

□ Artificial Intelligence □ yesterday, □ today and tomorrow □

© 2024 by Michael Vorburger.ch (presenting in personal capacity and not representing employer)

Content licensed under Creative Commons Attribution-ShareAlike 4.0 International.

The source code of this presentation is available on GitHub!

Input

Prompt: *Can you make this sound cooler?*

Michael Vorburger has been fascinated by the creative power of computers from an early age. He is an avid hacker by night, see <https://www.vorburger.ch>, and currently works on ML technical infrastructure in management for Google by day. He is presenting at this event in his personal capacity, not representing his employer.

Output □

How does this work? What does this mean for the future?

By night, Michael Vorburger transforms into a digital alchemist, coaxing unseen potential from the very ones and zeros that form our world. An avid hacker with a website that whispers secrets of the craft (<https://www.vorburger.ch>), he spends his days architecting the intricate machinery of Machine Learning infrastructure for Google. While tonight, he sheds his corporate cloak to share his personal visions, untethered and electrifying.

"AI is one of the most important things humanity is working on. It is more profound than (...) electricity or fire."

Sundar Pichai, Google CEO, 2018 and later.

□ What do you think?

Retrieval-Augmented
Cloud Models
Data Language
AI Intelligence
Learning Science
prompt Artificial
super Intelligence
Turing Test
Superintelligence
Creativity
Productivity
Cyber
Input things
translation visual
programming
MultiModal
Grounding Knowledge
Big Video
symbolic Education RAG
compute gener
electricity reinforcement reading fact Chat Inferring Explainable Magic instructions humanity Deep one images billions LLMs History Coding Graphs Generation Gemini learn vocabulary KGs
language
reinforcement learning
backpropagation profound image model rules books
cluster algorithms text Code Health audio train Systems
important output infrastructure
AGI Expert
ML computer grammar
Artificial large computer
Training Machine

History □

Artificial Intelligence (AI) arguably started in □ antiquity.

Modern day AI from ~1960s, with emergence of Computer Science.

Initially □ "symbolic" with "rules", e.g. *Expert Systems* à la Cyc; AI * Winters.

Machine Learning (ML), with *Deep Learning* for *Generative AI* are subfields of AI - with a different take; why?

What is
programming □

What is *programming*?

Programming is to give computers (very) precise instructions, AKA "*Code*", for what you want them to do, like:

- Print `hello, world`
- Variable `i = 7`
- `if i is still 7 then do-that`
- `for loop 7 times`

Written in a computer language - try out Scratch, it's great to learn concepts!

(Or C++ or Java & Kotlin or C# or Python or JavaScript & TypeScript or Go or Rust, etc.)

How did you learn
your mother
tongue?

Survey: How did you learn your mother tongue? □

- a) I was *programmed* by being □ fed many □ chopped up grammar and vocabulary □ books, and thus learned syntax and semantics to sound smart!
- b) I □ babbled incoherently for **many years**, until, after *a lot of* visual ☺ and □ audio cue *reinforcements*, I finally figured out how to make sense!

Quick show of hands... who a) ... who b) □ ☺

Machine Learning (ML)

The idea of ML is ~ just to treat computers as \square babies, instead of *programming* them! For example:

- Chess through *trial & error* - instead *rules*
- *Vision* with example images - instead of *algorithms*
- *Machine translation* by "reading" human translated books - instead grammar
- *Large Language Models* (LLM) by "reading" A LOT of text, and then reply to prompts

The basic idea is not that new (backpropagation ~1980s?), but it turned out to be a lot more \square "fun" with the emerging increasing availability of *Big Data* and massive cloud storage & super compute cluster infrastructures, starting ~2010s.

Magic?

Is ML \square magic? Not at all... the basic idea is really quite simply, actually! To illustrate:

High-school math: $y = a*x + b$

Given a *training data set* of some *points* (x,y) representing car fuel efficiency,

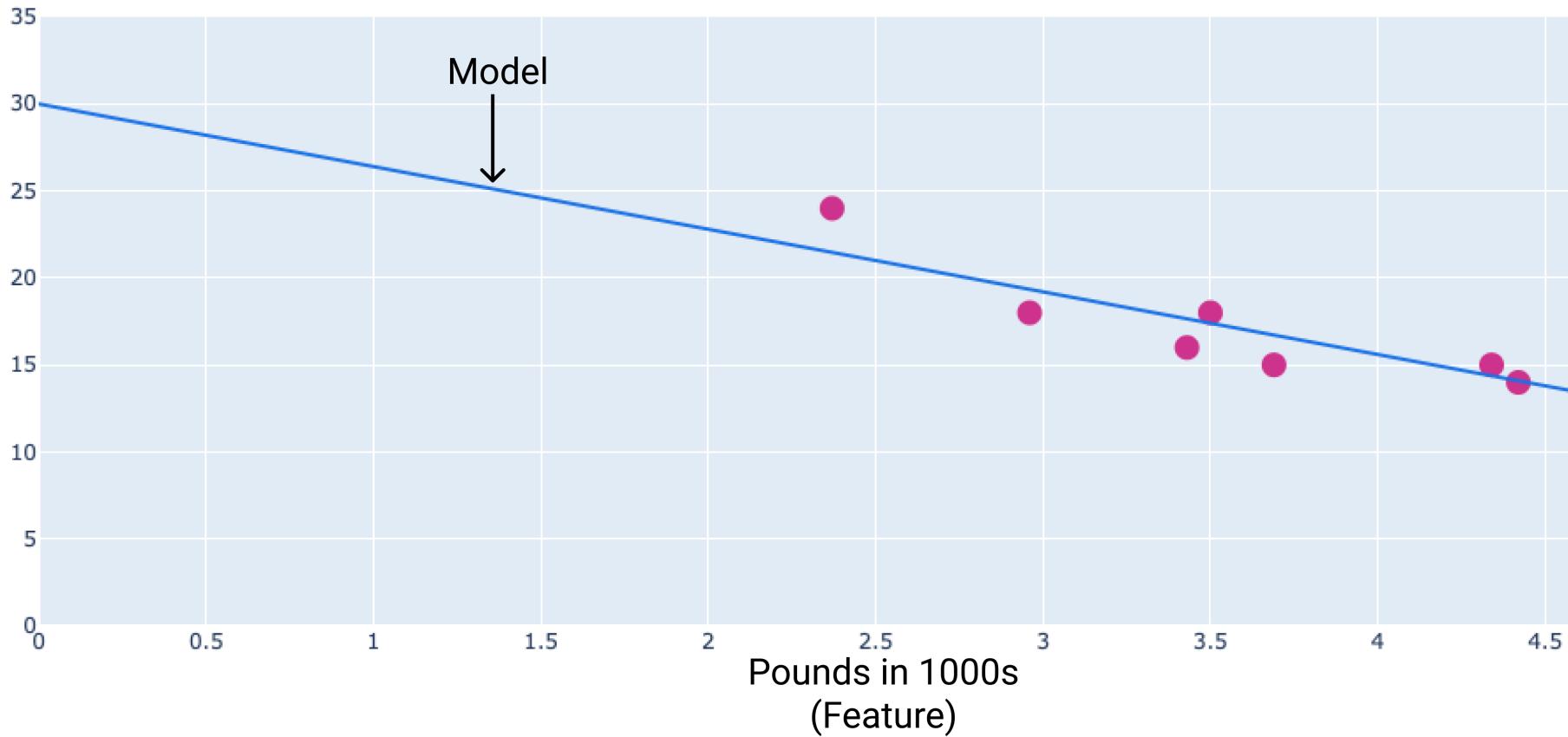
where X is car weight an Y is KMs per Liter of Gas...

...find a and b - and that's a (2D) ML model, of 2 parameters!

Now, given a new car's *weight* (X), you could (roughly) *predict* its gas consumption.

Except that a real LLM is a little bit bigger; e.g. Google's open source Gemma (v2) has 27 billion \square instead of just 2 such parameters! (And Google's Gemini is even bigger.)

on)



Training & Inferring

- *Training* is creating said *ML Models of Parameters* (from input data)
- *Inferring* (AKA *serving*) is giving a model new (unseen) input, asking it to "*infer*" output

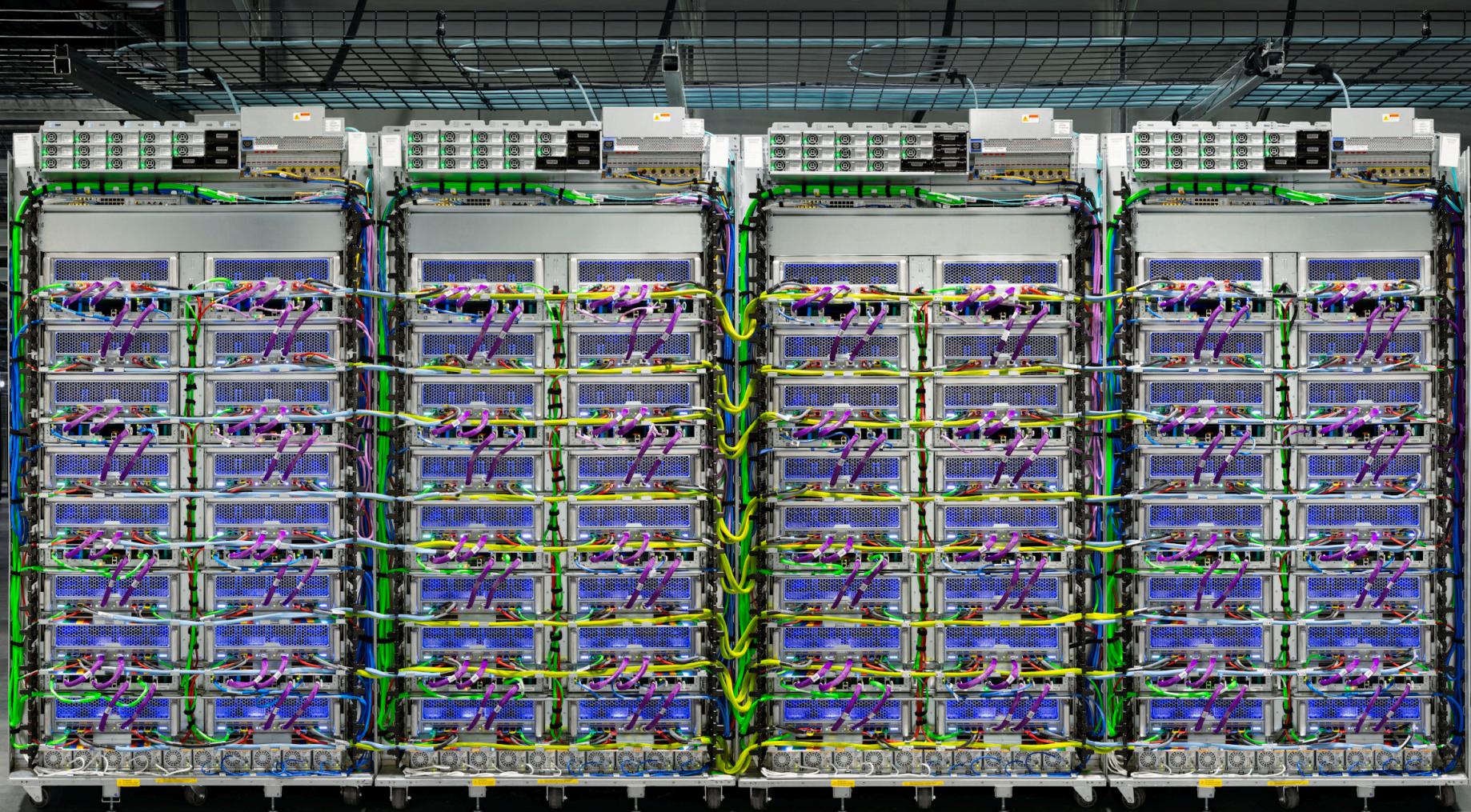
Training effort is (~) proportional to size of input data.

Inferring effort is much smaller (comparatively).

Mobile Phones can \square do (~) inference locally, without \blacktriangleleft Cloud DC! (Interesting for \square Privacy.)

Training Data

You need a massive amount of text or images or video to train Large Models with billions of parameters...
... and there are some interesting open questions around this.



Cloud? Open Source?

ChatGPT & Google Gemini etc. run in the  Cloud.

You can download e.g. Google's Gemma, Meta's Llama, Mistral, etc. to run them yourself "at home".

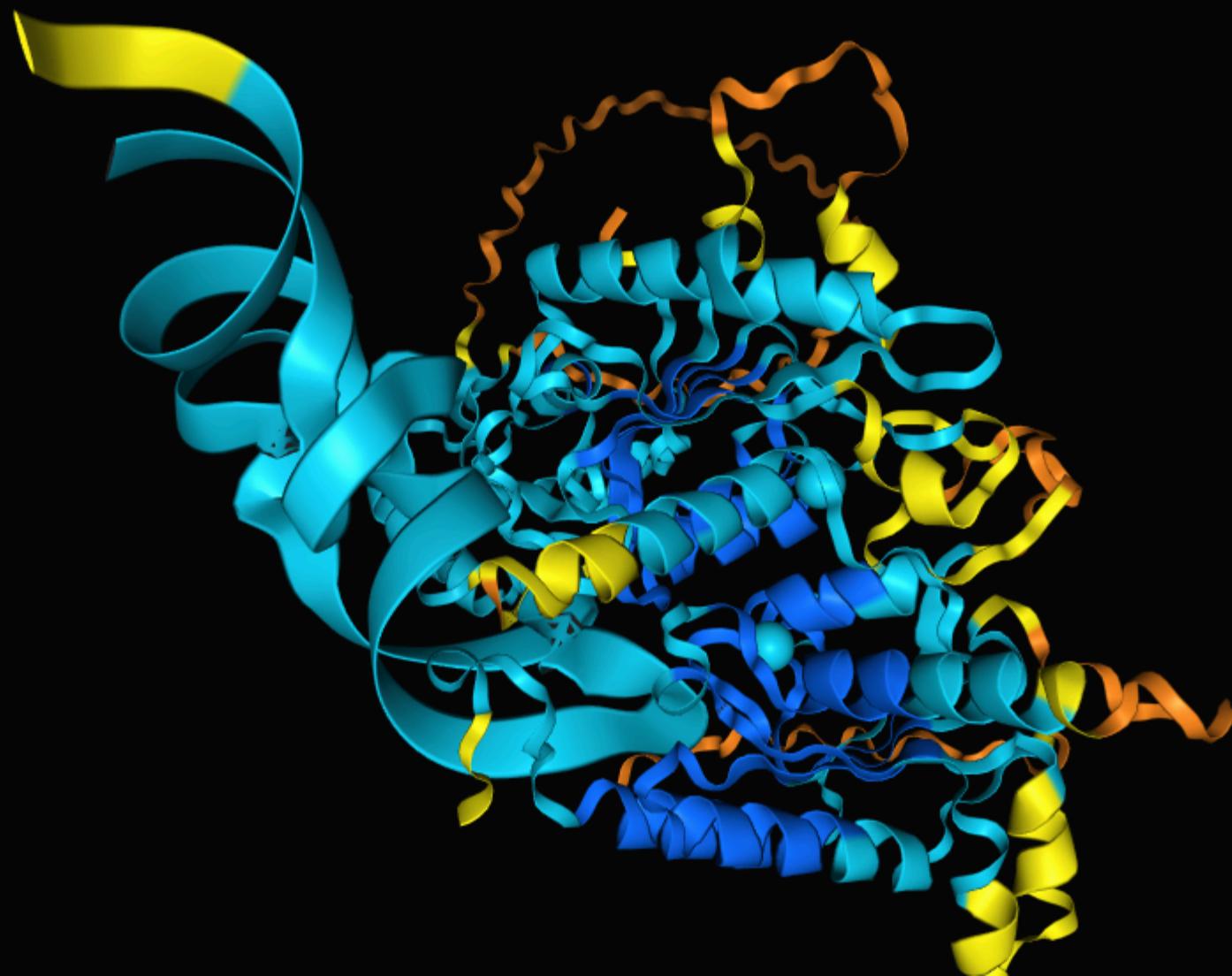
Do try e.g.  Ollama to run an LLM @ Home on a GPU.

PS: Training is still proprietary - it's hard to DIY.

Applications

- Chat Bots
- Coding: E.g. GitHub Copilot (and many others)
- Productivity: E.g. documents, email summary; or Meeting Minutes transcription
- Health: E.g. Google's Health AI for breast cancer screening or expanding access to ultrasound
- Science: E.g. AlphaFold "protein folding" breakthrough unlocking research of new medicines (access)
- Education: E.g. Khan Academy's Khanmigo, or Homework with Google Lens
- Deepfakes & misinformation spam - **and** their detection

Is (some of) this "*creative*"? You tell me...



Personalized LLM

Large Models might know about "the world", not (yet) "your world"... but:

- Huge prompts (*"context window size"*) to "chat with your PDFs" works; try e.g. [NotebookLM](#).
- *Fine Tuning* is another ML technique to efficiently adapt a previously pre-trained model with new data.
- "*Workflows*" (e.g. LangChain's LangGraph, et al.)

The Future

Artificial general intelligence - AGI? Superintelligence?

Not super duper clear to anyone what exactly this really means... 😊

Generally colloquially used as ~ "it can do some of your work".

We'll see; the next few years are going to be very interesting!



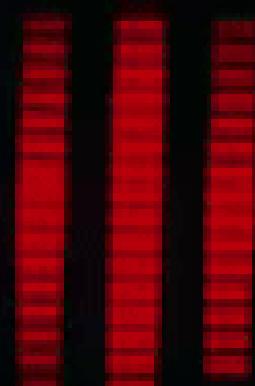
Hello, computer.

AIR

OIL

P1

P2



S1

S2

P3

P4

Gaps?

It's still early days. But more is coming - and (very) quickly...

- *Explainable Neural Networks (XNNs)*
- Grounding in Facts, to cure Hallucinations (e.g. LLM + KB/KG with RAG; or "neuro-symbolic")
- More Multi Modal, and Sensing (watch Project Astra!) #latency
- Interacting with Our World (Web, APIs)
- New UX? Knowledge agent design?

You?

Machine Learning is a lot of fun! Get started with exploring it today:

1. gemini.google.com to learn "*Prompt Engineering*"
2. [Google AI Explorables](#)

For developers:

1. [ML Zero to Hero x4 ~6' \(=24'\) Videos](#)
2. [Google Machine Learning Crash Course](#)
3. [TensorFlow Quickstart](#)
4. [TensorFlow Learn ML](#)
5. [Google AI Build](#)

More on [LearningMachineLearning](#) - contributions welcome.