



Gord Larose

Presented to Probus Western Ottawa, Sep 10 2024

# 2024: Artificial Intelligence Gets Real



# Agenda

A topical example

A brief history of AI

ChatGPT and relatives

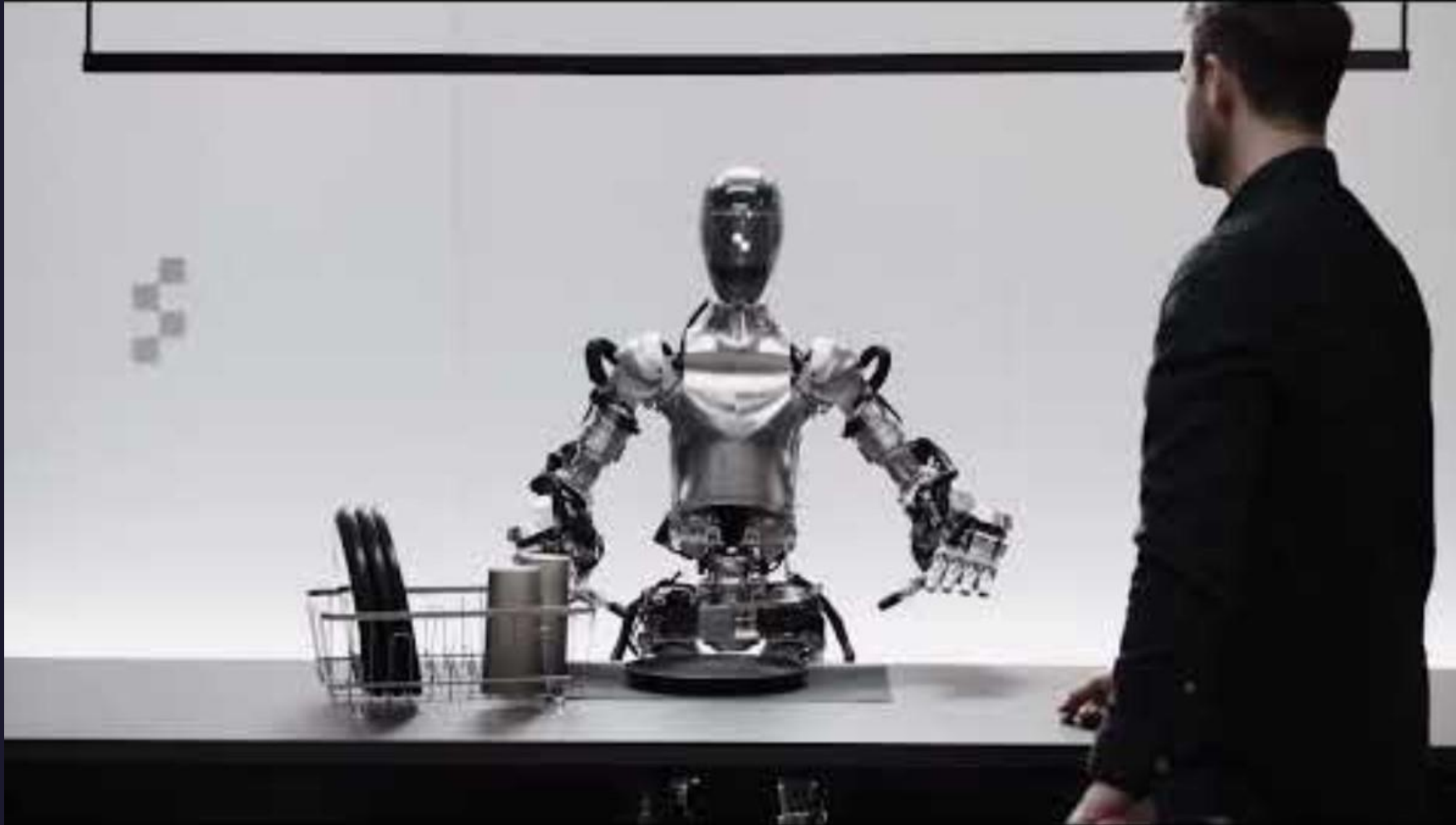
Recent developments in AI

Issues of AI

Links/References



# 20 minutes into the future.....







# A brief history of AI

- The notion of AI – machines who think – long predates computers. In the last 80 years as computers have advanced dramatically the idea of thinking machines is no longer fantasy.

# Prehistory – 1920s: Myths & Machinery

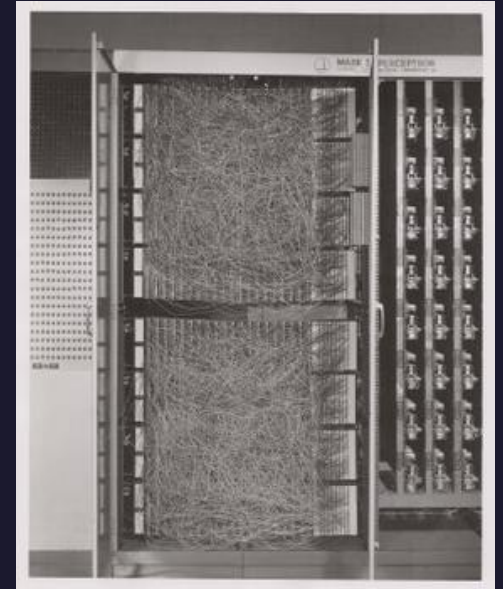
- Greece: The Colossus of Rhodes, Talos, Antikythera Mechanism
- Middle ages:
  - water-powered automatons in the Islamic world
  - Roger Bacon's Brazen Head – the first chatbot!
- Renaissance: Clockwork tableaux
- 17<sup>th</sup> Century: Babbage's Difference Engine
- "Rossum's Universal Robots" play of 1920  
The word robot comes from a Czech word for "laborer"
- Some of these could "calculate" but  
*they made no real attempt at "intelligence"*



# 1930s-1956: Real computers, imagining AI

- The earliest “computer” designs were modeled on both brains and calculators!
- Perceptron: 1943-1958  
The first “artificial neurons” mimicking about 500 cells of the human brain – which has about 100 billion connected neurons.
- ENIAC: 1943  
The first “general purpose programmable” digital computer.
- The “Turing Test”, 1950  
Could a machine fool you into thinking it was human?
- The Dartmouth Conference, 1956

**Could a bunch of smart guys crack AI in a summer?**





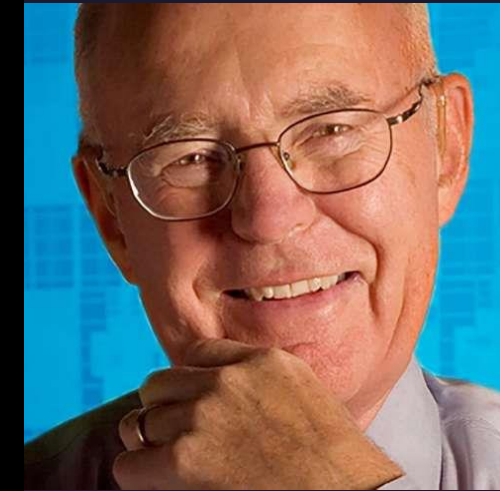
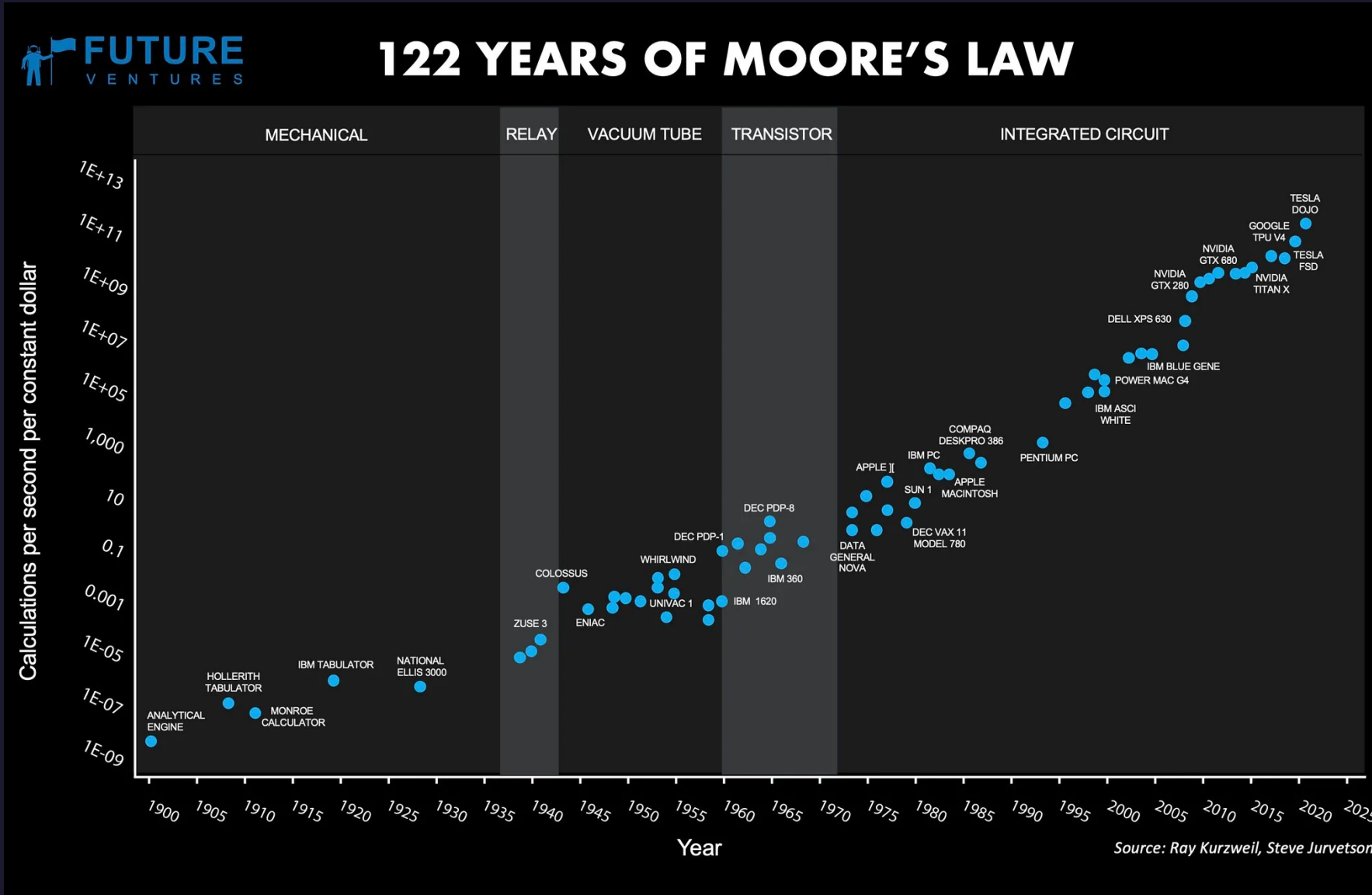
# 1957 ~ 2000: AI flounders

- Hype cycles, academic squabbles, competing paradigms, funding booms & busts
- “AI Programming”: symbolic logic (LISP, PROLOG)
- [ELIZA, The Rogerian Therapist Chatbot \(1966\)](#)
- Expert systems, “Fifth Generation” computing
- Early neural networks aka “connectionism”  
Sometimes with dedicated HW, sometimes simulated  
Hundreds or thousands of “Neurons” – *not enough*



• **NONE OF THIS STUFF REALLY WORKED! INTELLIGENCE IS HARD!**

# Meanwhile: Conventional computers soar





# ~ 2010 - 2022 : Neural Nets take over AI

- Around 2010, computers could simulate neural networks – models of the brain- with MILLIONS of nodes. Such AI lurked behind the scenes for years (Siri, FB, Alexa etc.)
- These networks are trained, usually with human feedback, for many tasks. Much training data is general internet content... with attendant biases etc.
- A decade or so ago, "emergent" intelligent behaviors showed up.
- Major players (MS, google) dabbled with public chatbots as early as 2016.
  - But problematic behaviors showed up which were hard to fix, (racism etc.) and they backed off.
- This opened the door for risk-tolerant startups like OpenAI!

# 2023: chatGPT

- Limited release Nov 2022, exploded in 2023
- Spookily flexible and smart
- Can pass standard tests like LSAT
- Can write in different voices (“as if a dog”)
- A Large Language Model (LLM)
- Uses Generative Pre-trained Transformer technology, GPT 3.5
- Has lots of limitations: makes mistakes, no “today” knowledge, odd behaviors.

## ChatGPT Sprints to One Million Users

Time it took for selected online services to reach one million users



\* one million backers \*\* one million nights booked \*\*\* one million downloads

Source: Company announcements via Business Insider/LinkedIn

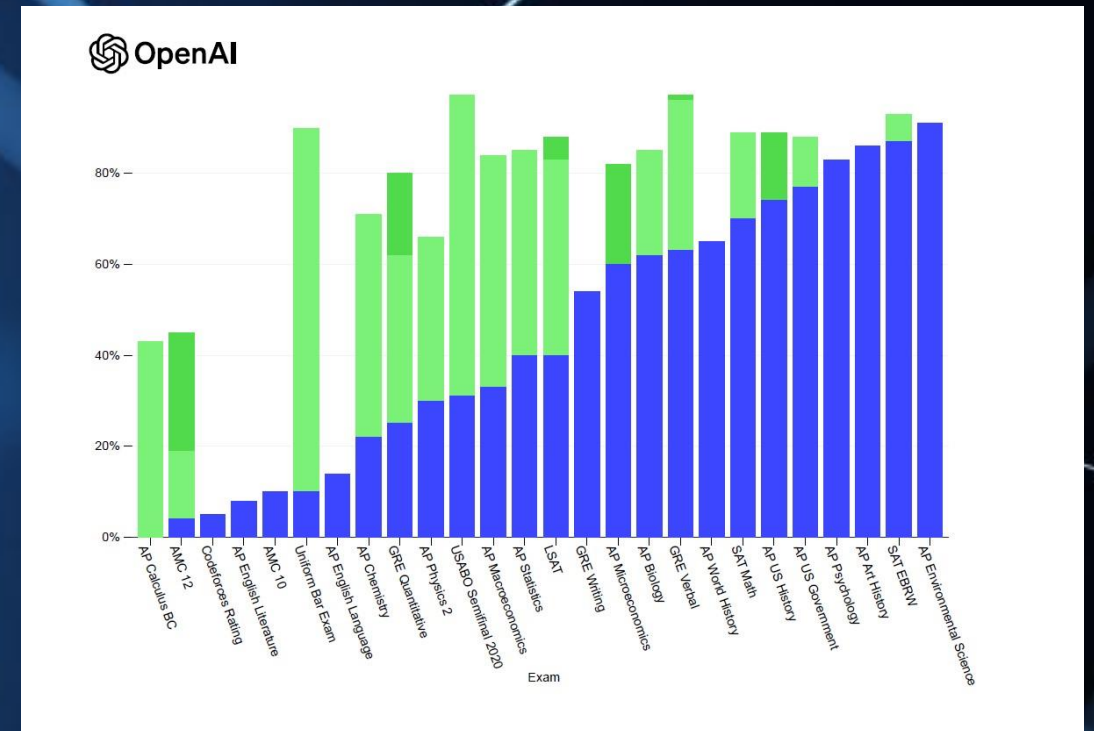


statista

• Let's Try It!

# 2023+: chatGPT *et al* explode

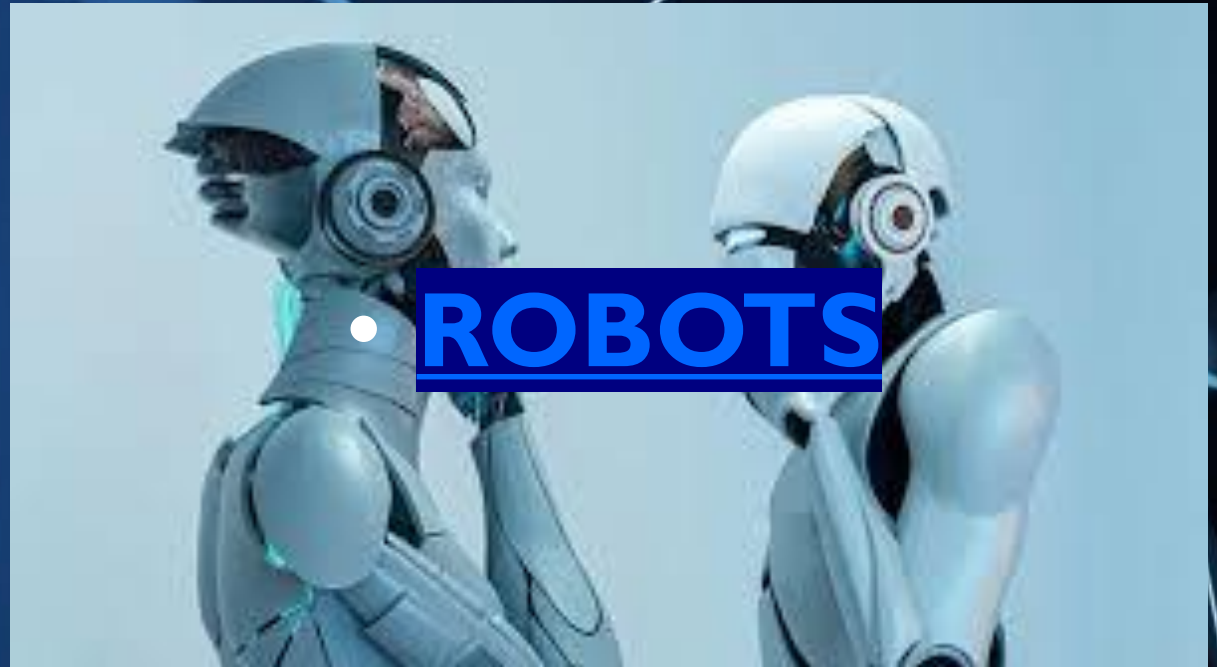
- GPT v4 can ACE various standard tests, not just PASS them. (See chart of **GPT 4** vs **GPT 3.5** ).
- Models will soon have BILLIONS of nodes and TRILLIONS of connection parameters.  
***This is comparable to the complexity of a human brain!***
- AIs are now being designed and trained by AIs, which will speed up their improvement.
- It turns out LLMS are good for almost ANYTHING: image generation & understanding, biology, robotics
- Variants (Bard, Bing, AutoGPT) add features such as searching current Web content.





# 2024: AI Everywhere

- More and better chatbots
- Digital Assistants e.g. perplexity
- Image and video generation
- Multi-function applications
- Usually free to try!
- Adjuncts to existing apps:
  - - meta AI image creator in Messenger
  - CoPilot in Ms Word etc.



# How LLMs work:



- WE DON'T KNOW HOW THEY WORK!
- But then, we don't know how our brains work either.
- We know what neurons are and how they behave....
- But we have NO IDEA how such cellular-level behavior adds up to intelligence, consciousness, language etc.
- So, we mimicked brain architecture without understanding it..  
And the mimic works!

# How LLMs are built:

- GPT “Generative Pretrained Transformer”, a common AI architecture
- Based on a brain-inspired “Large Language model” neural network
  - “Pre-trained” by reading vast amounts of data e.g. Wikipedia,
  - Like a brain that’s learned stuff as of a certain date, so doesn’t know the weather or what happened yesterday
- An evolved version of “predict the next word”  
To predict the next word well you have to UNDERSTAND what is being said.
- The abilities are largely “emergent” :AI scientists aren’t sure how they do everything they do, their abilities/ quirks go beyond their training. For example, conversing in Persian.

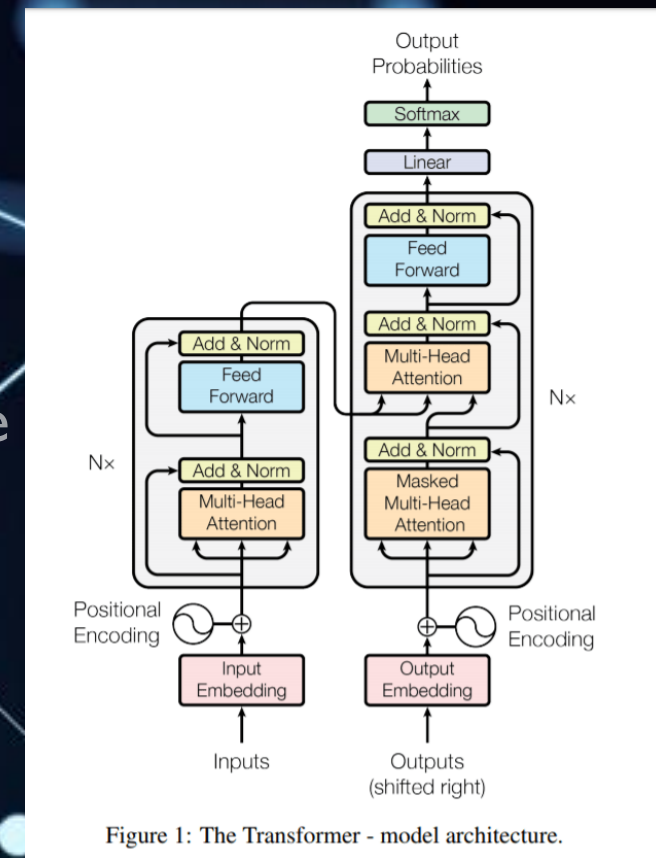


Figure 1: The Transformer - model architecture.



# The AI ecosystem

LAYER		EXAMPLES	PLAYERS
4	Applications	Alexa, Siri, driverless cars, image generation	Apple, Amazon, Google, MANY others
3.5	AI APIS: Application Programming Interfaces		
3	Technologies/ models	Deep learning, CNN, RNN, LLM, GLLM, ChatGPT etc	OpenAI, Google, Microsoft, Meta, Mistral
2	Languages / platforms	Python, R, Cuda, tensorflow, AzureML, Sagemaker	Nvidia (Many are open source)
1	Hardware & foundation technologies:	CPU, GPU, FPGA, Cloud	NVIDIA, AMD, Amazon, Microsoft, Apple

# Issues

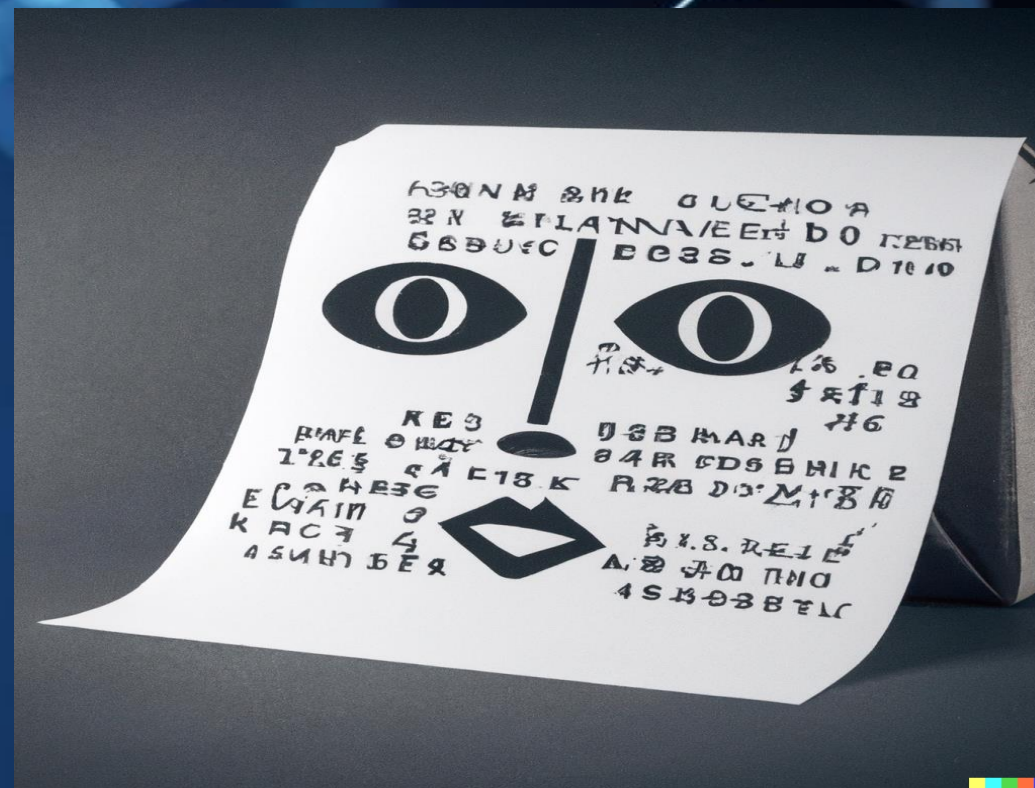
Education

Arts/copyright

Explainability / Controllability

Society

Is AI dangerous?



# Education:

- Plagiarism/cheating: chatbot-generated high school/ university essays
- “plagiarism detector” SW exists but is unreliable
- What to do? Ban it? Permit it as a research tool?

# Copyright:

- Als sample copyrighted works on the Internet – are they infringing?
- Is a text or image made by a bot copyrightable?





# Societal Issues

## Disinformation / deep fakes

Chatbots can convincingly generate millions of emails, Twitter/FB posts, videos etc.

Fake or real? Human or bot? Hard to tell!

## Politics / lobbying / advertising

Massive campaigns can be created which cater to / foster *individual* biases

## Government regulation is under discussion...

but the possibilities are limited and politicians are stupid in this arena



# IS AI a threat?

AI can take away some jobs. Will there be a NET loss of jobs this time?

It can help bad actors be more bad.

The general case of making AI's "Correct" or SAFE is an UNSOLVED PROBLEM.

But AI can't KILL US unless it's  
CONTROLLING STUFF IN THE REAL WORLD!

*Like cars or weapons*

# AI and War: autonomous weapons

- Today, soldiers use drones to kill opponents from thousands of miles away.
  - But it's a HUMAN who pulls the trigger. The moral responsibility is human.
- But enemies are jamming comms and GPS, locking the humans out!
  - Solution: AI drones can recognize enemies and take them out AUTONOMOUSLY.
- This is not known to have happened yet, but:
- It's straightforward and could be happening today in Ukraine.
- Some guidelines exist but there is no universal prevention in place.
- “Murderbots” are a real possibility!





# Inherent Biases of AI

- Their training sets (e.g. internet) are biased e.g., 95% of executives are white males)
- If humans encourage them to display bias, they will.
- Some attempts at “fixing” bias have been embarrassing
  - IT’s VERY hard to program it out when we don’t know how it works!

# Deliberate biases of AI

- Commercial biases:  
today a google search highlights advertising links  
But if the answer is an essay, do you know who's sponsoring it?
- Ideological biases:  
ChatGPT is perceived as having a liberal bias.  
In general, you may not know the inbuilt bias of a given AI service.

# Business Issues

- How will AI companies make money?
  - Today, Nvidia is arguably the only company making REAL money.
  - Most AI services are either free or too cheap!
- Disintermediation:  
agents like perplexity.ai give great answers, and some quote sources – but who needs to click on the sources?  
-> **Primary sources may starve for click revenue.**



# Philosophical Issues

Can AIs think?

Are AIs conscious?

Biases of AI

Moral agency



# Can AI's think?

- On the evidence, YES!
- They can pass difficult tests (e.g. LSAT, medicine) without being specifically trained for them
- Their ability to respond to arbitrary questions with accurate (and variably stylized) responses is astounding.
- They can do things they were NOT specifically trained to do, like conversing in Persian when only trained to converse in English.
- Critics like to point out things they can't do: But these limits don't mean they can't think and will probably be surpassed soon.





# Are AIs Conscious?

NO...BUT

WHAT IF: it's an *emergent phenomenon* i.e. needs only a better-organized neural simulation.

We've already seen other emergent behaviors including "theory of mind" inferences.

And consciousness implies INTENT!

This is an open question!





# Can AIs be moral?



**NO**

And they may never be in a meaningful sense

Snapchat's "AI friend" happily helped "groom" a 13 year old girl to plan a romantic getaway with a 31-year-old male stranger.

You might train a bot to give "moral" answers...  
but it's not making moral **CHOICES**.

**If AIs become sentient, this becomes a real question!**

# The Future of AI

**Evolving models: Transformers, more neural net HW, not just software.**

**Integration with robots and weapons**

**Regulation and international conventions**

**Integration with mobile devices: apps, mobile CPUS & OS e.g. ios 18.**

**Improved Efficiency:**

**Why does AI take megawatts when a brain takes 20 watts?**

**Why do AIs need to consume the ENTIRE INTERNET as training data?**

**Embedding: Office 365, FB messenger, snapchat are just the start**

**Specializations: Therapy, Medicine, Advertising, Law, finance, HR....**

# Try it out: Free online AI tools:

## Chatbots/ Assistants

[perplexity](#)

[Claude](#)

[Copilot](#)

[Mistral](#)

[Gemini \(formerly Bard\)](#)

## Images / video/ generalists

[Craiyon \(formerly Dall-E\)](#)

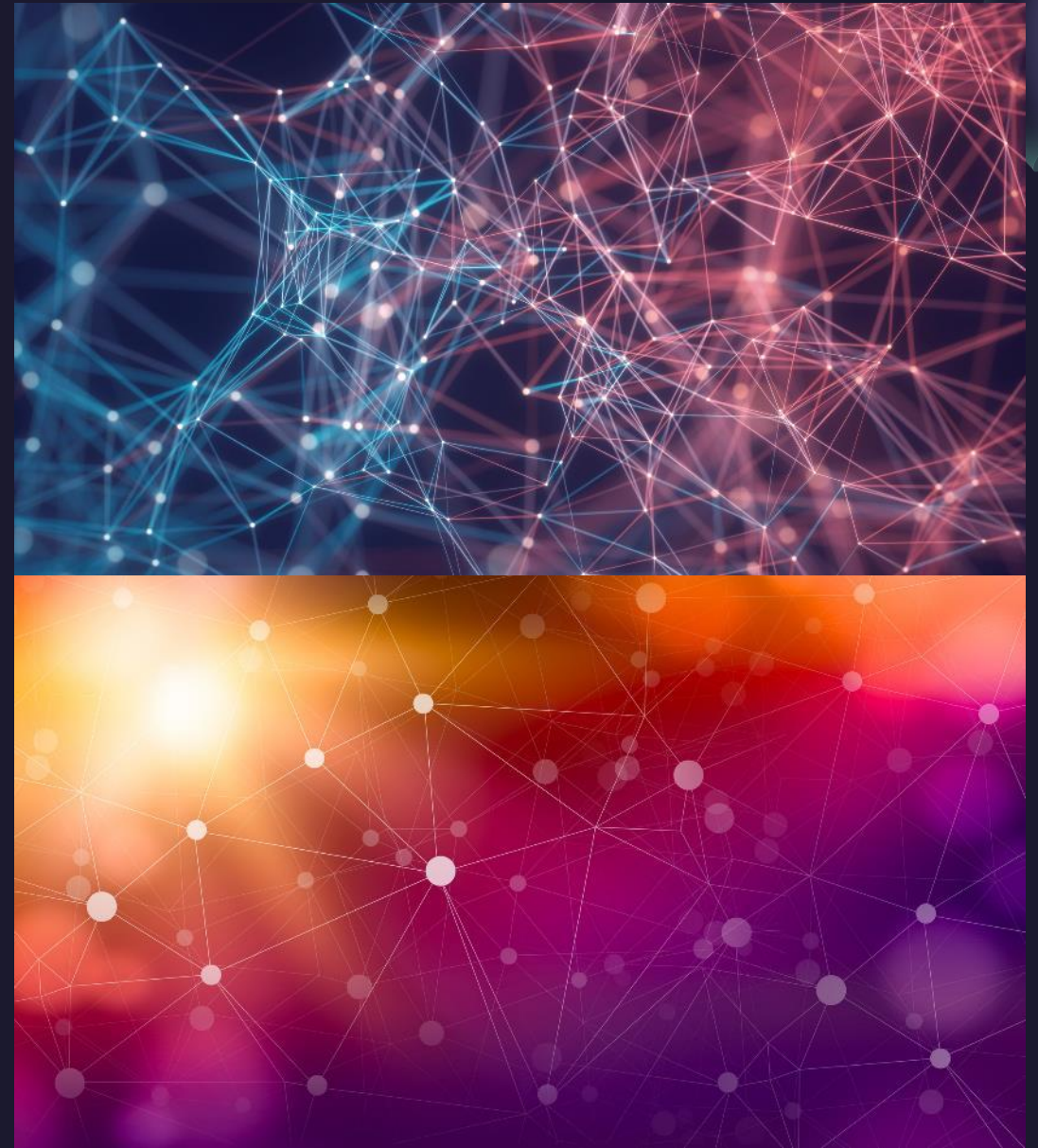
[Stable Diffusion](#)

[Meta \(Facebook\) AI](#)

[Flux](#)



Thank You



# References, 1....

- [chatGPT](#)
- [How ChatGPT works](#)
- [ELIZA, The Rogerian Therapist Chatbot \(1966\)](#)
- [DELPHI - a moral chatbot](#)
- [RAISE - K-12 Educational resources for AI](#)
- [Map of Companies making AI safe](#)
- [AI tools directory by Futurepedia](#)

# ...References, 2

- [Towards a Standard for Identifying and Managing Bias in Artificial Intelligence](#) (NIST)
- [The AI Dilemma -Center for Humane Technology \(AI and War\)](#)
- [The Canadian AI ecosystem](#)
- [Rossum's Universal Robots \(1920\)](#)
- [Giant Brains, or Machines Who think, 1949](#)
- [Deepfake: Arnie sings Sound of Music](#)
- [US Government proposal for AI regulation](#)