backed by Silicon Valley venture capital. It's daylight robbery," a new open letter she co-drafted states.

The trick, of course, is that Silicon Valley routinely calls theft "disruption" – and too often gets away with it. We know this move: charge ahead into lawless territory; claim the old rules don't apply to your new tech; scream that regulation will only help China – all while you get your facts solidly on the ground. By the time we all get over the novelty of these new toys and start taking stock of the social, political and economic wreckage, the tech is already so ubiquitous that the courts and policymakers throw up their hands.

We saw it with Google's book and art scanning. With Musk's space colonization. With Uber's assault on the taxi industry. With Airbnb's attack on the rental market. With Facebook's promiscuity with our data. Don't ask for permission, the disruptors like to say, ask for forgiveness. (And lubricate the asks with generous campaign contributions.)

In The Age of Surveillance Capitalism, Shoshana Zuboff meticulously details how Google's Street View maps steamrolled over privacy norms by sending its camera-bedecked cars out to photograph our public roadways and the exteriors of our homes. By the time the lawsuits defending privacy rights rolled around, Street View was already so ubiquitous on our devices (and so cool, and so convenient …) that few courts outside Germany were willing to intervene.

Now the same thing that happened to the exterior of our homes is happening to our words, our images, our songs, our entire digital lives. All are currently being seized and used to train the machines to simulate thinking and creativity. These companies must know they are engaged in theft, or at least that a strong case can be made that they are. They are just hoping that the old playbook works one more time – that the scale of the heist is already so large and unfolding with such speed that courts and policymakers will once again throw up their hands in the face of the supposed inevitability of it all.

It's also why their hallucinations about all the wonderful things that AI will do for humanity are so important. Because those lofty claims disguise this mass theft as a gift – at the same time as they help rationalize AI's undeniable perils.

By now, most of us have heard about the survey that asked AI researchers and developers to estimate the probability that advanced AI systems will cause "human extinction or similarly permanent and severe disempowerment of the human species". Chillingly, the median response was that there was a 10% chance.
How does one rationalize going to work and pushing out tools that carry such existential risks? Often, the reason given is that these systems also carry huge potential upsides – except that these upsides are, for the most part, hallucinatory. Let's dig into a few of the wilder ones.

Hallucination #1: AI will solve the climate crisis

Almost invariably topping the lists of AI upsides is the claim that these systems will somehow solve the climate crisis. We have heard this from everyone from the World Economic Forum to the Council on Foreign Relations to Boston Consulting Group, which explains that AI "can be used to support all stakeholders in taking a more informed and data-driven approach to combating carbon emissions and building a greener society. It can also be employed to reweight global climate efforts toward the most at-risk regions." The former Google CEO Eric Schmidt summed up the case when he told the Atlantic that AI's risks were worth taking, because "If you think about the biggest problems in the world, they are all really hard – climate change, human organizations, and so forth. And so, I always want people to be smarter."
According to this logic, the failure to "solve" big problems like climate change is due to a deficit of smarts. Never mind that smart people, heavy with PhDs and Nobel prizes, have been telling our governments for decades what needs to happen to get out of this mess: slash our emissions, leave carbon in the ground, tackle the overconsumption of the rich and the underconsumption of the poor because no energy source is free of ecological costs.

The reason this very smart counsel has been ignored is not due to a reading comprehension problem, or because we somehow need machines to do our thinking for us. It's because doing what the climate crisis demands of us would strand trillions of dollars of fossil fuel assets, while challenging the consumption-based growth model at the heart of our interconnected economies. The climate crisis is not, in fact, a mystery or a riddle we haven't yet solved due to insufficiently robust data sets. We know what it would take, but it's not a quick fix – it's a paradigm shift. Waiting for machines to spit out a more palatable and/or profitable answer is not a cure for this crisis, it's one more symptom of it.
Clear away the hallucinations and it looks far more likely that AI will be brought to market in ways that actively deepen the climate crisis. First, the giant servers that make instant essays and artworks from chatbots possible are an enormous and growing source of carbon emissions. Second, as companies like Coca-Cola start making huge investments to use generative AI to sell more products, it's becoming all too clear that this new tech will be used in the same ways as the last generation of digital tools: that what begins with lofty promises about spreading freedom and democracy ends up micro targeting ads at us so that we buy more useless, carbon-spewing stuff.
And there is a third factor, this one a little harder to pin down. The more our media channels are flooded with deep fakes and clones of various kinds, the more we have the feeling of sinking into informational quicksand. Geoffrey Hinton, often referred to as "the godfather of AI" because the neural net he developed more than a decade ago forms the building blocks of today's large language models, understands this well. He just quit a senior role at Google so that he could speak freely about the risks of the technology he helped create, including, as he told the New York Times, the risk that people will "not be able to know what is true anymore".

This is highly relevant to the claim that AI will help battle the climate crisis. Because when we are mistrustful of everything we read and see in our increasingly uncanny media

environment, we become even less equipped to solve pressing collective problems. The crisis of trust predates ChatGPT, of course, but there is no question that a proliferation of deep fakes will be accompanied by an exponential increase in already thriving conspiracy cultures. So what difference will it make if AI comes up with technological and scientific breakthroughs? If the fabric of shared reality is unravelling in our hands, we will find ourselves unable to respond with any coherence at all.

Hallucination #2: AI will deliver wise governance

This hallucination summons a near future in which politicians and bureaucrats, drawing on the vast aggregated intelligence of AI systems, are able "to see patterns of need and develop evidence-based programs" that have greater benefits to their constituents . That claim comes from a paper published by the Boston Consulting Group's foundation, but it is being echoed inside many thinktanks and management consultancies. And it's telling that these particular companies – the firms hired by governments and other corporations to identify costs savings, often by firing large numbers of workers – have been quickest to jump on the AI bandwagon. PwC (formerly PricewaterhouseCoopers) just announced a $1bn investment, and Bain & Company as well as Deloitte are reportedly enthusiastic about using these tools to make their clients more "efficient".

As with the climate claims, it is necessary to ask: is the reason politicians impose cruel and ineffective policies that they suffer from a lack of evidence? An inability to "see patterns," as the BCG paper suggests? Do they not understand the human costs of starving public healthcare amid pandemics, or of failing to invest in non-market housing when tents fill our urban parks, or of approving new fossil fuel infrastructure while temperatures soar? Do they need AI to make them "smarter", to use Schmidt's term – or are they precisely smart enough to know who is going to underwrite their next campaign, or, if they stray, bankroll their rivals?

It would be awfully nice if AI really could sever the link between corporate money and reckless policy making – but that link has everything to do with why companies like Google and Microsoft have been allowed to release their chatbots to the public despite the avalanche of warnings and known risks. Schmidt and others have been on a years-long lobbying campaign telling both parties in Washington that if they aren't free to barrel ahead with generative AI, unburdened by serious regulation, then western powers will be left in the dust by China. Last year, the top tech companies spent a record $70m to lobby Washington – more than the oil and gas sector – and that sum, Bloomberg News notes, is on top of the millions spent "on their wide array of trade groups, non-profits and thinktanks".

And yet despite their intimate knowledge of precisely how money shapes policy in our national capitals, when you listen to Sam Altman, the CEO of OpenAI – maker of ChatGPT – talk about the best-case scenarios for his products, all of this seems to be forgotten. Instead, he seems to be hallucinating a world entirely unlike our own, one in which politicians and industry make decisions based on the best data and would never put countless lives at risk for profit and geopolitical advantage. Which brings us to another hallucination.

Hallucination #3: tech giants can be trusted not to break the world

Asked if he is worried about the frantic gold rush ChatGPT has already unleashed, Altman said he is, but added sanguinely: "Hopefully it will all work out." Of his fellow tech CEOs – the ones competing to rush out their rival chatbots – he said: "I think the better angels are going to win out."

Better angels? At Google? I'm pretty sure the company fired most of those because they were publishing critical papers about AI, or calling the company out on racism and sexual harassment in the workplace. More "better angels" have quit in alarm, most recently Hinton. That's because, contrary to the hallucinations of the people profiting most from AI, Google does not make decisions based on what's best for the world – it makes decisions based on what's best for Alphabet's shareholders, who do not want to miss the latest bubble, not when Microsoft, Meta and Apple are already all in.

Hallucination #4: AI will liberate us from drudgery

If Silicon Valley's benevolent hallucinations seem plausible to many, there is a simple reason for that. Generative AI is currently in what we might think of as its faux-socialism stage. This is part of a now familiar Silicon Valley playbook. First, create an attractive product (a search engine, a mapping tool, a social network, a video platform, a ride share …); give it away for free or almost free for a few years, with no discernible viable business model ("Play around with the bots," they tell us, "see what fun things you can create!"); make lots of lofty claims about how you are only doing it because you want to create a "town square" or an "information commons" or "connect the people", all while spreading freedom and democracy (and not being "evil"). Then watch as people get hooked using these free tools and your competitors declare bankruptcy. Once the field is clear, introduce the targeted ads, the constant surveillance, the police and military contracts, the black-box data sales and the escalating subscription fees.

Many lives and sectors have been decimated by earlier iterations of this playbook, from taxi drivers to rental markets to local newspapers. With the AI revolution, these kinds of losses could look like rounding errors, with teachers, coders, visual artists, journalists, translators, musicians, care workers and so many others facing the prospect of having their incomes replaced by glitchy code.

Don't worry, the AI enthusiasts hallucinate – it will be wonderful. Who likes work anyway? Generative AI won't be the end of employment, we are told, only "boring work" – with chatbots helpfully doing all the soul-destroying, repetitive tasks and humans merely supervising them. Altman, for his part, sees a future where work "can be a broader concept, not something you have to do to be able to eat, but something you do as a creative expression and a way to find fulfillment and happiness".
That's an exciting vision of a more beautiful, leisurely life, one many leftists share (including Karl Marx's son-in-law, Paul Lafargue, who wrote a manifesto titled The Right To Be Lazy). But we leftists also know that if earning money is to no longer be life's driving imperative, then there must be other ways to meet our creaturely needs for shelter and sustenance. A world without crappy jobs means that rent has to be free, and healthcare has to be free, and every person has to have inalienable economic rights. And then suddenly we aren't talking about AI at all – we're talking about socialism.
Because we do not live in the Star Trek-inspired rational, humanist world that Altman seems to be hallucinating. We live under capitalism, and under that system, the effects of flooding the market with technologies that can plausibly perform the economic tasks of countless working people is not that those people are suddenly free to become philosophers and artists. It means that those people will find themselves staring into the abyss – with actual artists among the first to fall.

That is the message of Crabapple's open letter, which calls on "artists, publishers, journalists, editors and journalism union leaders to take a pledge for human values against the use of generative-AI images" and "commit to supporting editorial art made by people, not server farms". The letter, now signed by hundreds of artists, journalists and others, states that all but the most elite artists find their work "at risk of extinction". And according to Hinton, the "godfather of AI", there is no reason to believe that the threat

won't spread. The chatbots take "away the drudge work" but "it might take away more than that".

Crabapple and her co-authors write: "Generative AI art is vampirical, feasting on past generations of artwork even as it sucks the lifeblood from living artists." But there are ways to resist: we can refuse to use these products and organize to demand that our employers and governments reject them as well. A letter from prominent scholars of AI ethics, including Timnit Gebru who was fired by Google in 2020 for challenging workplace discrimination, lays out some of the regulatory tools that governments can introduce immediately – including full transparency about what data sets are being used to train the models. The authors write: "Not only should it always be clear when we are encountering synthetic media, but organizations building these systems should also be required to document and disclose the training data and model architectures .... We should be building machines that work for us, instead of 'adapting' society to be machine readable and writable."

Though tech companies would like us to believe that it is already too late to roll back this human-replacing, mass-mimicry product there are highly relevant legal and regulatory precedents that can be enforced. For instance, the US Federal Trade Commission (FTC) forced Cambridge Analytica, as well as Everalbum, the owner of a photo app, to destroy entire algorithms found to have been trained on illegitimately appropriated data and scraped photos. In its early days, the Biden administration made many bold claims about regulating big tech, including cracking down on the theft of personal data to build proprietary algorithms. With a presidential election fast approaching, now would be a good time to make good on those promises – and avert the next set of mass layoffs before they happen.

A world of deep fakes, mimicry loops and worsening inequality is not an inevitability. It's a set of policy choices. We can regulate the current form of vampiric chatbots out of existence – and begin to build the world in which AI's most exciting promises would be more than Silicon Valley hallucinations.

Because we trained the machines. All of us. But we never gave our consent. They fed on humanity's collective ingenuity, inspiration and revelations (along with our more venal traits). These models are enclosure and appropriation machines, devouring and privatizing our individual lives as well as our collective intellectual and artistic inheritances. And their goal never was to solve climate change or make our governments more responsible or our daily lives more leisurely. It was always to profit off mass immiseration, which, under capitalism, is the glaring and logical consequence of replacing human functions with bots.

Is all of this overly dramatic? A stuffy and reflexive resistance to exciting innovation? Why expect the worse? Altman reassures us: "Nobody wants to destroy the world." Perhaps not. But as the ever-worsening climate and extinction crises show us every day, plenty of powerful people and institutions seem to be just fine knowing that they are helping to destroy the stability of the world's life-support systems, so long as they can keep making record profits that they believe will protect them and their families from the worst effects. Altman, like many creatures of Silicon Valley, is himself a prepper: back in 2016, he boasted: "I have guns, gold, potassium iodide, antibiotics, batteries, water, gas masks from the Israeli Defense Force and a big patch of land in Big Sur I can fly to."

I'm pretty sure those facts say a lot more about what Altman actually believes about the future he is helping unleash than whatever flowery hallucinations he is choosing to share in press interviews.

- Naomi Klein is a Guardian US columnist and contributing writer. She is the bestselling author of No Logo and The Shock Doctrine and Professor of Climate Justice and Co-director of the Centre for Climate Justice at the University of British Columbia

- Source: https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein