The Rhetoric of Machine Learning[1]

(talk presented at Persuasive Algorithms? A symposium on the rhetoric of generative AI)

12 November 2024

Robert C. Williamson,
University of Tübingen
https://fm.ls/bob

I'm an engineer / scientist / mathematician who works on theoretical questions arising machine learning. I am not a rhetorician. However, I have found that looking at machine learning through the lens of rhetoric is remarkably refreshing and generative (ha!), and I will now present some of the thoughts that have thus arisen.

ML is Rhetoric

Let me start with a mild disagreement. The conference is on persuasive *algorithms*; I will argue it is not the *algorithms* that should be the focus of our attention. And the conference focusses on "generative" AI. I will argue that there is a strong and essential rhetorical component to even the simplest of AI methods, and by understanding this better we may gain insight into the newest technologies.

Machine Learning (I will stick to this more specific term rather than "AI") is a *technology*. But unlike many technologies it is a *cognitive* technology (just as Erik Havelock[2] and Jack Goody have described written language). It thinks *on our behalf*. Not on its *own*. We *delegate* certain tasks to it, and then let it persuade us of its results. We offer it data, and then pay attention to what it "generates" (predictions or forecasts or explanations). In the same way that science[3], economics, and statistics have all been fruitfully viewed from the perspective of rhetoric (argument intended to persuade) I think it is instructive to examine the style of "rhetoric of machine learning".

Two simple points to start:

---

[1] This is essentially the text of the talk I delivered at the conference. I hope (eventually!) to turn this into a slightly longer and more carefully constructed essay, but I am happy to share the text as is noting its provenance … in a purely written work I would probably hedge a little more on some things!

[2] Eric A. Havelock. *The Muse Learns to Write: Reflections on Orality and Literacy from Antiquity to the Present.* Yale University Press, 1986; , Eric A. Havelock, *The literate revolution in Greece and its cultural consequences.* Princeton University Press 1982.; Jack Goody. Technologies of the intellect: Writing and the written word. In *The Power of the Written Tradition*, pages 132–151. Smithsonian Institution Press, 2000.

[3] Walter B. Weimer,. "Science as a rhetorical transaction: Toward a nonjustificational conception of rhetoric." *Philosophy & Rhetoric* 10.1 (1977): 1-29.

1) Like the law, and science for that matter, the rhetorical style of ML is primarily that of "*anti-rhetoric*[4]," which is to say it explicitly denies that it is rhetoric. It is not (so the consensus would insist) trying to *persuade* anyone, but rather it is just revealing what the data showed on its own — the "intrinsic structure of the data". It is described as being "data-driven" and this is held to be a very fine thing as it avoids the biases that people have. It is "objective" and reveals the world as it really is. Well so it is claimed…
2) All arguments start somewhere. And that place is the *fact*. Rather than construing the fact as given (after all its etymology comes from the latin word "to *make*"!), students of rhetoric (in particular Chaim Perelman[5]) define the fact as the *beginning* of the argument and, crucially, *that which one chooses not to question.* For ML, data is its fact[6]. It is "given" (funnily enough, the root of data is *dare*, to give). And it is (largely) taken for granted. Many LLMs do not even declare the data they were trained upon (or like Llama, just refer to the "pile", an unspecified collection of pirated texts; perhaps on their lawyer's advice given the piracy involved!)

An argument is a chain (or more generally a web) of claims underpinned by a warrant (which certifies that the argument is adequately sound and valid). The traditional view of ML only pays attention to the final step of the argument, and reduces it to a simple scalar score on a benchmark which serves as the sole, and inadequate warrant — not unlike the baffling naivety of bothering to even look at a text or video without some credible notion of its provenance.

The new dream of method

The logical positivists fantasied that if they got their protocol sentences and the like in order they could develop an unbiased method of revealing the truth about the world. They have largely disappeared from philosophy … because they now work for Machine Learning companies.

I once read somewhere: "If you go to a conference on physics, you hear discussions about physics (how the world works). If you go to a sociology conference you hear debates about method."

If you go to a *ML* conference *all* you see is methods. They are given the much grander name of "algorithm", but methods they are nevertheless. The development of new methods is the taken-for-granted goal of most ML research. (The *ends* are paid little attention to.)

And how do you *compare* such methods? You run them on a "benchmark" data "set". This way of working has overtaken the field: a couple of decades ago there were few benchmarks. Now we drown in them. And they are so loved because they allow the construction of leaderboards so

---

[4] Gerald B. Wetlaufer, Rhetoric and Its Denial in Legal Discourse, *Virginia Law Review,* Nov., 1990, Vol. 76, No. 8 (Nov., 1990), pp. 1545-1597

[5] Chaim Perelman and Lucie Olbrechts-Tyteca. *The New Rhetoric – A Treatise on Argumentation.* University of Notre Dame Press, 1969.

[6] Confer Daniel Rosenberg, Data Before the Fact, in *Raw Data is an Oxymoron,* pages 15-40, MIT Press 2013.

everyone can see whose algorithm is “best” (perhaps this serves the same purpose as Merton concluded priority claims in science do — to provide *validation* to the authors).

I perceive several problems with this perspective.

1) The benchmark is presumed to be *representative* (there’s that “factyness”again)
2) The performance is (usually, not always) judged in terms of predictive accuracy *alone*
3) The benchmarks often are made from “convenience samples” (data you find just lying around) and little or no effort is made to justify they use with methods that assume certain properties of the data (see later re randomness)

But my biggest concern is not with “benchmark” but with data “*set.”* A set is a thing. A pervasive presumption is that data is actually a *thing*. After all, one can just download the thing onto your computer. What on earth could data be if it were not a thing?

A *process*[7].

Contrast the approach to data of much of ML with that of a careful empirical scientist. The ML user literally will just download the data. All that matters is the numbers. Questions of its provenance, reliability, choice of categorical labels are usually ignored as inconsequential.

So what is an alternative?

Consider data as a *process*, or to use a metaphor of Latour, as a “*reversible black box*[8]”. A common complaint about algorithms is that they are “black boxes”. I wish they really were. The engineer’s black box comes with data sheets and a manual that tells you how it will behave. But data can also be framed as a black box, and its manual would serve as its warrant.

All this matters because there is *no such thing* as the “intrinsic structure of the data” notwithstanding the commonality of this phrase! (Try a google, or scholar search for the quoted phrase to see its pervasiveness). Data is *always* the end result of a chain of operations, and without knowing them, all you have is a string of bits with no “intrinsic structure” from which it is impossible to make a persuasive argument.

The comfort of randomness

I have learned that the scientific paper has the rhetorical structure of an *enthymeme* (ML folks are not the only ones with impenetrable jargon!), whereby a major premise is left entirely implicit — the scientific “background knowledge”. So too is the rhetorical structure of ML an enthymeme, or if not, a major premise is at least swept under the carpet and not examined.

---

[7] Confer Robert Williamson, Process and Purpose not Thing and Technique, *Harvard Data Science Review* 2(3), 2020

[8] Bruno Latour, *Pandora’s hope: essays on the reality of science studies.* Harvard university press, 1999

The major premise I have in mind is that the data is “drawn independently from some probability distribution.” This is important for much ML because its methods are built on the mathematics of probability theory. Indeed, *if* one’s data is so “drawn”, then there are many mathematical results one can rely upon.

The difficulty is that much of the data used by ML is most certainly not so “drawn.” Rather it is a “convenience” sample of whatever could be found lying around. (This matters a lot: Facebook’s vaccine hesitancy study had a sample size of 250,000, but this was as informative as an actual “random sample” of size 10^9.)

There is also the subtle point whereby “drawn from a fixed but unknown distribution” is widely used (again google the quoted phrase) *even though you will not find, in any statistics text an explanation of this process.* That is, you will not find a complete mathematical description of how you “draw” a sample from a distribution!

Why is this strange assumption so popular? Because via the simple incantation of “drawn iid” *one need not worry about the long chain that created your data* — you can get on with playing with your models — and models are sexy, right? (Funnily enough a few years ago the world’s premier ML conference forbade someone demoing work on the use of ML in fashion design from bringing (fashion) models into the conference venue to show off their results due to concern about the “optics”. I observed that their torrent of papers would dry up if they enforced this rule forbidding models consistently!)

Why does this matter? After all everyone does it.

Because it greatly affects the results! If one tries to more deeply understand this “drawn independently” notion one is led to try and understand *randomness*. Widely used as an undefined primitive that is presumed to nevertheless have a clear meaning, it *is* actually amenable to analysis. When one does so, one finds there is actually no single universal “randomness” — such a notion is logically void. What we find instead is an infinite family of different types of randomness. But, you might say, surely this is just hairsplitting? Why not just pick a canonical one and move on?

Because it matters, and matters *ethically*. It turns out (see the paper by Rabanus Derr and myself available on fm.ls[10]) that randomness and fairness are essentially the *same thing*: if you accept that the choice of a “protected attribute” (e.g. race or sex) is a consequential (and no doubt essentially contested choice), then you have to draw the same conclusion regarding randomness because the very definition of fairness in ML reduces to a particular kind of independence, which

---

[9] Valerie C.Bradley, Shiro Kuriwaki, Michael Isakov, Dino Sejdinovic, Xiao-Li Meng, and Seth Flaxman. "Unrepresentative big surveys significantly overestimated US vaccine uptake." Nature 600, no. 7890 (2021): 695-700.

[10] https://fm.ls/publications/fairness-and-randomness-machine-learning-statistical-independence-and-relativization

is a (very strong) type of randomness. And this “independence” is on very shaky ground: it is used to justify the move from individual samples to a distribution, but it is defined in terms of the *very thing that it is used to justify — the distribution*. (More sophisticated notions of randomness are defined, e.g. due to von Mises and Vovk, more reasonably, in terms of the data itself, rather than the theoretical entity of the distribution).

These subtleties cause much grief even for very simple prediction problems (who is best suited to get a college placement offer?). What they mean for the more complex uses of LLMs defies my imagination.

One vs Many

Another fundamental problem with ML is that of the one and the many. ML is built on statistics, and statistics works with aggregates. We count. But what is that do we count? We count according to predefined categories — this many white people, this many black. Then we can compare our counts. That is statistics.

In one of the earliest books on the theory of statistics (John Venn’s *The Logic of Chance*[11]) he so recognised the importance of the choice of categories (you can’t accumulate counts if you do not what you are counting) that he coined the notion of a “natural kind” providing much work for future philosophers (who are all barking up the wrong tree I would say).

It matters not that you have the best algorithm (according to your simple leaderboard) if you have not warranted the choice of categories that you used when feeding the algorithm. Why is that? Because while ML algorithms work with aggregates, they typically have their consequences on individuals, who can be placed into many *different* aggregates — this is often called ``the reference class problem” as if it is a problem that can be solved. (To be sure, sometimes one cares of an aggregate, but in the most emotionally and politically charged cases, it is individuals). Perhaps old style soviet social planners can *exclusively* concern themselves with the welfare of the aggregate. But essentially all ethics is about individuals.

I claim that (astonishingly) the connection between the aggregate and the individual is largely ignored in almost all uses of ML and it is taken for granted that reasoning at the aggregate level suffices. (Confer Casey Mock’s talk this morning: a large part of the problem with bureaucracies is that they tend not to care about the individual, but rather only aggregate performance indicators … like the predictive accuracy leaderboards of ML methods.)

Some statisticians are aware of the problem, under their name for it “the ecological fallacy” — the fallacy being reasoning from the group to the individual: what can you conclude about *me* if the conference organisers tell you that 50% of the attendees are vegetarian? *Nothing*. Interestingly, this is simply the dual to the problem of randomness (which assumes away the difficulties of going from the individual to the aggregate).

---

[11] John Venn. *The Logic of Chance, an essay on the theory of probability*. MacMillan and Co., London, 1876.

But the single move done in most ML is to deal with this simply by use of the phrase “probability of” — this individual “has a probability of X” of some outcome, and all of the difficulties are buried in the undefined semantics of that single misunderstood word.

An appeal to spells

The dual problems of individual-to-aggregate and aggregate-to-individual both rely on what one could reasonably call a *spell*: one simply has to state “the data is drawn iid from some distribution” or that “the `probability’ that the patient will suffer is x” and all is well.

This is obviously far easier than deeply interrogating the data and constructing proper warrants. And yes, there *are* mathematical theories that go beyond probability and offer the promise (if not yet the practice) of being able to deal with the non-use of such spells — this is a compelling research topic!

“Generative” AI

What about “generative” AI? The name is odd isn’t it? After all, existing ML methods “generated” predictions or forecasts. And like them, GenAI is based on a statistical view of the world. As leading ML scholar Tom Dietterich has observed[12], LLMs are not a knowledge base, but a *statistical model of a knowledge base.* And how do LLMs “generate” their outputs? By sampling (!) from their posterior distribution! An LLM is very complex model of its inputs. But it is a *probabilistic* model. In other words it is literally a *probability distribution*. And the outputs it “generates” are the most likely ones under that model — a fancy version of autocomplete.

Yes, after the investment of countless billions of dollars and petajoules of energy we have built a cliché generator…

LLMs offer next to nothing in terms of warrants for their arguments. They do not know what data they ingested and they will not tell you. They are a library made from all the books found at a recycling plant, with pages removed from their bindings, all stirred together, and used without taking any account of their provenance. The apparent knowledge that they have is simply a statistical summary of the texts for which they have counted relative frequencies of different phrases. Thus my concerns re classical ML and its cavalier treatment of data apply forcefully to LLMs.

What is to be done?

If you treat ML as rhetoric then you should take care of the solidity of its arguments and warrants, and that has to be for the whole chain, not just the final step. Fortunately there is much that could be done differently. I offer some suggestions:

[12] https://www.youtube.com/watch?v=e8vg1vin78U

- **Don't view data not as a fact or even a thing, but as a process or protocol.** Keep its traces. *Show all your working!* And pass it on. View data as "capta[13]" (taken) or "sublata[14]" (lifted up)
- **Don't sweep the challenges of randomness, statistical stability and the ecological fallacy under the carpet.** Recognise that randomness is contextual, and document it as part of your warrant. Be especially wary of convenience samples. You would not put some random thing you found on the road in your mouth (unless you were 2 years old); why would you behave like a 2 year old with your LLM? And seriously grapple with the problem of going from the aggregate back to the individual.
- **Move on from benchmarks.** Their power (that they scalarise everything) is the biggest weakness. One could argue that the problems of AI powered bureaucracy are nothing to do with "algorithms" and all to do with the objective function being optimised.
- **Devalorise algorithms** — move on from the obsession with methods, and focus on the results of the *use* of methods. Decades ago it was observed by Rob Holte[15] that actually the simplest and "dumbest" ML methods worked just as well as the fanciest new ones, as long as one paid attention to the data. Perhaps a similar insight is waiting for us, and we will eventually have effective SLMs (small language models)
- **Make better black boxes,** that are reversible, and come with a manual. All complex engineering only works via the device of black boxes. So it is not a problem that LLMs are opaque, but they do need data sheets and manuals. And you still need to understand what they were built with (who would go in a skyscraper built with concrete which was not tested with every single pour?)
- **Expunge phrases** such as "data-driven", "the intrinsic structure of data", "drawn iid from an unknown distribution" and other comforting fictions that we collectively use to avoid grappling with the hard problems of ML.
- **Rather than viewing technology as a *thing*, or worse as an *agent*, think of it as a *delegation*** (Latour observed how we delegate to speedbumps[16] our complex moral views regarding the responsibility of car drivers to moderate their speed). We, unlike Socrates, are comfortable with the technology of written language (he was sure it would lead to mental degradation in the young). But just remember, when you delegate, you are still responsible…
- Above all else, **provide some room for disagreement.** Make ML, like all good arguments, *controvertible.*

---

[13] Rob Kitchin, *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage, 2014.

[14] Bruno Latour, Ibid

[15] https://webdocs.cs.ualberta.ca/~holte/Publications/simple_rules.pdf

[16] Bruno Latour, *Ibid.*