



Arbitration

A Few Useful Things about Generative AI (With Arbitrators in Mind)

John Paul MacDuffie

Wharton School, University of Pennsylvania



National Academy of Arbitrators Conference

Boston, MA

May 8, 2024



Definitions

Human Intelligence

I think therefore I am...

Artificial Intelligence (AI)

Computer systems capable of performing complex tasks historically only a human could do, e.g., reasoning, making decisions, or solving problems.

Machine Learning (ML)

The subsection of AI that teaches a system to make a prediction based on data it's trained on. [Auto-complete for texts is a simple ML model.]

Generative AI

ML models that “generate” brand-new output, e.g., text, photos, videos, code, by making predictions based on data they have been trained on

Large Language Models

LLMs use unsupervised machine learning and are trained on massive amounts of text to learn how human language works.

Specific Tools

Tools based on LLMs: Open AI’s Chat-GPT & GPT-4; Google’s Gemini; Anthropic’s Claude; Microsoft’s Co-Pilot; Facebook’s Meta AI

Definitions

Human Intelligence

I think therefore I am...

Artificial Intelligence (AI)

Computer systems capable of performing complex tasks historically only a human could do, e.g., reasoning, making decisions, or solving problems.

Machine Learning (ML)

The subsection of AI that teaches a system to make a prediction based on data it's trained on. [Auto-complete for texts is a simple ML model.]

Generative AI

ML models that “generate” brand-new output, e.g., text, photos, videos, code, by making predictions based on data they have been trained on

Large Language Models

LLMs use unsupervised machine learning and are trained on massive amounts of text to learn how human language works.

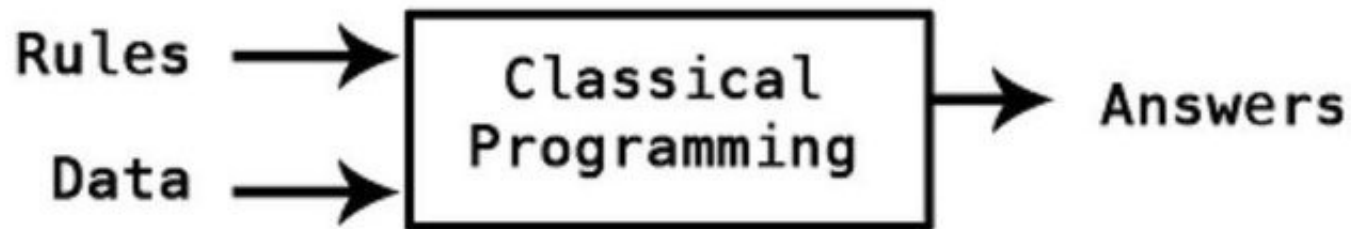
Specific Tools

Tools based on LLMs: Open AI’s Chat-GPT & GPT-4; Google’s Gemini; Anthropic’s Claude; Microsoft’s Co-Pilot; Facebook’s Meta AI

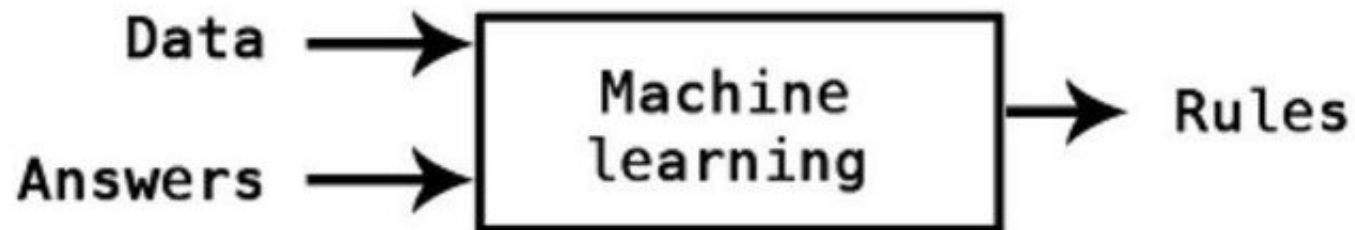
General Truths about Gen AI & Machine Learning

- How GenAI differs from conventional AI (and most software coding)
- What Machine Learning makes feasible
- Improving very rapidly when models fed with more data and “compute”
- No one understands exactly how it works (*AI experts know this best*)
- Wants to please...
- ... yet “hallucinates”...
- and is a master bullshitter

What Make Machine Learning Different?

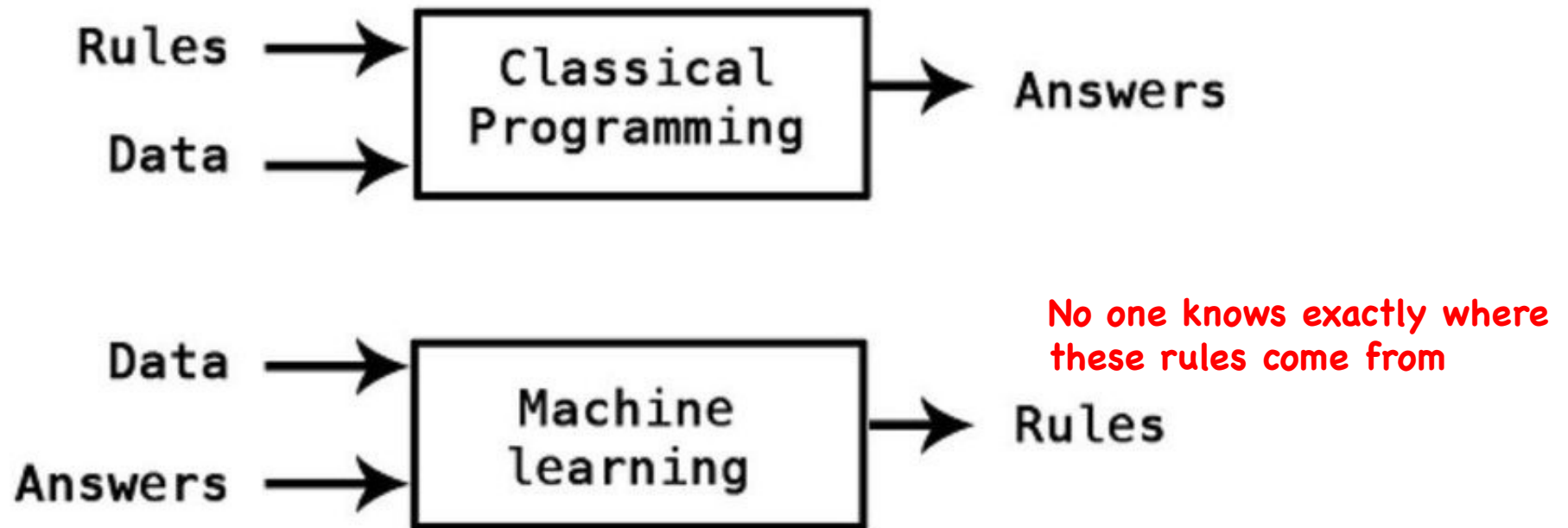


Rules show how answers are reached



Source: [François Chollet](#) from
Benedict Evens (2018)

What Make Machine Learning Different?



Data can be problematic... a constraint or source of bias

"Answers" are tricky - i.e. when models optimized for one outcome are applied for a different outcome

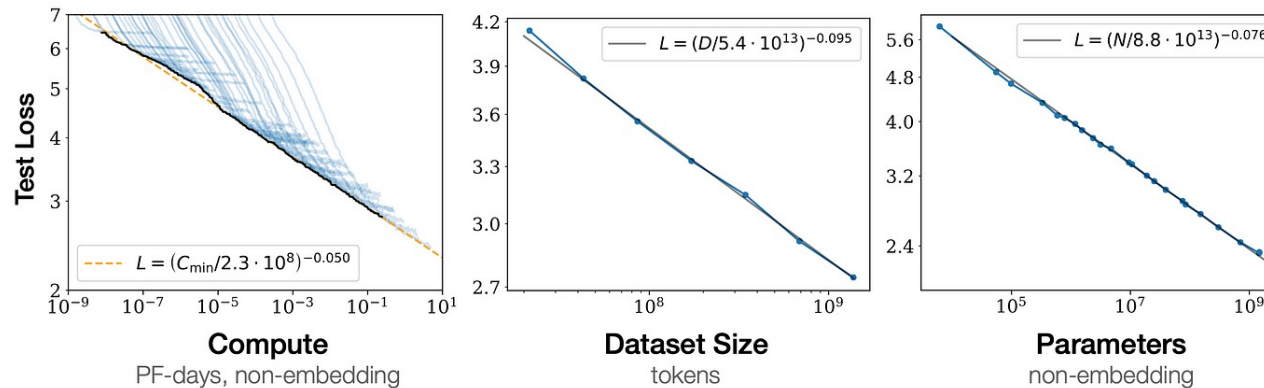
Source: [François Chollet](#) from Benedict Evens (2018)

What Machine Learning Makes Feasible

- Applied to Semi-Routine Tasks That We Consider Uniquely Human
 - Writing!
- Applied to Routine Tasks That Could Never Be Automated Before
 - Picking up irregularly shaped items from a haphazard pile
 - Spotting wrinkles in the fabric on a car seat
- Asking New Questions of Data You Already Have
 - “Listen to all the customer service calls and pick out the really angry/exasperated ones”
- “Reading” Data We Couldn’t Read Before
 - Images and Video



Improving very rapidly – tech trajectory vs. application



- Exponential improvements, aka “Scaling Laws”, when adding data and compute
 - Humans have trouble envisioning exponential change
 - E.g. 1 COVID case in Month 0 → 1000 cases in Month 1 → 1 million cases in Month 2 ...
- Technology steadily improving behind the scenes – the world sees huge jumps when new versions are released or new applications are demonstrated

Quirks of Gen AI (1)

- No one understands how it works
 - The “rules” created by applying machine learning to data to optimize a particular “answer” are **opaque**
 - Impossible to know how the model got to those specific rules
- Wants to please, i.e. gives us strings of words that we perceive as:
 - Legitimate (well-written)
 - Persuasive (presents an appealing logic and flow)
 - Pleasing (responsive to our prompts, good at dialogue)

Quirks of Gen AI (2)

- “Hallucinates”
 - When generating a legitimate and persuasive strings of words to please the prompter, it makes stuff up...
 - ...and can't tell you that is has done so
- Master bullshitter – using the definition of Harry Frankfurt (2005)
 - “Bullshit is speech intended to persuade without regard for truth.”
 - “The liar cares about the truth and attempts to hide it; the bullshitter doesn't care if what they say is true or false”

AI and the “Jagged Frontier”



- Ezra Klein: “You’ve talked about AI’s ‘jagged frontier,’ which means the technology is good at performing some tasks and bad at performing others, even when those tasks seem to be the same level of difficulty. Do you have any tips for determining whether or not AI will be useful for a given task?”
- Ethan Mollick: “That is what makes it so interesting. The answer is generally no. Any heuristics I give you will be overwhelmed by weirdness inside the system.”
- **“So in your field you're going to have to figure it out.** It's too complicated. Nobody knows in advance. There isn't a really good heuristic other than: **Use it to find out.**” *Emphasis added.*

AI and Creativity

- “Selection is AI’s Achilles heel.”
- “The technology can give us lots of ideas, but it cannot judge which are better.”



Christian Terwiesch

Early Ideas on Shaping GenAI's Impact (1)

- Strategy: Large Language Models (LLM) are the “Mass Market” for GenAI – Leaving Niches for Small LMs (SLM)
 - Non-public data from past arbitration decisions (not accessed by GPT-4, Gemini, etc) could provide input for a SLM organized to support the parties engaged in arbitration.
 - SLMs: more focused machine learning from high-quality specialized data; less compute-intensive; how to control access?
- Risk: “The Boring Apocalypse”
 - Efficiency orientation drives use of GenAI to draft documents based on bullet point prompts
 - Recipients of those documents use GenAI to summarize the content → a new list of bullet points
 - Rinse and repeat... with huge loss of information, nuance, insight

Early Ideas on Shaping GenAI's Impact (2)

- Opportunity: Follow the Lead of the Screenwriters
 - Recent contract after Writers Guild of America (WGA) against movie and TV producers put limits on use of GenAI
 - Writers can choose to use GenAI but can't be required to do so – and can prevent their past work from being used as data to train future models
 - A particular concern – avoid “first draft” use of GenAI - risk of constrained scope for starting point of creative process - rather use it to revise

Early Ideas on Shaping GenAI's Impact (3)

- Perspective: No “Technological Determinism” – Human agency in deciding how AI models are built and applied, what data to use
 - Time is now to be proactive in anticipating potential risks and challenges and setting parameters and policies to influence future trajectory for arbitrators
- Conclusion:
 - Contexts like arbitration -- in which specialized expertise is crucial, protecting confidentiality is essential, and liability for mistakes is a big risk -- are perfect for designing complementary roles for humans and AI

