

Turning off the gaslight: the best of scholarly critical thinking as a response to genAI's dystopias

Jonathan A. Poritz

jonathan@poritz.net
poritz.net/jonathan



21 February 2025
Education After the Algorithm



This slide deck, except where otherwise indicated, is by [Jonathan Poritz](https://poritz.net/jonathan) and is released under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/). This version: 20 Feb 2025 00:12GMT
These slides, also in editable form, are available at poritz.net/jonathan/share/Totg/.

This just might do nobody any good

This just might do nobody any good. At the end of this discourse a few people may accuse this reporter of fouling his own comfortable nest, and your organization may be accused of having given hospitality to heretical and even dangerous thoughts. But I am persuaded that the elaborate structure of networks, advertising agencies and sponsors will not be shaken or altered. It is my desire, if not my duty, to try to talk to you journeymen with some candor about what is happening to radio and television in this generous and capacious land.

Edward R. Murrow's¹ *Wires and Lights in a Box* speech to the Radio-Television News Directors Association, 15 October 1958²

¹ Murrow was a reporter for CBS News whose *See It Now* show "criticized McCarthyism and the Red Scare, contributing, if not leading, to the political downfall of Senator Joseph McCarthy," in [Wikipedia's description](#).

² You can read [the text](#), listen to [an audio recording of Murrow giving the speech](#), or watch [the actor David Strathairn deliver part of the speech](#) as he played Murrow in the surprisingly excellent 2005 movie "[Good night, and Good Luck](#)," directed by George Clooney. The RTDNA is now the [RTDNA](#).

Evidence in empiricism

A precept of critical thinking ($\approx$ “the scientific method” $\approx$ “empiricism”) states

Extraordinary claims require extraordinary evidence

Carl Sagan, in his 1979 book *Broca's Brain*

or, pretty much the same but 231 years earlier:

...it is a general maxim worthy of our attention “That no testimony is sufficient to establish a miracle, unless the testimony be of such a kind, that its falsehood would be more miraculous, than the fact, which it endeavors to establish...”

David Hume, in the tenth section, *Of Miracles*, of his book
An Enquiry Concerning Human Understanding, 1748

(Hume may have been one of the founders of empiricism, but Sagan was far pithier!)

The extraordinary claims of the AI hype cycle have been for utopia or apocalypse, a complete transformation of work and science and education, and, fundamentally, that some performance vs scale curve would continue on its merry climb as long as the investment funds kept coming. The evidence has been an *uncanny mountain range*.

Extraordinary claims: AI hype, part 1

We are told that “Artificial Intelligence”

- is here
- is right around the corner
- is inevitable³
- is a supremely valuable investment (*give me the money!*)
- makes individuals that use it far more productive than those who do not
- ... so will be used in all workplaces
- ... and so not teaching it would so disadvantage our students as to constitute a form of academic malpractice
- gives nations that control it so much more power and such an economic advantage that they will completely dominate geopolitically
- ... so for a nation (or block, like the EU) not to use all possible government power – supplying generous funding, harsh protectionism, and the creation of friendly legislative, regulatory, and judicial environments, etc. – in support of their domestic AI industry would end with that nation slipping into obscurity

³When we accept something like this during the rise of fascism, it is called *anticipatory obedience*.

AI hype, part 2

... and that AI...

- will enable everyone to become a world-class writer and artist
- will enable people with disabilities to have full and equal access to everything on the Internet (human-like speech-to-text, detailed description of images, etc.)
- will correctly diagnose disease (highly accurate reading of X-rays!)
- ...and then cure all diseases (new miracle drugs!)
- will invent new energy sources (fusion, anyone?)
- ...and then completely “solve” the climate crisis,
- will fix homelessness and help governments plan their way past all social ills
- will end intercultural misunderstanding by translating between *all* human languages
- *i.e.*, will usher in a technoutopia

All for the low, low price of

\$7 Trillion

and unimpeded access to *everything ever put online*, including

- all of YouTube, Reddit, GitHub, LinkedIn, Facebook, and other social media
- every student assignment ever submitted through Turnitin and Canvas or any other LMS
- every Google Doc (even ones never made public), Microsoft Cloud document, *etc.*
- the complete text of every message ever sent or received by Gmail
- in short: every document or media file that a computer user ever put in a cloud-based platform or on a website *even if robots.txt forbids this scraping*
- regardless of the copyright status of any of these digital objects

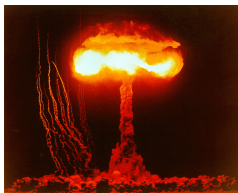
Fearmongering

Oh, and BTW, without that investment and access, there is a serious chance that AI will *escape human control* and either

set off all the nuclear weapons

or

turn all of the surface of the planet, including all of us humans, into paperclips.



4



5

⁴ [Upshot-Knothole Climax 001](#) by [National Nuclear Security Administration / Nevada Site Office](#) is in the public domain in the United States, as it was made by an employee of the U.S. federal government as part of that person's official duties.

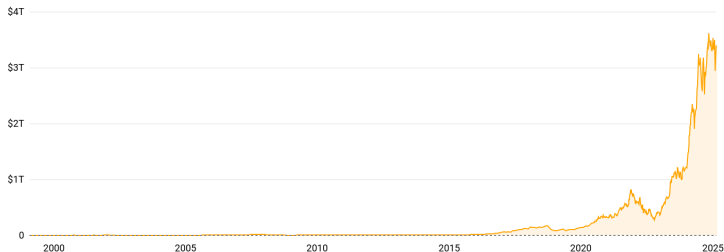
⁵ [Büroklammern – 2021 – 6481](#) by [Dietmar Rabich](#) was released under [CC BY-SA 4.0](#)

The real cost of AI

Independent estimates suggest that AI

- requires about half a liter of clean water per query (to cool the server farms)
- requires as much energy per query as it would take to charge a cell phone
- ... hence requiring mothballed coal-fired power plants to be returned to service
- leading to thousands of deaths per year in communities near server farms⁶
- is losing Microsoft \$20 per user per month on it's "Copilot" product
- ... and similarly for all major AI companies – total in the billions!
- Taiwan-based NVIDIA makes almost all chips for AI hardware

Market cap history of NVIDIA from 1999 to 2025



⁶ see [The Unpaid Toll: Quantifying the Public Health Impact of AI.](#)

What is really going on? through the lens of *Cui bono*?

With so much at stake, it seems important to validate these claims, understand what is really going on, and to make the best policy and practical choices.

When getting advice about vital questions of public policy, it makes sense to keep in mind Cicero's question: "*Cui bono*?"

I trust the people in this room, or people in rooms like this, to do that work. After all, *"Institutions of higher education are conducted for the common good and not to further the interest of either the individual teacher or the institution as a whole. The common good depends upon the free search for truth and its free exposition."*⁷

At least I trust them most of the time .. but blind faith is perhaps misplaced. After all, *"It is difficult to get a [person] to understand something, when [their] salary depends upon [their] not understanding it!"*⁸

⁷ The AAUP's [1940 Statement of Principles on Academic Freedom and Tenure](#), which was built on the [1915 Declaration of Principles on Academic Freedom and Academic Tenure](#).

⁸ From *I, Candidate for Governor: And How I Got Licked*, by Upton Sinclair, University of California Press, 1934; edited to be gender-neutral.

What are we talking about? Some history

A first step of an impartial, thoughtful analysis of AI and related policy would be to *define the term*.⁹

Philosophers have long had trouble with the concept of other minds – there's that entire subdiscipline *Philosophy of the Mind*, which includes such lovely ideas as that of the **philosopher's zombie**:

*"Unlike the ones in films or witchcraft, they are exactly like us in all physical respects but without conscious experiences: by definition there is 'nothing it is like' to be a zombie. Yet zombies behave just like us, and some even spend a lot of time discussing consciousness."*¹⁰

Alan Turing, the mathematician who founded computer science, proposed a test of whether machines could "think," which is now known by his name. One of the great lessons of the last few years of commercial AI products might simply be that the Turing Test was always an absolutely terrible idea.¹¹

⁹ Or maybe I just think that way because mathematicians love definitions and axioms.

¹⁰ **Zombies** from [The Stanford Encyclopedia of Philosophy](#).

¹¹ Actually, this shouldn't really be a surprise, since Turing's original test was explicitly about testing for gender identity, and it was proposed by a closeted gay man who, when he was forcefully outed, was driven to suicide by the abusive punishment of a homophobic society.

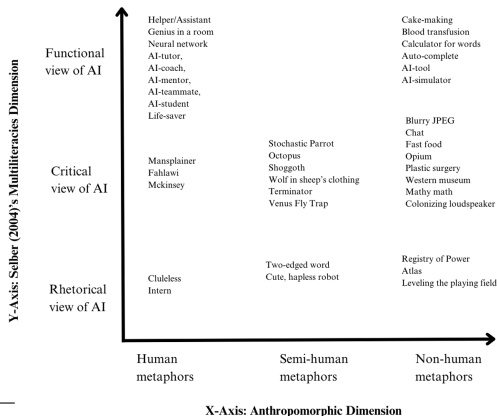
What is AI *like*?

But something different clearly happened in the last few years. Some major technical advance enabled programs to start to pass the Turing test in a more credible way than ever before. Unfortunately, that technical advance is shrouded in complex terminology and diagrams in the computer science literature, which tends to drive away legislators, policy makers, and nearly everyone else.

12

One approach to enabling conversations with non-technical folks about highly technical topics is to use **metaphors**. A marvelous recent paper collected some of the metaphors being used recently in discussions of AI, and even usefully placed them in groups on two different axes of classification:

Metaphors, of course, tell us what something is *like*, not what it *is*



¹² from Assistant, Parrot, or Colonizing Loudspeaker? ChatGPT Metaphors for Developing Critical AI Literacies. by Gupta, A., Atef, Y., Mills, A. and Bali, M., 2024, Open Praxis, 16(1), p.37–53. DOI: <https://doi.org/10.55982/openpraxis.16.1.631>

What *is* AI, without metaphors

The recent technical innovation which is used to justify the hype above is based on something called a **Large Language Model [LLM]** or **Generative Pre-trained Transformer [GPT]**.

LLMs consist of a data structure called a “neural network”¹³ and two accompanying algorithms:

- one which processes (hundreds of) TB (or more) of *training data* - text and data usually copied from the internet (“scraped”) - to create (hundreds of) MB of values in the data structure (“model parameters”);
- another which uses the data structure to transform sequences of *input tokens* into sequences of *output tokens* in a way that maximizes the probability of the output given the input and the training data.

A “token” is a letter or word or image pixel, *etc.* Note the input tokens might be words in one language and the output words in another language, for machine translation, or outputs might be pixels, for image generation.

¹³in a very poor analogy with biology

Aside on the Neural Network Data Structure/Algorithm

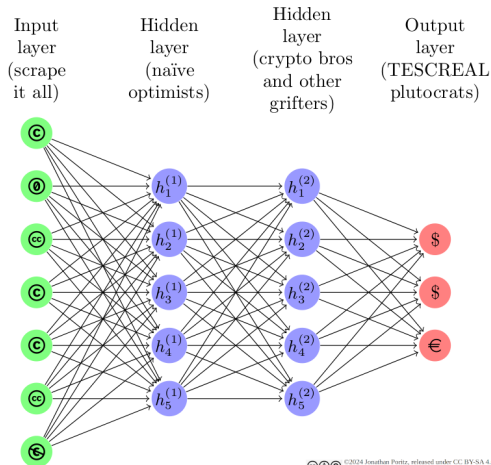
The neural network data structure looks something like this:

Input values are imagined to be put in each of the input layer nodes (those disks on the left).

Arrows show where values are used in simple formulæ to create values in each of the hidden layer nodes

The formal “output” of the model is a label (maybe a token) associated to whichever of the output layer nodes has the largest calculated value.

Real-World Neural Network Architecture



Aside on *Attention is all you need*

This data structure has been around for a while, with some success in particular applications, but no one thought it was about to escape its creators and turn us all into paperclips.

Then a computer science paper called [Attention is All You Need](#) appeared in 2017, describing a particular way to use the data structure, together with some small additions that have been dubbed the "attention mechanism" (in a very poor analogy with "paying attention" as human minds do) in a particular way that allows relatively efficient scaling to extremely large versions of the data structure, trained on enormous datasets.

A "neural network" with the attention mechanism architecture is called a **transformer**, and when it is populated with parameter values created by processing some training corpus, it is said to be **pre-trained**: hence, GPTs.

Despite more than 150,000 citations in the scientific literature, the *Attention is all you need* paper does not present a deep new idea, or new framing of a deep old idea. Rather, it is a proposal for an incremental engineering change in how neural networks are used in applications which seem to produce very pleasing sequences of output tokens.

So:

$$NNs + \langle \text{attention} \rangle + \langle \text{huge corpora of training data} \rangle$$

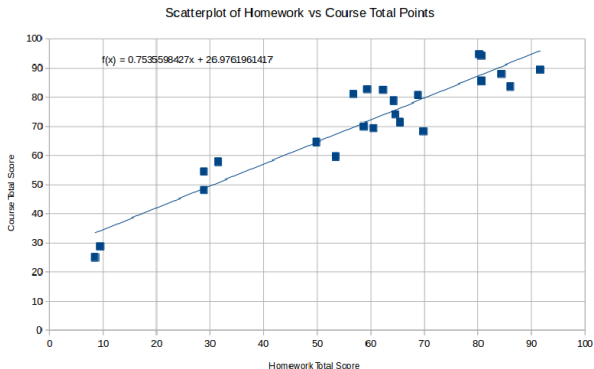
~~~~~

$\langle \text{algorithm that produces pleasing, statistically likely sequences of output tokens} \rangle$

## In other words, a statistical model

Processing sample data to create parameters from which one can calculate output for new inputs is exactly what we call a *statistical model*.

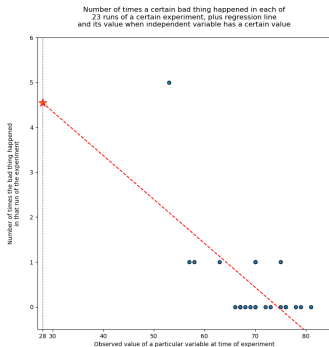
For example, you may have seen a statistical model called a “regression line.”



This *scatterplot* comes from an OER I wrote a few years ago. One calculates the coefficients of the equation of the line from the sample data and with this equation one can compute output values (“ $y = f(x)$ ”) from new input values (“ $x$ ”).

# Statistical models can do some wonderful things

Don't get me wrong, I love statistical models. They are very powerful. E.g.,



this one

might have prevented



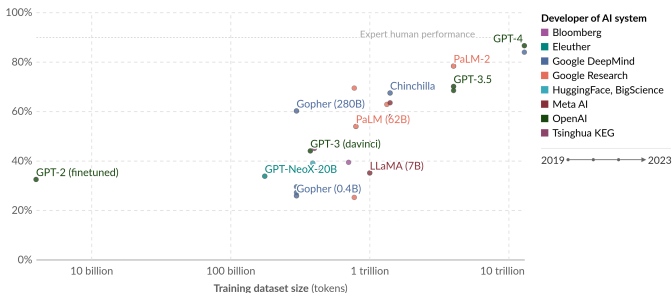
# Statistical models are not Skynet, nor do they have agency

Every sentence that uses AI as the subject of a sentence with an active verb is *prima facie* ridiculous:

- When Google's AI-created search-results summary suggested to include a quarter of a cup of glue in a pizza recipe to prevent the cheese from slipping off when you pick up a slice, it was not “hallucinating.” Instead, it was outputting a statistically likely collection of tokens (words) in the context of its training data and the givens “pizza recipe” and “prevent slipping,”
- An AI will never “start a nuclear war”: a statistical model has no agency, the very idea is ridiculous. A bad human actor who used hacking techniques and nuclear codes extracted by a clever choice of prompt from an LLM which was trained with a little too much information might launch a nuclear weapon, but that’s pretty far from Skynet.
- An LLM cannot “summarize a scholarly article”: statistical models do not understand scholarly articles, that would be like saying that boats are good at swimming ... it’s a context error. Although an LLM *might* assemble sequences of words which are statistically likely in the context of the sequence of words in the article, and that could be a useful output in particular situations.

## Another scatterplot: The famous “scaling law”

Artificial intelligence: Performance on knowledge tests vs. dataset size



Our World  
in Data

15

This hasn't been updated in a bit more than a year, in part because these performance scaling laws seem to be running into some sort of wall. See, e.g, Gary Marcus's [The new AI scaling law shell game - by Gary Marcus](#)

Nevertheless, the implied regression line here is taken to indicate exponential growth, and therefore an “AGI Singularity” predicted by some in 2025, by others in the next few years.

<sup>15</sup> From [Our World in Data](#), released under [CC BY 4.0](#)

## AGI, TESCOREAL, and eugenics

That “AGI” stands for “Artificial General Intelligence,” for which there have been given various definitions by researchers which mostly amount to human or more-than-human levels of creativity, problem-solving, agency, autonomy, and ability to improve itself (“learn”) – basically, Deep Thought or HAL 9000 on a good day, and Skynet on a bad day.

Timnit Gebru and Émile Torres have placed the pursuit of AGI in the context of a group of ideologies they call “the TESCOREAL bundle,” which are basically a form of technoutopia version of something like a Calvinist (chauvinist) eschatology with deep roots in eugenics. Read [their paper](#), it’s disgusting and important. [TESCREAL stands for “transhumanism, Extropianism, singularitarianism, cosmism, Rationalism, Effective Altruism, and longtermism,” all related ideologies which seem very influential for the current AI companies. For example, the firing and re-hiring of OpenAI CEO Sam Altman in 2023 seems to have revolved around competing interpretations of some application of Effective Altruism their work!]

Even to mention these issues may seem to be stooping to making an *ad hominem* argument. Maybe. But how do we balance the benefit of automated help writing or grading quiz problems in large lecture classes against empowering rich eugenicists? or the timeliness of the trains against a little fascism, as Italians had to figure out under Mussolini?

## More terminology: dividing AI into different categories

There is a tendency to use the term “Artificial Intelligence” these days for a whole range of different types of computer programs.

Mostly (but not always), one uses this term to talk about programs which have come about since LLMs/GPTs were invented – in opposition to “old-school AI” (which **did** try to encode meaning and not merely statistical patterns of tokens!).

I’m more interested in drawing a distinction between situations where huge corpora of human creativity are compressed and the resulting GPT produces outputs which I think is *justifiable* and *useful*. I propose these terms:

- **translational AI** (or maybe **transformational AI**) is one where the informational content<sup>16</sup> of the prompt is of a similar size<sup>17</sup> as that of the output.
- **plagiaristic AI** (or maybe **decompressive AI**, because it is about decompressing the human creativity of the training corpus lossily compressed into the model parameters).

I could see translational AI as being more permissible under that idea that it constitutes a human making a derivative work of the prompt using a tool, with some idea of scholarly norms and IP law, both with regards to the prompt, coming into play. And I could see decompressive AI as owing more debt to the human creators of training corpus via some reasoning that even finely diced harm/responsibility, at scale, must have consequences.

---

<sup>16</sup> Probably as measure by something like [Kolmogorov complexity](#)

<sup>17</sup> This clearly needs some additional precision....

## Takeaway 1: An important general question about education

We have to ask ourselves

***Is education a complete failure?***

Possible qualifiers:

- Higher education?
- Education in WEIRD nations?
- Liberal Arts or vocational/technical/engineering education?
- In the last generation? Since Reagan/Thatcher (*i.e.*, since the rise of neoliberalism)?

Certainly academia's response to AI seems to me to be a complete failure, in that we seem to be performing so much anticipatory obedience without even trying to do critical thinking, without standing up for the principle of giving credit to our sources, without fighting for education in the service of the common good.

## Takeaway 2: Many important specific questions about AI

Whether you are subject or object of the AI hype cycle, consider questions like these:

- Do you care about the climate?
- Do you care about violating the fundamental academic value of giving credit?
- Are you happy giving more money and power to powerful and wealthy fascists?
- Have you actually seen extraordinary evidence of the extraordinary claims?
- Have you done your homework to persist doggedly in learning enough to be able to understand the technology regardless of how intimidating it is?
- Are you reasoning about AI as it really is or via some weak and biased metaphor?
- Do you care that institutionalizing AI devalues human creativity and labor?<sup>18</sup>
- Are you OK with journalists, educators, and other creatives being fired en masse?
- Is a proposed AI “solution” really something that could be done (and done better!) by humans, if politics and economics were really supporting human flourishing rather than the further concentration of wealth and power?
- Is a proposed AI “solution” actually itself a little slice of dystopia, if you look at it without the starry eyes of a tech enthusiast?

---

<sup>18</sup> by remixing it at scale and selling it without acknowledging or compensating the creative and laboring humans

## Further references

### Books:

- **The AI Con: How to Fight Big Tech's Hype and Create the Future We Want** by Emily M. Bender and Alex Hanna, forthcoming
- **Unmasking AI: My Mission to Protect What Is Human in a World of Machines** by Joy Buolamwini, Random House (2023)
- **Blood in the Machine: The Origins of the Rebellion Against Big Tech**, Brian Merchant, Little, Brown and Company (2023)
- **Race After Technology: Abolitionist Tools for the New Jim Code**, Ruha Benjamin, Polity (2019)
- **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy**, Cathy O'Neil, Penguin Random House (2016)
- **Algorithms of Oppression: How Search Engines Reinforce Racism**, Safiya Noble, NYU Press (2018)
- **Resisting AI: An Anti-fascist Approach to Artificial Intelligence**, Dan McQuillan, Bristol University Press (2022)

### Articles:

- *Data centers drain Ireland's power grid, add their own fossil fuel-powered generators – Pivot to AI*
- *OpenAI says it's "impossible" to create useful AI models without copyrighted material*
- *Zuckerberg approved Meta's use of 'pirated' books to train AI models, authors claim – Sarah Silverman and others file court case claiming CEO approved use of dataset despite warnings*
- *Knowing less about AI makes people more open to having it in their lives – new research*
- *Ex-Google CEO: AI startups can steal IP, hire lawyers to "clean up the mess" — The Verge*
- *AI solve housing crisis at DuckDuckGo*
- *AI is Making Housing Discrimination Easier Than Ever Before — Kreisman Initiative for Housing Law & Policy at UChicago*
- *AI has the potential to 'cure all diseases', says DeepMind chief*
- *Sorry, AI won't "fix" climate change – OpenAI's Sam Altman claims AI will deliver an "Intelligence Age," but tech breakthroughs alone can't solve global warming.*
- *Former Google CEO says climate goals are not meetable, so we might as well drop climate conservation — unshackle AI companies so AI can solve global warming*
- *[2412.06288] The Unpaid Toll: Quantifying the Public Health Impact of AI*
- *Microsoft and OpenAI have a financial definition of AGI: Report — TechCrunch*
- *The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers*

## Contact Info

Email: [jonathan@poritz.net](mailto:jonathan@poritz.net)

Fediverse: [@poritzj@mastodon.social](https://@poritzj@mastodon.social)

These slides:

[poritz.net/j/share/Totg.pdf](https://poritz.net/j/share/Totg.pdf)

and all files for remix:

[poritz.net/j/share/Totg/](https://poritz.net/j/share/Totg/)

Or else use —————→

to go to my homepage, where you can follow  
links to all of my shared materials.



Shameless promotion:

Look for new (started 20 January 2025) weekly podcast

**The Revolution Will Not Be Televised:  
The Theory and Practice of Resistance**

jodintly by [Kathy Giuffre](#) and [JP](#)