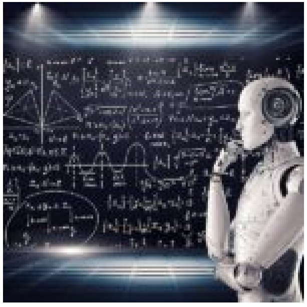
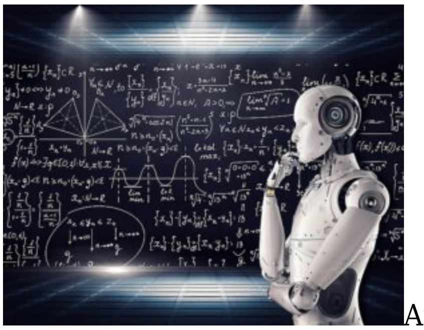
Assessing and Regulating Artificial Intelligence

August 9, 2023





Artificial Intelligence (AI) can be defined as machines generating human-like outcomes for tasks that we humans do with our brains.

AI began roughly six decades ago as a scientific research program meant to enlarge our understanding of human mental processes. To that end, it failed. Over time, an engineering aspect rose to prominence to accomplish tasks that matter. To that end too, results were spotty and limited for decades. Then about 25 years ago some steps began succeeding, and about seven years ago, an innovation opened new prospects, and into the present various incarnations of what's called neural networks and large language models have dramatically changed the scene.

The pace continues. What is tumultuous news one day is replaced with new tumultuous news in months, weeks, or even days. We wonder, what is happening, why is it important, and what ought leftists concerned to build a better world do about it?

As I write, machines paint pictures, compose music, diagnose diseases, and research and prepare legal opinions. They write technical manuals, news reports, essays, stories, and perhaps even novels. Machines code software and design buildings. They ace an incredible diversity of exams. Right now, in most states, machines could pass the bar exam and become lawyers. For all I know they have probably passed medical licensing exams as well. Machines now provide or will soon provide mental health counseling, elder care, personal support, and even intimate companionship. Machines converse, find patterns, design products, and solve complex problems (like protein folding). You can now feed an AI an entire book, not just a short prompt, but say 75,000 words, and in seconds ask it for an outline or summary, or for other reactions. Next week, what will it be? You can feed it 100,000 or 200,000 words and get a translation in minutes? Similarly, you can now feed in 6 hours of audio, and get a transcription. And, hot off the wires, perhaps most consequentially, AIs can now collaborate and even prompt one another.

Prompt an AI for an article on some topic, perhaps inflation or Covid, written in some style, perhaps humorously like Jon Stewart, and almost instantly you get what you asked for. Prompt an AI to undertake some kind of major project—solve the current housing crisis in California or end global warming—and AI can break the large-scale project into an agenda of component tasks. Then it can undertake the various tasks including prompting other programs and AIs for help when needed. More, when a step proves impossible for your favorite AI and its helpers, it will try to develop a new approach, change its agenda, and proceed anew until success. Note, once it embarks on the big project that you assign it, this new feature means it has considerable autonomy. It can prompt itself. It can prompt other Als. It can use other programs. A few years back, that kind of thing was thought to be ten, twenty, or more years off. Now it is here. As you read this article, who knows what new things AIs will be doing?

Some say AI news is all hype. They say AI is feeble. It is failure prone. But the current version of AI is barely an infant. And in any event, how many humans can create very complex pictures, compose music, read and summarize reports, and write and program better than even today's fledgling AIs, much less do these things in seconds? Very, very few.

Does current AI make mistakes? Definitely, including many humdingers. Then again, so do humans. And in any event, what matters is AI's trajectory, not its current status. Anecdotes about weird failures made yesterday are amusing. Assessments of next year's likely abilities are a wholly different matter. Criticizing today's occasional blunders is like laughing at a baby gurgling while ignoring that this particular baby can age years in minutes. I suspect GPT-2 wouldn't have known a legal bar exam from a broom. GPT-3 took a bar exam and scored in the bottom 10 percent making lots of mistakes that legal scholars likely laughed at. A year later GPT- 4 scored in the top 10 percent. Who's laughing now?

Maybe we should watch the trajectory. Maybe we shouldn't get stuck on a snapshot of a gurgling moment. Perhaps we shouldn't ignore that the AI's bar exam results were not compared to random humans plucked off the street. They were compared to law school graduates. What will GPT-5 score next year? And what will happen to its number of errors, however many it is still making, when a stone's throw down the road one neural net will routinely send results to a second neural net or to a fact checker program to grade, and then the first will correct errors reported back by the second and only then deliver its updated results to us? Will it be better than 99 percent of law students? Will all its current silly and easily fact-checkable errors disappear?

I hypothesized about the above scenario with those same words a couple of weeks prior to writing this essay. I did so to try to put anecdotal stories about AI hallucinations in context. A couple of scholars saying AI is all hype told me it would be years if ever before we saw such a self-checking capability. Now I can watch a YouTube video of an OpenAI tech officer displaying exactly that process with a version of GPT-4 still not public. It should be clear that we ought to examine the trajectory, not laugh at snapshots of gurgling moments.

So, what is this new AI that is no longer knocking at our door, but already conversing with us from within numerous programs we use?

The big picture is that AI achieves human-like outcomes for lots of tasks we do with our brains. More, it does this incredibly faster and often much better than us. And far from having plateaued, at least so far AI does more things better each month, even each week, at what is at the moment an incredible pace of innovation followed closely by proliferation into mass use. Yes, this is all just profit-seeking firms struggling to stay ahead of other profit-seeking firms, each using the breakthrough, breakneck technical assets of AI today to have their own still more profitable assets tomorrow. Indeed, these facilities have almost overnight grown from unique and mind-boggling lab adventures to tools dispersed for personal or group use by anyone anywhere via applications available world-wide including search engines, social media platforms, word processors, text-to-picture tools, text-to-video tools, text-to-software tools, prompts-to-text platforms, and so on. These are all within your reach, wherever you may be. If this proliferation stays unregulated, will the growth of capacity and dispersal of access continue for years? Or will the former hit a wall and very soon slow and stop? Will the latter be regulated? No one knows. But the giant investments say both capacity and dispersal will proceed.

How should we who seek another world think about what is happening? Should we even bother thinking about it? After all AI is just a technology that we use or not, like any other technology. You can use a hammer to build. You can use a hammer to destroy. You can use an AI to build. You can use an AI to destroy. The technology isn't the issue. The capitalist context in which it emerges and is used is the issue. For a hammer this is true. For biological or nuclear weapons, not so much. What about for AI?

AI used for good might help with a cure for cancer or Alzheimer's. It might help with procedures to reverse global warming and avert resource depletion. It might take over rote, tedious, dangerous tasks to free human time for human creativity and loving. AI, we are told, might enlighten and uplift. Truth or hype? We don't know for sure yet.

AI used for ill, however, will without doubt magnify surveillance. It will spew lies and manipulate so massively as to make Trumpian machinations to date seem minuscule. It will guide warfare, subvert elections, and sell garbage piled upon garbage. All that, even if it doesn't take over. Truth or hype? It is all already happening.

More subtly, even when used with good intent, even when popularly pursued and welcomed, AI can have some very serious unintended consequences. It can replace workers, or just make each worker in certain realms way more productive, leading not to a shorter work week for all, but to pink slips and growing unemployment for many, while in turn weakening the bargaining power of those still employed, leading to increased exploitation for them too. Less obviously, in a society of loneliness and fragmentation, AI as day care worker, therapist, personal assistant, agenda planner, teacher, composer, and writer may introduce and tout itself as a wondrous aide for humanity that hungers for its powers, even as over time it usurps functions that make humans who we are. Excluding us from those functions, it may leave us less human. It may increasingly replace human conversation and even human intimacy with machine variants.

The above three paragraphs could each be hugely expanded, but let's not belabor what ought to be obvious. There are positive and perhaps even remarkably positive possibilities. There are negative and perhaps even remarkably negative possibilities. Which will it be?

Can we look into the innards of large language models and neural networks to see what's coming? Can we find the answer there? Actually, what we would see there is a huge collection of numbers, let's call it a trillion numbers, arranged in complex patterns. When you prompt GPT-4 your words too are translated into numbers. The trillion numbers in the AI act on your relatively few prompt numbers. The AI spews out a result. Or it spews out an agenda of steps that it then undertakes. No one can predict exactly what it will spew at us. We can't, nor can its creators. No one can even say, in advance, what new "emergent capacities" it will display each time we increase the number of numbers it has, or tweak their relations, or speed up the hardware. Yes, in the past it made sense to say computer programs were just doing things that human programmers gave them the instructions to do more quickly and more accurately than people could do those things but not in a qualitatively different way. Now even the programmers have no idea what the programs will do. So, no, we can't find the answer inside the machine. To find the answer we must look at the institutions within which AI is utilized and at the ends to which it is put. And we know about those institutions—overwhelmingly they are corporations, government, and for that matter communities and even families that all manifest or suffer the pressures of class, race, gender, and power dynamics in our abysmal societies.

So we come to what matters for policy: What are the short- and long-run consequences that are already happening or that without regulation are highly likely to happen? What’s potentially good? What’s likely bad?

First, I should acknowledge that there is a big unknown lurking over this entire essay and how we assess AI. That is, will it keep getting more “intelligent” or will it hit a wall? Will more nodes and numbers and clever alterations continue to diminish errors and yield ever more functionality? Or, will there come a point with the neural network approach—perhaps even soon—when scaling up the numbers provides diminishing returns? We don’t know what is coming because it depends on the degree to which AIs can or will keep getting more powerful.

So what is potentially good and what is likely bad about AI? At one extreme, and in the long run (which some say is only a decade or two, or even less), we hear from thousands of engineers, scientists, and even officials who work with, who program, and who otherwise utilize or produce AI, nightmare predictions about AI enslaving or terminating humanity. Really responsible, capable, informed people are now routinely worrying that these material entities will dominate all humans, if not eliminate us.

At the other extreme, from equally informed, involved, and embedded folks, we hear about AI creating a virtual utopia on earth by creating cures for everything from cancer to dementia to who knows what, plus eliminating drudge work and thereby facilitating enlarged human creativity. Sometimes, and I suspect pretty often, the same person, for example the CEO of OpenAI or a chief tech officer at Google, not to mention mad Musk, says both outcomes are possible and we have to find a way to get only the positive result.

In the short run, we can ourselves easily see prospects for false voice recordings and phony pictures and videos flooding not just social media, but also mainstream media, alternative media, and even legal proceedings. We can see prospects for massive, ubiquitous intentional fraud, individual or mass manipulation, mass intense surveillance, and new forms of violence all controlled by AIs that are in turn controlled by corporations that seek profit (think Facebook), by governments that seek control and power (think your own government), or even by smaller scale entities (think Proud Boys or even distasteful individuals) who seek joyful havoc or group or personal advantage. The dismal picture includes AI-generated porn starring you with your own family unable to tell it isn’t you. It includes AI-generated speeches by candidates that no one can tell aren’t real. Think Trumpian dirty tricks and lies but vastly smarter and vastly more effective. If an AI can help find a chemical compound to cure cancer, it can no doubt help find one highly effective at killing people. And less ominously, you can have a personal assistant that makes Siri and Alexa look as dumb as a toad.

And then there is the question of jobs. It very much appears that AI can or will soon be able to fully do many tasks in place of humans or at the very least be able to dramatically augment the productivity of humans doing those tasks. Top-level programmers report that by using GPT-4 they can already double their output. Law firms report similar and even greater gains. How long until AIs translate better than translators? The good side of this is attaining desired economic output with fewer labor hours, and thus, for example, potentially adopting a shorter work week with full income for all, or even with more equitable incomes. The bad side of this is that without sustained pressure corporations will instead keep some employees working as much as now, but with twice the output. They will pay those retained employees reduced income, and pink-slip the rest into unemployment.

Consider, for example, that there are roughly 400,000 paralegals in the United States. Suppose by 2024 AI enables each paralegal to do twice as much work per hour. Suppose paralegals in 2023 work 50 hours a week. In 2024, do law firms retain them all, maintain their full pay, and have them each work 25 hours per week? Or do law firms retain half of them at 50 hours a week and full salary, while firing the other half? And then with 200,000 unemployed paralegals reducing the bargaining power of those who still have a job due to fear of being replaced, do the law firms further reduce pay and enlarge required output and the work week of those retained, while they fire still more paralegals? With no effective regulations or system change, we know profit-seeking will rule, and we know the outcome of that. And this is not just about paralegals, of course. AI can deliver personal aides to educate, to translate, to deliver day care, to diagnose and medicate, to write manuals, to maintain financial records, to conduct correspondence, to make and deliver product orders, to compose music, to write stories, to create films, and even to design buildings. It will be amusing to hear the Beatles, Elvis, Joplin, Fitzgerald, and Sinatra sing new songs, but will it be good for future music and for potential musicians? Try to imagine the havoc in the industry from that, or from when Apple or whoever uses AI to create new AI "artists" crowding out human artists? With no powerful regulations, with profit in command, is there any doubt about whether AI will bring something nearly utopian or impose something highly dystopian? The above enumeration could go on. In the past week, and as far as I am aware not even contemplated a month before, there is a firm now training AI in managerial functions, financial functions, policy-making functions, and so on. Or, if there isn't, might there be next week? Minutes ago, I listened to a 71-year-old Paul McCartney singing one of his recent songs, morph into an AI-created young McCartney singing it, morph into young John Lennon singing it. Next week, what? With what consequences?

Before moving on from crystal balling the future, we might also consider some unintended consequences of trying to do good with AI. Short of worst-case nefarious agendas, what will be the impact of AI doing tasks that we welcome it to do but that are part and parcel of our being human? Let's even suppose AIs do these functions as well as we do, as compared to just well enough for corporations to utilize them in our place. I repeat these questions because attention to this issue seems utterly absent, even in the cacophony of pundit pronouncements about AI. It is quite similar to the way that concerns that social media might dull our wits and produce antisocial attitudes was absent from most assessments of Facebook and Twitter until they were long since ubiquitous—except AI may be to social media what big bombs are to little bullets.

Day care for children? Companion care for the elderly? Psychological and medical counseling for the ill? Planning daily agendas for us all? Teaching us all? Cooking for everyone? Intimate conversation for the lonely? Sounds promising, doesn't it? But if AIs do these things what happens to our capacity to do them? If AIs crowd us out of such human-defining activities, are they becoming like people, or are we becoming like machines?

Try conversing with even current AIs. I would wager that before long you will move from referring to it as "it"—meaning just a massive pile of numbers—to referring to it as he or she, or literally by its name. On Discord, just months into the AI tsunami, Chatbot Clyde showed up. AI everywhere? That's our trajectory. AIs teaching, counseling, care taking, note taking, agenda setting, drawing, designing, medicating, and what all—and you do what? Are you uplifted and liberated from prior responsibilities as you watch movies that AIs make? As you eat food that AIs prepare? As you read stories that AIs write? As you do the few errands left to you, but which AIs organize? Even assume (against what we know of our capitalist economy) that income is handled well. Even assume that remaining work for humans is allocated well. You want something, you ask an AI for it. Clyde or Bard or whatever delivers. Some will say that is fantastic. My private slave. Oh boy! If AI development doesn't hit a wall, this appears to be the non-nefarious utopian scenario. And I guess it does look utopian to the eyes of some beholders, but to my eyes it looks highly dystopian. This all feels quite like the way social media promised great things and then partially as unintended consequence and partially as welcomed evil, social media did quite horrible things, except that in this analogy AI’s trajectory of harm looks to me like social media on steroids cubed.

So what do we do? We need to invoke ecology’s “precautionary principle.” Facing innovations that have potential to cause great harm, we need to emphasize caution. We need to pause and review before we leap into proliferation that may prove disastrous. We need to take preventive action in the face of uncertainty. We need to put the burden of proof on the proponents of a risky activity. We need to increase public participation in decision making. Boiled down, we need to look before we leap.

To conceive a sensible response to the emergence of steadily more powerful AI doesn’t require genius. We should pump the brakes. Hard. As even people in the industry have advised, but with little serious follow-up, and as even governments have advised but not yet required, we need to at least opt for a moratorium. During the ensuing hiatus, we need to establish regulatory mechanisms. We need to set rules. We need to establish means of enforcement able to ward off dangers as well as to advantageously benefit from possibilities. How about this for a simple take on one aspect. We already outlaw with very serious consequences producing counterfeit money, and/or passing it on. The prospect of fake money struck banks and the rich in general and governments too as seriously dangerous. So, they imposed serious regulations. The prospect of false AI products masquerading as real threatens to make a shambles of elections. That upsets some people in power, though it is welcomed by others. It upsets me and you too, I bet. How about severe penalties for false news and indeed for AI products that are not clearly labelled as such?

This isn’t rocket science. Suppose a third of the engineers involved in building a new airplane, like those involved with AI, said there was a ten percent chance the plane would crash. Would you climb in, buckle up, and celebrate the miracle of your hoped-for journey? Or would you jam on the brakes?

A moratorium is simple to propose—indeed, one and by the time you read this I bet two or more, have already been proposed, domestically and internationally. But this has occurred without serious follow-up that I am aware of—despite that in our world, a moratorium will be hard to achieve. In our world, owners and investors seek profits regardless of wider implications for others. Pushed by market competition and by short term agendas, they proceed full speed ahead. Their feet avoid brakes. Their feet pound gas. And off we go on a suicide ride. Yet unusually and indicative of the seriousness of the situation, hundreds and even thousands of central actors inside AI firms are concerned/scared enough to issue warnings. And even so, we know that markets are unlikely to heed them. Indeed, even those urging caution are unlikely to retain good sense. Investors will mutter about risk and safety, but they will barrel on. And the thousands of central actors will in most cases, other things equal, succumb to the pressure.

So, can we win time to look before corporate suicide pilots leap? If human needs are to replace competitive, profit-seeking corporate insanity regarding further development and deployment of AI and how we use AI, or, for that matter, regarding the deployment and use of everything else corporations impose on us, we who have our heads screwed on properly will have to make demands and exert very, very serious pressure to win our demands.