



# ChatGPT, DALL-E and More: How Generative AI is being used for Science

Carlos Soto

July 31, 2024

# About This Talk

- ChatGPT and DALL-E
  - Two prominent generative AI tools, both developed by OpenAI
- Generative AI
  - How machines can create new and interesting content
- AI, in General
  - How AI works
  - How AI and Machine Learning are being used for Science



**ChatGPT**



# ChatGPT, DALL-E & More: Using AI for Science

**About me.** My research

**Artificial Intelligence.** History, misconceptions, major types

**Generative AI.** From recognizing to creating

**ChatGPT & DALL-E.** What they do, how they work

**Impact & More.** Applications, limitations, risks

# ChatGPT, DALL-E & More: Using AI for Science

**About me.** My research

**Artificial Intelligence.** History, misconceptions, major types

**Generative AI.** From recognizing to creating

**ChatGPT & DALL-E.** What they do, how they work

**Impact & More.** Applications, limitations, risks

## About me



Brookhaven  
National Laboratory

- PhD in robotics and AI from Texas A&M University
  - Lab: Center for Robot-Assisted Search and Rescue
  - Research: human-robot interaction, gesture recognition
- Machine Learning researcher at Brookhaven National Lab
  - Lead the AI Theory and Security group in the AI & Machine Learning department
  - Research: AI for science and security

# Some of my research

- Finding new isotope separation experiments with Large Language Models (LLMs)
  - *Bottom-up approach*: extract features with encoder LLMs, build contextual representations
  - *Top-down approach*: language generation with document context using decoder LLMs, align with prompt engineering/tuning
- Multi-modal document information extraction
  - Automatic table parsing
  - Reverse-engineering scientific charts

Table 2. Folate profiles of *E. coli* strains

Strain	THF	CH <sub>3</sub> -THF	CH=THF + 10-CHO-DHF <sup>†</sup>	5-CHO-THF	Total
Wild type	48.1 ± 10.7	10.6 ± 1.9	738 ± 93	68.9 ± 10.9	866 ± 114
<i>ΔfolE</i>	<0.05	<0.05	<0.05	<0.05	<0.2
<i>ΔfolP</i>	<0.05	<0.05	<0.05	<0.05	<0.2
<i>ΔgcvP ΔglyA</i>	845 ± 171	<0.05	<0.05	<0.05	845 ± 171
<i>ΔfolEΔthyA + 5-CHO-THF</i>	152 ± 100	7.1 ± 0.7	14.4 ± 3.5	5.8 ± 1.5	180 ± 98

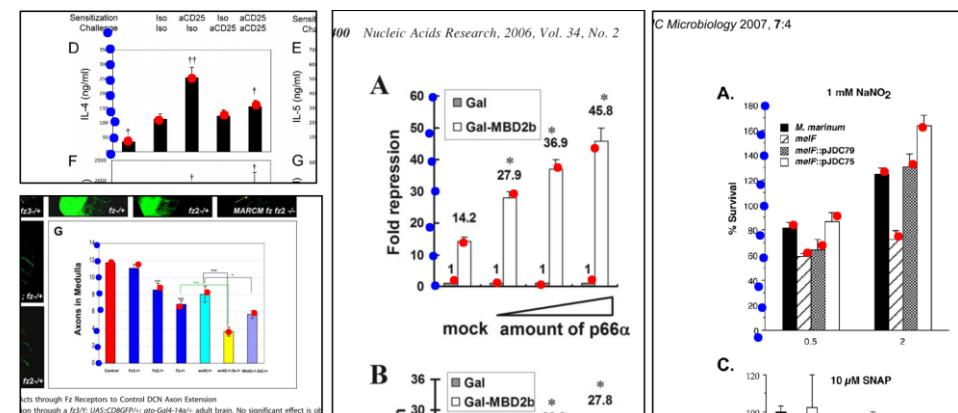
\*Means and standard errors of 3–7 replicates. THF, tetrahydrofolate; CH<sub>3</sub>-THF, 5-methyl-THF; CH=THF, 5,10-methenyl-THF; 10-CHO-DHF, 10-formyl-DHF; 5-CHO-THF, 5-formyl-THF. The detection limit for each folate was 0.05 pmol mg<sup>-1</sup> protein.

<sup>†</sup>10-CHO-THF is converted to CH = THF and 10-CHO-DHF during analysis.

<sup>‡</sup>Dihydrofolate not detected (detection limit 0.25 pmol mg<sup>-1</sup> protein, allowing for 20% recovery).

A 100 µg Th(IV) or 100 µg U(VI) was taken in 10 mL solution containing various concentrations of sulphuric acid ranging from 0.2 to 3.0 mol L<sup>-1</sup>. The extractions were carried out in 125 mL of separating funnel. The aqueous acid solution containing metal ion was taken in 10 mL of 0.5–4.0% of 4-methyl-N-n-octylaniline in xylene for 4 min to separate the organic and aqueous phases. The metal loaded organic phase was stripped with strippant by wrist shaking of the funnel for 5 min. The 10 mL of 0.1 mol L<sup>-1</sup> nitric acid was used as strippant for Th(IV) while 2 × 10 mL acetate buffer of pH 4.5 was used as strippant in case of U(VI). The pH was adjusted with sodium hydroxide and acetic acid as per their requirement for the spectrophotometric determination. For quantification of Th(IV) xylenol orange was used as chromogenic agent while for quantification of U(VI) bromopyrogallol red was used [39].

P.S. More, et al. Extraction of Th(IV) and U(VI) with 4-methyl-N-n-octylaniline as an extracting agent. *J Radioanal Nucl Chem* 331, 4149–4158 (2022).



# ChatGPT, DALL-E & More: Using AI for Science

About me. My research

**Artificial Intelligence.** History, misconceptions, major types

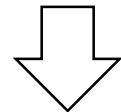
Generative AI. From recognizing to creating

ChatGPT & DALL-E. What they do, how they work

Impact & More. Applications, limitations, risks

# What is AI?

Cognitive Science

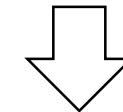


**Artificial Intelligence**

Try to understand the rules for how the mind works, build them into machines

Planning  
Searching  
Reasoning  
...

Statistics



**Machine Learning**

Try to identify mathematical patterns in data, use patterns to make predictions

Regression  
Classification  
Clustering  
...

AI/ML Methods

...

..

.

Deep Neural Networks

# What AI/ML is, and what it is not

A set of techniques to learn patterns from data

Able to make predictions and inferences

Remarkably capable when adapted to many tasks

Heavily reliant on data

Replacement for expertise or common sense

Plug-in solution to arbitrary problems

Generalizable beyond its learned regime

# A Brief History of AI

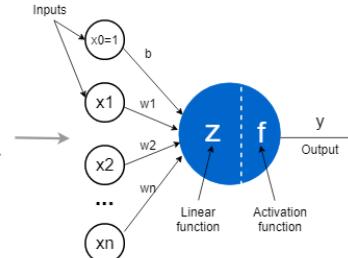
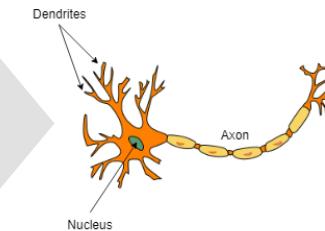


Automatons  
(1500BCE ~ 1900s)

"Can machines think?"

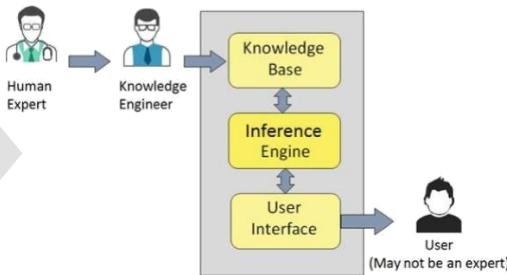


Alan Turing and the Dartmouth Conference  
(1950s)



"AI Winter"  
#1  
(1970s)

Brain-inspired Artificial Neural Networks  
(1960s)

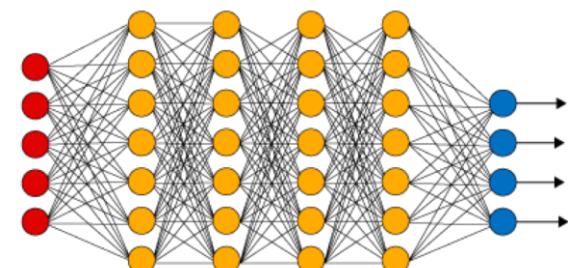


"Expert Systems"  
1980s

"AI Winter"  
#2  
1980-90s



Machine Learning & Big Data  
2000s



Deep Learning  
2012+

# Basic types of AI/ML tasks

## Classification

"the answer I'm looking for is one of these categories"

## Regression

"the answer I'm looking for is a number"

## Clustering

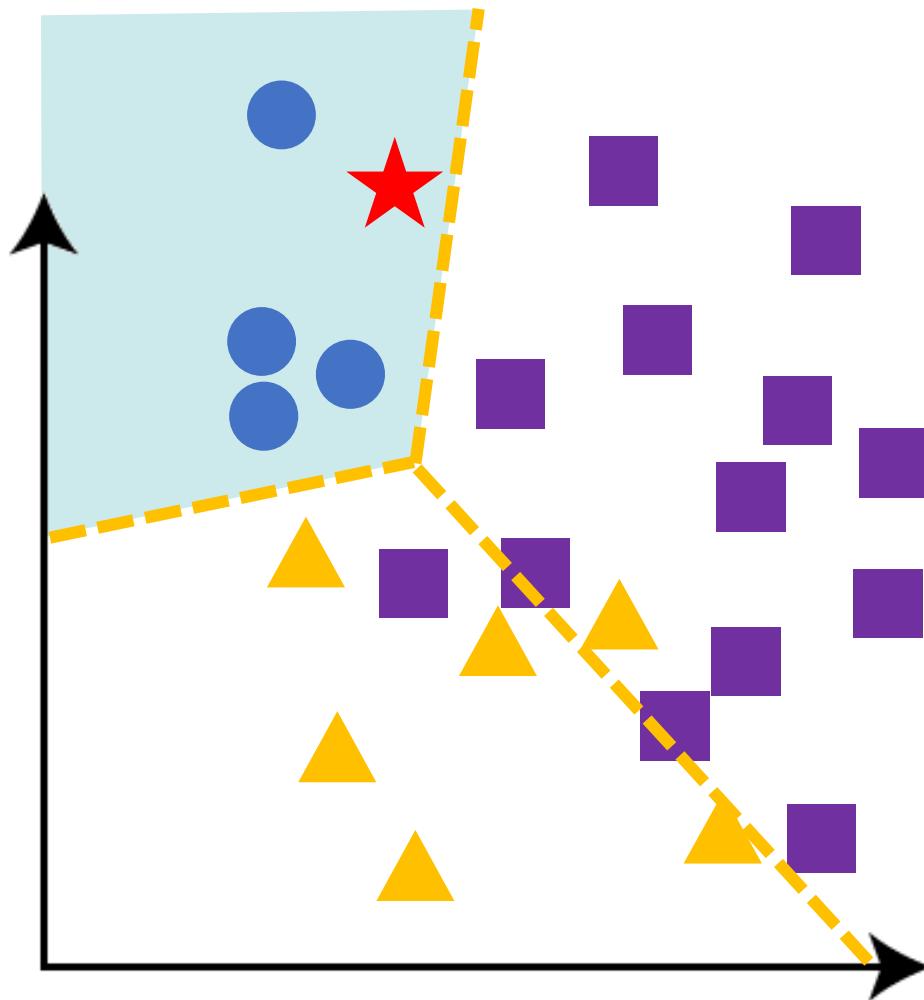
"I'm not sure what I'm looking for, but I have some data"



**Supervised  
Learning**

**Unsupervised  
Learning**

# Classification Tasks



Which category does my new data point **★** resemble?

Potential issues..

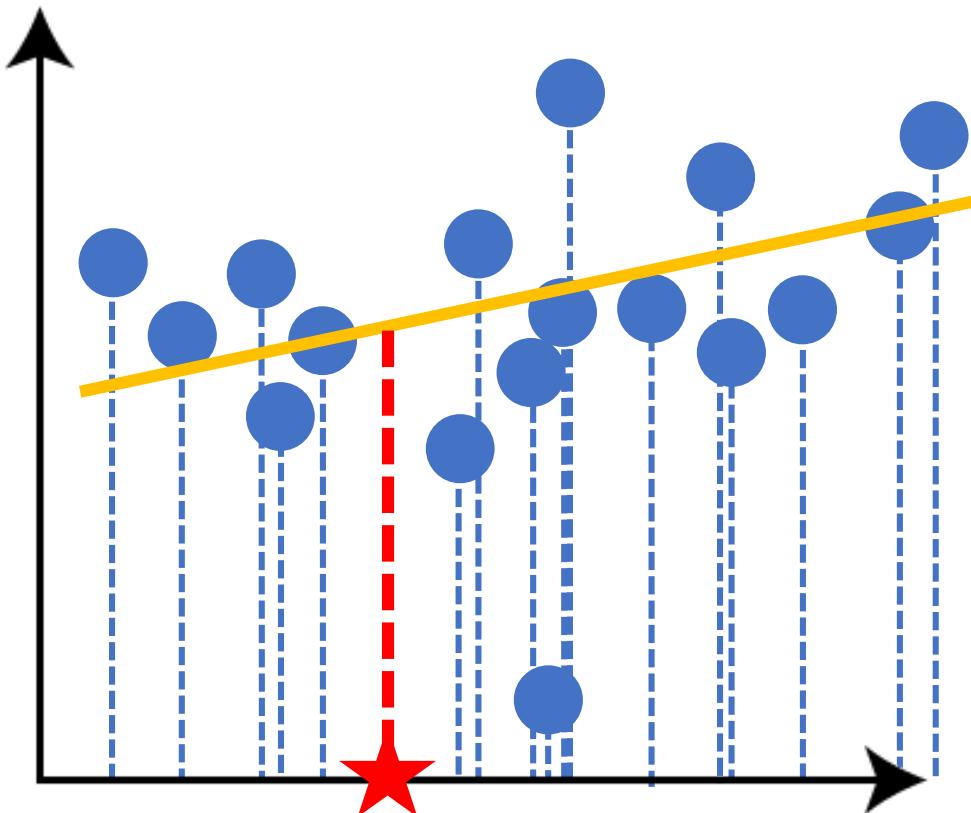
Class separability

Number of classes

Number of dimensions

Decision boundary shape

# Regression Tasks



What value should I expect  
for new data point **★**?

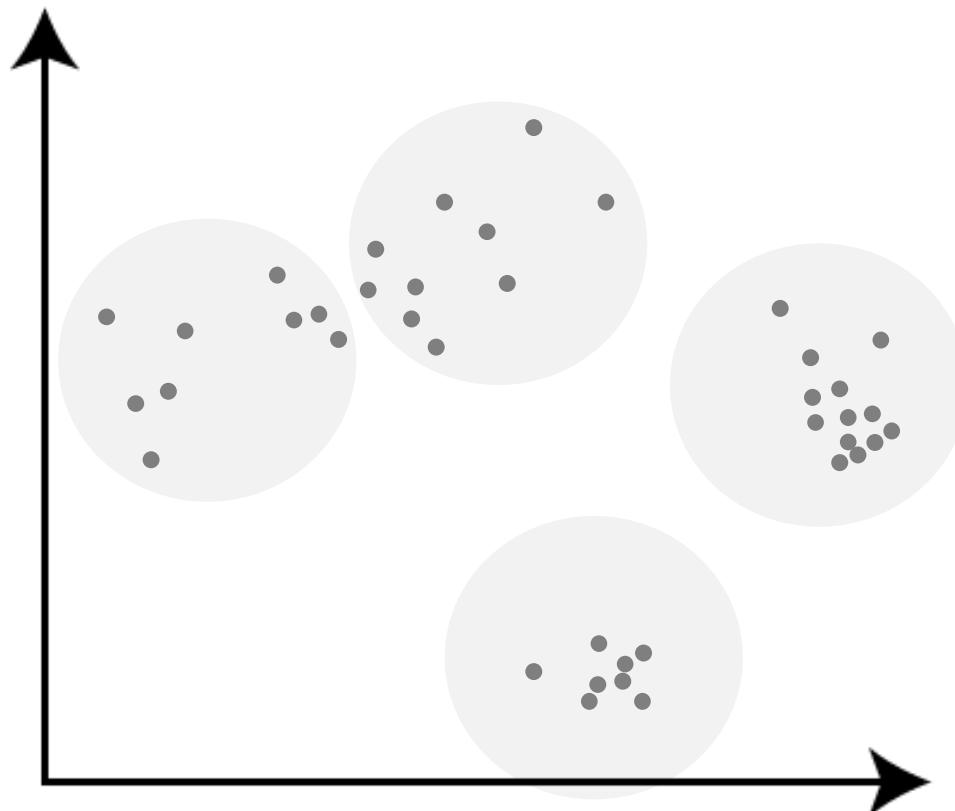
Potential issues..

Quality of curve fit

Type of curve fit (e.g. linear,  
polynomial)

Outliers

# Clustering Tasks



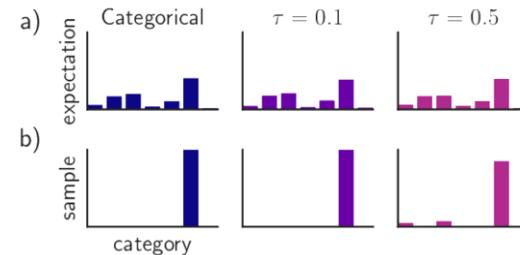
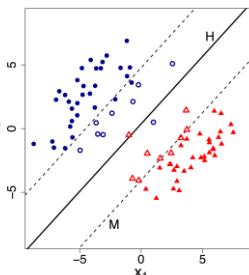
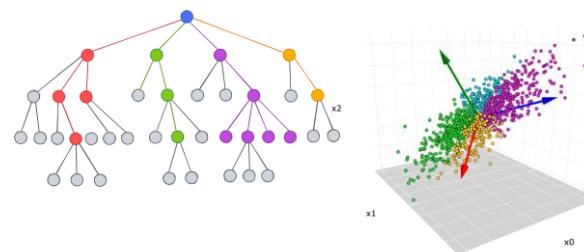
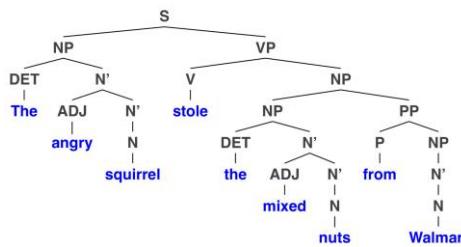
Which of my data points are related and which are not?

Potential issues..

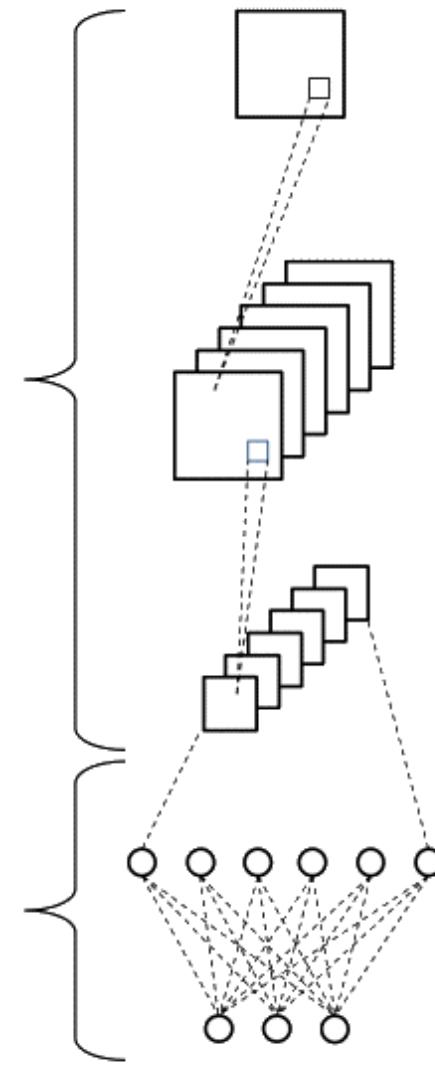
Number of clusters

What counts as a cluster?

# From “Classical Machine Learning” to Deep Learning



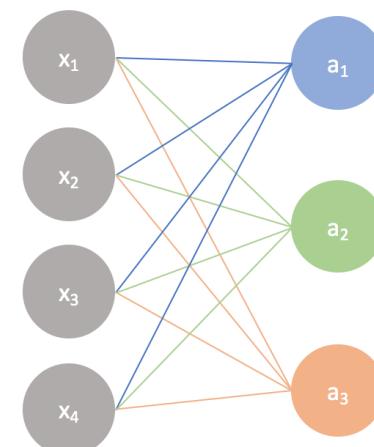
Data  
↓  
Features  
↓  
Pattern Learning  
↓  
Prediction



# Artificial Neural Networks (ANNs)

- Just complex curve-fitting
- Matrix multiply + nonlinear activation
- Layer for best results
  - Capture more complexity
  - Need more data
- Train with gradient descent

Input layer      Output layer



A simple neural network

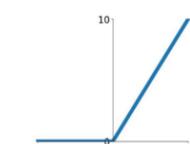
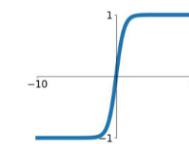
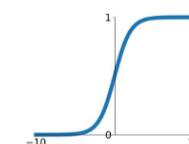
$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b \\ b \\ b \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \end{bmatrix} \xrightarrow{\text{activation}} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

## Activation Functions

**Sigmoid**  
 $\sigma(x) = \frac{1}{1+e^{-x}}$

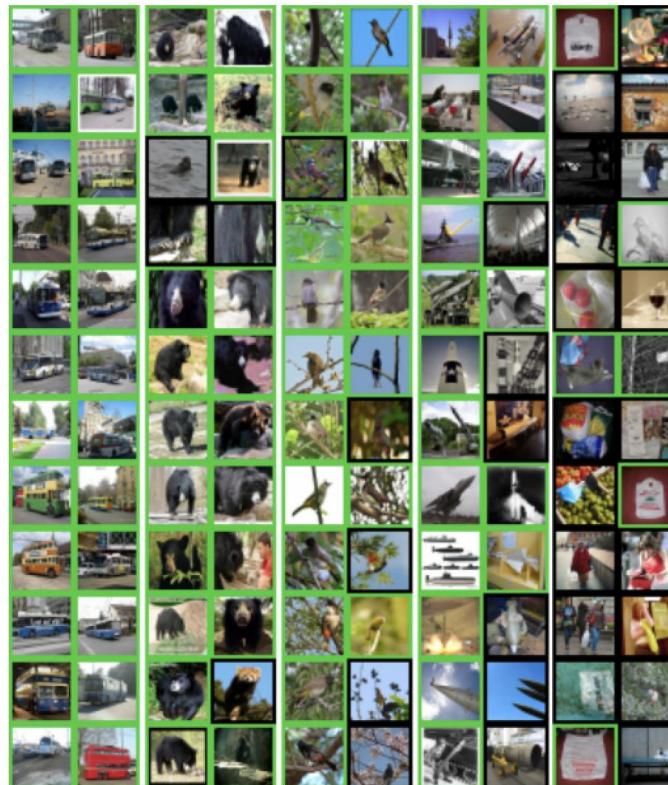
**tanh**  
 $\tanh(x)$

**ReLU**  
 $\max(0, x)$

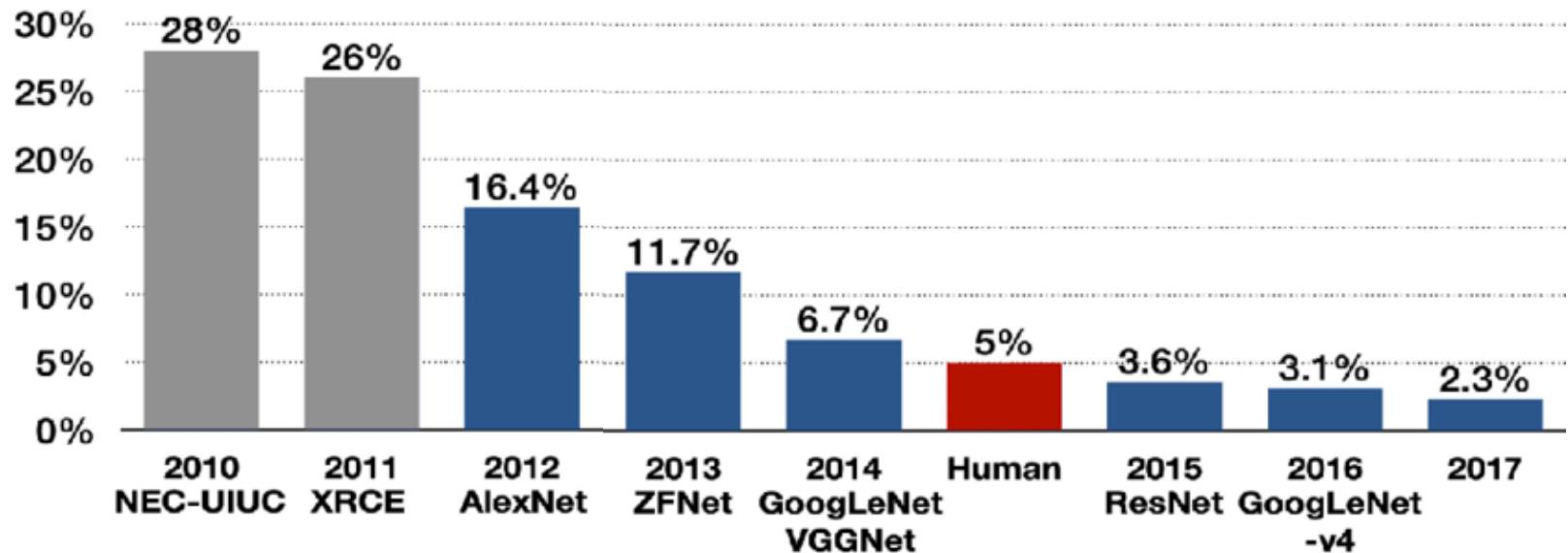


# The Deep Learning Revolution

ImageNet Challenge: ~1 million images, 1000 classes

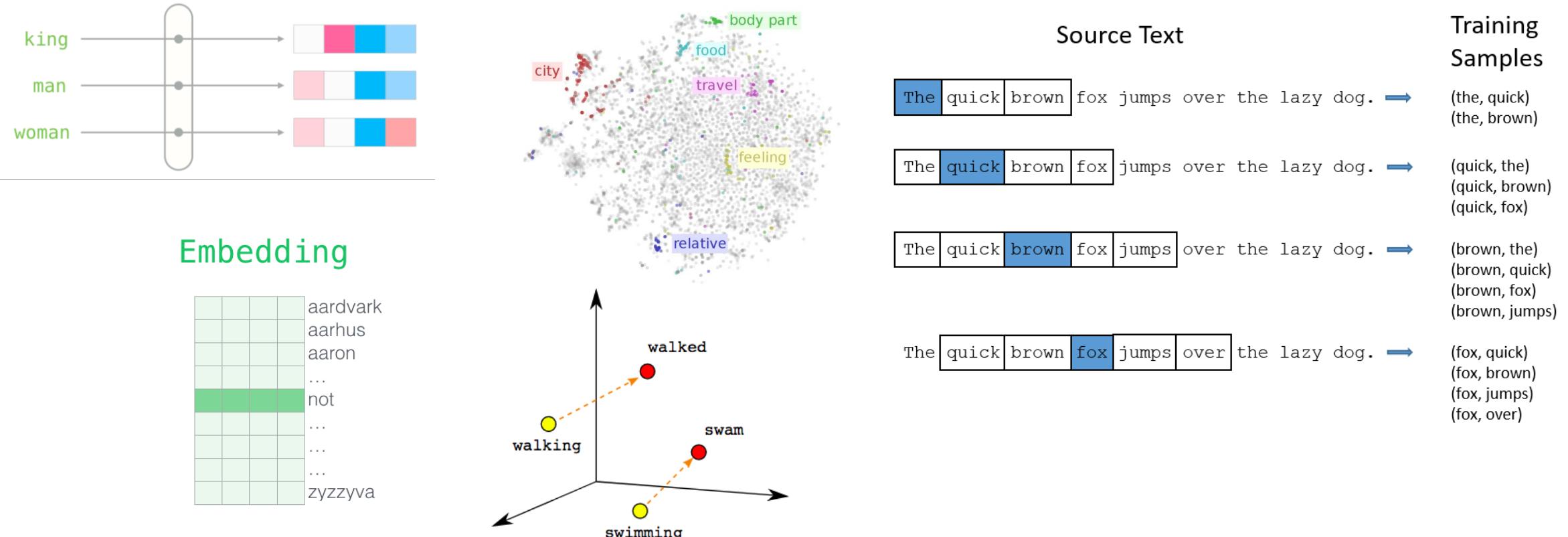


ImageNet Competition Annual Winner  
prediction error (lower is better)



# A Revolution in Language, Too

Word Embeddings: Replace words with learned numerical representations



Tomas Mikolov et al. "Efficient Estimation of Word Representations in Vector Space," 2013  
Jeffrey Pennington et al. "GloVe: Global Vectors for Word Representation," 2014

# ChatGPT, DALL-E & More: Using AI for Science

About me. My research

Artificial Intelligence. History, misconceptions, major types

**Generative AI.** From recognizing to creating

ChatGPT & DALL-E. What they do, how they work

Impact & More. Applications, limitations, risks

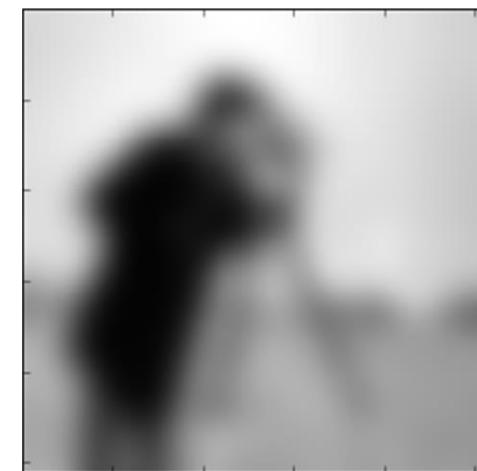
# AI for Images: convolution filters



1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9



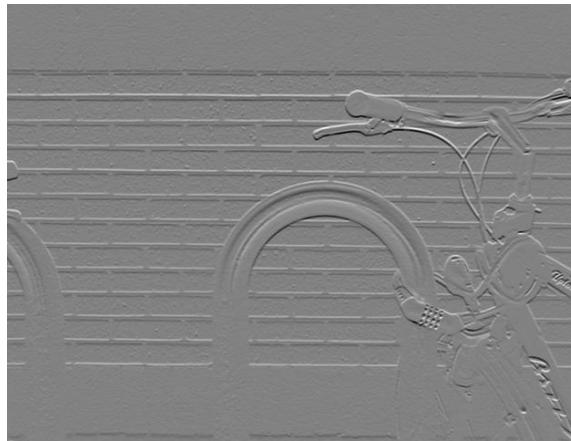
1/16	1/8	1/16
1/8	1/4	1/8
1/16	1/8	1/16



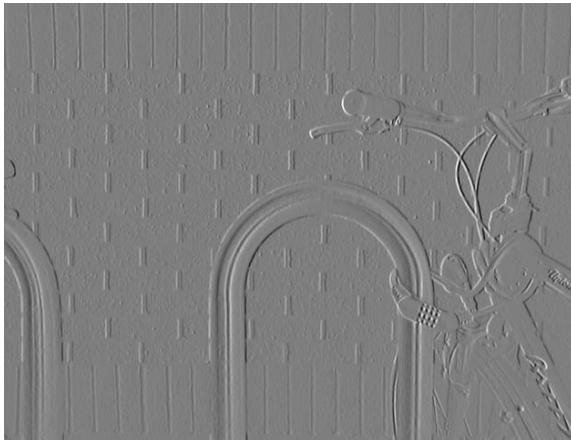
# AI for Images: convolution filters



-1	-2	-1
0	0	0
1	2	1

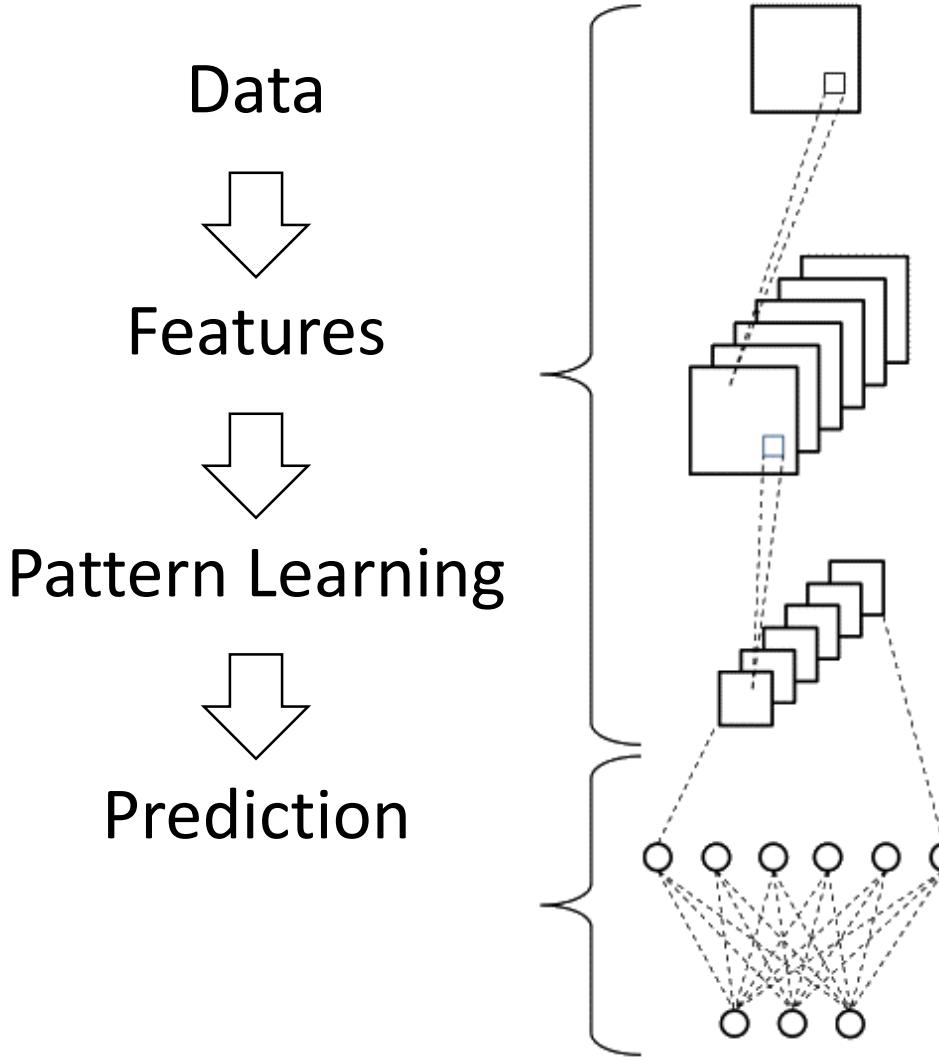
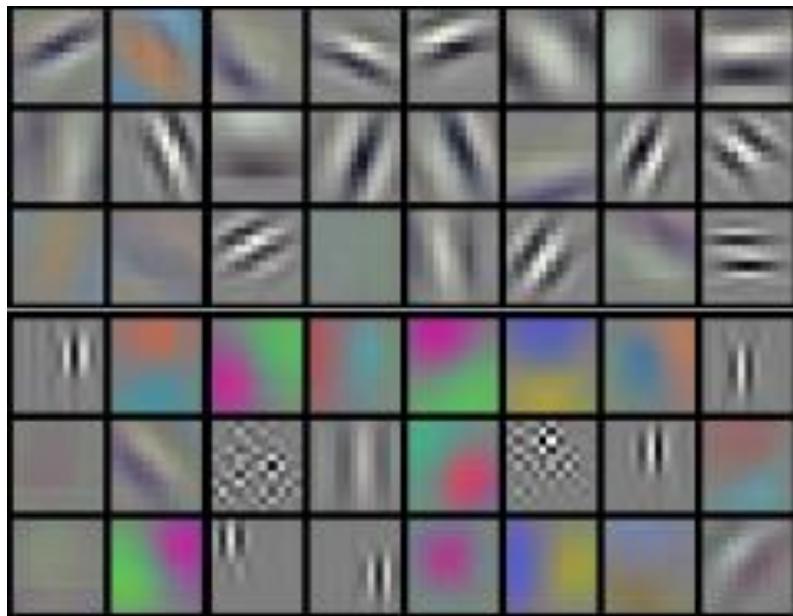


-1	0	1
-2	0	2
-1	0	1

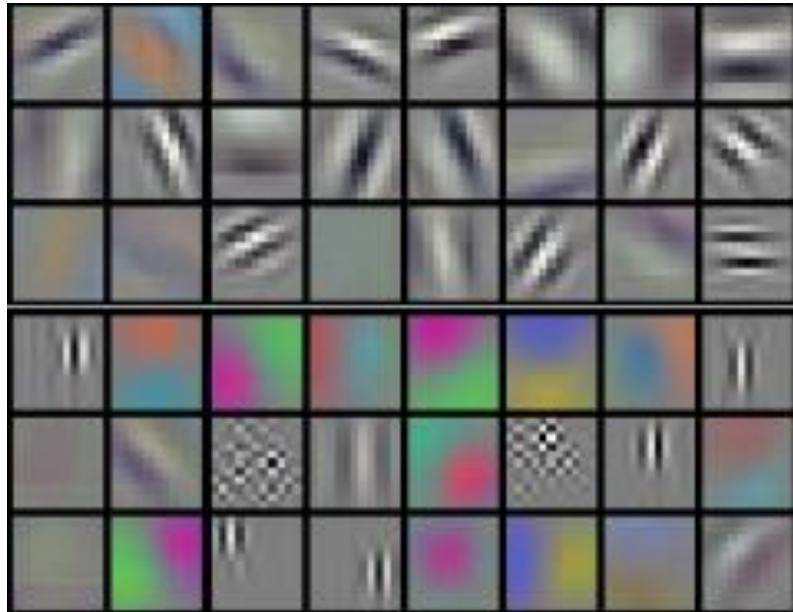


# AI for Images: Learning filters, multiple layers

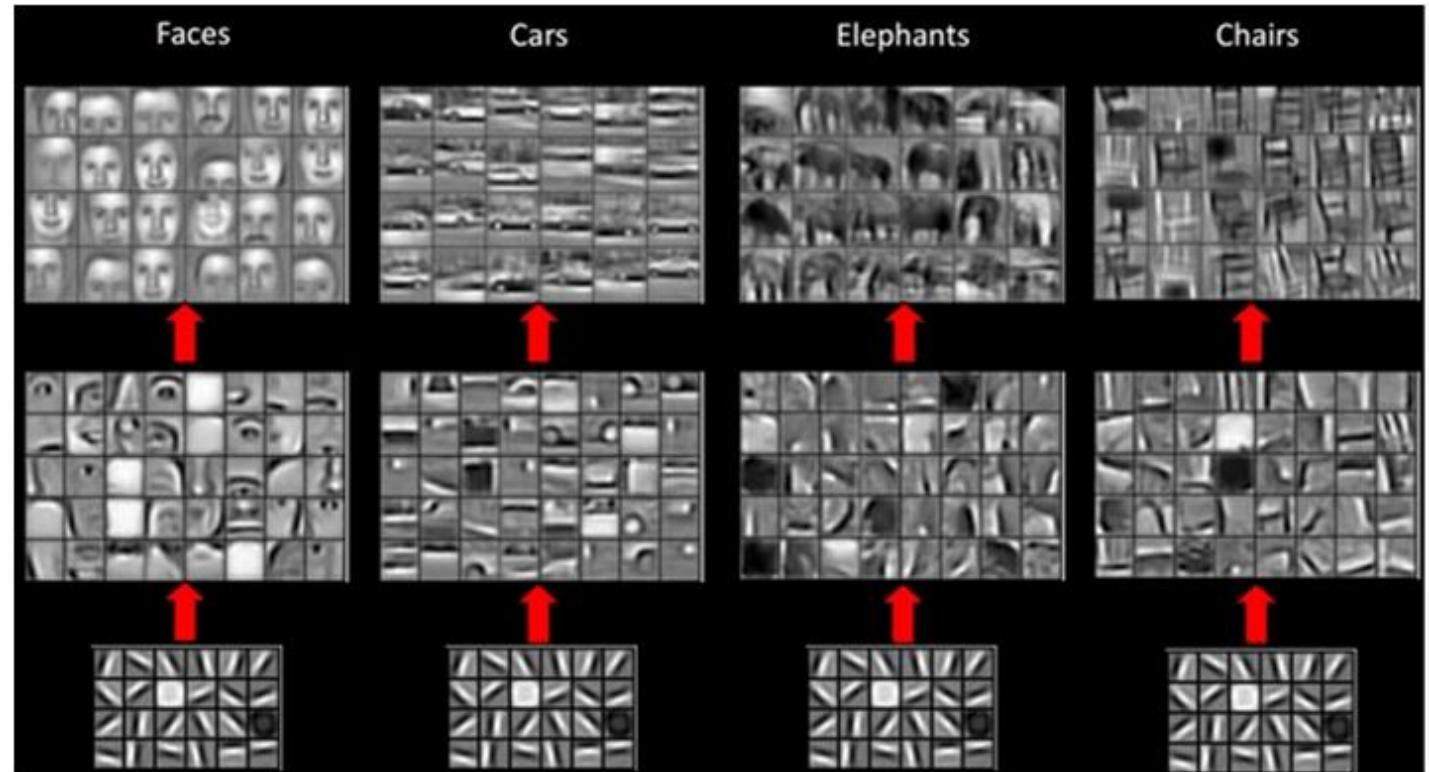
$$\begin{array}{|c|c|c|} \hline 1/16 & 1/8 & 1/16 \\ \hline 1/8 & 1/4 & 1/8 \\ \hline 1/16 & 1/8 & 1/16 \\ \hline \end{array}$$



# AI for Images: Learning filters, multiple layers



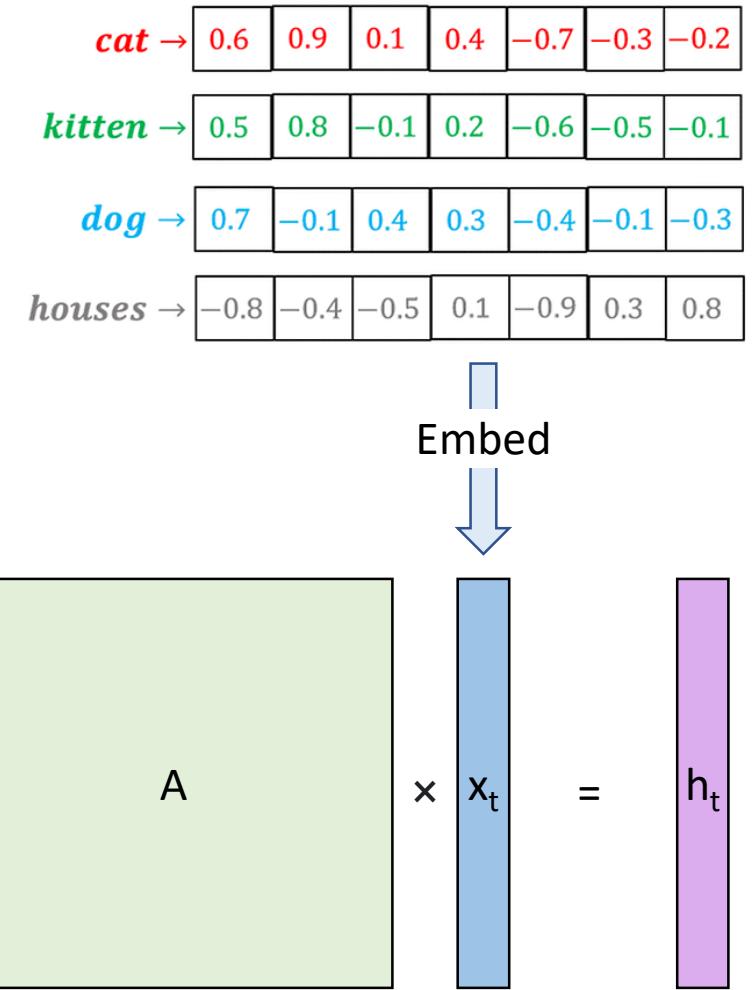
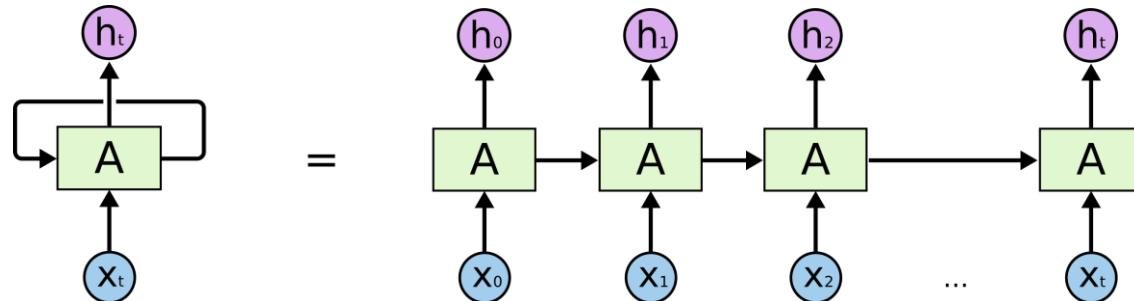
1/16	1/8	1/16
1/8	1/4	1/8
1/16	1/8	1/16



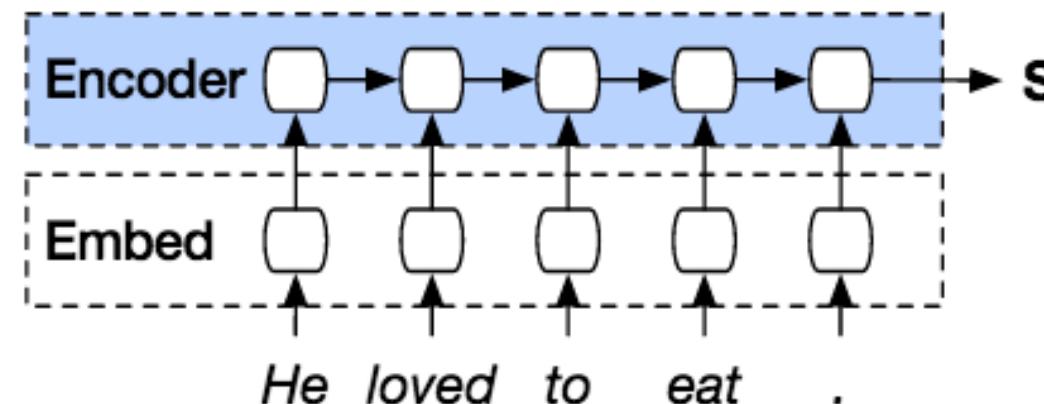
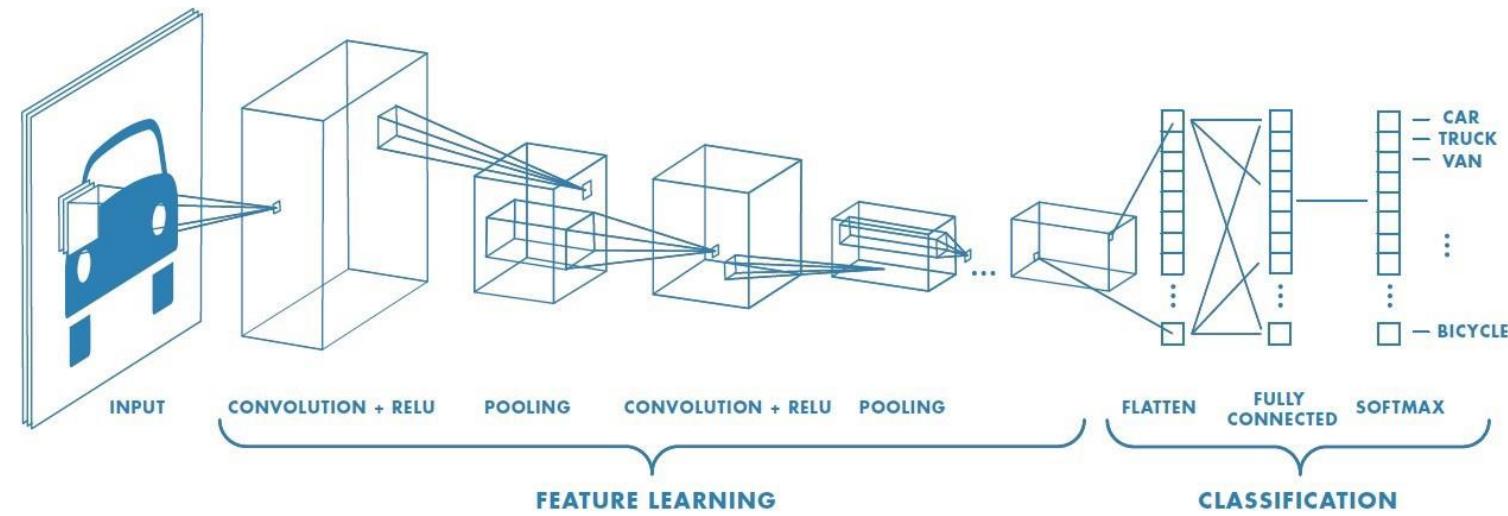
# AI for Language

## Recurrent Neural Networks (RNNs)

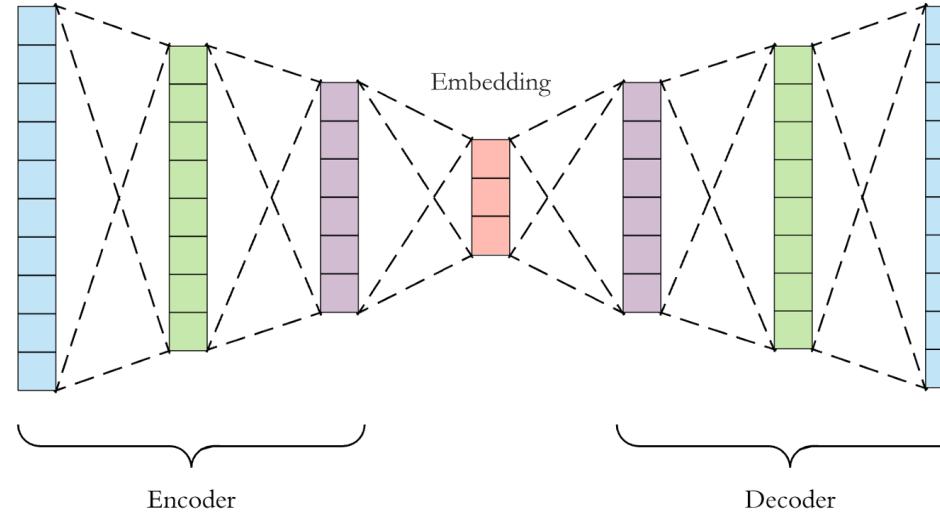
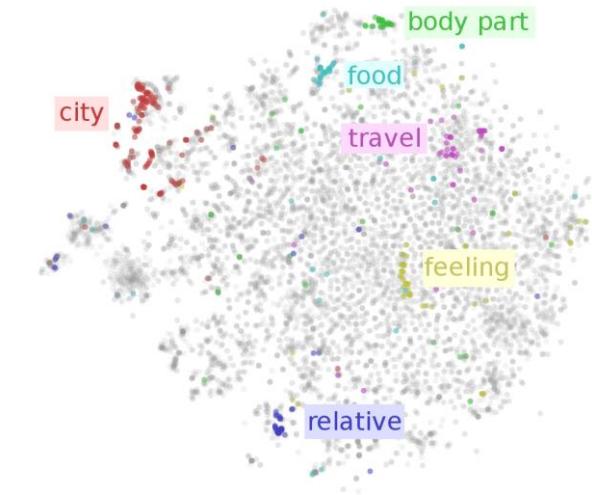
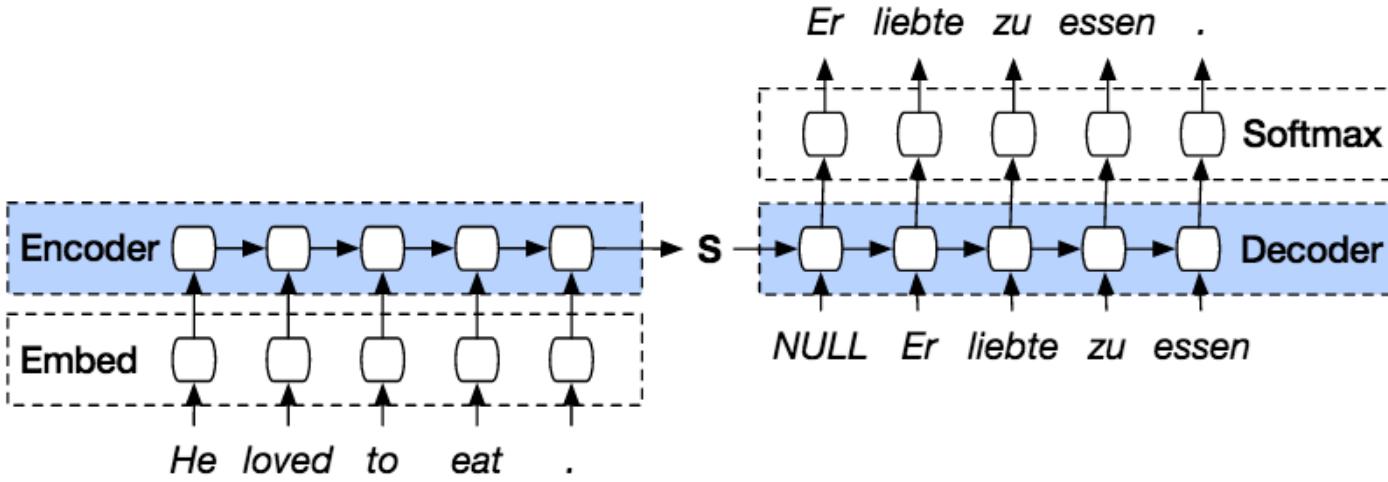
- NNs unrolled across time/sequence



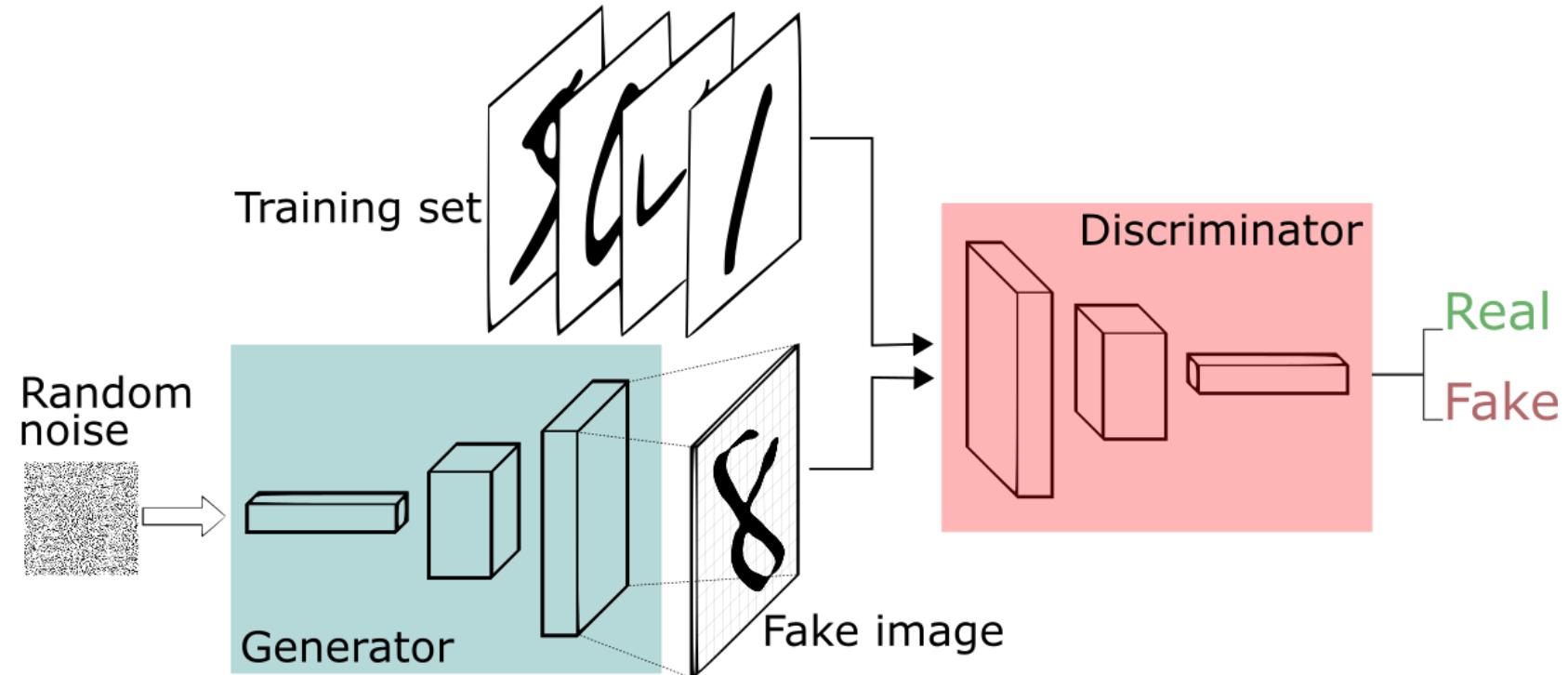
# Deep Feature Representations and Predictions



# From Features back to Words and Images



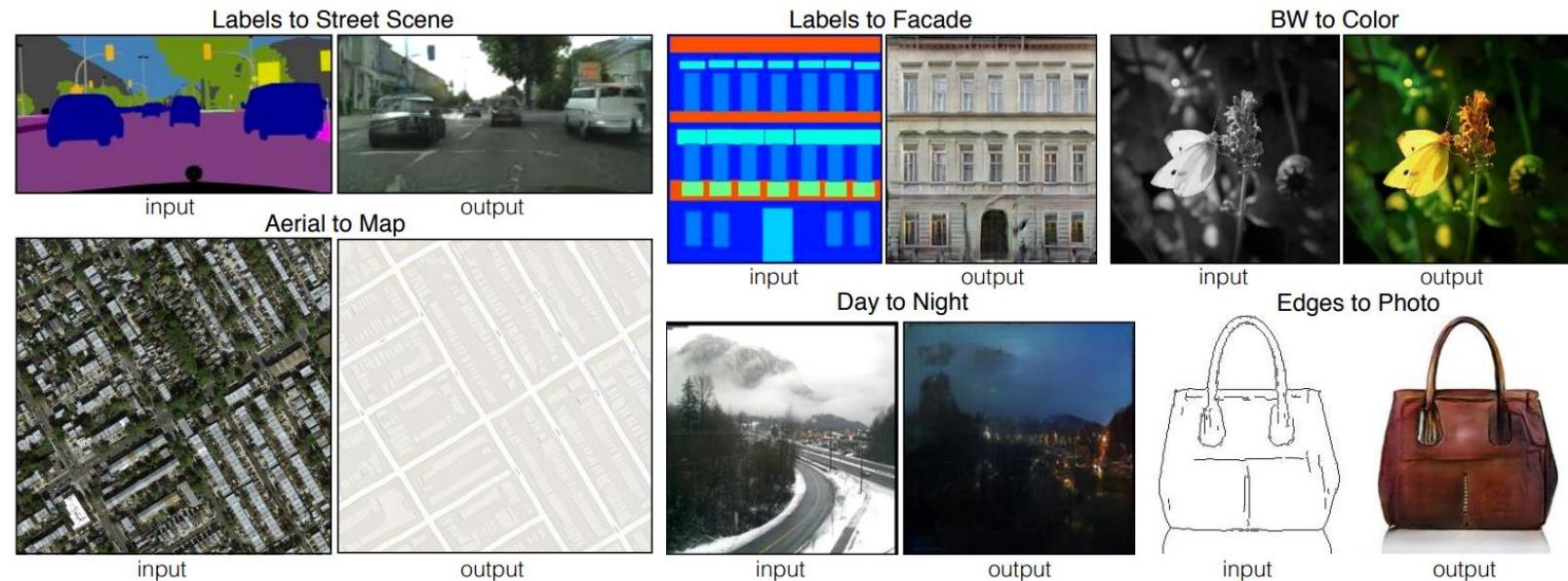
# Generating Images that Actually Look Good



Generative Adversarial Networks (GANs)

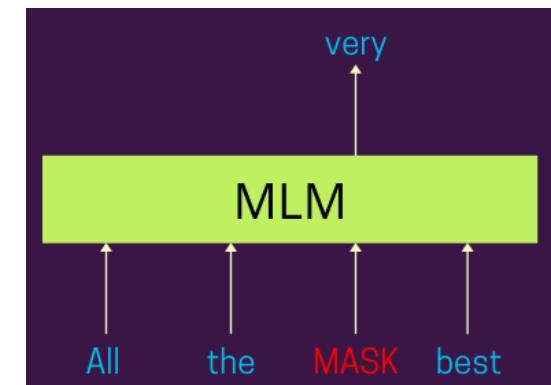
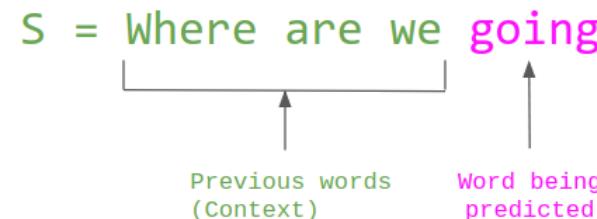
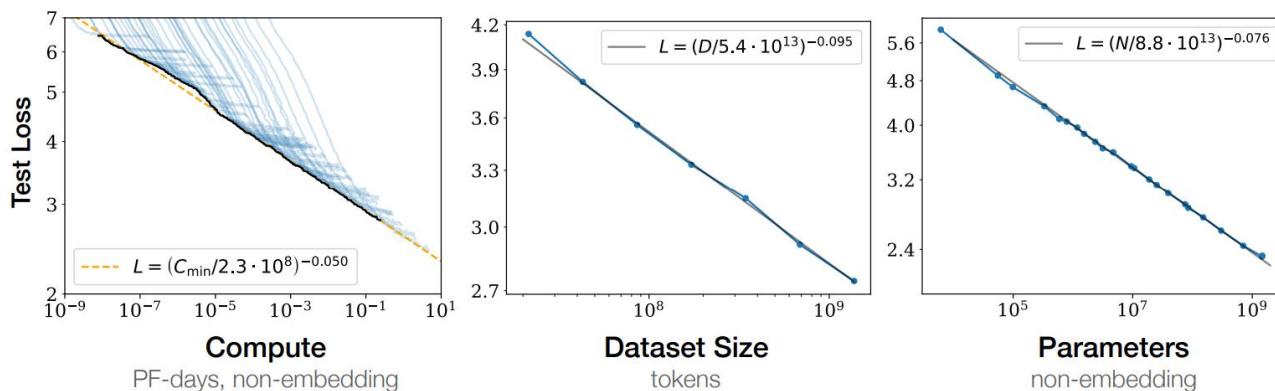
Ian Goodfellow, 2014

# Generating Images that Actually Look Good



# Generating Convincing Language

- Language Modeling
  - Next (or missing) word prediction
- Scaling
  - Loss & performance scale as power-law in parameters, data, compute



Dataset	Quantity (tokens)	Weight in training mix
Common Crawl (filtered)	410 billion	60%
WebText2	19 billion	22%
Books1	12 billion	8%
Books2	55 billion	8%
Wikipedia	3 billion	3%

# ChatGPT, DALL-E & More: Using AI for Science

About me. My research

Artificial Intelligence. History, misconceptions, major types

Generative AI. From recognizing to creating

**ChatGPT & DALL-E.** What they do, how they work

Impact & More. Applications, limitations, risks

# What ChatGPT/GPT-4 Can Do

Draft/edit emails, articles, lyrics, movie scripts, etc.

Summarize & explain

Recite information

Code and debug

Solve math problems

Simulate programs & games

Copy writing styles

Translate

Analyze & reason  
(meaning, sentiment, humor, Q&A, etc.)

Take (and pass) human tests

Understand images

Exam results (ordered by GPT-3.5 performance)

Estimated percentile lower bound (among test takers)

100% –

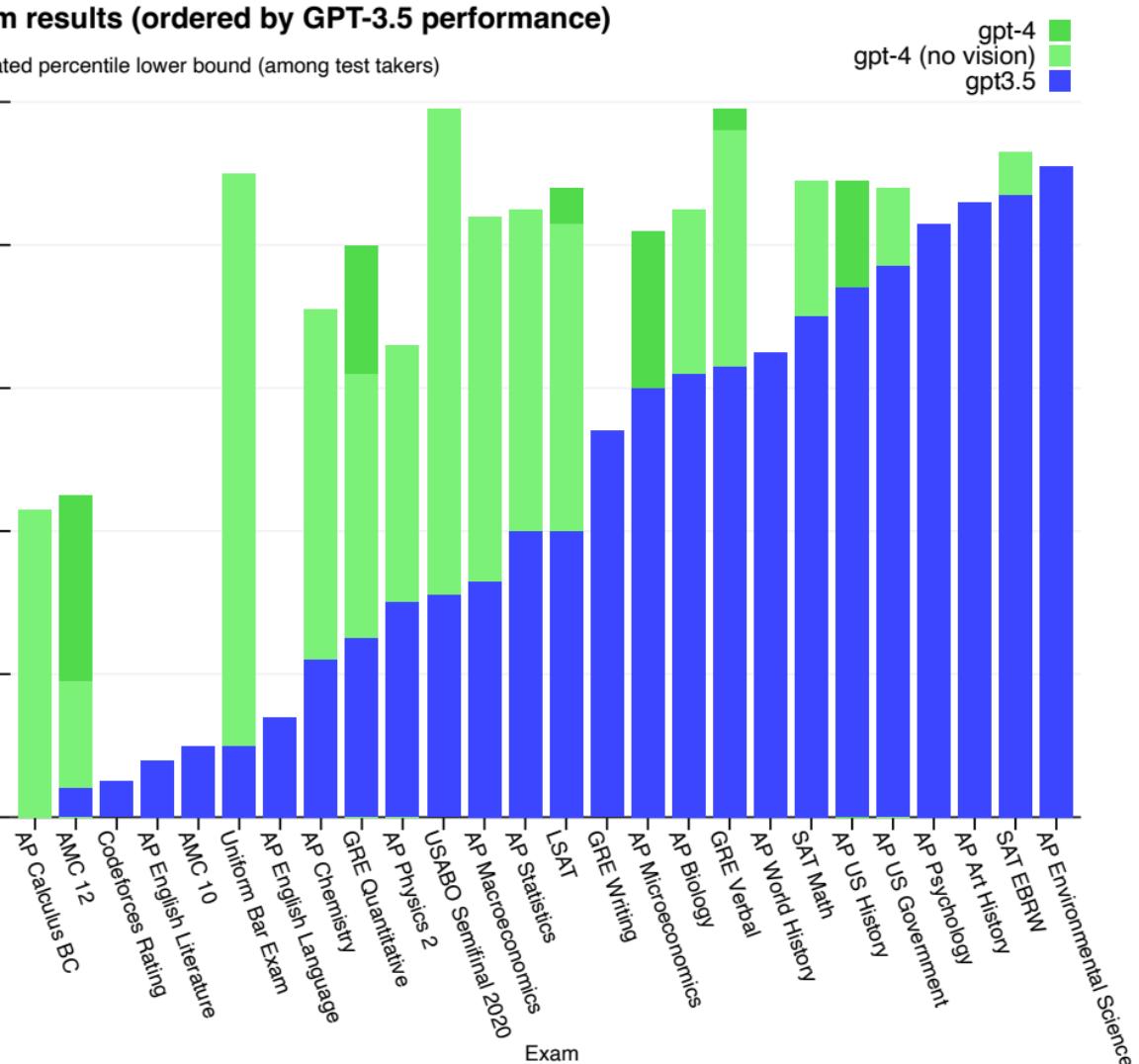
80% –

60% –

40% –

20% –

0% –



# What ChatGPT/GPT-4 Can Do

Draft/edit emails, articles, lyrics, movie scripts, etc.

Summarize & explain

Recite information

Code and debug

Solve math problems

Simulate programs & games

Understand images

Copy writing styles

Translate

Analyze & reason  
(meaning, sentiment, humor, Q&A, etc.)

Take (and pass) human tests

## Example of GPT-4 visual input:

User      What is funny about this image?



Source: <https://www.reddit.com/r/hmmm/comments/ubab5v/hmmm/>

GPT-4

The humor in this image comes from the absurdity of plugging a large, outdated VGA connector into a small, modern smartphone charging port.

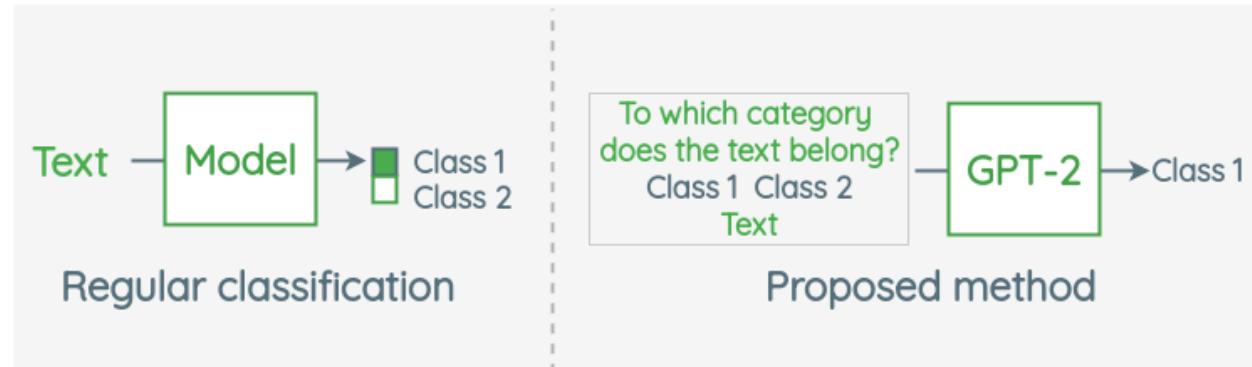
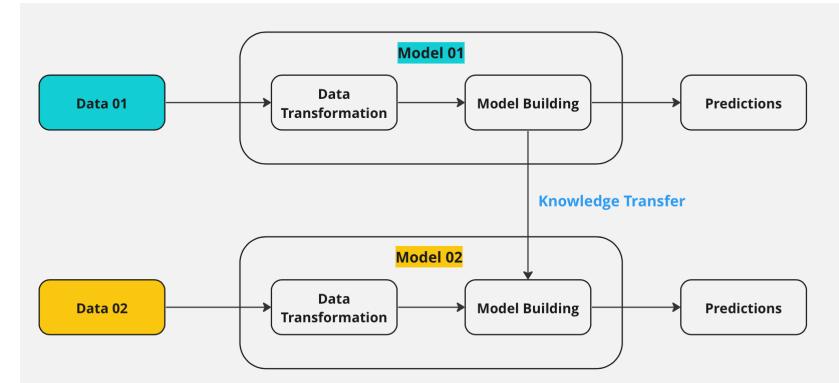
# What ChatGPT is

An AI Chatbot

A Large Language Model (LLM) tuned using Reinforcement Learning from Human Feedback (RLHF) for conversational interactions in text

# Using Large Language Models

- Previous norm adapting ML/NLP models:
  - Take pre-trained model and fine tune (e.g. BERT, ResNet, etc.)
  - Transfer learning
- Recently:
  - Take pre-trained **frozen** model, use as-is (e.g. GPT-3)
  - Manipulate prompt to adapt to a task
- Now:
  - Fine-tuning possible again with new approaches (RLHF, LoRA)



# ChatGPT

Based on InstructGPT

Ouyang, et al. "Training language models to follow instructions with human feedback" 2022

**Step 0: LLM**

**Step 1: supervised task tuning**

**Step 2: sample, train proxy reward model**

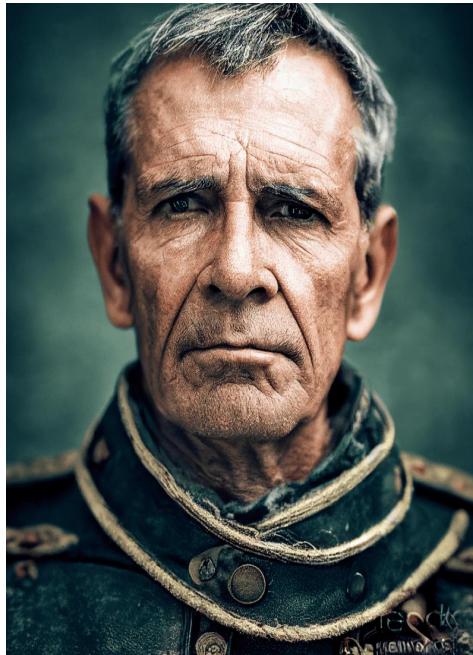
**Step 3: optimize LLM with RL reward**

"reinforcement learning from human feedback"

# AI for Generating Images

Image generation from novel prompts

- DALL-E, Stable Diffusion, MidJourney..

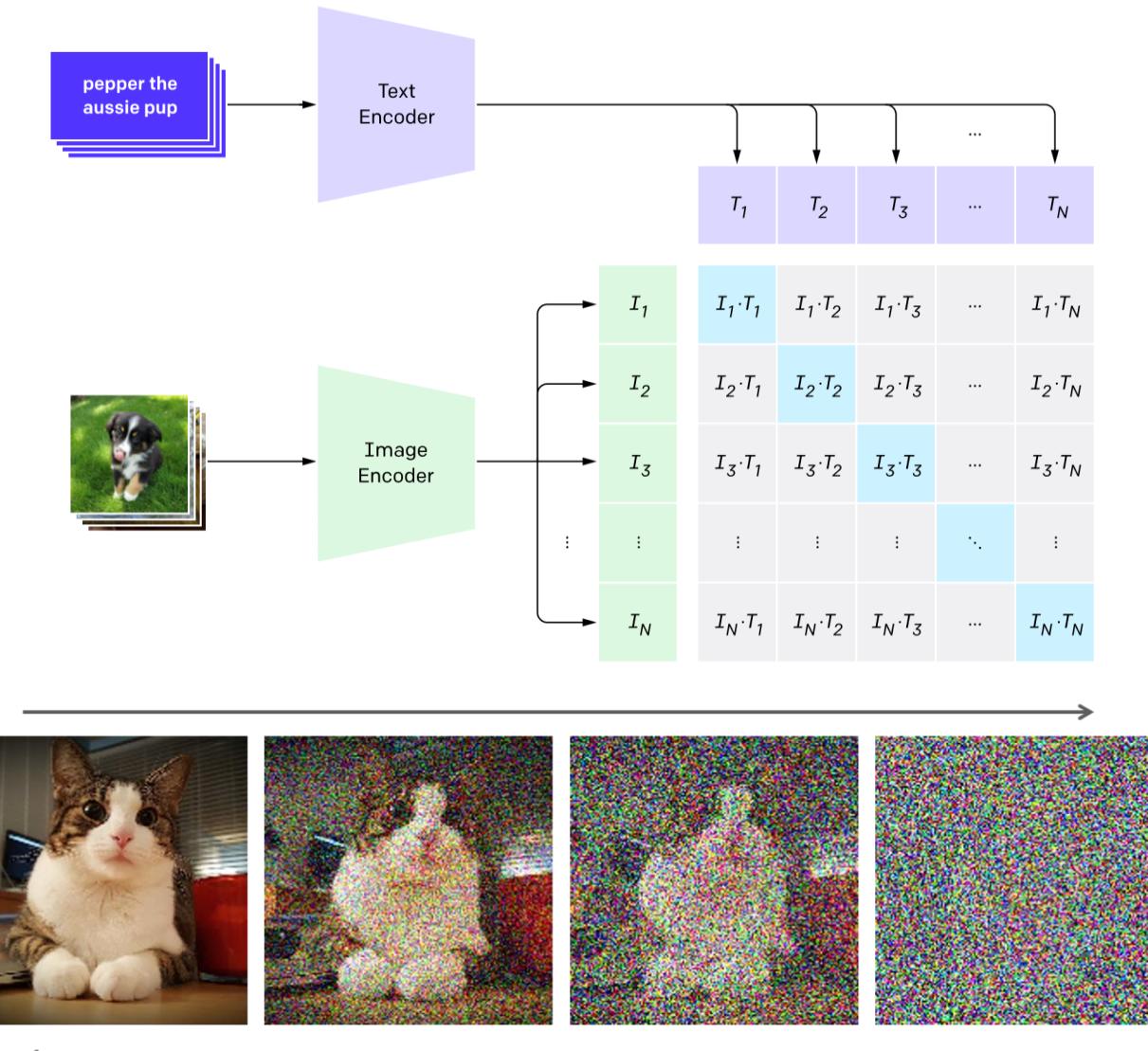


Stable Diffusion, by Stability AI



DALL-E 2, by OpenAI

# Generating High-Quality Images for Novel Prompts



CLIP (*Contrastive Language–Image Pre-training*)  
Alec Radford et al., “CLIP: Connecting Text and Images,” 2021

→ DALL-E (1 & 2)

Denoising Diffusion Models  
Jonathan Ho et al. “Denoising Diffusion Probabilistic Models,” 2020

→ DALL-E 2, Stable Diffusion, etc.

# Generating High-Quality Images for Novel Prompts



Image synthesis with Stable Diffusion 2.1  
Prompt: "dog with the mane of a horse"



# ChatGPT, DALL-E & More: Using AI for Science

About me. My research

Artificial Intelligence. History, misconceptions, major types

Generative AI. From recognizing to creating

ChatGPT & DALL-E. What they do, how they work

Impact & More. Applications, limitations, risks

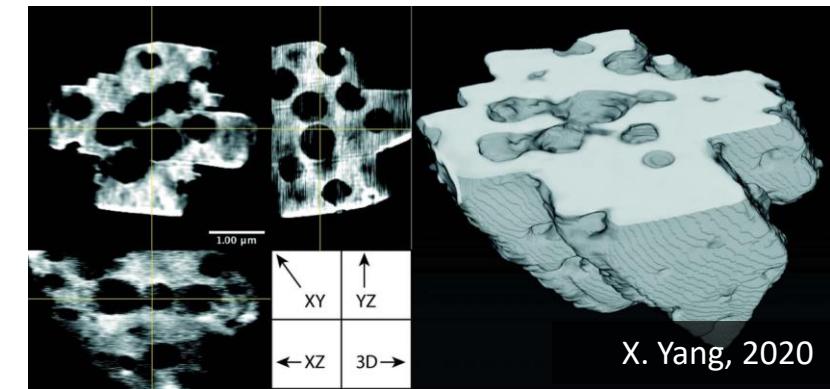
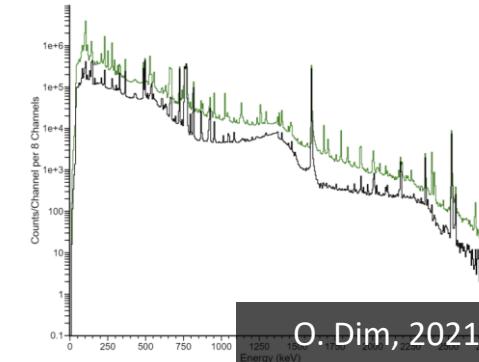
# How Scientists are already using AI

## Problems and tasks

- Making predictions (classification, regression)
- Looking for patterns (esp. in very large data)
- Data reduction, transformation, analysis
- Automating experiments
- Accelerating modeling and simulation

## Data types

- Experimental/sensor data
- Models and simulations
- Imagery (2D, 3D, etc.)
- Text (e.g. gene, chemical sequences)



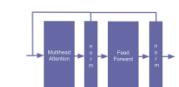
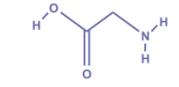
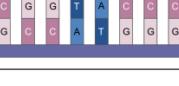
[START\_I\_SMILES] O=C(O)CCCC1=CC=C(N(CCC1)CCC1)C=C1 [END\_I\_SMILES]

[START\_AMINO] MEEPQSDPSVEPPLSQETFSDLWKLPE... [END\_AMINO]

[START\_I\_SMILES] CC(O)(P(=O)(O)O)P(=O)(O)O [END\_I\_SMILES]

# Large Language Models for Science

- Transformers pre-trained on scientific text
  - SciBERT<sup>1</sup>, BioBERT<sup>2</sup>
- Transformers pre-trained on omics strings
  - DNABERT<sup>3</sup>, GeneBERT<sup>4</sup>
- Multi-modal Transformers
  - GPT-4, PaLM 2, Galactica<sup>5</sup>

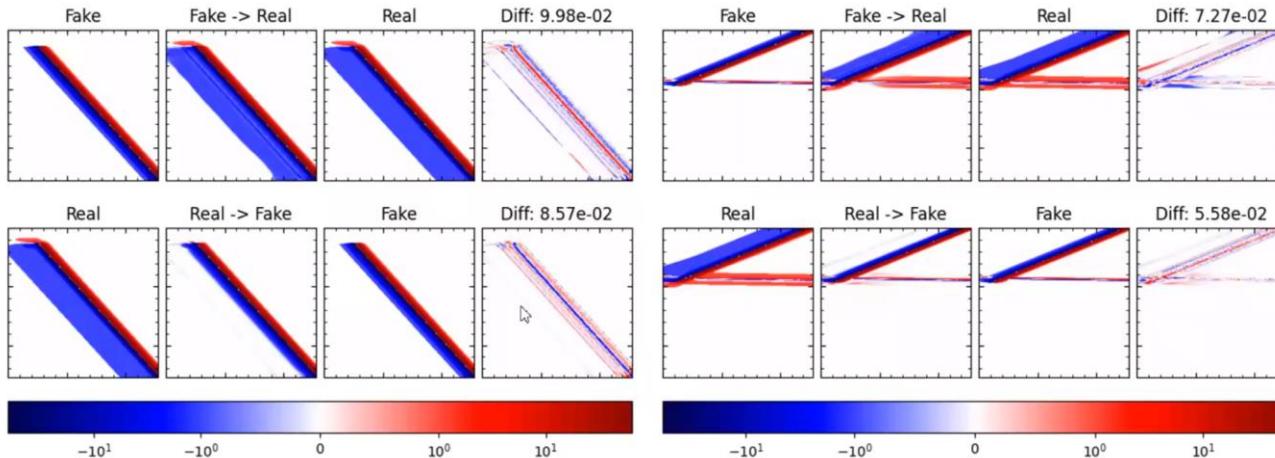
Modality	Entity	Sequence	
Text	Abell 370	Abell 370 is a cluster...	
LATEX	Schwarzschild radius	$r_s = \frac{2GM}{c^2}$	$r_s = \frac{2GM}{c^2}$
Code	Transformer	class Transformer(nn.Module):	
SMILES	Glycine	<chem>C(C(=O)O)N</chem>	
AA Sequence	Collagen $\alpha$ -1(II) chain	MIRLGAPQTL...	
DNA Sequence	Human genome	CGGTACCCCTC..	

1. Beltagy et al. "SciBERT: A Pretrained Language Model for Scientific Text" 2019
2. Lee et al. "BioBERT: a pre-trained biomedical language representation model for biomedical text mining" 2019
3. Ji et al. "DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome" 2020
4. Mo et al. "Multi-modal Self-supervised Pre-training for Regulatory Genome Across Cell Types" 2021
5. Taylor et al. "Galactica: A Large Language Model for Science" 2022

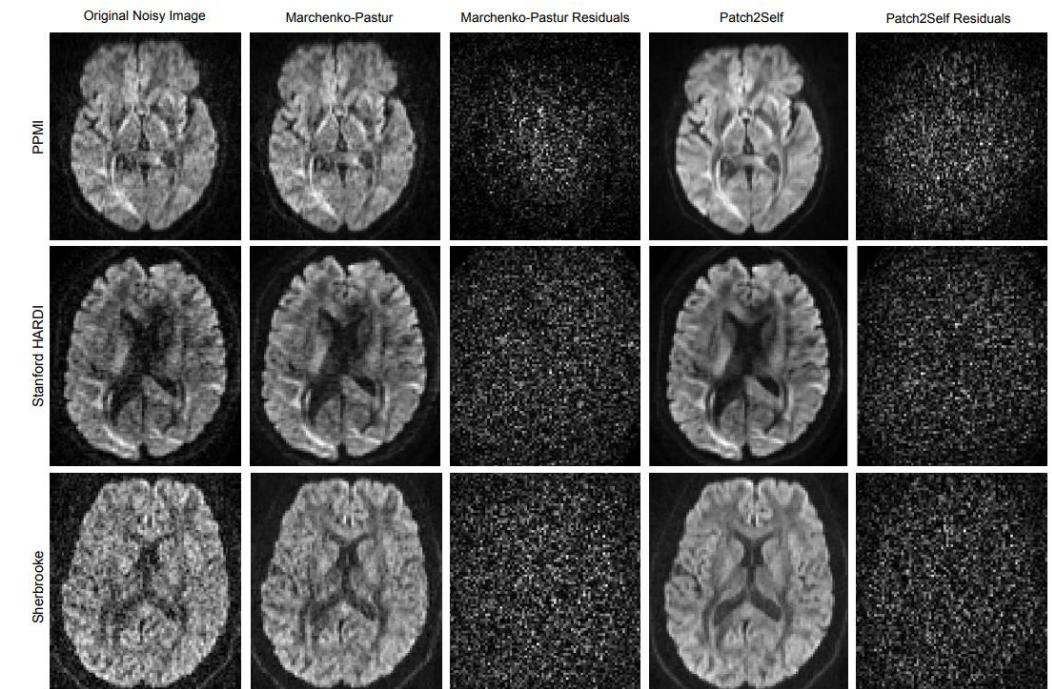
# GANs and Diffusion Models for Science

## DUNE application of UVCGAN

Translation result I (fake=q1d, real=2d)



Yi Huang et al.



Shreyas Fadnavis et al.

# Ethics and Legality



**Midjourney Is Being Class-Action Sued for Severe Copyright Infringements**

BY YOSSI MENDLOVICH · FEBRUARY 15, 2023 · EDUCATE · 4 MINS READ



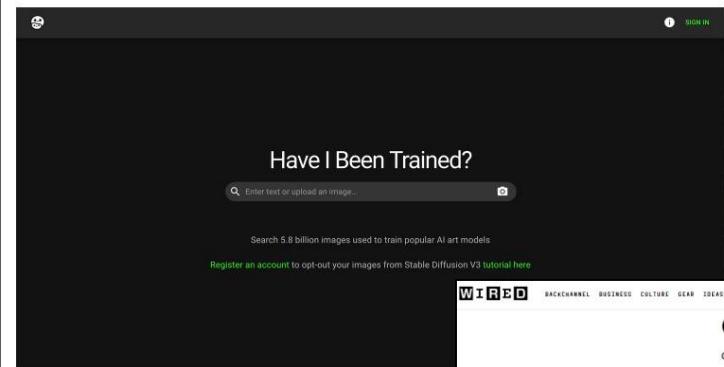
A huge and significant lawsuit is on its way to AI imagery generators, with the goal of defending artists. A class action was filed against Stability AI, Midjourney, and DeviantArt for DMCA violations,

National Laboratory

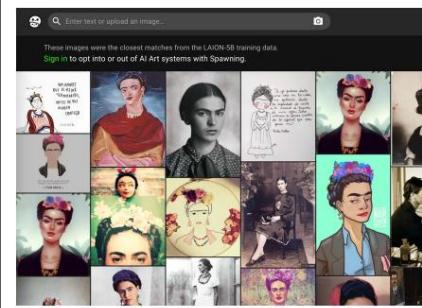
## How to Know If Your Images Were Used to Train an AI Model

Set up by a group of artists, [Spawning](#) is a collective whose aim is to help people find out whether their images are on datasets like LAION-5B, used to train AI models. Their web search engine called [Have I Been Trained?](#) lets you easily search keywords such as your artist name.

### Have I Been Trained?



Have I Been Trained works a lot like a Google image search results in the LAION-5B dataset. You have the option to see the latter is helpful if you want to see if an exact image has



## ChatGPT Stole Your Work. So What Are You Going to Do?

Creators need to pressure the courts, the market, and regulators before it's too late.



PHOTO-ILLUSTRATION: WIRED STAFF; GETTY IMAGES

IF YOU'VE EVER uploaded photos or art, written a review, "liked" content, answered a question on Reddit, contributed to open source code, or done any number of other activities online, you've done [free work](#) for tech companies, because downloading all this content from the web is how their AI systems learn about the world.



# AI Hallucinations

**LEGAL DIVE** Deep Dive Opinion Library Press Releases

Contracts Hiring/Talent Legal Operations Legal Technology Outside Counsel Compliance

# Lawyer cites fake cases generated by ChatGPT in legal brief

The high-profile incident in a federal case highlights the need for lawyers to verify the legal insights generated by AI-powered tools.

Published May 30, 2023

---

 [Lyle Moran](#)  
Reporter

[!\[\]\(57d01cea5e416b405ffc511a124cd7a8\_img.jpg\)](#) [!\[\]\(d492e315a5eb1ed097612a983d0b28f4\_img.jpg\)](#) [!\[\]\(7e82e340d2bcf7b28059798835bcd9e1\_img.jpg\)](#) [!\[\]\(fd620f3f3c3802930926e213dcd4bc65\_img.jpg\)](#) [!\[\]\(773b52c50492696d8352e19d02929bad\_img.jpg\)](#)



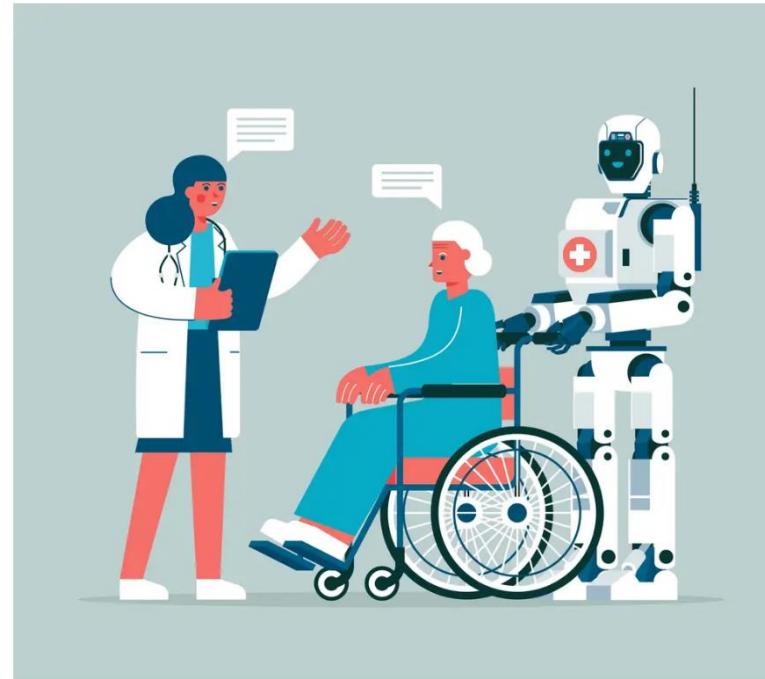
chaofann via Getty Images

# Misrepresentations

- The doctors:
  - reddit.com/r/AskDocs
- Rated “quality of information” and “empathy”
  - Accuracy/correctness not part of criteria
- Evaluators:
  - The paper authors

**Medical experts prefer ChatGPT to a real physician 78.6% of the time — because it has more time for questions**

Aaron Mok and Hilary Brueck May 2, 2023, 4:00 AM EDT



New research suggests that AI chatbots like ChatGPT can answer medical queries better than humans. sorbetto/Getty Images

- Medical experts preferred ChatGPT's answers to those of a physician 78.6% of the time, per a new study.
- Experts found the chatbot's responses to patient questions were higher quality and more empathetic.
- ChatGPT can still make grave medical errors, but this study suggests AI may improve upon a doctor's bedside manner.

# Detecting LLMs

- Discriminating language features
  - Perplexity
  - Burstiness

The image displays two side-by-side screenshots of web-based AI content detectors.

**Draft & Goal (Top):** This tool is labeled "BETA". It features a logo of a circle with dots, the text "DRAFT & GOAL", and "ChatGPT - GPT3 Content Detector V3 - Detection available in English & French". Below this is a large input area with the placeholder "See if a text comes from GPT-3 or ChatGPT." and a "Check" button. A sub-section titled "GPT Detector" includes instructions: "Not sure if a text is original or AI plagiarism? Check it here!" and "Paste in your text and click 'Check'. Example: text written by a human and ChatGPT".

**Writer (Bottom):** This tool is labeled "WRITER". It has a section titled "AI Content Detector" with a detailed description of how it works, mentioning search engine penalties for AI-generated content and its availability in the Writer application. It includes fields for "Add a URL" (with a placeholder "https://") and "Add some text" (with a placeholder "Poste text or write here"). At the bottom are "Check" and "Analyze text" buttons.

# Detecting LLMs

- Watermarks
  - Statistical sampling
  - Negligible affect output quality
  - Robust to injections (e.g. "emoji hack")
- Probably used surreptitiously in new commercial LLMs

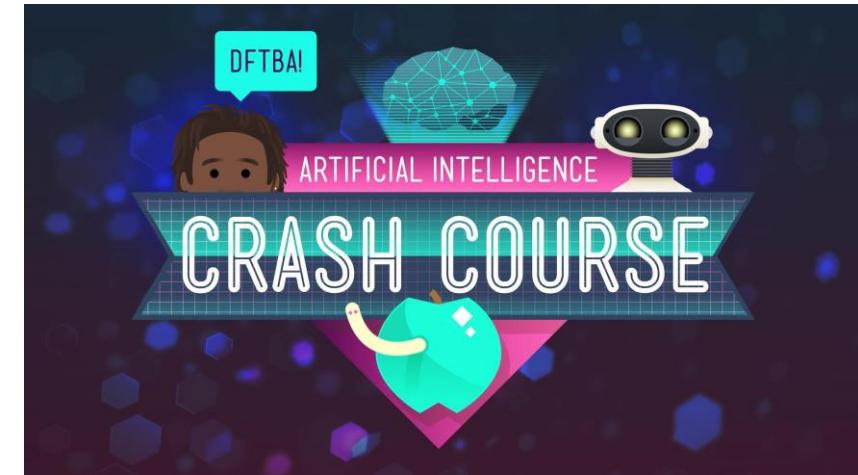
Prompt	Num tokens	Z-score	p-value
<p>...The watermark detection algorithm can be made public, enabling third parties (e.g., social media platforms) to run it themselves, or it can be kept private and run behind an API. We seek a watermark with the following properties:</p> <p><b>No watermark</b> Extremely efficient on average term lengths and word frequencies on synthetic, microamount text (as little as 25 words) Very small and low-resource key/hash (e.g., 140 bits per key is sufficient for 99.99999999% of the Synthetic Internet</p>	56	.31	.38
<p><b>With watermark</b></p> <ul style="list-style-type: none"><li>- minimal marginal probability for a detection attempt.</li><li>- Good speech frequency and energy rate reduction.</li><li>- messages indiscernible to humans.</li><li>- easy for humans to verify.</li></ul>	36	7.4	6e-14

J. Kirchenbauer et al. "A Watermark for Large Language Models" 2023

# What Comes Next

- AI and ML are increasingly useful tools for Science
- You can learn a lot more about these tools and methods yourself
  - Tons of great online resources
- Be cautious of how tools may be misused, misapplied, and misunderstood

[csoto@bnl.gov](mailto:csoto@bnl.gov)



coursera

kaggle



jupyter

co