

"Artificial Intelligence": Practical, Legal, and Ethical Considerations for Academia in a Global Context

Jonathan A. Poritz

jonathan@poritz.net
poritz.net/jonathan



10 September 2024
CSUP



This slide deck, except where otherwise indicated, is by [Jonathan Poritz](#) and is released under a [Creative Commons Attribution-ShareAlike 4.0 International License](#). This version: 11 Sep 2024 21:39MDT
These slides, also in editable form, are available at poritz.net/jonathan/share/AIPLEC4A/.

“AI” hype, part 1

We are told that “Artificial Intelligence”

- is here
- is right around the corner
- is inevitable
- 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 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

“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
- *i.e.*, will usher in a technoutopia

All for the low, low price of

\$7 Trillion

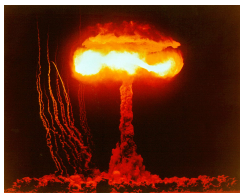
and unimpeded access to *everything ever put online*, including

- 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 shared), 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 instead of keeping on their own device
- 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



1

or

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



2

¹ [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.

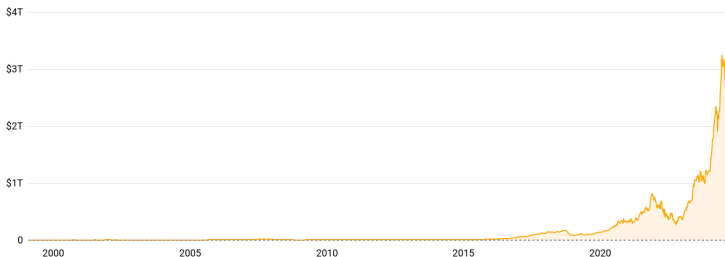
² [Büroklammern – 2021 – 6481](#) by [Dietmar Rabich](#) was released under [CC BY-SA 4.0](#)

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
- is losing Microsoft \$20 per user per month on it's “Copilot” product
- ... and similarly for all major AI companies, so they are all losing tens of billions of dollars (at least) per year
- Taiwan-based NVIDIA makes almost all chips for “AI” hardware

Market cap history of NVIDIA from 1999 to 2024



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."*³

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!"*⁴

³The 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](#).

⁴From *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*.⁵

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.”*⁶

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.⁷

⁵ Or maybe I just think that way because mathematicians love definitions and axioms.

⁶ **Zombies** from [The Stanford Encyclopedia of Philosophy](#).

⁷ 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.

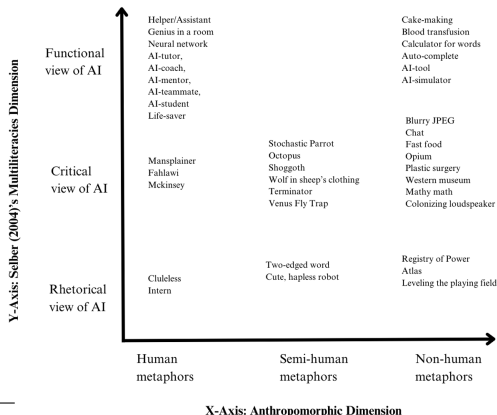
What is “AI” like?

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.

8

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*



⁸ 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”? (part 1)

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”⁹ 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.

⁹ in a very poor analogy with biology

Aside on the Neural Network Data Structure/Algorithm

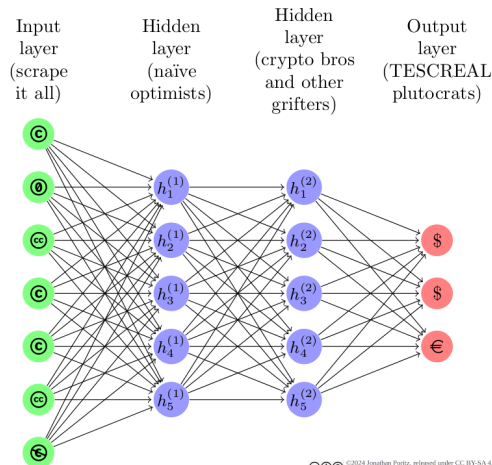
The neural network data structure looks something like this:

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 132,000 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:

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

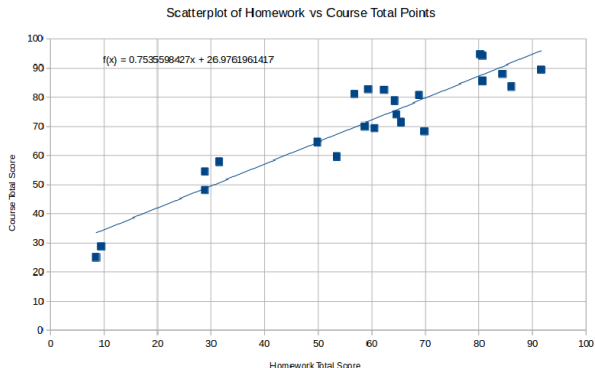
~>

algorithm that produces pleasing, statistically likely sequences of output tokens

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*.

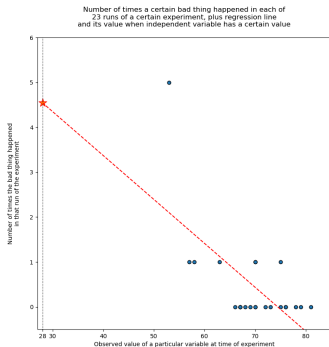
For example, you may have seen a statistical model called a “regression line.”



This *scatterplot* comes from an OER I wrote a few years ago for a stats class here at CSUP. 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.,



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.

Legal aspects: training LLMs

When copying and processing the vast quantities of data used to train LLMs, are “AI” companies violating the rights of the creators of all of those works?

The law that might be violated is **copyright** – little-known fact: almost everything on the internet is under all-rights-reserved copyright, and you cannot legally copy it (or sell it or create derivative works from it, etc.) without the rightsholder’s permission.

But copyright does not apply to **ideas or facts**, but rather to **original expressions** (of those ideas or facts, often).

Some have argued that the training of LLM constitutes merely extracting some facts – the statistics of the sequence of input tokens – from works in their training corpora. I’m not sure how that holds up when copies of entire works from the training data often show up in LLM output.

Others have argued that training is *fair use* (a doctrine in the US which allows for transformative uses of others’ works without their permission, subject to certain restrictions in a famous “four-step test”). But I don’t see how training LLMs can be fair if journalists and illustrators are getting fired...

Note: **IAmNotALawyer** and the above does not constitute legal advice! Additionally, my skepticism is not widely shared among legal experts (who, I would like to point out, clearly do not understand the tech).

Legal aspects: LLM outputs

Copyright springs into existence (in the US) when someone creates an “original work of authorship fixed in a tangible medium of expression.” But who is the author and is it original enough when the work in question is the output of an LLM?

The US Copyright Office has **decided** this in one case, which, it is thought, may be a strong precedent. Not a great precedent, IMHO.

The reasoning of
Copyright Office for
the graphic novel
Zarya of the Dawn...

[cover of [Zarya of the Dawn](#),
created by [Kris Kashtanova](#)] with
Midjourney; in the public domain,
at least in the USA, for now.]



...was based on
that of the famous
Monkey Selfie

[Self-portrait by the depicted *Macaca
nigra* female via [Wikimedia Commons](#),
in the public domain, at least in
USA, at least for now.]



in that the Copyright Office found the locus of creativity in the LLM's software and said, based on the Monkey Selfie case, that non-human entities, no matter how creative, could not own copyrights.

From my perspective on LLMs as statistical models, this is absurd, and the creativity surely lies in a combination of the creators of the training data and in the LLM user's creativity in using this tool to make a remix. ...But, again, **IAmNotALawyer**.

Legal aspects: legislative actions

There are a number of bills pending in the California state legislature, but there has been such complete regulatory capture by the rich tech companies that I don't have much hope about these.

The European Union passed something recently which is called the "EU AI Act," which has fairly strong

- protections of the rights of creators (to opt out of their work being used in LLM training corpora, for example),
- preventative measures for draconian uses of LLMs (e.g., no use by police or in schools for things like facial recognition, which is known to be very highly racially biased¹¹)
- transparency measures about their training corpora, model parameters, etc.

And yet... the EU AI Act defines "AI" as software which shows "nearly human levels of agency and creativity" – so, from my perspective, the Act applies to exactly nothing (today, at least) in the real world, outside a tech company's PR campaign!

Although maybe the EU parliament is playing 4-dimensional chess, to get big tech to admit its PR is nonsense or else suffer under the restrictions of the AI Act....

¹¹ E.g., see **Algorithms of Oppression** by Safia Noble, **Race After Technology** by Ruha Benjamin, and **Unmasking AI** by Joy Buolamwini

You may be wondering how the discussion in this presentation would describe the statements of “AI”-promoters like Sam Altman, Tim Cook, Bill Gates, Elon Musk, Sundar Pichai, and Mark Zuckerberg:

They're all lying.

Well, maybe some of them are just stupid.

And others might be suffering under Upton Sinclair's salary-induced inability to understand.

Ethical considerations

(This should not be the last slide, damn it!)

The climate damage is a problem for me.

The theft of other humans' creativity is also a problem.

The abandonment of the greatest academic norm of acknowledging one's sources is deeply disturbing.

But in some sense the worst part of all of this is that it sells an idea that there is no such thing as expertise, that you can “fake it ‘till you make it” with the help of technology and it's perfectly fine, that an OER author can type the right prompt into ChatGPT and it will produce a textbook as good as one an actual expert would write, that an LLM-driven chatbot with extended training corpus that includes some math books can be as good a tutor as can I, with my decades of experience teaching math....

This all seems to me to be deeply corrosive to foundational structures of truth-seeking and -verification which our society desperately needs at this moment in history.

Contact Info

Email: jonathan@poritz.net

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

These slides:

poritz.net/j/share/AIPLEC4A.pdf

and all files for remix:

poritz.net/j/share/AIPLEC4A/

Or else use —————→

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

