Vincent W thoughts

What is the point of ML? What is it learning? Based on a fixed set of training data, learn to predict the label of unseen test data. They learn this by assuming this information is in a vacuum and nothing outside (unpredictable) can arise. Example: we assess data on bubblegum placement on shelves in a NYC CVS. Running a model, we can find there is a strong correlation between placement by cash register and purchase. Intuitively, this is nudged to be implemented in every branch. However, when the LA store does so, it finds that nothing really changed, and profit even decreased a bit. This is because we trained our machine on data on correlates in data, not on what would happen if you change things in the store. Because once you change the store, the "distribution" changes, and the model (trained on the previous distribution) breaks. Obviously, this is the classic Correlation ≠ Causation.

We can not be selling these models based on "doing" that sounds plausible. We know in theory, they are not. Instead, we must rethink how to achieve causation through ML. Especially if much of the business model is rooted in changing behavior through correlates. If that is the point of ML, what should be learning then? Specifically, we must ask the question: What is the right success concept for causal discovery methods. This requires inductive inference. With Strong AI, we get guaranteed bounds on error over all possibilities deduced from a distribution. In causal inferences or counter-factual inferences we modify the system and extrapolate way beyond the data. So now here comes Socrates: What could it even be that you think you're doing [with these algorithms]. So we must stop the "doing", pause, and ask what would it be to do it? Aka "Why?". And if you look at the gap in theory right now, you realize, we don't even know.

So let's approach with the (epistemic) modal logic: if ML models "learning" is necessary, then it is also possible. One of two possible approaches I can think of is Transfer Learning: applying knowledge stored in one domain towards another. Using insights from a sample to other samples (within or not within the population). Example: we know use a model's semantics (fundamental patterns/attributes it uses to identify data) from a chimpanzee sample is opposable thumbs, two eyes and hair; can we apply it to gorillas, 💩 or humans? The key is being able test models on test data without training on pictures of that same data, and yet still find underlying semantic commonalities (not necessarily correlates) that explain the new data from the bottom

up. This is how we humans generally deal with Transfer Learning; we look for the semantic underpinning (first principle) of a new object. We then equivocate it to something pre-registered that we know. And build up as much as we can infer about this new object from there. We can extrapolate chimps, gorillas and humans all share a trait up to where the genus of the (bipedal featherless chicken) species diverge. And of the traits that are shared (eyes, mouth, poopy hole) we can infer the purpose caused by these correlates (see, eat, finger). The "why" in Transfer Learning, and thus Machine Learning, is to understand the semantic similarity of objects we can and can't identify. Like dogs and pens or time travel and 4D knots. The other possibility of why we need ML is Self-Supervised learning. I'm too lazy to flesh it out the idea bc its late but basically, if we can understand and point out the attention mechanisms of ML algorithms & where/why/how they create the semantics/patterns, then we can apply them to the understand of how we humans learn, and how best we can/should learn. Rather than telling robots to think like us (which they can't/shouldnt because we were evolved to worry about survival and creativity solutions), they tell us how to think like them. Thus the "why" being: to find out what is learning, and how best to learn. Note that this is still problematic because we know by feeding it structured data, it has our human biases engrained. Rather than optimizing the algorithms around the "accuracy" parameter or another loss functions we humans deem important, we should instead let it run completely unsupervised on unlabeled data. See what it picks up, and why/how it did so, and what/where can it find the same pattern in other objects. Basically, the stuff philosophy is traditionally focused on; substance, causation, identity, essence, etc. is essential to understanding ML's purpose. Unfortunately, ML's current purpose is to yield profit, especially since the strongest and "best" algorithms are backed by industrial titans.

Took me a while, sorry. Take a while to read/respond if u want. Maybe audio message if u don't wanna type. If it's too much effort no worries either