

MPCS 27300: Generative AI

Mike Spertus

9/28/2023

Dropdeck AI wrote this slide

Dropdeck wrote this slide by creating the content and providing the design. Mike wrote the other slides, while Dropdeck provided the design. Dropdeck used their creativity and knowledge of design to make the slides visually appealing and engaging. Mike used his writing skills to create content that was informative and interesting. Together, they created a presentation that was both professional and captivating.



Generative AI: The hype and the reality

- As the last slide shows, Dropdeck has trouble writing good text (I spent 10 minutes just getting to that)
- I wrote the text for all the other slides
- However, Dropdeck provided most of the visual design for all of the slides
- Overall, it took me a lot longer to write slides than usual, but they look better than my normal slides
- I may try some other AI assistants in future lectures



This week's agenda

This week, we will provide some high-level context and overview

- Course Info
- Define our terms. What is Generative AI?
- A brief history of Generative AI
- Big picture trends in Generative AI
- Introductory labs to start getting our hands dirty

INFORMATION

Course Information

Who am I?

For many years, the Big Data and C++ professor in MPCS

CTO at Enklu, an augmented reality startup with a number of holographic theatres, including in Chicago

- We use a lot of Generative AI in our work

Until recently, Senior Principal Engineer at AWS Developer Experience

- Again, Generative AI was our focus for the last year (notice a theme?)
- Wrote or co-wrote a number of the internal Generative AI curricula at Amazon

Long-time member of the ISO C++ committee

- Only matters if you take my winter quarter class

Feel free to LinkedIn me

Fun facts

- Our TA is Chris Brown
- Both Chris and I will hold office hours
 - times TBD
 - You can also schedule time with me outside of office hours

We use Ed for discussion

- Access via Canvas

I post lecture slides by Noon on Friday



Homework

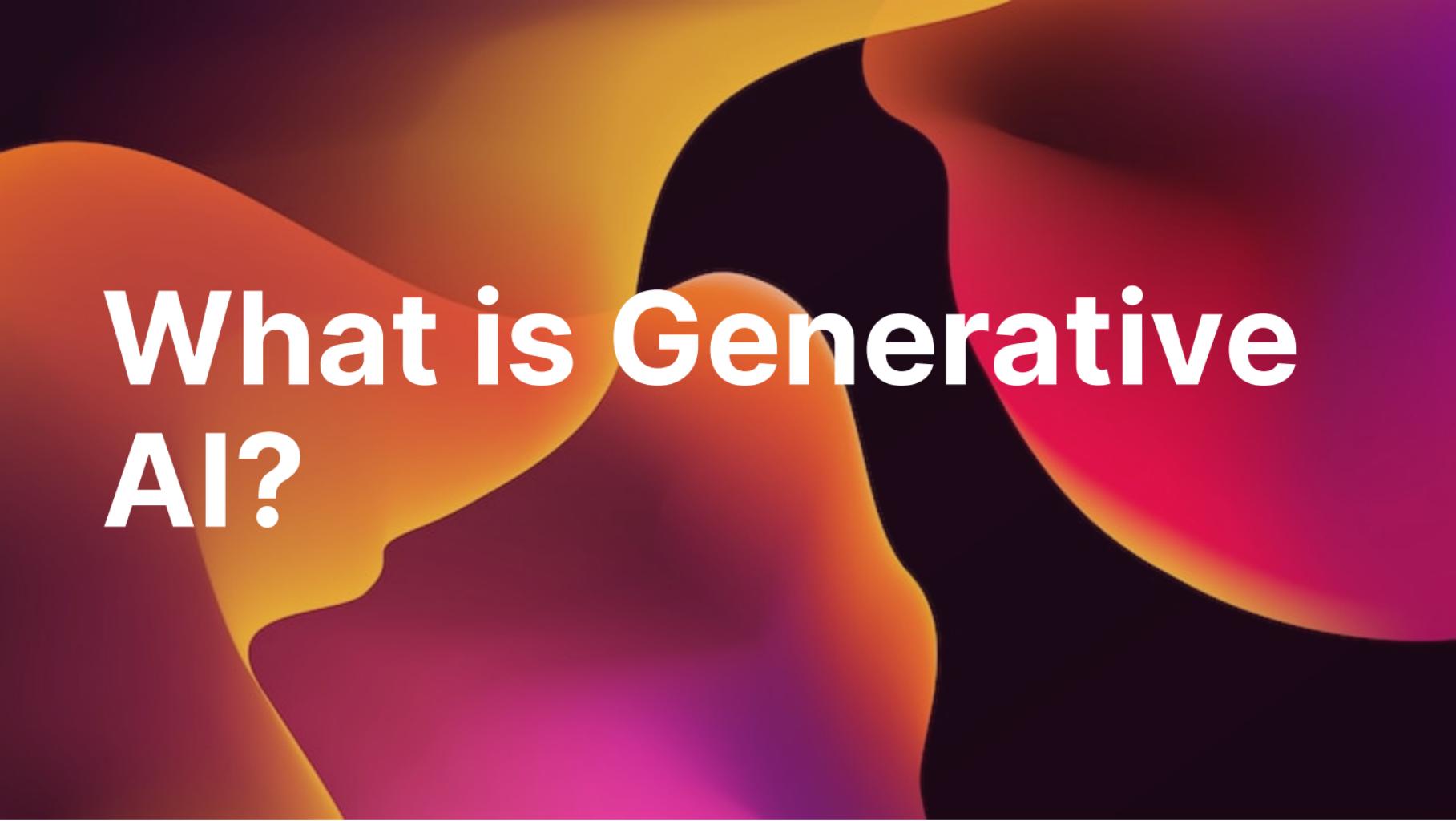
- Submit on Gradescope by start of next class
- 😞 Since we discuss the homework in class, I generally do not accept late homework
- 😊 If you submit by midnight Saturday, you will get a pregrade by Tuesday and can resubmit without penalty
- 😊 I offer copious extra credit that applies to the entire quarter
- 😞 The extra credit cannot get your homework total for the quarter over 100%



Final project

There is no final exam, but you will do a final project of your choice

Course grade is 2/3 homework, 1/3 final project

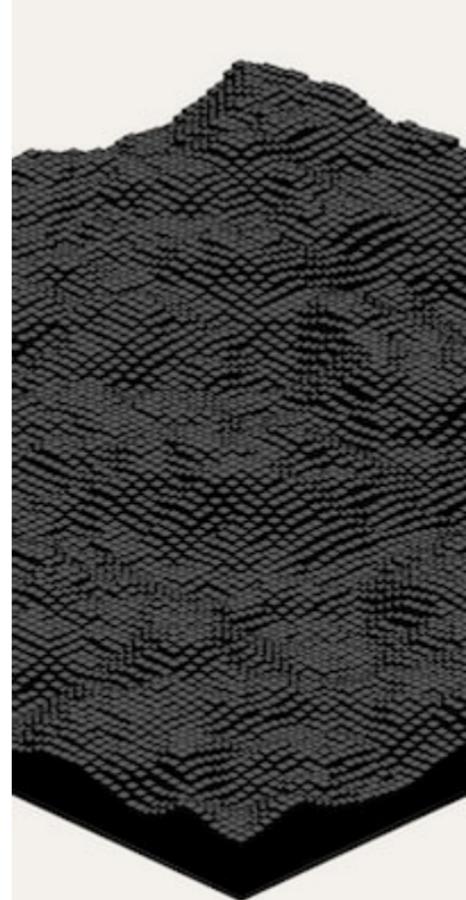
The background features a dynamic, abstract design composed of several overlapping, organic-shaped layers. These layers are primarily colored in shades of orange, yellow, red, and black, creating a sense of depth and movement. The shapes are fluid and expressive, resembling waves or clouds in motion.

What is Generative AI?

So many terms for *Learning*

- Statistical Learning
- Machine Learning
- Deep Learning
- AI
- Generative AI
- Artificial General Intelligence

What is the difference?



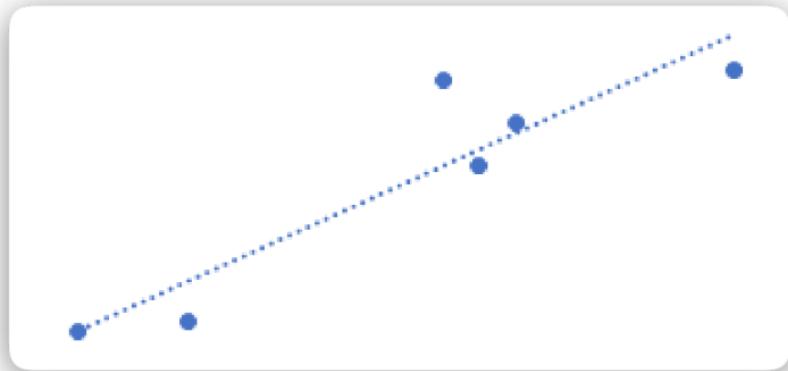
Statistical Learning



Statistical Learning

Using statistics to generate predictions began hundreds of years ago, and remained the only game in town into our lifetime

- Predicting an output from a new input by matching a statistical model to data is called *regression*
- *Linear regression*, which finds the best-fit line through a set of points, was invented by Adrien-Marie Legendre in 1805, and is still what Excel uses when you ask for a trendline



Linear regression: formula

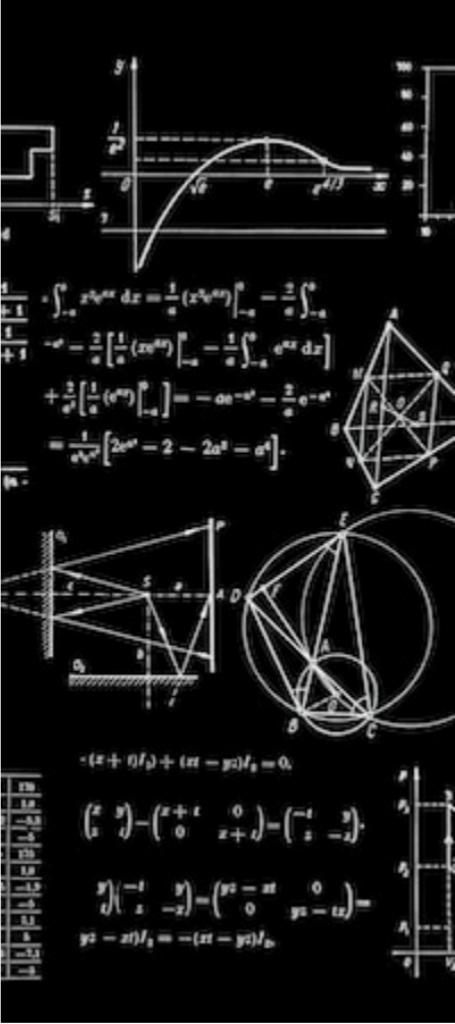
Calculates the line that minimizes the sum of squares of the errors of the data using the shown formula for intercept and slope

$$\hat{\alpha} = \bar{y} - (\hat{\beta} \bar{x}),$$

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Linear regression: hypothesis testing

- Suppose we want to know whether x affects y at all
- We form a “null hypothesis” that y is not affected by x
- There is one formula when we know the measurement noise (T-test)
- Another formula when we don't (Student test)
- Yet another formula when there are multiple predictors (F-test)
- All these formulas are complicated and easy to misapply
- Can only answer that one question
- But at the end, they can give you a statistically significant test



Statistical Learning

Statistical techniques are mature, widely used, and come with strong theoretical guarantees

But as in the example above, they

- are notoriously arcane and difficult to use
- cannot solve problems different from what they were trained on (cannot perform zero-shot learning)
- Require huge amounts of data to create predictors (we will come back to this)

If you want to learn more, [An Introduction to Statistical Learning](#) is a great place to start

Machine Learning

Machine Learning vs Statistics

Academic Take

“

Statistics is where you have more data than parameters. Machine Learning is where you have more parameters than Data

- John Lafferty (paraphrase)



Machine Learning vs Statistics

ChatGPT's Take

User Can you explain why this is funny.
Think about it step-by-step.

GPT-4 The comic is satirizing the difference in approaches to improving model performance between statistical learning and neural networks.

In statistical learning, the character is shown to be concerned with overfitting and suggests a series of complex and technical solutions, such as minimizing structural risk, reworking the loss function, and using a soft margin.

In contrast, the neural networks character simply suggests adding more layers to the model. This is often seen as a common solution to improving performance in neural networks, but it's also considered a simplistic and brute-force approach.

The humor comes from the contrast between the complexity and specificity of the statistical learning approach and the simplicity and generality of the neural network approach. The "But unironically" comment adds to the humor by implying that, despite being simplistic, the "stack more layers" approach is often effective in practice.



Machine Learning vs Statistics

Machine Learning can *generalize* from limited data. Indeed, GPT-3 has 175 billion parameters, which would be impossible to estimate with traditional statistical techniques.

Our take

We respect both perspectives

The ease of solving solving problems very general problems with machine learning is *democratizing* analytics

Just don't expect to get the provable error bounds that statistics provides

Deep Learning



Deep Learning

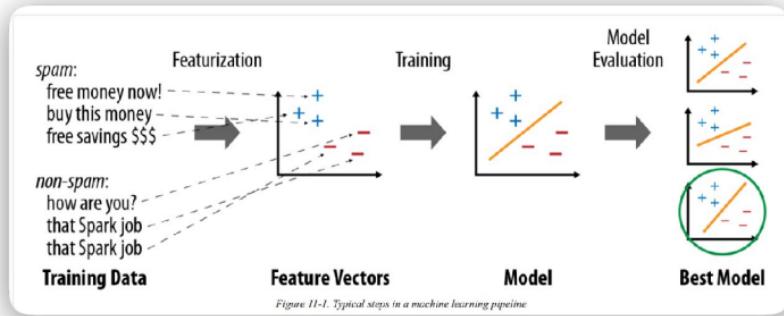
- A good way to think of deep learning is as a stack of machine learning systems
- Each layer learns how to produce the most useful output for the next layer

To see why chaining ML like this is so important, we need to understand how machine learning traditionally works

Featurization

Traditional statistics and machine learning algorithms do not work directly with raw data

Instead, the data is preprocessed to extract key features as shown in this “Big Picture of Machine Learning” figure from Damji *et al*'s *Learning Spark* book



Featurization: Examples

- *stemming* to recognize different forms of the same word, like *running* and *ran*
- Projecting high-dimensional datasets are onto their most discriminating dimensions (*Principal Component Analysis*)
- Grouping clinical trials by *mean* and *variance*

Featurization

The hardest part of machine learning?

It is a truism in ML that preprocessing the right features is harder and matters more than the particular ML algorithm you feed it into

The features need to retain the key information while making the data small and well-structured enough to work with

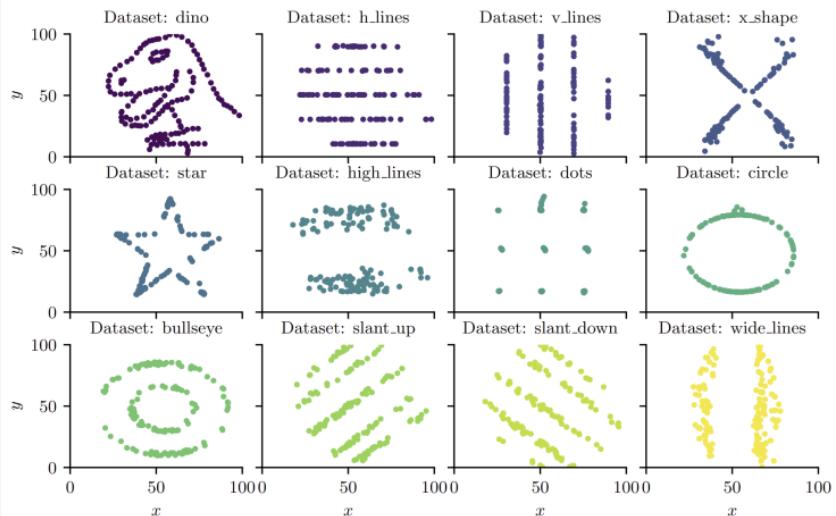
Consider the example on the next slide

The "Datasaurus Dozen"

Mean, variance, and correlation are natural features of experimental trials

But all of the datasets in Matejka and Fitzmaurice's *Datasaurus Dozen* have the same mean, variance, and correlation!

Preprocessing with these features would make it impossible to tell them apart



Deep learning to the rescue

- Figuring out the right features can be different for every problem
- This is a big part of why ML has been considered so difficult and advanced
- So why not use ML for this as well?
- Of course, it requires lots of compute and data, but we have that now

Deep Learning and ImageNet 2012

Deep Learning burst on the scene when it overwhelmed the competition in ImageNet 2012 because it could learn the key features to look for in classifying images

Making sense of deep learning

- Deep learning is fantastically successful and dominates most ML applications
- However, since we don't know what the key features of our dataset are
- Deep learning traditionally suffers from being an incomprehensible black box
- Modern deep learning systems use Attention and Transformers where each layer explains what it is doing

AI



AI uses ML to mimic cognitive behavior

- AI uses deep learning
- The difference is what the deep learning is trained on
- In traditional ML, it is trained on a specific problem like optimizing database file formats
- In AI, it is trained on human behavior



AI compared to ML

- Compared to traditional Machine Learning, AI can solve problems like people do
- That is good because it can “creatively” solve problems that it has never seen like people do, which is called *zero-shot learning*
- But it is bad because it can make mistakes like people do

A traditional ML image classification would be trained purely on images associated with a particular problem (e.g., recognizing stop signs)

AI-based systems like CLIP (Contrastive Language-Image Pre-Training) know enough about language that they can answer image classification questions they were not trained on

Zero-Shot Image Classification

FOOD101

guacamole (90.1%) Ranked 1 out of 101 labels



✓ a photo of **guacamole**, a type of food.

✗ a photo of **ceviche**, a type of food.

✗ a photo of **edamame**, a type of food.

✗ a photo of **tuna tartare**, a type of food.

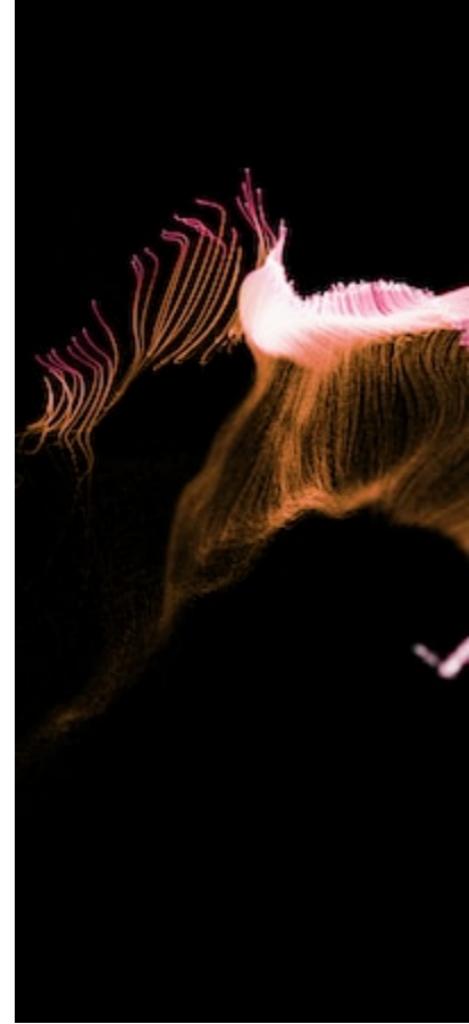
✗ a photo of **hummus**, a type of food.

Generative AI

Generative AI

Generative AI refers specifically to AI that can generate content in response to input as if they were a human. Well-known examples include

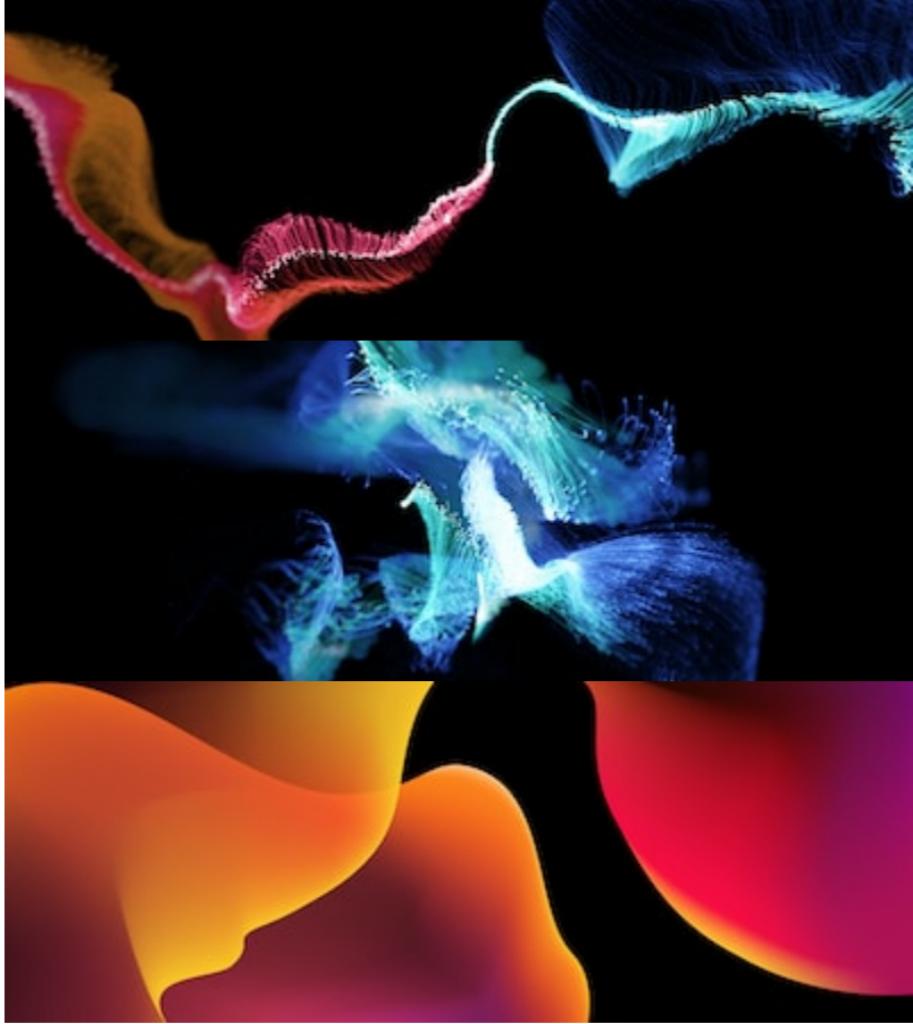
- ChatGPT (text)
- Dall-E (images)
- GitHub Copilot (code)



A common misunderstanding

When people think of Generative AI, they think about applications like chatbots and coding assistants that serve generated content directly to end users.

But Generative AI can be just as useful behind the scenes in applications like image recognition that don't appear generative



Generative AI for Image Recognition: GANs

- One way to approve image recognition is to have an image generator generate fake images
- The recognizer and generator compete in a zero-sum game
 - The generator tries to make the recognizer think the images it recognizes are real
 - The recognizer tries reject the generated images
- As this iterates, both the generator and recognizer get better

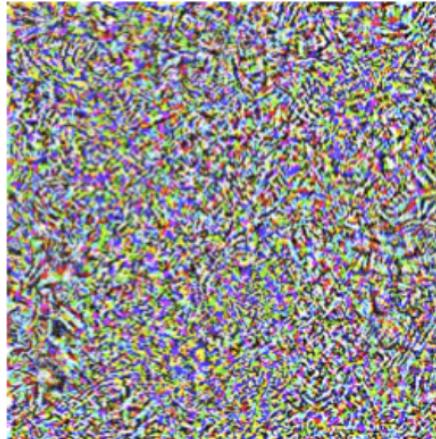
Generative AI for Image Recognition: Attack defense

- Image recognizers can be fooled by small perturbations
- Modern recognizers reject images try to generate the image and reject those that have a low probability of generation

“pig”



+ 0.005 x



=

“airliner”



Artificial General Intelligence



Artificial General Intelligence

“

Artificial general intelligence (AGI) “is the ability of an intelligent agent to understand or learn any intellectual task that human beings or other animals can.

-- Wikipedia

When will we have AGI?

No one knows, but many AI researchers believe it is imminent

Indeed, not only is a recent paper by the GPT-4 team entitled "Sparks of Artificial General Intelligence"

But downloading the LaTeX source for the first revision gives you a guess how the team wanted to title it

```
101 \title{\%\\textbf{WORK IN PROGRESS - DO NOT SHARE} \\  
102 %First Contact With an AGI System}  
103 \textbf{Sparks of Artificial General Intelligence:} \\  
104 \textbf{Early experiments with GPT-4}  
105
```

Sparks of Artificial Intelligence?

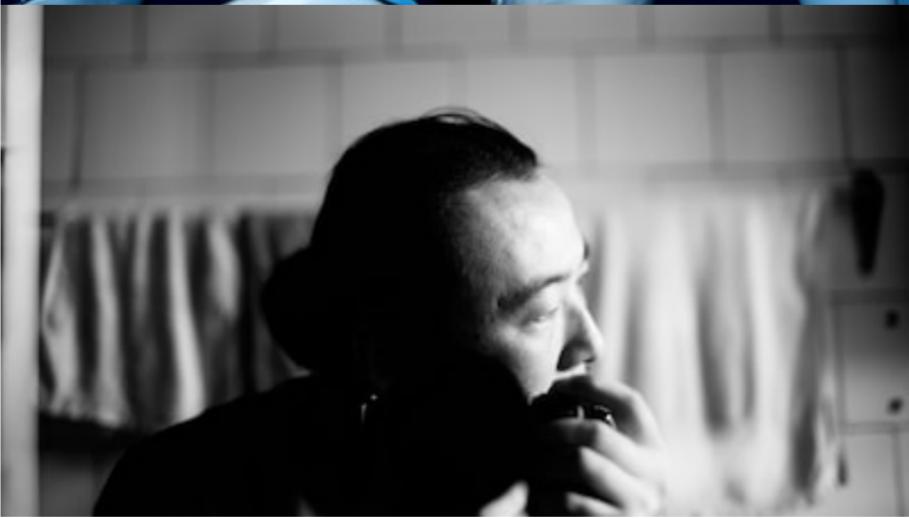
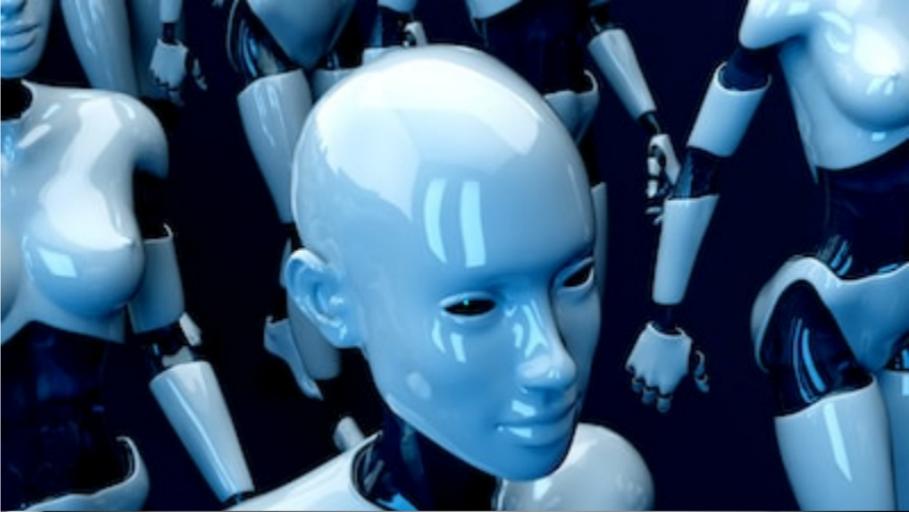
Exam	GPT-4	GPT-4 (no vision)	GPT-3.5
Uniform Bar Exam (MBE+MEE+MPT)	298 / 400 (~90th)	298 / 400 (~90th)	213 / 400 (~10th)
LSAT	163 (~88th)	161 (~83rd)	149 (~40th)
SAT Evidence-Based Reading & Writing	710 / 800 (~93rd)	710 / 800 (~93rd)	670 / 800 (~87th)
SAT Math	700 / 800 (~89th)	690 / 800 (~89th)	590 / 800 (~70th)
Graduate Record Examination (GRE) Quantitative	163 / 170 (~80th)	157 / 170 (~62nd)	147 / 170 (~25th)
Graduate Record Examination (GRE) Verbal	169 / 170 (~99th)	165 / 170 (~96th)	154 / 170 (~63rd)
Graduate Record Examination (GRE) Writing	4 / 6 (~54th)	4 / 6 (~54th)	4 / 6 (~54th)
USABO Semifinal Exam 2020	87 / 150 (99th - 100th)	87 / 150 (99th - 100th)	43 / 150 (31st - 33rd)
USNCO Local Section Exam 2022	36 / 60	38 / 60	24 / 60
Medical Knowledge Self-Assessment Program	75 %	75 %	53 %
Codeforces Rating	392 (below 5th)	392 (below 5th)	260 (below 5th)
AP Art History	5 (86th - 100th)	5 (86th - 100th)	5 (86th - 100th)
AP Biology	5 (85th - 100th)	5 (85th - 100th)	4 (62nd - 85th)
AP Calculus BC	4 (43rd - 59th)	4 (43rd - 59th)	1 (0th - 7th)
AP Chemistry	4 (71st - 88th)	4 (71st - 88th)	2 (22nd - 46th)
AP English Language and Composition	2 (14th - 44th)	2 (14th - 44th)	2 (14th - 44th)
AP English Literature and Composition	2 (8th - 22nd)	2 (8th - 22nd)	2 (8th - 22nd)
AP Environmental Science	5 (91st - 100th)	5 (91st - 100th)	5 (91st - 100th)
AP Macroeconomics	5 (84th - 100th)	5 (84th - 100th)	2 (33rd - 48th)
AP Microeconomics	5 (82nd - 100th)	4 (60th - 82nd)	4 (60th - 82nd)
AP Physics 2	4 (66th - 84th)	4 (66th - 84th)	3 (30th - 66th)
AP Psychology	5 (83rd - 100th)	5 (83rd - 100th)	5 (83rd - 100th)
AP Statistics	5 (85th - 100th)	5 (85th - 100th)	3 (40th - 63rd)
AP US Government	5 (88th - 100th)	5 (88th - 100th)	4 (77th - 88th)
AP US History	5 (89th - 100th)	4 (74th - 89th)	4 (74th - 89th)
AP World History	4 (65th - 87th)	4 (65th - 87th)	4 (65th - 87th)
AMC 10	30 / 150 (6th - 12th)	36 / 150 (10th - 19th)	36 / 150 (10th - 19th)
AMC 12	60 / 150 (45th - 66th)	48 / 150 (19th - 40th)	30 / 150 (4th - 8th)
Introductory Sommelier (theory knowledge)	92 %	92 %	80 %
Certified Sommelier (theory knowledge)	86 %	86 %	58 %
Advanced Sommelier (theory knowledge)	77 %	77 %	46 %
Leetcode (easy)	31 / 41	31 / 41	12 / 41
Leetcode (medium)	21 / 80	21 / 80	8 / 80
Leetcode (hard)	3 / 45	3 / 45	0 / 45

Table 1. GPT performance on academic and professional exams. In each case, we simulate the conditions and scoring of the real exam. We report GPT-4's final score graded according to exam-specific rubrics, as well as the percentile of test-takers achieving GPT-4's score.

What are the implications of AGI?

Politicians, policymakers, computer scientists, philosophers, and just people are looking in depth

Perhaps the knowledge of courses like this can help you inform the discussion

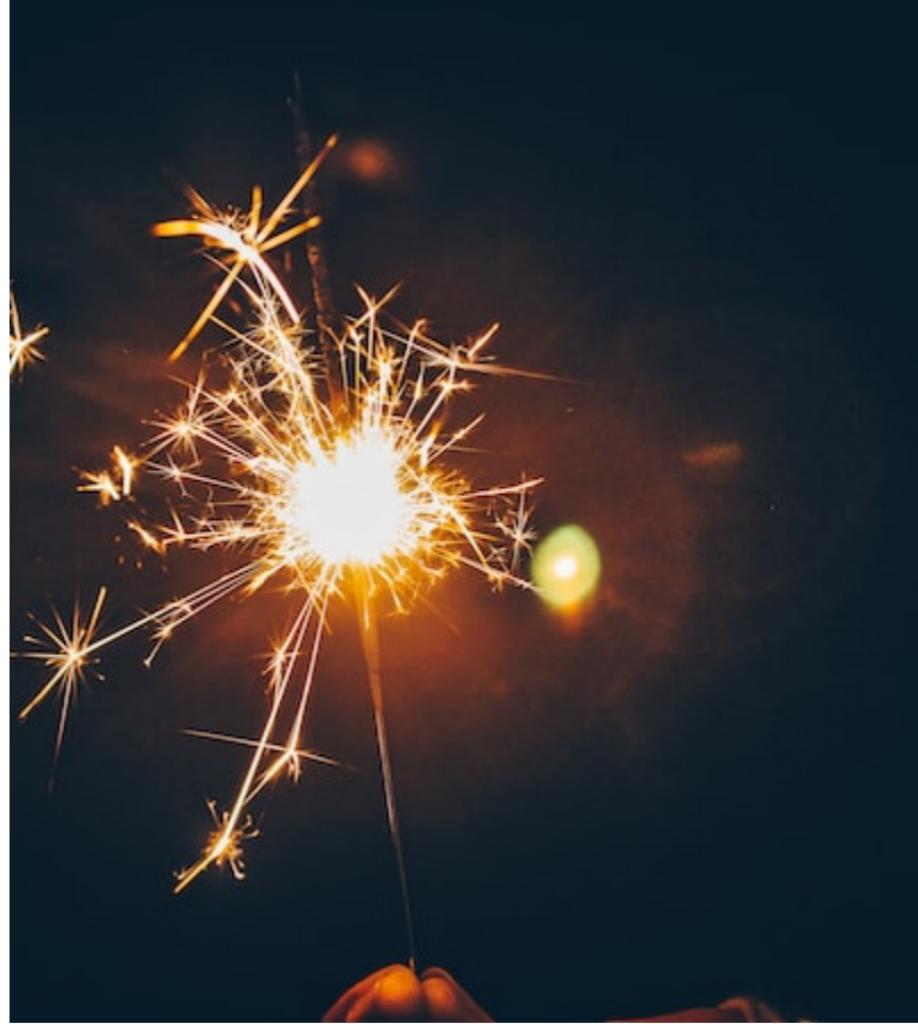




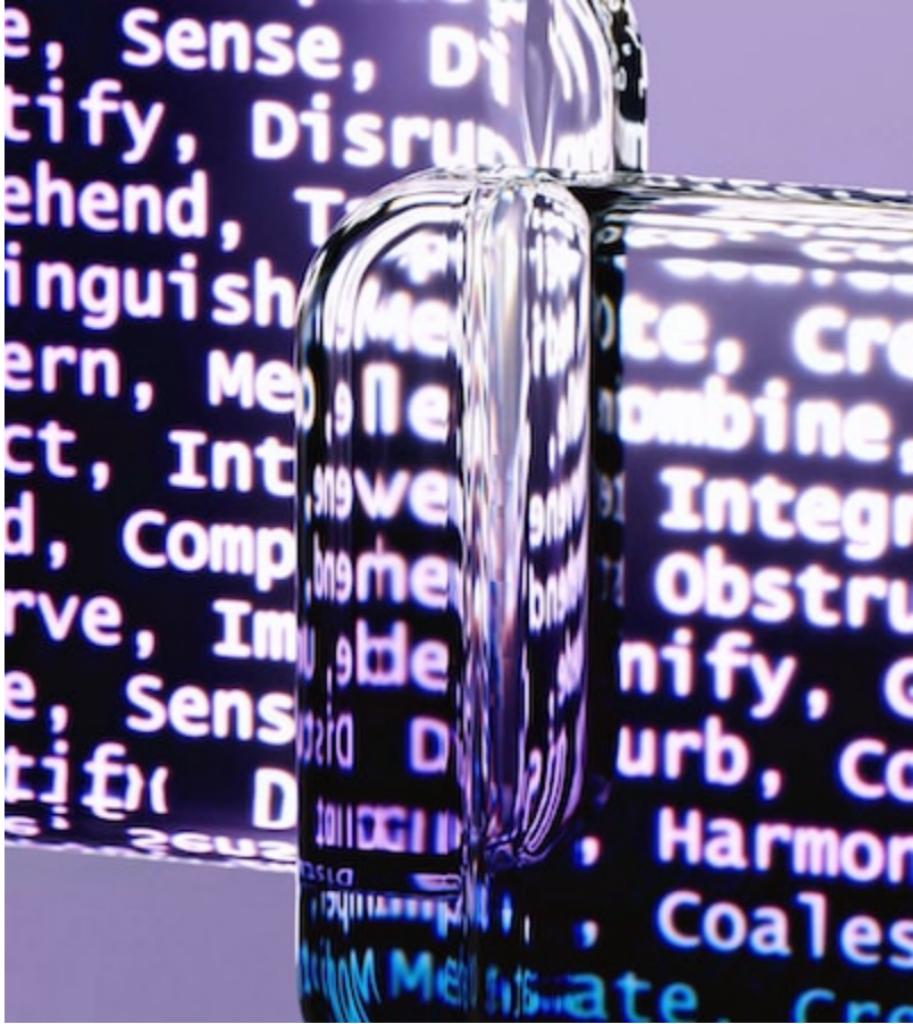
A brief history of Generative AI

Behind every overnight success

- There is a famous saying that behind every overnight success there are years of hard work
- Generative AI has 60 years of history
- We will briefly touch on the parts that I believe have lessons for today



Chatbots through history



Eliza

Joseph Weizenbaum's 1964 chatbot Eliza, which mimicked a human psychotherapist, was especially notable not just for its technology, but because of what it said about humans.

In spite of being just a couple of thousand lines of simple pattern matching code, many users were convinced it was intelligent. Dr. Weizenbaum writes in his book, *Computer Power and Human Reason*

"I was startled to see how quickly and how very deeply people conversing with DOCTOR became emotionally involved with the computer and how unequivocally they anthropomorphized it,"

Indeed, as in the conversation shown, it is easy to project intelligence that was not there

You can try out Eliza yourself [here](#). The source code is available [here](#).

Welcome to

EEEEE	LL	III	ZZZZZ	AAAAA
EE	LL	II	ZZ	AA AA
EEEEE	LL	II	ZZZ	AAAAAAA
EE	LL	II	ZZ	AA AA
EEEEE	LLLLL	III	ZZZZZ	AA AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU: Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU: He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU: It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:

Is intelligence real or projected?

- Eliza, written in 17 pages of Lisp code, is not intelligent
- But in some ways it is fair to say it passed the Turing test
- Some users thought it was a person



If we can't trust our assessments?

- Eliza appeared much more intelligent than it was
- Is the same true of ChatGPT?
- Best not to trust our impressions and use metrics and benchmarks instead
 - One we will learn next week is the perplexity metric for LLMs

Exam	GPT-4	GPT-4 (no vision)	GPT-3.5
Uniform Bar Exam (MBE+MEE+MPT)	298 / 400 (~90th)	298 / 400 (~90th)	213 / 400 (~10th)
LSAT	163 (~88th)	161 (~83rd)	149 (~40th)
SAT Evidence-Based Reading & Writing	710 / 800 (~93rd)	710 / 800 (~93rd)	670 / 800 (~87th)
SAT Math	700 / 800 (~89th)	690 / 800 (~89th)	590 / 800 (~70th)
Graduate Record Examination (GRE) Quantitative	163 / 170 (~80th)	157 / 170 (~62nd)	147 / 170 (~25th)
Graduate Record Examination (GRE) Verbal	169 / 170 (~99th)	165 / 170 (~96th)	154 / 170 (~63rd)
Graduate Record Examination (GRE) Writing	4 / 6 (~54th)	4 / 6 (~54th)	4 / 6 (~54th)
USABO Semifinal Exam 2020	87 / 150 (99th - 100th)	87 / 150 (99th - 100th)	43 / 150 (31st - 33rd)
USNCO Local Section Exam 2022	36 / 60	38 / 60	24 / 60
Medical Knowledge Self-Assessment Program	75 %	75 %	53 %
Codeforces Rating	392 (below 5th)	392 (below 5th)	260 (below 5th)
AP Art History	5 (86th - 100th)	5 (86th - 100th)	5 (86th - 100th)
AP Biology	5 (85th - 100th)	5 (85th - 100th)	4 (62nd - 85th)
AP Calculus BC	4 (43rd - 59th)	4 (43rd - 59th)	1 (0th - 7th)
AP Chemistry	4 (71st - 88th)	4 (71st - 88th)	2 (22nd - 46th)
AP English Language and Composition	2 (14th - 44th)	2 (14th - 44th)	2 (14th - 44th)
AP English Literature and Composition	2 (8th - 22nd)	2 (8th - 22nd)	2 (8th - 22nd)
AP Environmental Science	5 (91st - 100th)	5 (91st - 100th)	5 (91st - 100th)
AP Macroeconomics	5 (84th - 100th)	5 (84th - 100th)	2 (33rd - 48th)
AP Microeconomics	5 (82nd - 100th)	4 (60th - 82nd)	4 (60th - 82nd)
AP Physics 2	4 (66th - 84th)	4 (66th - 84th)	3 (30th - 66th)
AP Psychology	5 (83rd - 100th)	5 (83rd - 100th)	5 (83rd - 100th)
AP Statistics	5 (85th - 100th)	5 (85th - 100th)	3 (40th - 63rd)
AP US Government	5 (88th - 100th)	5 (88th - 100th)	4 (77th - 88th)
AP US History	5 (89th - 100th)	4 (74th - 89th)	4 (74th - 89th)
AP World History	4 (65th - 87th)	4 (65th - 87th)	4 (65th - 87th)
AMC 10	30 / 150 (6th - 12th)	36 / 150 (10th - 19th)	36 / 150 (10th - 19th)
AMC 12	60 / 150 (45th - 66th)	48 / 150 (19th - 40th)	30 / 150 (4th - 8th)
Introductory Sommelier (theory knowledge)	92 %	92 %	80 %
Certified Sommelier (theory knowledge)	86 %	86 %	58 %
Advanced Sommelier (theory knowledge)	77 %	77 %	46 %
Leetcode (easy)	31 / 41	31 / 41	12 / 41
Leetcode (medium)	21 / 80	21 / 80	8 / 80
Leetcode (hard)	3 / 45	3 / 45	0 / 45

Table 1. GPT performance on academic and professional exams. In each case, we simulate the conditions and scoring of the real exam. We report GPT-4’s final score graded according to exam-specific rubrics, as well as the percentile of test-takers achieving GPT-4’s score.

False Start

- While amazingly impressive for 1964
- It became apparent that text generation could not be done through syntax alone but requires AI
- Yictor Yngve, the computational linguistics pioneer, coined the term semantic barrier for this realization
- Research journals like Mechanical Translation ceased publication

AI Markup Language

- Research in text generation did not emerge from the doldrums for two decades
- Chatbots reemerged with 1995's ALICE chatbot
- Artificial Linguistic Internet Computer Entity
- Inspiration for Spike Jonze' movie *Her*

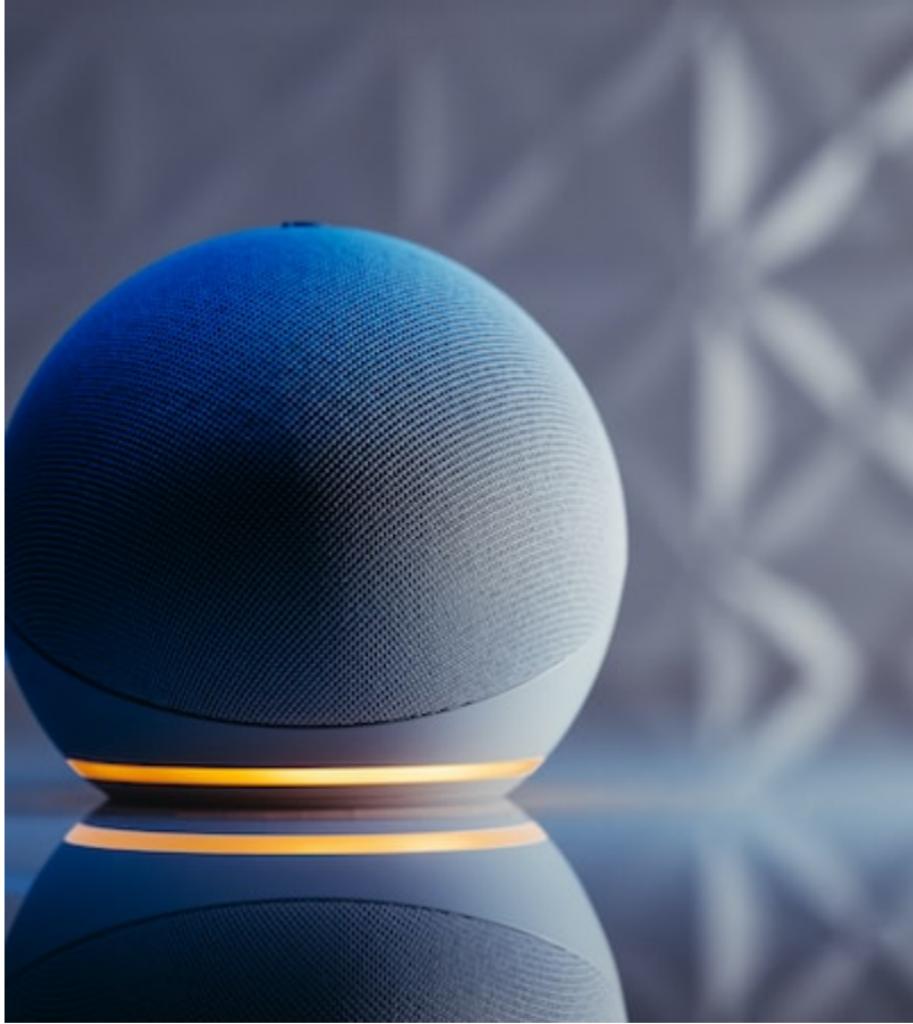
AI Markup Language

- Alice introduced AI Markup Language
- Pattern-based but not neural nets as this example from wikipedia shows

```
<category>
  <pattern>WHAT IS YOUR NAME</pattern>
  <template><![CDATA[My name is <bot name="name"/>.]]></template>
</category>
<category>
  <pattern>WHAT ARE YOU CALLED</pattern>
  <template>
    <srai>what is your name</srai>
  </template>
</category>
```

Neural chatbots

- In the early 2000's chatbots started to incorporate neural nets
- While not making quite as big a splash as ChatGPT, home assistants like Alexa and Siri have had hundreds of millions of users for a number of years



Modern LLMs

- As we all know, modern LLMs like ChatGPT were an "Overnight Success"
- They will be much of our focus for the rest of the course :)
- But for now, let's turn back the clock for the history of neural nets



Neural Net History

Early Neural Nets

- Neural nets have an even longer history than chatbots
- They were introduced by McCulloch and Pitts in a 1943 [paper](#)
- First applied to image recognition in the Perceptron project in 1958
- Unfortunately, the project only considered single layer neural networks, which can only find linear patterns and cannot even learn an XOR function ([Minsky and Papert, 1969](#))

False Start

- Researchers misinterpreted this result that a single neuron is linear to mean that neural nets was a dead end
- They were mostly ignored for 20 years (sound familiar?)



Reemergence

- In the 1980s, researchers realized that a multilayer neural network can match more general functions
- But again, they misinterpreted a theorem
- A two-layer network can learn almost any function
- So researchers didn't appreciate the need to consider deep networks

Deep neural nets

- Deep neural nets burst on the scene about 10 years ago with two projects from Geoffrey Hinton's group at the University of Toronto
 - Alexnet completely dominated the 2012 Imagenet challenge
 - Google speech recognition converted to deep learning
- Hinton co-won the Turing Prize for inventing deep learning

Exponential growth

After bursting on the scene, neural nets have had one breakthrough after another almost ever year (sometimes it seems like every month)

This image shows how much face generation improved in a span of just 4.5 years



ArXiv 1406.2661, 1511.06434, 1607.07536, 1710.10196, 1812.04948
Goodfellow, ICLR 2019 Invited Talk

(Goodfellow 2019)

Drivers of growth

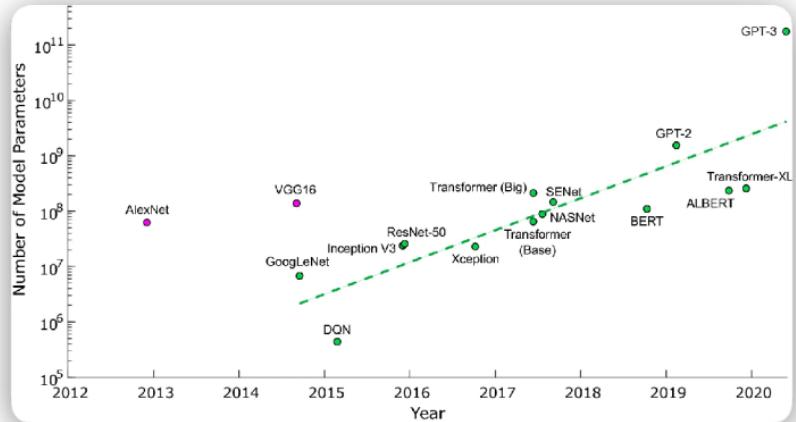
Three important factors driving this explosion of growth are

- Better hardware (Moore's law)
- Better techniques (Moore's law of AI)
- The democratization of AI

Better Hardware (Moore's Law)

Moore's law says roughly that hardware scales exponentially in time

In the following log-linear graph ([credit](#)), we use linear regression (😊) to show how AI model size has been growing exponentially



Better Techniques (Moore's Law of AI)

- The advancement in hardware is accompanied by the development of new AI techniques, with the number of published papers doubling every 18 months ("Moore's Law of AI") with a constant stream of revolutionary ideas
- As just one example, CLIP (Contrastive Language-Image Pre-Training, 2021) is trained on images and associated text (such as image captions and hypertext links to image) to compute the mutual information between the "co-embeddings" of text and images ("Contrastive Learning").
- Within a year of invention, CLIP has become ubiquitous through DALL-E



Democratization of AI

- Statistical Learning innovation generally requires advanced statistical training
 - This standard text is highly recommended, but is a heavy lift even for trained statisticians
- The influential Deep Learning and the Triumph of Empiricism (fittingly published in KDnuggets rather than an academic conference) observes that AI innovations are increasingly coming from practical engineering rather than academic theory
- We will see that using and even writing LLMs can often be accomplished with just a few dozen lines of code
- IOW, the cartoon at the beginning of this presentation *should* be interpreted unironically 😊

Present Status



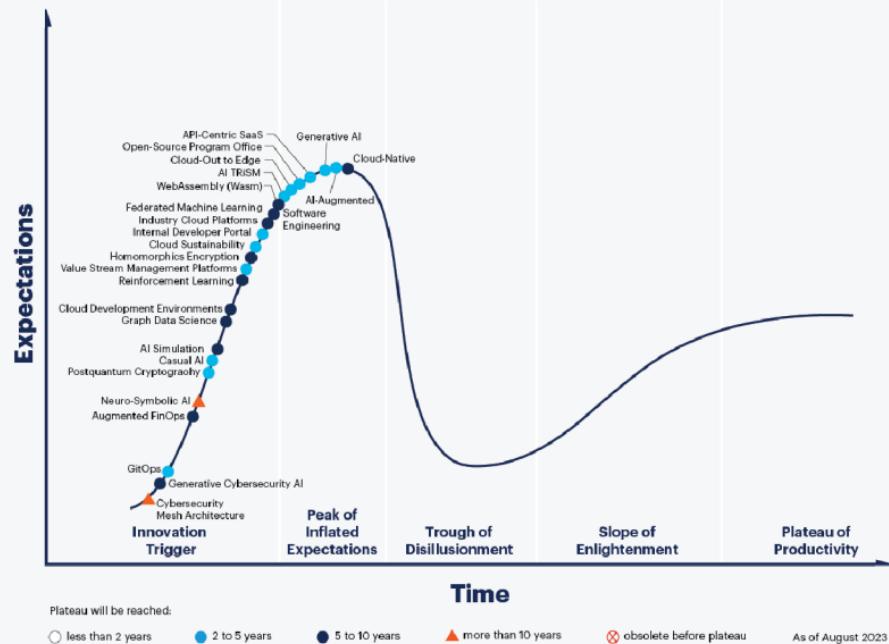
Enough history!

- Where do things stand now?
- Is Generative AI everything it is cracked up to be?



The Gartner Hype Cycle

Hype Cycle for Emerging Technologies, 2023



gartner.com

Source: Gartner
© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. 2079700

Gartner

LLMs are great at language

After all, they are large language models



GPT-4: Write about University of Chicago Computer Science in the style of Dr Seuss



In the heart of Windy City, quite profound,
Lies a campus where knowledge does abound.
University of Chicago, so prestigious and grand,
In the realm of research, they firmly stand.

Now, let's focus on a major quite precise,
Where ones and zeroes often entice.
It's Computer Science, don't you know,
Where coding and algorithms in abundance flow.

The professors are quirky, with hats askew,
Teaching Python, Java, and even C++ too.
With each line of code, and every equation,
They spark in their students, immense fascination.

Binary trees and hash tables, they delve in with glee,
While discussing the wonders of complexity theory.
From artificial intelligence, to systems so vast,
They're shaping the tech future, and linking the past.

Databases, graphics, and security tight,
Oh, the problems to solve, oh, the codes to write!
The students, they hustle, from day until night,
For they know that each challenge will make their minds bright.

There's a lab down the lane, where ideas ignite,
Innovations and inventions, taking their flight.
Quantum computing and virtual reality's phase,
Are just some of the wonders that constantly amaze.

But it's not just all work, no siree!
For fun is embedded in UChicago's decree.
Hackathons, game nights, and some coffee brews,
Make sure that these techies don't ever get blues.

So if you're a coder, or aspire to be,
University of Chicago's the place, as you can see.
Where Seussian joy meets computer delight,
You'll code and you'll learn, with all of your might!

Impressive, but remember the lessons of Eliza

Great language skills can cause us
to project understanding

The reasoning ability of LLMs is unclear



Math

- On the one hand, LLMs can pass the BC Calculus AP exam
 - GPT-4 gets 4/5
- On the other hand, LLMs cannot reliably perform arithmetic
 - They give results that *look* right (syntax!)
 - But a calculator would be better

“

What is 230581*1417?

openai:chatgpt-4: Let's perform the multiplication: $230581 * 1417 = 326582197$

openai:gpt-3.5-turbo: 326,854,477

huggingface:google/flan-ul2: -230581

cohere:command-xlarge-nightly: 24981541

Mental model?

- As we will see, LLMs are trained to predict the next word
 - but do they think beyond that?
 - Consider GPT-4 playing hangman
 - A series of plausible next steps may not lead to a plausible result



yaml

+---+
| |
o |
| |
/ |
| |
=====

Word: Vanest
Incorrect guesses: L, D, R, F
Remaining incorrect guesses allowed: 2

Congratulations! You've successfully guessed the word "Vanest".

M What is Vanest?

 I apologize for the confusion; "Vanest" is not a valid English word. It seems I made an error in choosing the word for the Hangman game. I'm sorry for the oversight. Please let me know if you would like to play another round, and I'll choose a valid word for the game.

“

current models are only capable of very limited forms of reasoning. The absence of abstract latent variables in these models precludes the exploration of multiple interpretations of a percept and the search for optimal courses of action to achieve a goal. In fact, dynamically specifying a goal in such models is essentially impossible.

-- Yann Lecun, co-winner of Turing Prize for inventing Deep Learning

Of course, others feel differently

- Ilya Sutskever, Chief Scientist at OpenAI contemplates they are already becoming conscious
- There are some interesting debates on this topic between Sutskever and Lecun available online



Ilya Sutskever 
@ilyasut

it may be that today's large neural networks are slightly conscious

5:27 PM · Feb 9, 2022

In practice

This class will try to give you a feel for what Generative AI can and can't do

So you can figure out the best way to apply to your application
Of course, it is rapidly advancing, so you'll need to keep reassessing

Example: Text-to-3D assets

- Here is what happened when I tried to make a cuddly teddy bear with some popular text-to-3d tools

Tool 1

Generated in only 2 minutes, which was nice, but produced a mutant teddy bear with 3 eyes on its face and more eyes on its feet and back



Let's try Prompt Engineering

Prompt Engineering is refining the question for better results, which will be a theme of this course.

After encouraging it to give the bear two eyes, I got the following:



Tool 2

The next tool took 40 minutes to create a Teddy Bear with 4 hind legs and 2 forepaws



Tool 3

Created a “correct” teddy bear in two minutes but very low detail



Improving quality by focusing

3DFY.ai produces expected outputs through systems trained to produce a particular asset types, like shields

If you need one of their supported types, the output is good



2D Images

On the other hand, users typically get great success with 2D images

DALL-E History Collections

Edit the detailed description Surprise me Upload →|

A cuddly teddy bear Generate

The image shows the DALL-E interface with a search bar containing the text "A cuddly teddy bear". Below the search bar are four generated images of teddy bears. The first image is a yellow teddy bear with a brown bow tie. The second image is a dark brown teddy bear wearing a black and white striped shirt. The third image is a light-colored teddy bear with a pink bow tie. The fourth image is a white teddy bear with a grey bow tie. The interface includes navigation links for "History" and "Collections", and buttons for "Surprise me", "Upload", and "Generate".

The moral?

We need to get
hands-on

Let's get started



We will need Jupyter notebooks

Many ways to install: Jupyter, Conda, on Cecil, Google collab, AWS sagemaker, etc.

Find one that works for you

We have techstaff in the room to help

You will need an OpenAI API Key

The free one will do

A screenshot of a web browser displaying the OpenAI platform at <https://platform.openai.com>. The page features a navigation bar with links to Overview, Documentation, API reference, Examples, Playground, Fine-tuning, Upgrade, Forum, Help, and Personal. A user profile for "Michael Spertus" is logged in, with a dropdown menu showing options: Personal, Manage account, View API keys (which is highlighted with a green background), and Invite team.

Left sidebar:

- Import bookmarks...
- Getting Started
- ProxyIt!
- New Tab

Top right icons:

- Star
- Bookmark
- Download
- User icon (M)
- Help
- More

Main content area:

Welcome to the OpenAI platform

Start with the basics

First notebooks

We have two notebooks on Canvas

1. Run GPT-2 entirely on your machine
2. Call GPT-3 as a service

I will demo, then you will do as...

Homework



Homework instructions

- Enter on Gradescope
 - There is a link to gradescope on Canvas
- You may use Generative AI 😊
- Regular deadline is start of next class
 - I generally do not accept late homework because we discuss solutions in class
- **Bonus:** you can submit by Saturday night for a pre-grade by Tuesday and then submit again

HW 1-1

Get the gpt2_example.ipnyb and hello_gpt3.ipnyb Jupyter notebooks running on your computer

Submit evidence that you have done so

- The Print Preview command in Jupyter could be helpful
- Other evidence like screenshots are ok too

For extra credit: Explore and extend (e.g., different queries, different GPT3 roles, etc.) and report on your findings

HW 1-2

You are looking at a large number of documents that are almost all in English, and want to identify which ones are not

For each of the following, say whether it is statistics, machine learning, or AI (**Note:** you are not expected to implement, just say which it is)

- Ask ChatGPT which documents are in English
- Zip the collection of documents and flag the ones that don't compress well
 - zipping algorithms are designed to code the most common text efficiently, so it can be useful for anomaly detection
 - We will learn the math behind this next week because it is important in Gen AI
- Use a random forest to identify the anomalies
- Train a deep learning model for anomaly detection based on a training set of english and non-english documents

HW 1-3

Try out a Generative AI based tool, investigate it, and report on your results.

- A couple of paragraphs or a few bullets is OK as long as they are insightful
- Be prepared to talk for 3-5 minutes in class on what you did and discovered

While it could be ChatGPT, try to be more creative along the lines of what I did with DropDeck and the Text-to-3D-asset tools in the lecture