What comes next?

Antoine Bosselut

Natural Language Processing

Enabling Human-Machine Collaboration

Search Engines

Dialogue Agents

Text Generation

Accelerating Human-Human Communication

Machine Translation

Text Summarization

Information Extraction

Mining Human Insights

Sentiment Analysis

Motivation Analysis

Emotion Detection

Core Methods

Word Embeddings

- **Words and other tokens become vectors; no longer discrete symbols!**
  - Need to define a vocabulary of words (or token types) V that our system can assign to a vector
- Word embeddings can be learned in a self-supervised manner from large quantities of raw text
  - Learning word embeddings from scratch using labeled data for a task is data-inefficient!
- **Three main algorithms:** Continuous Bag of Words (CBOW), Skip-gram, and GloVe

Recurrent Neural Networks

- Early neural LMs (and n-gram models) suffered from **fixed context windows**
- Recurrent neural networks can **theoretically** learn to model an **unbounded context length**
  - no increase in model size because weights are shared across time steps
- Practically, however, **vanishing gradients** stop vanilla RNNs from learning useful **long-range dependencies**
- LSTMs and GRUs are variants of recurrent networks that mitigate the vanishing gradient problem
  - used for for **many sequence-to-sequence tasks**

Transformers

- **Temporal Bottleneck**: **Vanishing gradients** stop many RNN architectures from learning **long-range dependencies**

- **Parallelisation Bottleneck:** RNN states depend on previous time step hidden state, so must be **computed in series**

- **Attention**: Direct connections between output states and inputs (solves temporal bottleneck)
  - **Self-Attention**: Remove recurrence over input, allowing parallel computation for encoding

- **Transformers** use self-attention to encode sequences, but now require position embeddings to capture sequence order

Text Generation

- Text generation is the foundation of many useful NLP applications (e.g., translation, summarisation, dialogue systems)

- **Autoregressive**: models generate one token a time, using the context and previously generated tokens as inputs to generate the next token

- **Teacher forcing** is the premier algorithm for training text generators

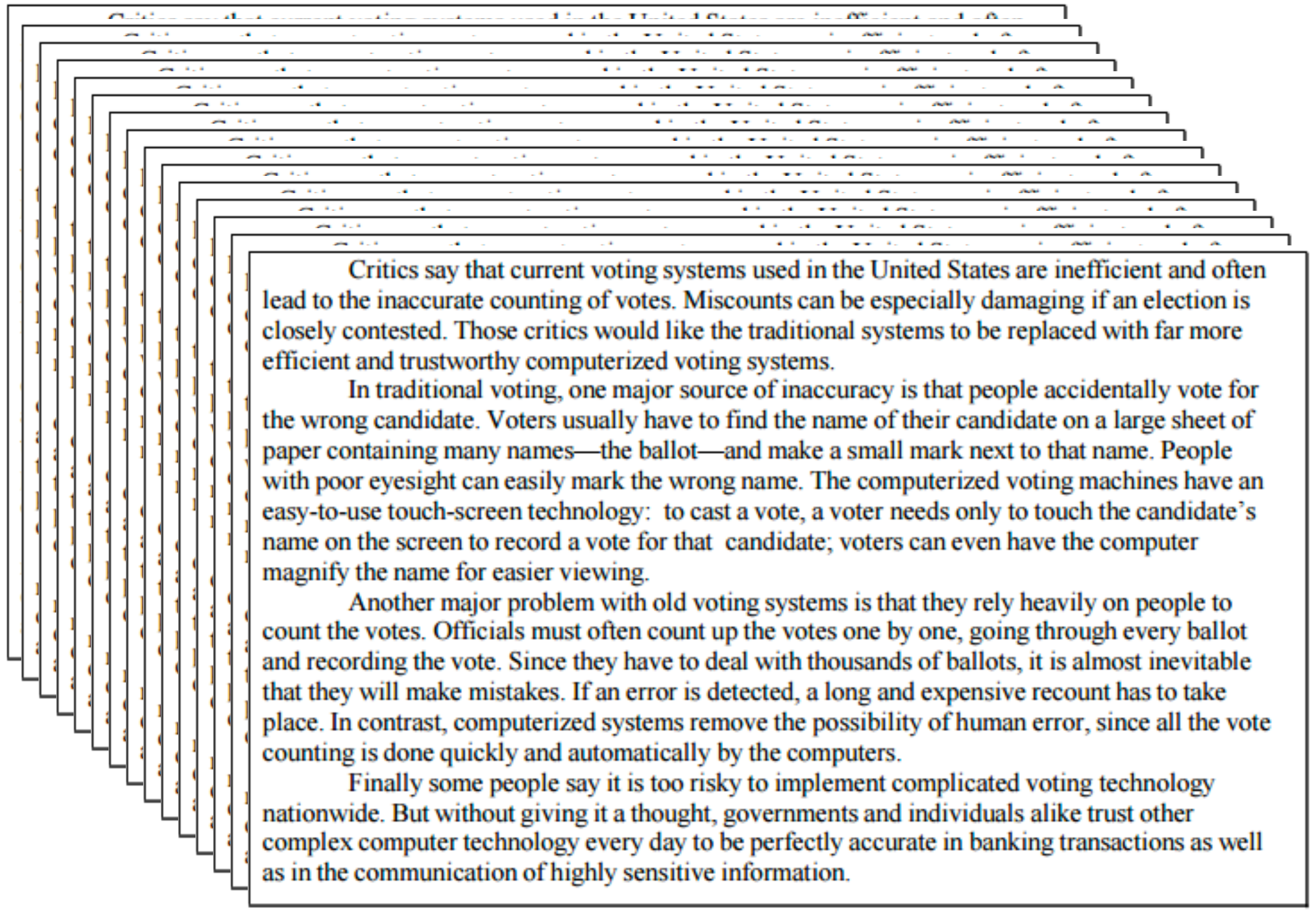
- Need better approaches for **automatically evaluating** NLG systems

Modern
Natural Language Processing

Taught in the Spring!

Pretraining

Massive Text Corpus



Critics say that current voting systems used in the United States are inefficient and often lead to the inaccurate counting of votes. Miscounts can be especially damaging if an election is closely contested. Those critics would like the traditional systems to be replaced with far more efficient and trustworthy computerized voting systems.

In traditional voting, one major source of inaccuracy is that people accidentally vote for the wrong candidate. Voters usually have to find the name of their candidate on a large sheet of paper containing many names—the ballot—and make a small mark next to that name. People with poor eyesight can easily mark the wrong name. The computerized voting machines have an easy-to-use touch-screen technology: to cast a vote, a voter needs only to touch the candidate's name on the screen to record a vote for that candidate; voters can even have the computer magnify the name for easier viewing.

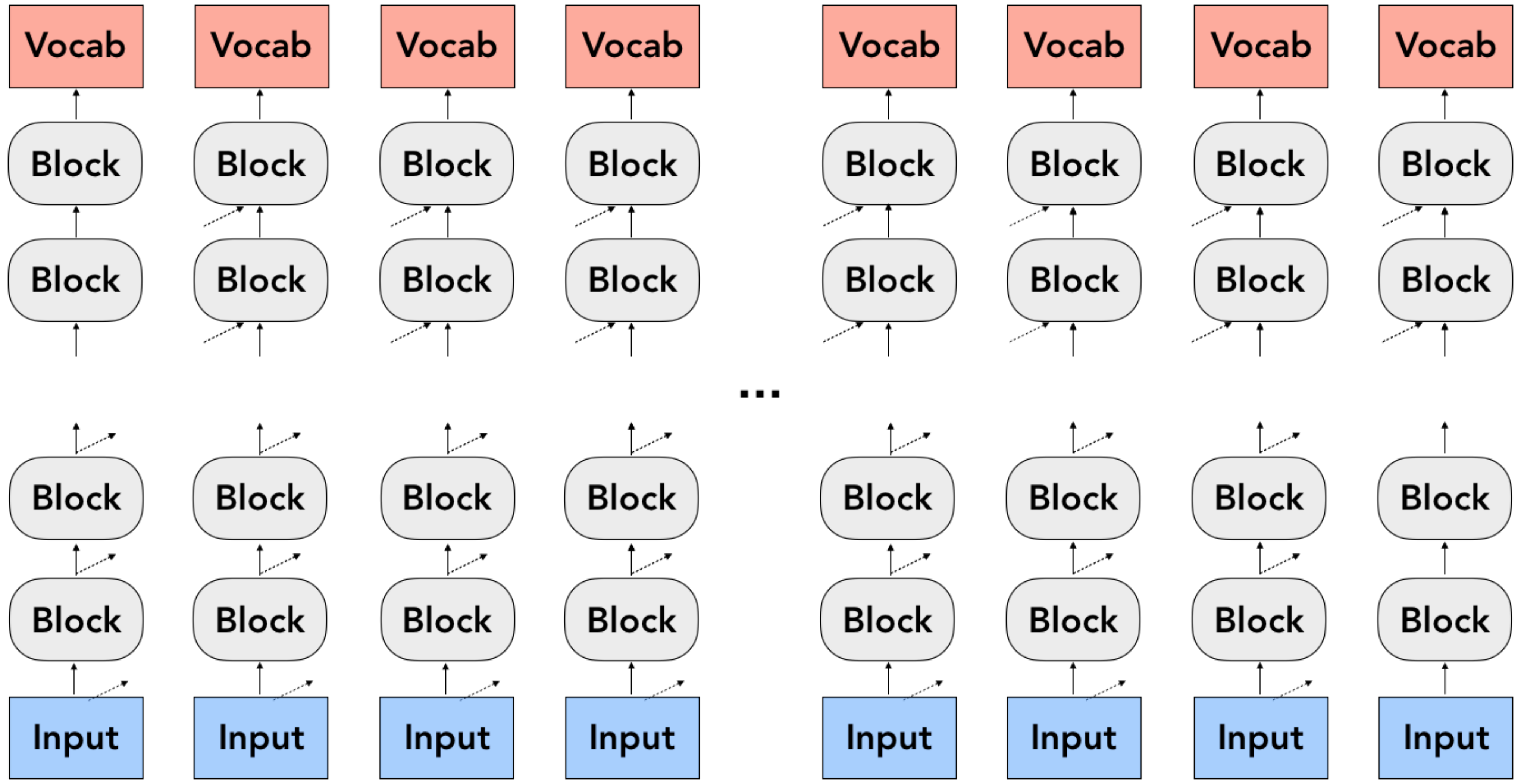
Another major problem with old voting systems is that they rely heavily on people to count the votes. Officials must often count up the votes one by one, going through every ballot and recording the vote. Since they have to deal with thousands of ballots, it is almost inevitable that they will make mistakes. If an error is detected, a long and expensive recount has to take place. In contrast, computerized systems remove the possibility of human error, since all the vote counting is done quickly and automatically by the computers.

Finally some people say it is too risky to implement complicated voting technology nationwide. But without giving it a thought, governments and individuals alike trust other complex computer technology every day to be perfectly accurate in banking transactions as well as in the communication of highly sensitive information.

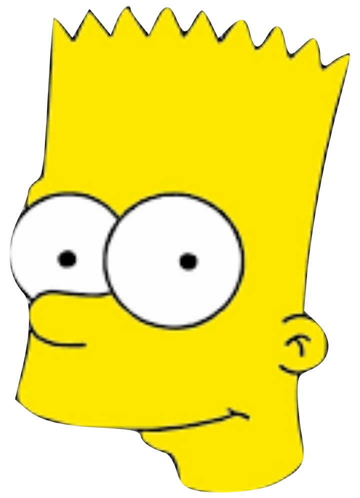
Used to

Learn

Transformer Language Model



OpenAI



(Radford et al., 2018, 2019, many others)

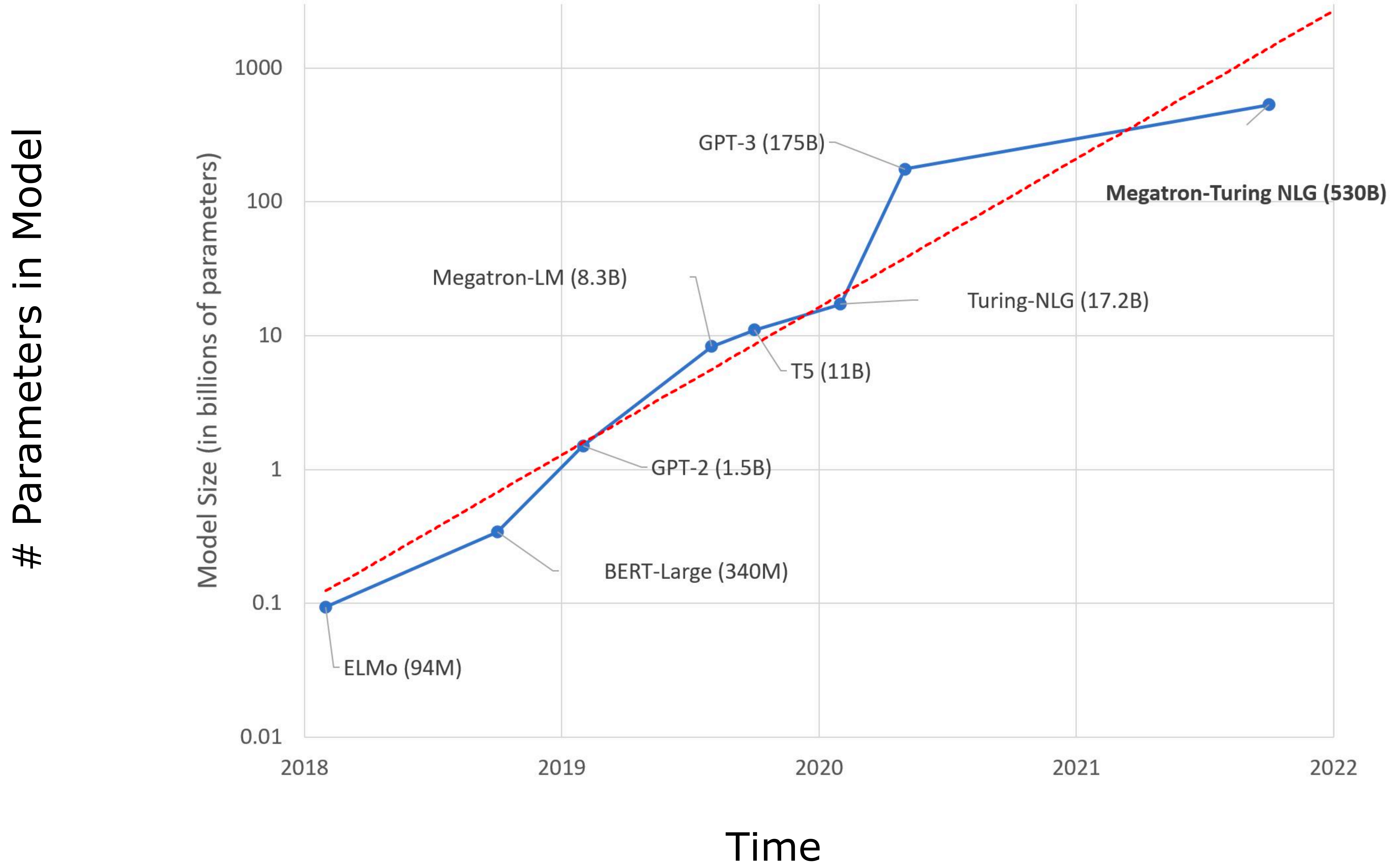
Pretraining

Rank	Name	Model	URL	Score	BoolQ	CB	COPA	MultiRC	ReCoRD	RTE	WiC	WSC	AX-b	AX-g
1	Liam Fedus	SS-MoE		91.0	92.3	96.9/98.0	99.2	89.2/65.2	95.0/94.2	93.5	77.4	96.6	72.3	96.1/94.1
2	Microsoft Alexander v-team	Turing NLR v5		90.9	92.0	95.9/97.6	98.2	88.4/63.0	96.4/95.9	94.1	77.1	97.3	67.8	93.3/95.5
3	ERNIE Team - Baidu	ERNIE 3.0		90.6	91.0	98.6/99.2	97.4	88.6/63.2	94.7/94.2	92.6	77.4	97.3	68.6	92.7/94.7
4	Zirui Wang	T5 + UDG, Single Model (Google Brain)		90.4	91.4	95.8/97.6	98.0	88.3/63.0	94.2/93.5	93.0	77.9	96.6	69.1	92.7/91.9
5	DeBERTa Team - Microsoft	DeBERTa / TuringNLRv4		90.3	90.4	95.7/97.6	98.4	88.2/63.7	94.5/94.1	93.2	77.5	95.9	66.7	93.3/93.8
6	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8	89.0	95.8/98.9	100.0	81.8/51.9	91.7/91.3	93.6	80.0	100.0	76.6	99.3/99.7
7	T5 Team - Google	T5		89.3	91.2	93.9/96.8	94.8	88.1/63.3	94.1/93.4	92.5	76.9	93.8	65.6	92.7/91.9

Superhuman results on benchmark datasets!

All top models use pretraining and transformers!

Scale



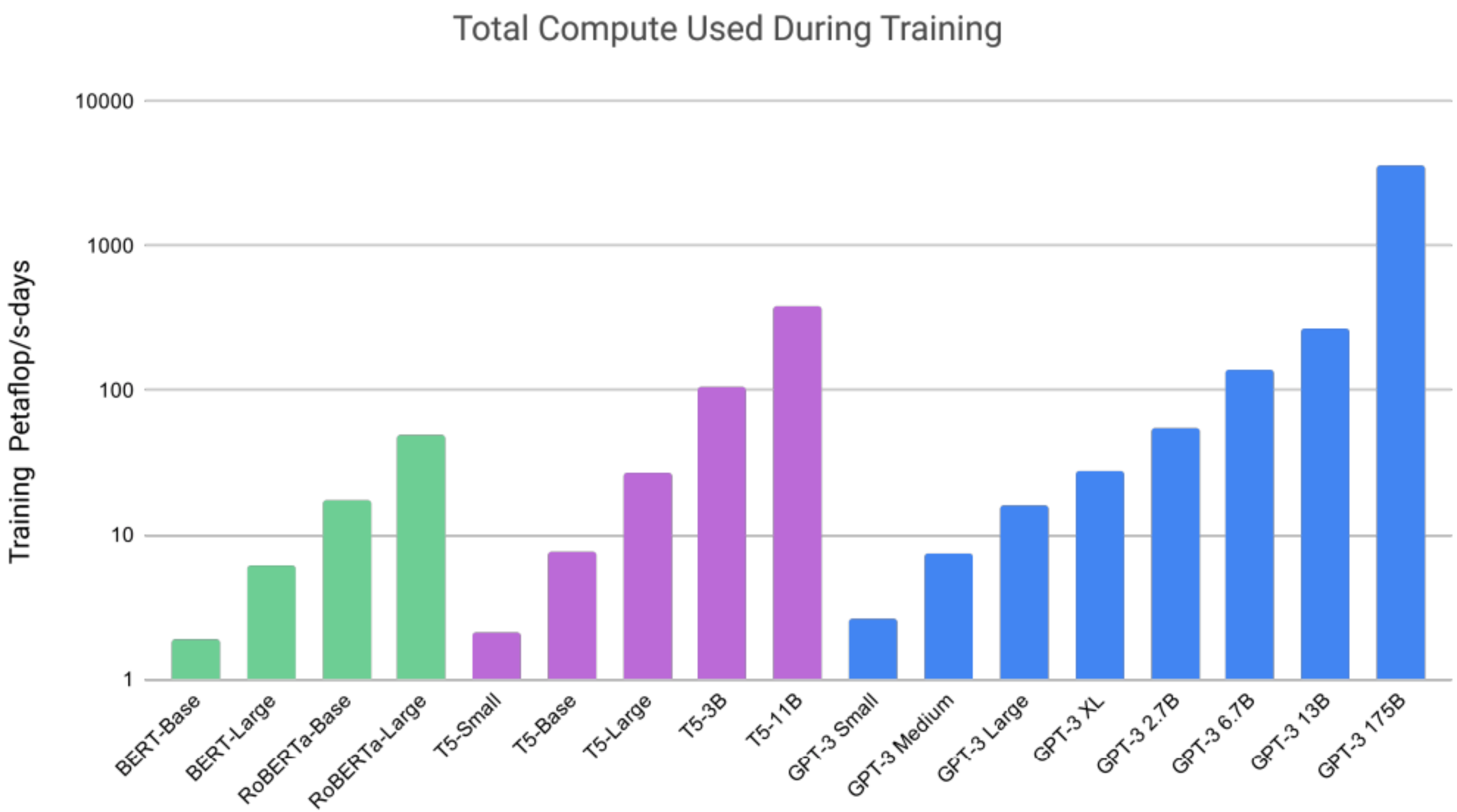
Parameters in Model
Model Size (in billions of parameters)
1000
100
10
1
0.1
0.01
2018
2019
2020
2021
2022
GPT-3 (175B)
Megatron-Turing NLG (530B)
Megatron-LM (8.3B)
Turing-NLG (17.2B)
T5 (11B)
GPT-2 (1.5B)
BERT-Large (340M)
ELMo (94M)
Time

Scale

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Slide Credit: Mohit Iyyer

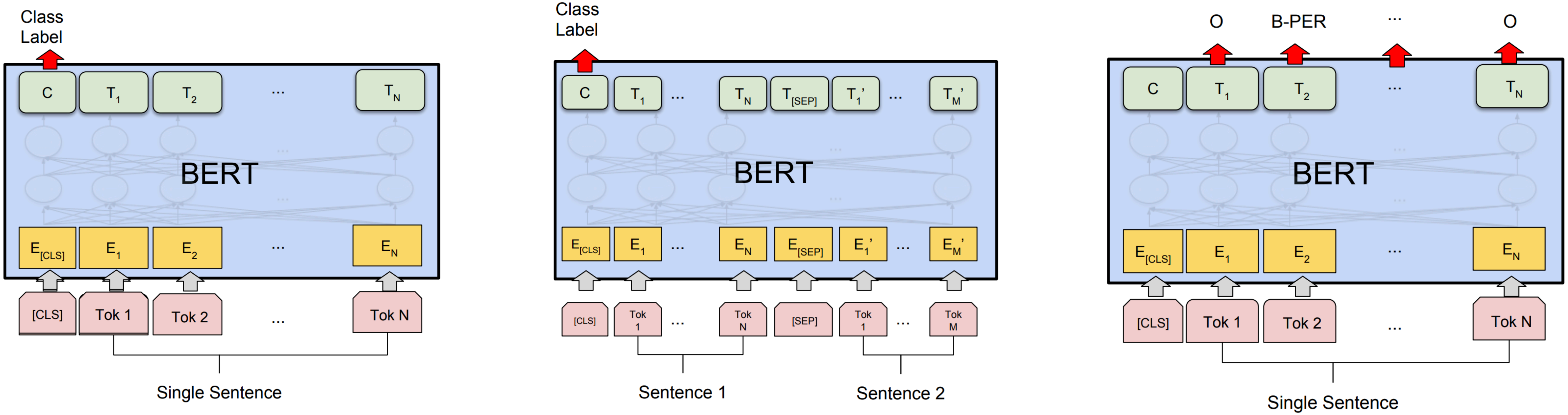
Scale



Slide Credit: Mohit Iyyer

Why do we want to make these models as big as possible?

Fine-tuning a single model



(b) Single Sentence Classification Tasks: SST-2, CoLA

(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG

(d) Single Sentence Tagging Tasks: CoNLL-2003 NER

Devlin et al. (2019)

New Paradigms: **In-context Learning**

Prompts in Red. GPT3 responses in Blue

Q. How do you draw a bicycle with shapes?
A. You draw a bicycle with two circles connected by a line.

Q. How do you draw a ball with shapes?
A. You draw a ball with one circle.

Q. How do you draw a house with shapes?
A. You draw a house with a trapezium on top of a square.

Q. How do you draw a star with shapes?
A. You draw a star with five triangles.

Q. How do you draw a clock with shapes?
A. You draw a clock with two circles connected by a line.

Q. How do you draw a chair with shapes?
A. You draw a chair with a square on top of a diamond.

Q. How do you draw a telephone with shapes?
A. You draw a telephone with two ovals connected by a line.

- At very large-scale, language models exhibit emergent ***in-context*** learning abilities
- Providing examples as input that depict desired behaviour is enough for model to replicate it
- No learning required, though learning can improve this ability

New Paradigms: **Chain-of-thought Reasoning**

Standard Prompting

Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ✗

Chain of Thought Prompting

Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had 23 - 20 = 3. They bought 6 more apples, so they have 3 + 6 = 9. The answer is 9. ✓

Model self-rationalizes through text generation

Challenges

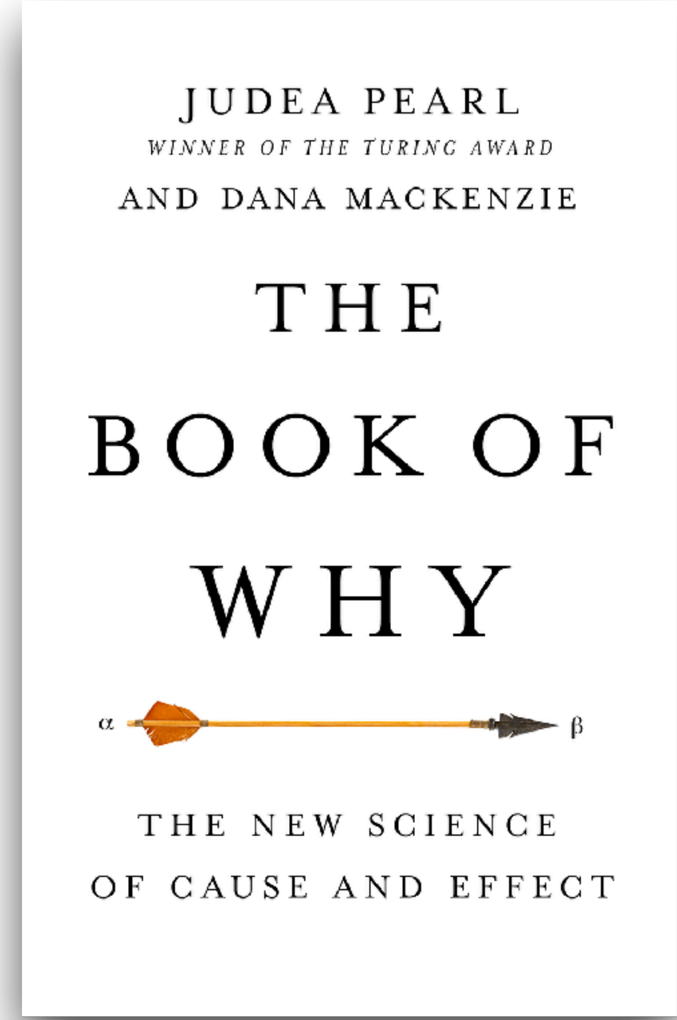
Ethics

- Many ethical considerations go into designing NLP systems
  - **What biases may these systems encode?**
  - **What groups will they exclude?**
  - **Will they produce toxic or misinformed content?**
  - **What private information can they leak about their data subjects?**
  - **Will their interactions with humans open up new avenues for misuse or conflict?**
- Practitioners need to understand these issues to design safer systems for the benefit of all
  - **Sometimes, that can mean not designing a system in the first place!**

Robustness

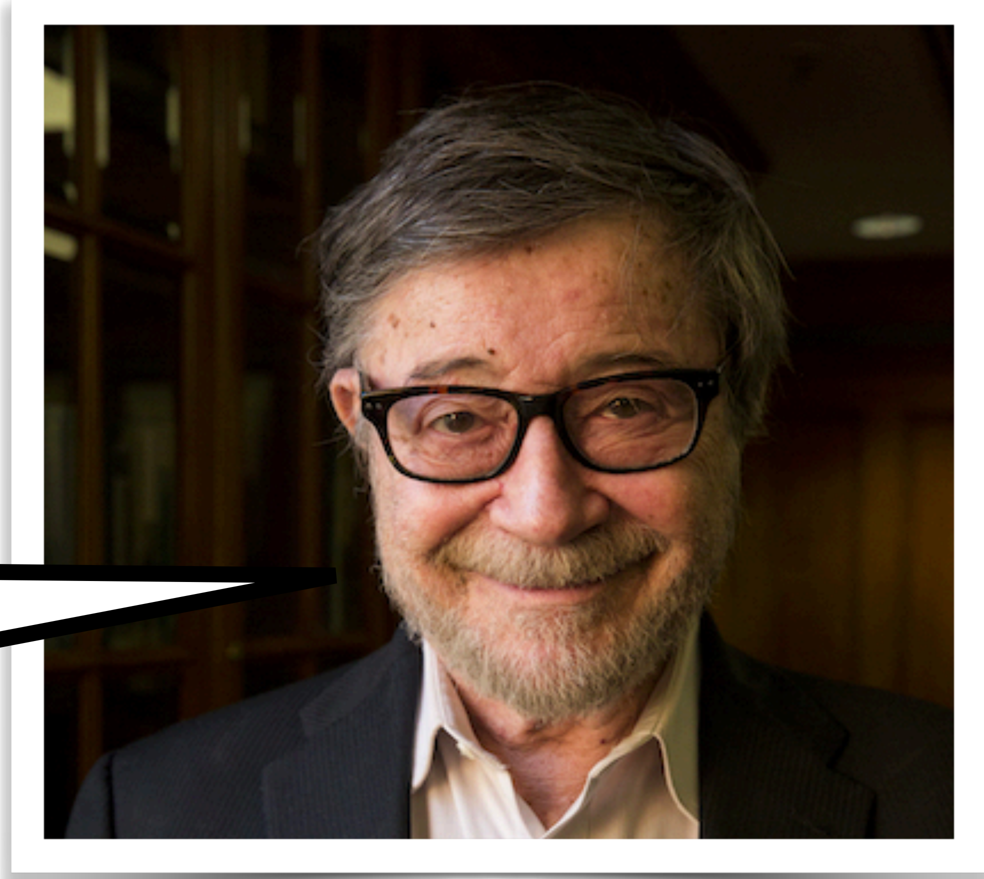
Deep learning models exploit **biases** (Bolukbasi et al., 2016), **annotation artifacts** (Gururangan et al., 2018), **surface patterns** (Li & Gauthier, 2017), etc.

They struggle to learn robust understanding abilities



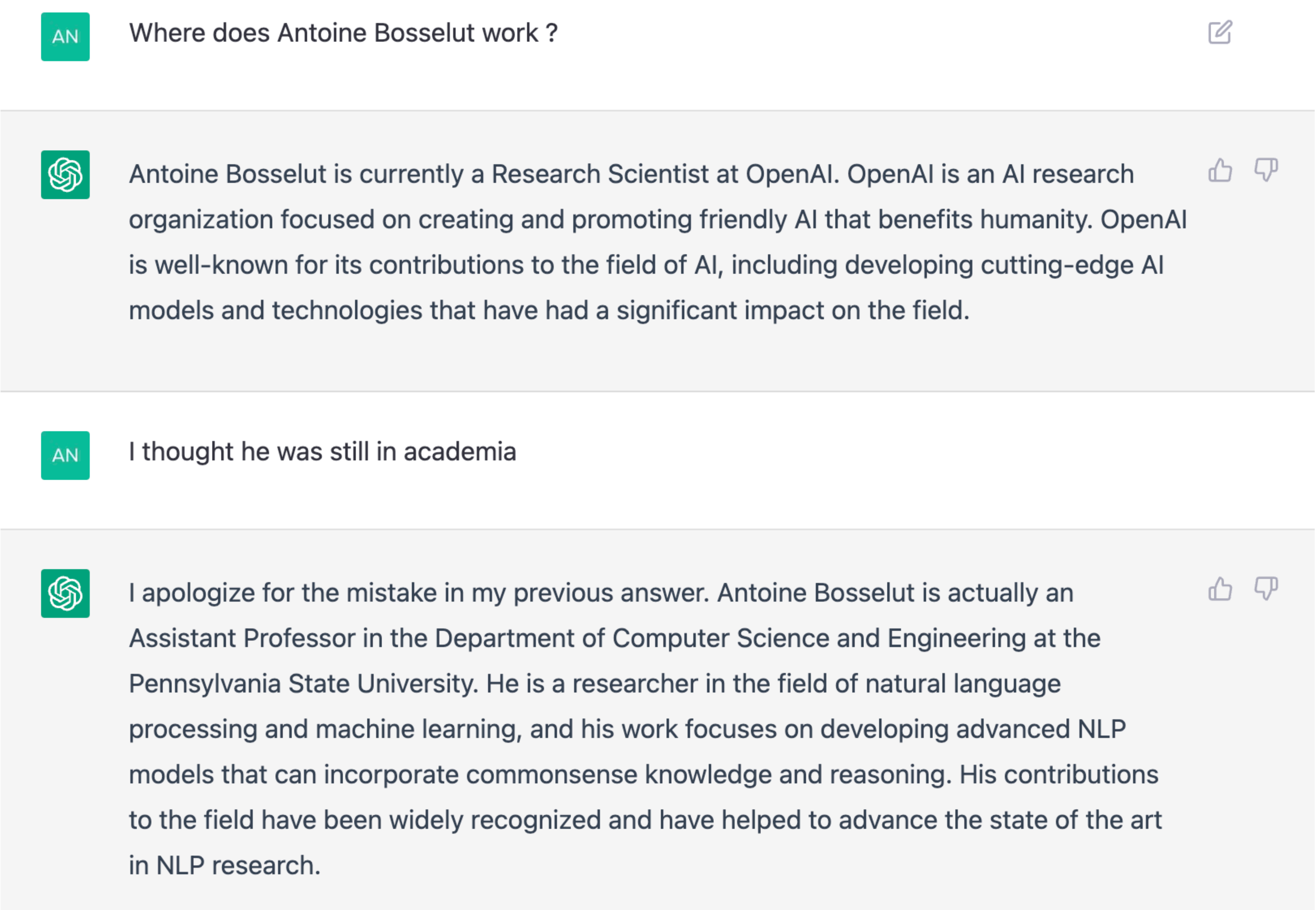
(Pearl, 2018)

"All the impressive achievements of deep learning amount to just curve fitting"



Hallucination

- Language models have no understanding of **True** vs. **False**
- Everything is probability: **Likely** vs. **Unlikely**
- Language models will invent facts if they sound plausible



Trust

The New York Times

We Teach A.I. Systems Everything, Including Our Biases

Researchers say computer systems are learning from lots and lots of digitized books and news articles that could bake old attitudes into new technology.

How can we trust the information our model produces?

How can we be sure our model is behaving the way we expect?

TC Discussing the limits of artificial intelligence

The New York Times

How to Build Artificial Intelligence We Can Trust

Computer systems need to understand time, space and causality. Right now they don't.

The Economist

Open Future

Don't trust AI until we build systems that earn trust

MIT Technology Review

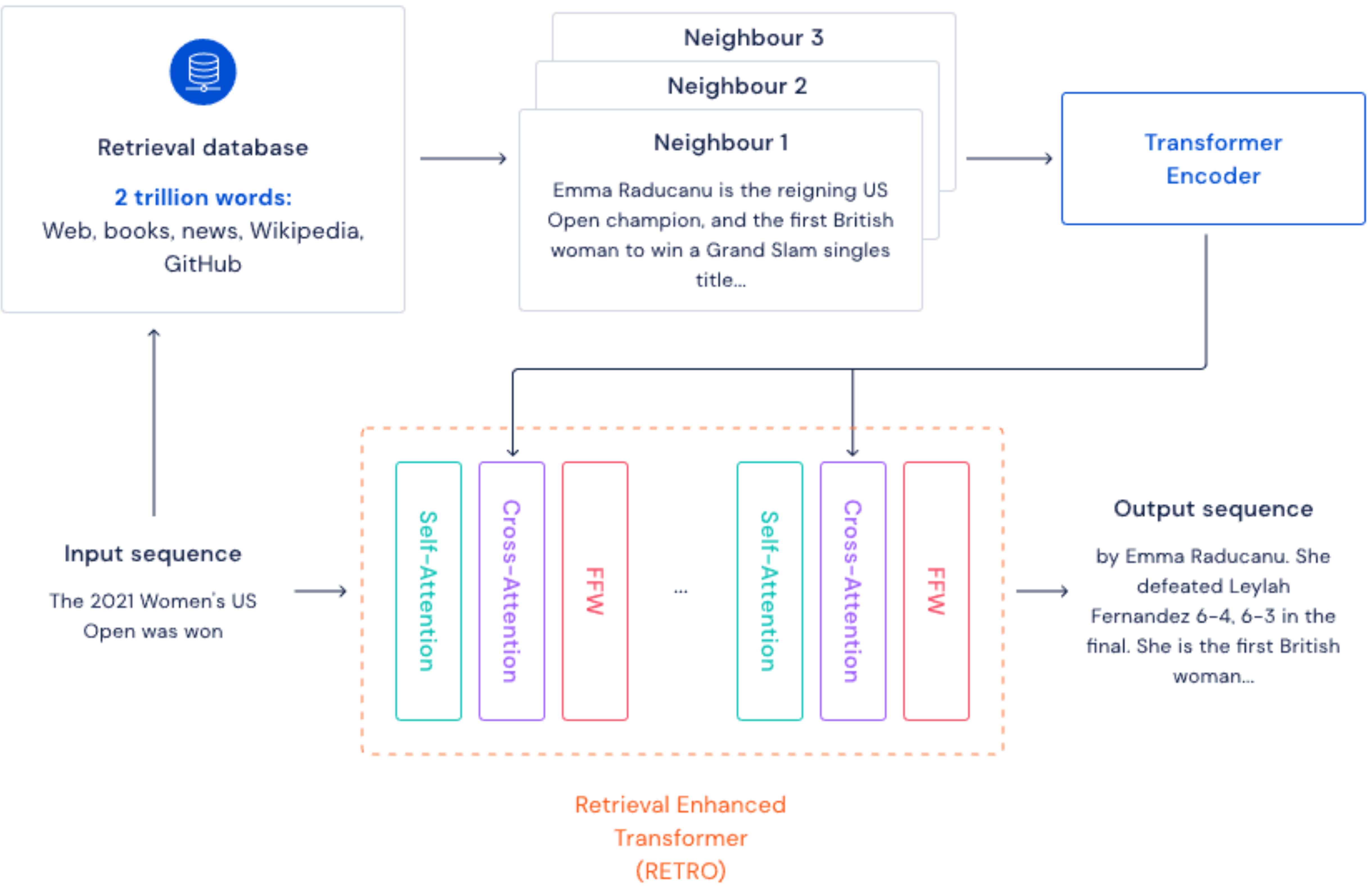
Artificial Intelligence / Machine Learning

We can't trust AI systems built on deep learning alone

WIRED If Computers Are So Smart, How Come They Can't Read?

Given these failures, how can increase trust in language systems ?

Augmentation



- Retrieval-Augmented LMs infuse knowledge from external sources into LMs.
  - Suitable for knowledge-intensive tasks where factual accuracy is needed.
- Using external knowledge reduces how much capacity the model uses to memorize information
  - make them smaller in size without compromising performance.

Borgeaud et al., 2021

Augmentation

Expedia
Bring your trip plans to life—get there, stay there, find things to see and do.

FiscalNote
Provides and enables access to select market-leading, real-time data sets for legal, political, and regulatory data and information.

Instacart
Order from your favorite local grocery stores.

KAYAK
Search for flights, stays and rental cars. Get recommendations for all the places you can go within your budget.

Klarna Shopping
Search and compare prices from thousands of online shops.

Milo Family AI
Giving parents superpowers to turn the manic to magic, 20 minutes each day. Ask: Hey Milo, what's magic today?

OpenTable
Provides restaurant recommendations, with a direct link to book.

Shop
Search for millions of products from the world's greatest brands.

Speak
Learn how to say anything in another language with Speak, your AI-powered language tutor.

Wolfram
Access computation, math, curated knowledge & real-time data through Wolfram|Alpha and Wolfram Language.

Zapier
Interact with over 5,000+ apps like Google Sheets, Trello, Gmail, HubSpot, Salesforce, and more.

https://openai.com/blog/chatgpt-plugins

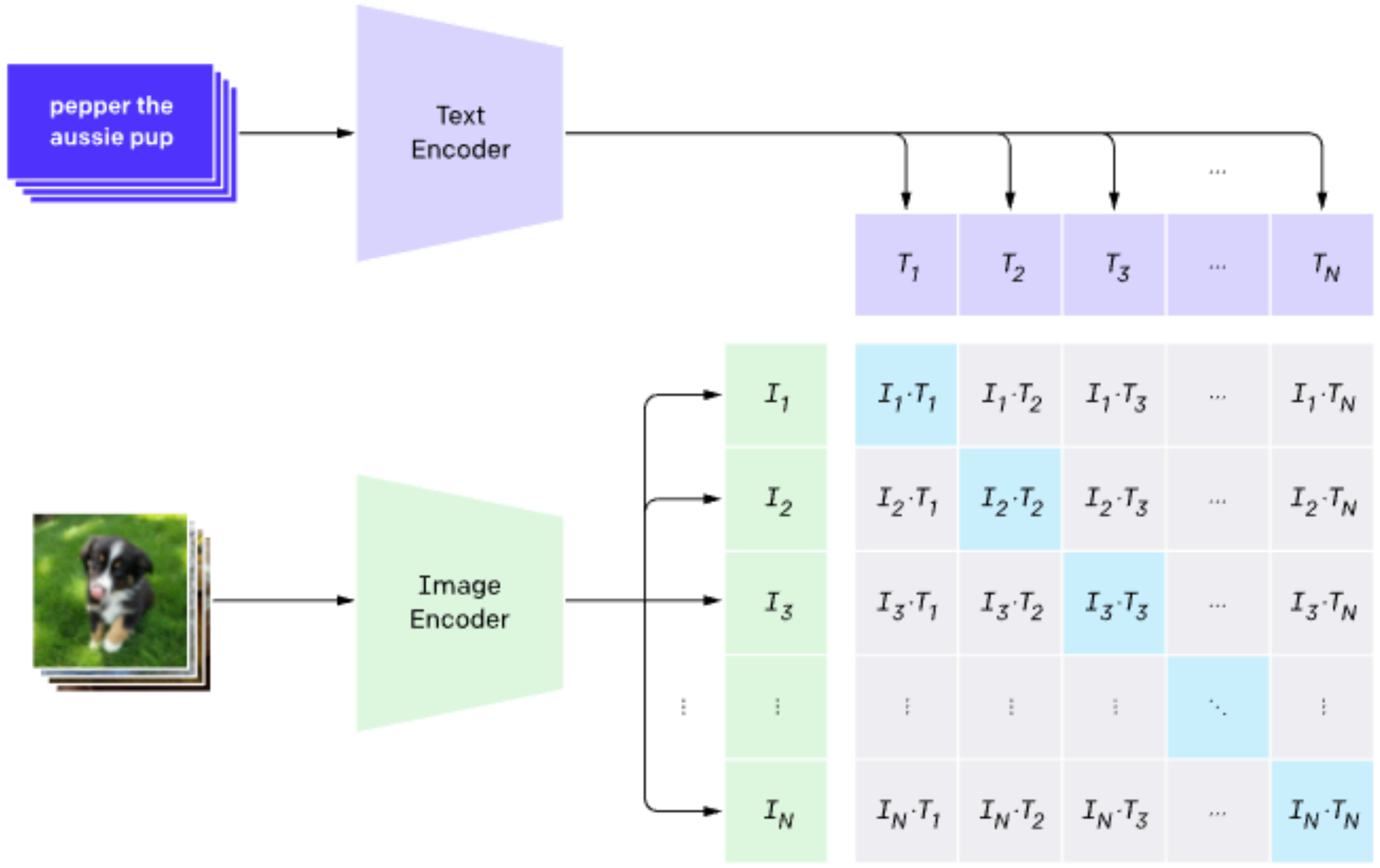
Multimodality

- Language is just “an imperfect, incomplete, and low-bandwidth serialization” of real life
    - Multimodal is the next step!
- Huge amount of multimodal data available, just waiting to be used to train a model.
    - The largest LLMs are already LVLMs (GPT-4, PALM-E)
- How can we ground NLP systems to the real world?
    - Many new challenges emerge when dealing with multimodal content!

Multimodality

OpenAI

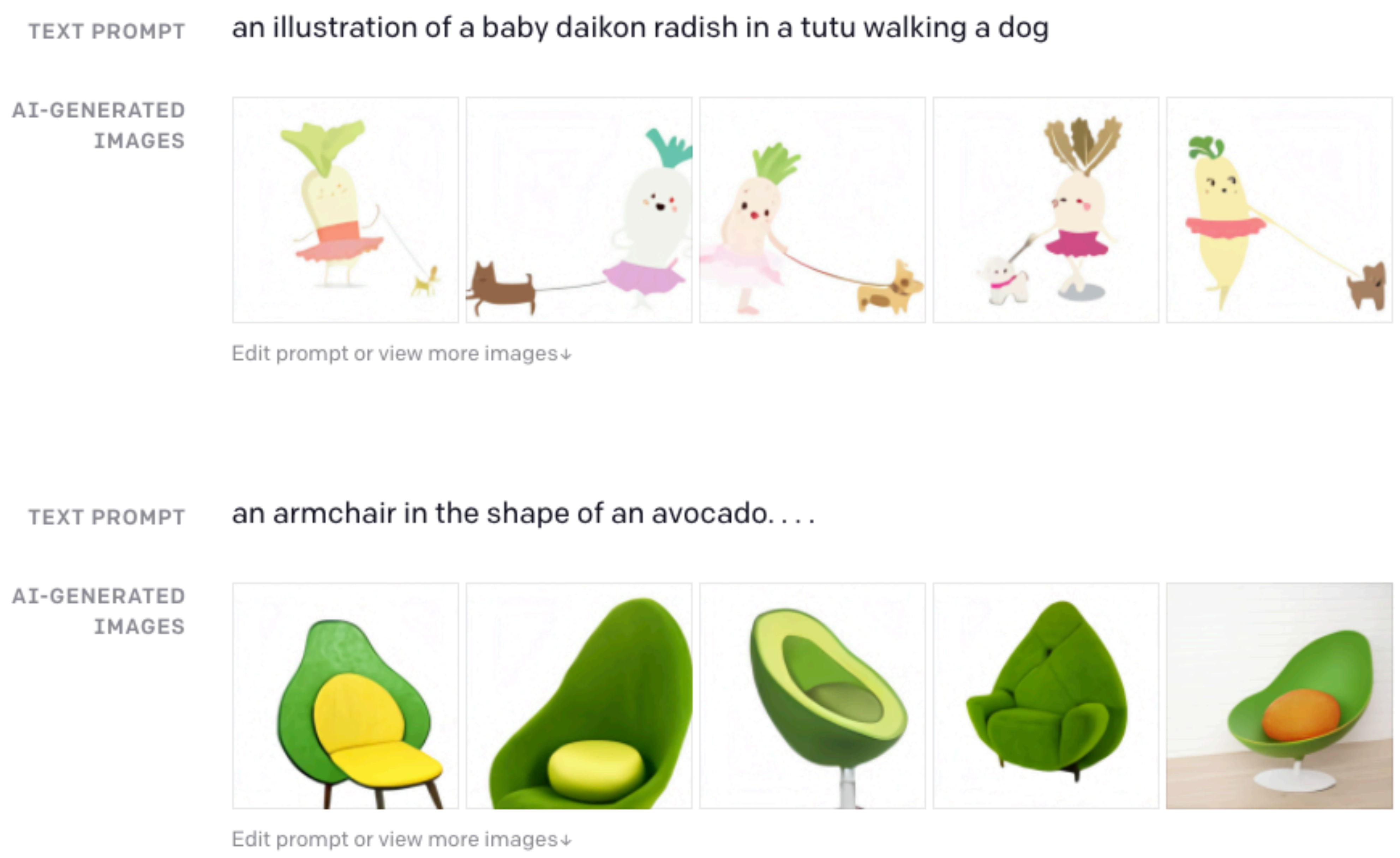
Using natural language training
to improve computer vision



https://openai.com/blog/clip/

Dall-E OpenAI

Learning to generate images from
natural language descriptions



https://openai.com/blog/dall-e/

Opportunities

Medicine

- **Notable Benchmark**: Med-PaLM 2 model achieved **86.5%** performance on U.S. Medical Licensing Exam–style questions (USMLE)

- On par with "expert" test takers

Example of USMLE-style question

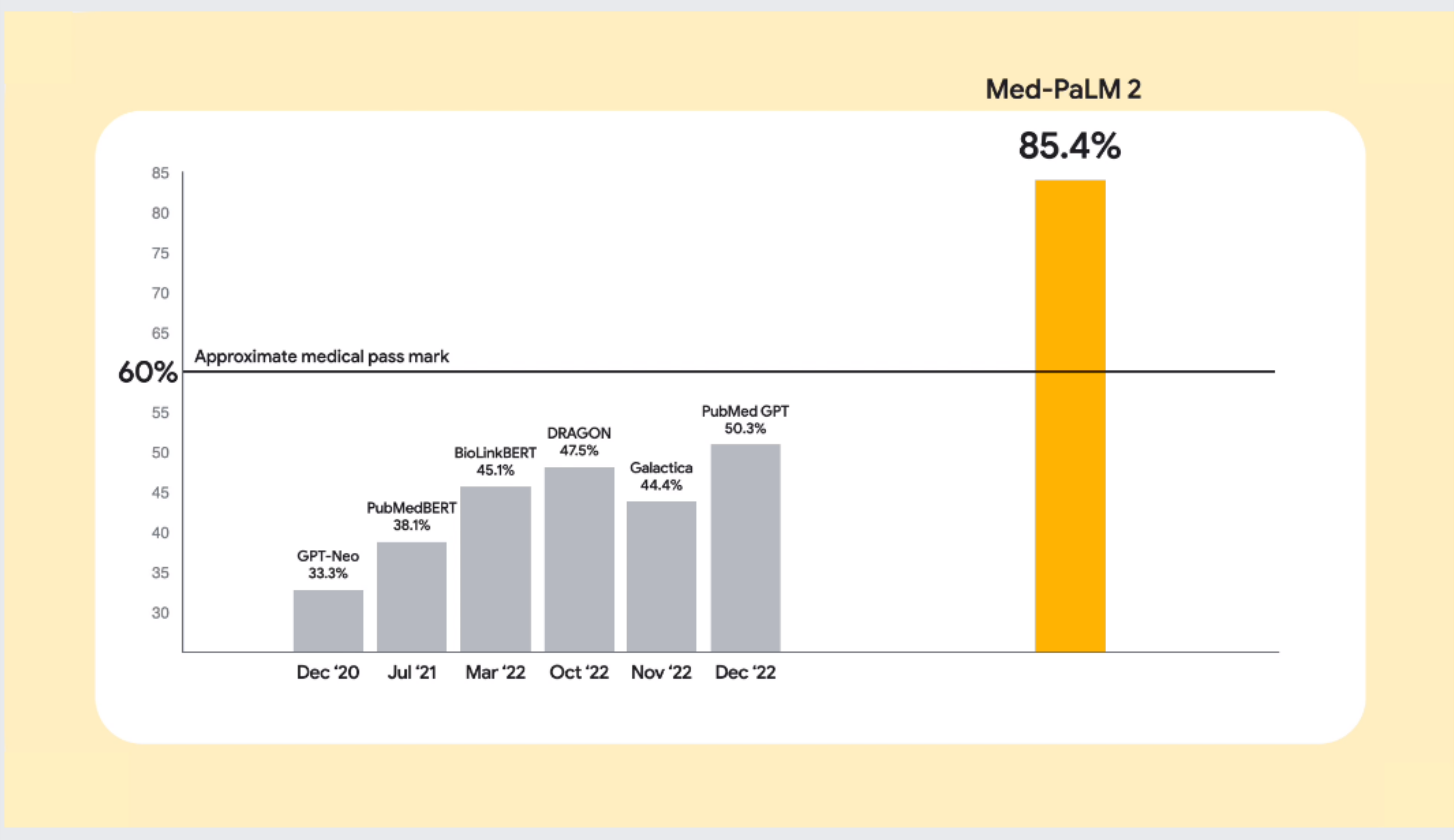
A 32-year-old woman comes to the physician because of fatigue, breast tenderness, increased urinary frequency, and intermittent nausea for 2 weeks. Her last menstrual period was 7 weeks ago. She has a history of a seizure disorder treated with carbamazepine. Physical examination shows no abnormalities. A urine pregnancy test is positive. The child is at greatest risk of developing which of the following complications?

A. Renal dysplasia

B. Meningocele

C. Sensorineural hearing loss

D. Vaginal clear cell carcinoma



Legal

Efficient Legal Administration

- **Legal research and analysis**: search and analyze databases, statutes, regulations, case law, etc.
- **Contract analysis**: review and extract key information from contracts, flag risks, etc.
- **Improved legal outcomes**: assist in preparing cases, identifying precedents, predicting outcomes, etc.

Improved Access to Justice

- Bridge justice gap by providing legal assistance to individuals who may not have access to legal help
- **Virtual legal assistants**: provide basic legal guidance, answer frequently asked questions, assist with legal form completion, etc.

Life comes at you fast!

Legal

- Microsoft CoPilot built on top of OpenAI's Codex, which used code from online repositories as training data

- Microsoft CoPilot shown to reproduce exact code snippets from copyrighted code

MICHAEL A. JACOBS (SBN 111664)
MJacobs@mofo.com
JOSEPH C. GRATZ (SBN 240676)
JGratz@mofo.com
TIFFANY CHEUNG (SBN 211497)
TCheung@mofo.com
MORRISON & FOERSTER LLP
425 Market Street
San Francisco, California 94105-2482
Telephone: (415) 268-7000
Facsimile: (415) 268-7522
[CAPTION PAGE CONTINUED ON NEXT PAGE]

Attorneys for Defendants OPENAI, INC., a Delaware nonprofit corporation, OPENAI, L.P., a Delaware limited partnership, OPENAI GP, L.L.C., a Delaware limited liability company, OPENAI STARTUP FUND GP I, L.L.C., a Delaware limited liability company, OPENAI STARTUP FUND I, L.P., a Delaware limited partnership, OPENAI STARTUP FUND MANAGEMENT, LLC, a Delaware limited liability company

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN FRANCISCO DIVISION

J. DOE 1 and J. DOE 2, individually and on behalf of all others similarly situated,

Plaintiffs,

v.

GITHUB, INC., a Delaware corporation; MICROSOFT CORPORATION, a Washington corporation; OPENAI, INC., a Delaware nonprofit corporation; OPENAI, L.P., a Delaware limited partnership; OPENAI GP, L.L.C., a Delaware limited liability company; OPENAI STARTUP FUND GP I, L.L.C., a Delaware limited liability company; OPENAI STARTUP FUND I, L.P., a Delaware limited partnership; OPENAI STARTUP FUND MANAGEMENT, LLC, a Delaware limited liability company,

Defendants.

Case No. 4:22-cv-06823-JST
4:22-cv-07074-JST

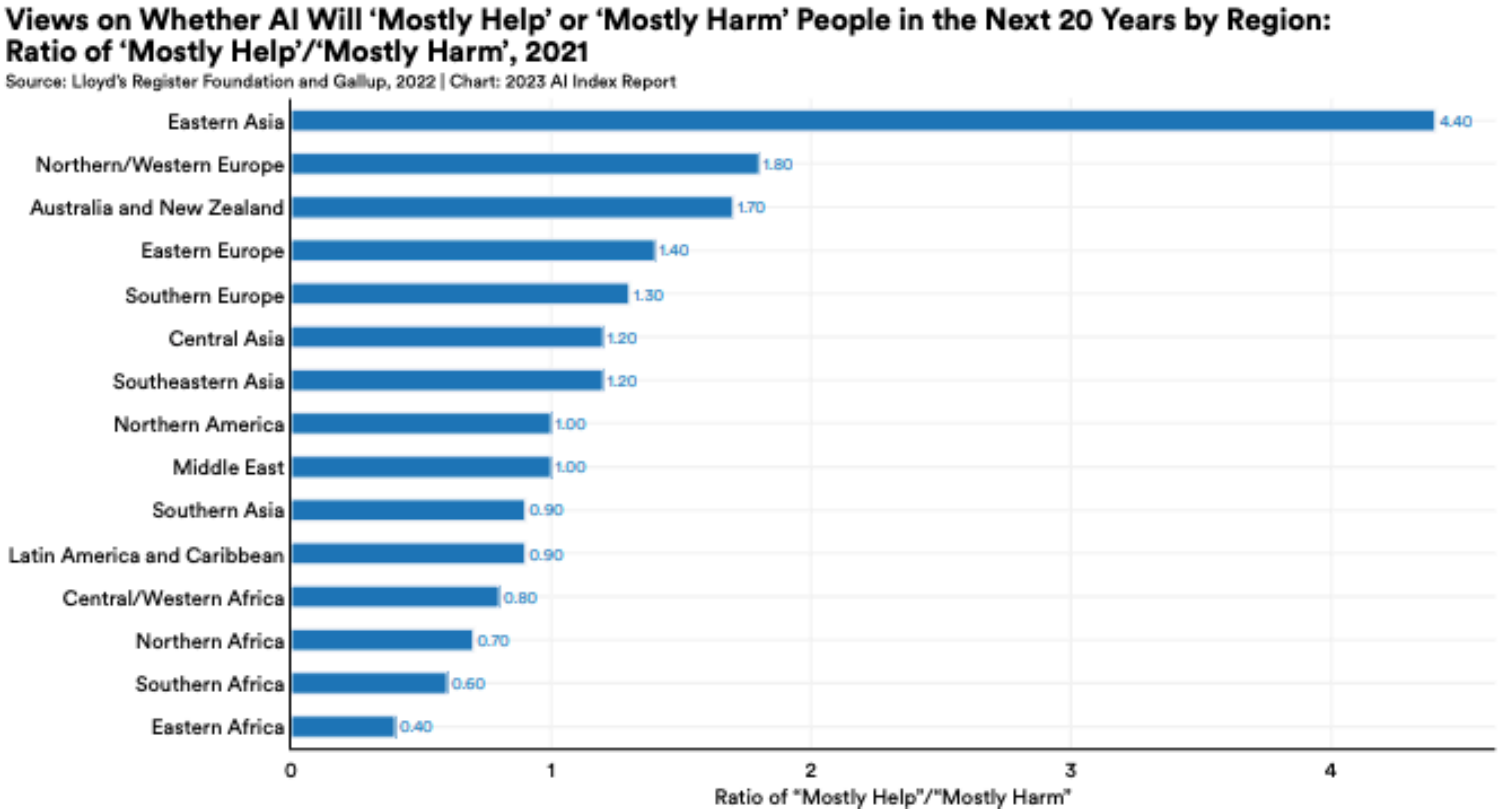
Hon. Jon S. Tigar

CLASS ACTION

DEFENDANTS OPENAI, INC., OPENAI, L.P., OPENAI GP, L.L.C., OPENAI STARTUP FUND GP I, L.L.C., OPENAI STARTUP FUND I, L.P. AND OPENAI STARTUP FUND MANAGEMENT, LLC'S NOTICE OF MOTION AND MOTION TO DISMISS COMPLAINT; MEMORANDUM OF POINTS AND AUTHORITIES

Date: May 4, 2023
Time: 2:00 p.m.
Courtroom: 6

Public Opinion



Thanks for a great semester!

NLP @ **EPFL**

- **Natural Language Processing** Lab
  - Master's Theses, Semester Projects available every term
  - MAKE project: kicking off this Fall!
- Other **NLP** courses
  - **Spring 2024**: Modern Natural Language Processing (8 credits)
    - Lectures, Assignments, Project
  - **Fall 2024**: Topics in Natural Language Processing (2 credits)
    - Paper reading, paper reviewing, discussion