

CSCI 5541: Natural Language Processing

Lecture 2: Introduction to NLP

Dongyeop Kang, University of Minnesota





UNIVERSITY OF MINNESOTA
Driven to Discover®


Announcement





□ Tutorial Recording and In-class Tutorials (next week)


- Computing basics (Sep 4, Drew)
 - ✓ Setting up environment for PyTorch and Transformers
- Tutorial on PyTorch and SciKit-learn Basics (Sep 11, Shirley)
- Tutorial on HuggingFace/vLLM (Sep 11, Shuyu)





Sep 4 Intro to NLP 
Tutorial on Notebook & Colab (Drew)   

Sep 9 Text Classification 
[HW2 out](#) 

Sep 11 Text Classification: Practices
Tutorial on Scikit-Learn and PyTorch (Shirley) 

- Scikit-Learn  
- Pytorch  

Tutorial on Finetuning & vLLM (Shuyu) 

- Huggingface  
- vLLM  










HW1 due (Sep 11 Thursday)



Announcement

- ❑ Deadline for course drop (tonight, Sep 4)
 - We currently have 9 students on original waitlist and additional 9 students
- ❑ Tomorrow morning, you will be invited to
 - class Slack,
 - MSI, and
 - CoLab Pro.

CSCI 5541 - Natural Language Processing

Component	Time	Location	Class List	Grades
Lecture (001)	T,Th 11:15 AM - 12:30 PM	 <u>Akerman Hall 319</u>  Twin Cities	 <u>62/64</u>  <u>0/32</u>	
Lecture (883)	T,Th 11:15 AM - 12:30 PM	 <u>Twin Cities Remote</u>  Twin Cities	 <u>0/0</u>	



Announcement

- ❑ If you miss the first class, please check out the course details in the lecture slides
- ❑ Share your interests and project ideas in **#random** channel and actively look for your teammates. Team formation is due on Sep 18.
- ❑ If you are enrolled but not invited to Slack until this Friday, please send me an email.
- ❑ HW1 out today (**Due: Sep 11 Thursday**) with an updated rubric
 - ❑ https://dykang.github.io/classes/csci5541/F25/hw/csci5541f25_hw1.pdf
- ❑ OH out on course website



Outline

- ❑ What is NLP?
- ❑ Does ChatGPT solve every NLP problem?
- ❑ Language consists of many levels of structure
- ❑ What makes language so difficult to process?
- ❑ How to process language?
- ❑ Recent Developments (2019-2024)
- ❑ Limits of LLMs and the Financial Incentives of GenAI

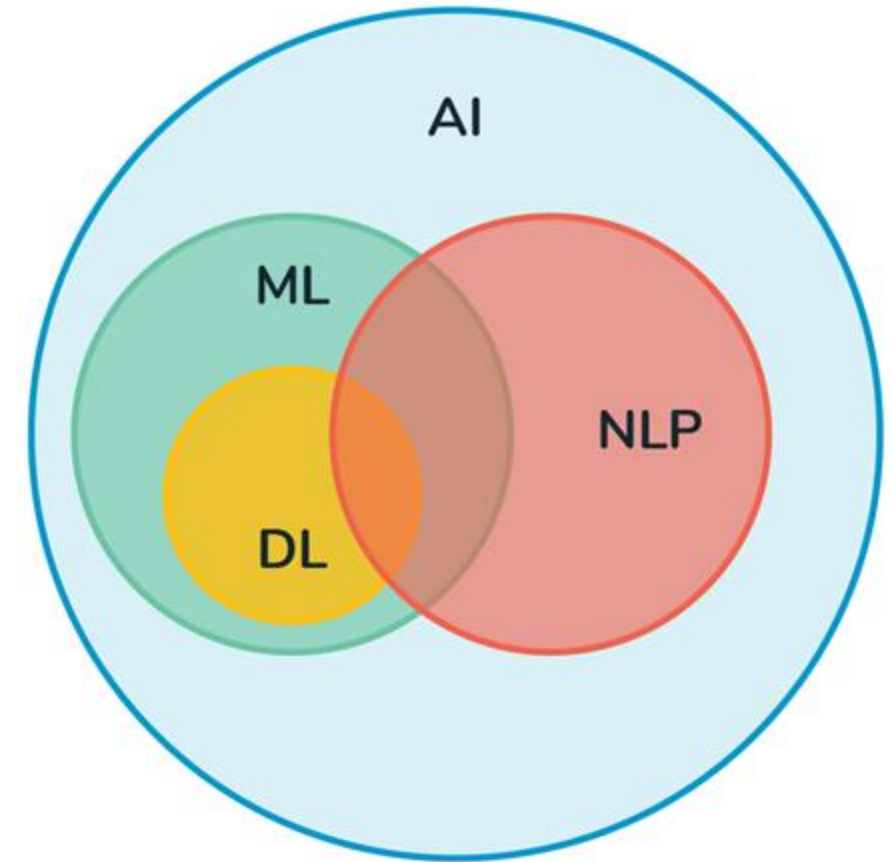


NLP is interdisciplinary

- ❑ Linguistics
- ❑ Artificial Intelligence
- ❑ Machine Learning (2000-present)

Recently,

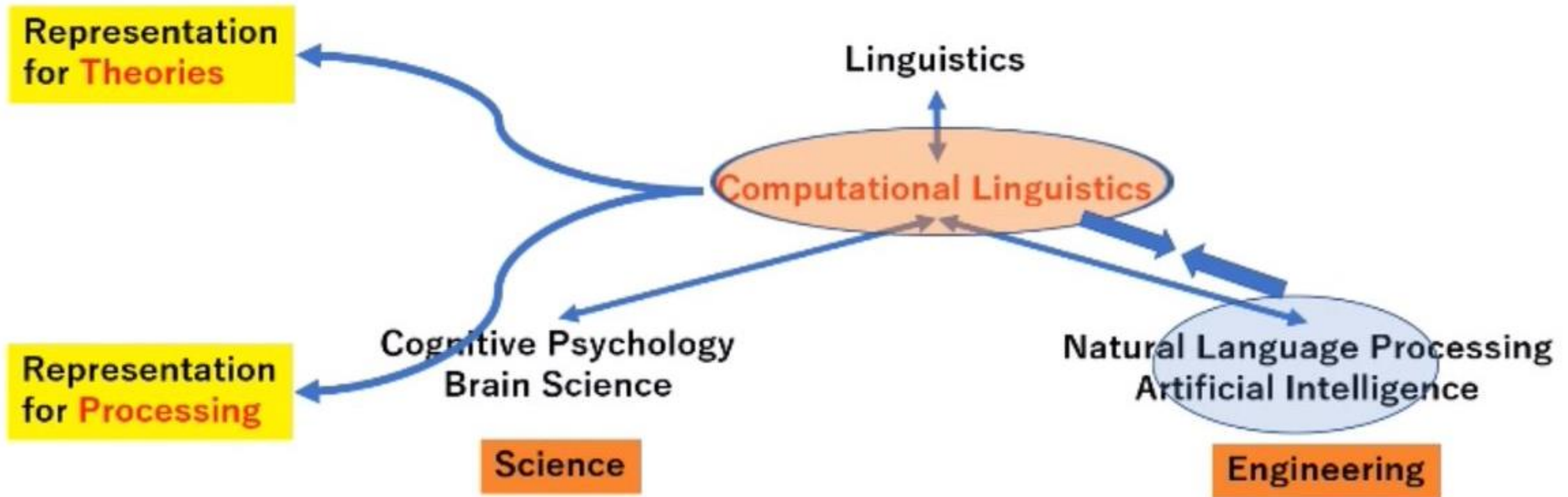
- ❑ Social Science and Humanities
- ❑ Human-computer Interaction
- ❑ Education
- ❑ Robotics
- ❑ Cognitive Science / Brain Science / Neuroscience
- ❑ Psychology
- ❑ Law / Medical / Biology
- ❑ ..



NLP vs (Computational) Linguistics

- ❑ **Linguistics** involve the nature of *linguistic representations and linguistic knowledge*, and how linguistic knowledge is acquired and deployed in comprehension of language.
- ❑ **Computational linguistics** asks *what humans are computing and how*, by *mathematically defining* classes of linguistic representations and *formal grammars* to capture the range of phenomena in human languages.
- ❑ **NLP** is the art of *solving engineering problems* that need to analyze (or generate) natural language text. The metric is whether you got good solutions on the engineering problem. After all, their goal is not a full theory but rather the simplest, most efficient approach that will get the job done.





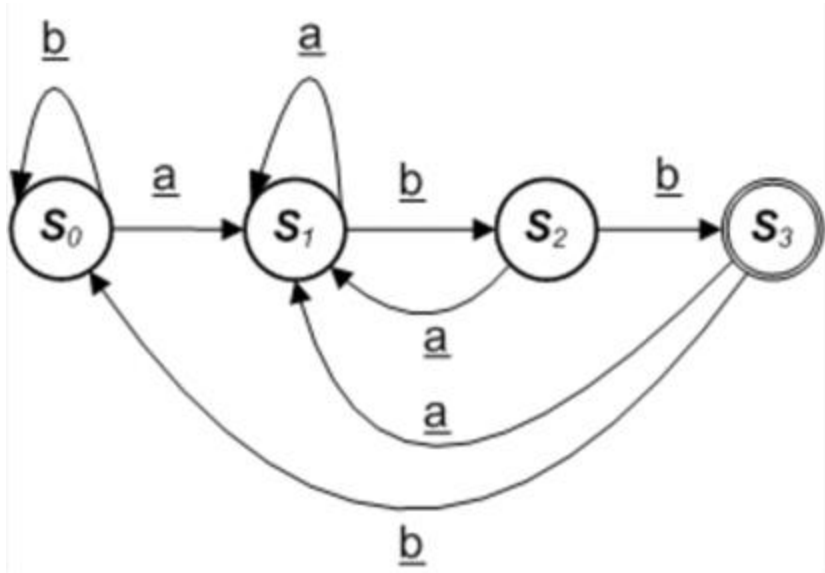
<https://twitter.com/radamihalcea/status/1422892875218628616>



Linguistic Theories



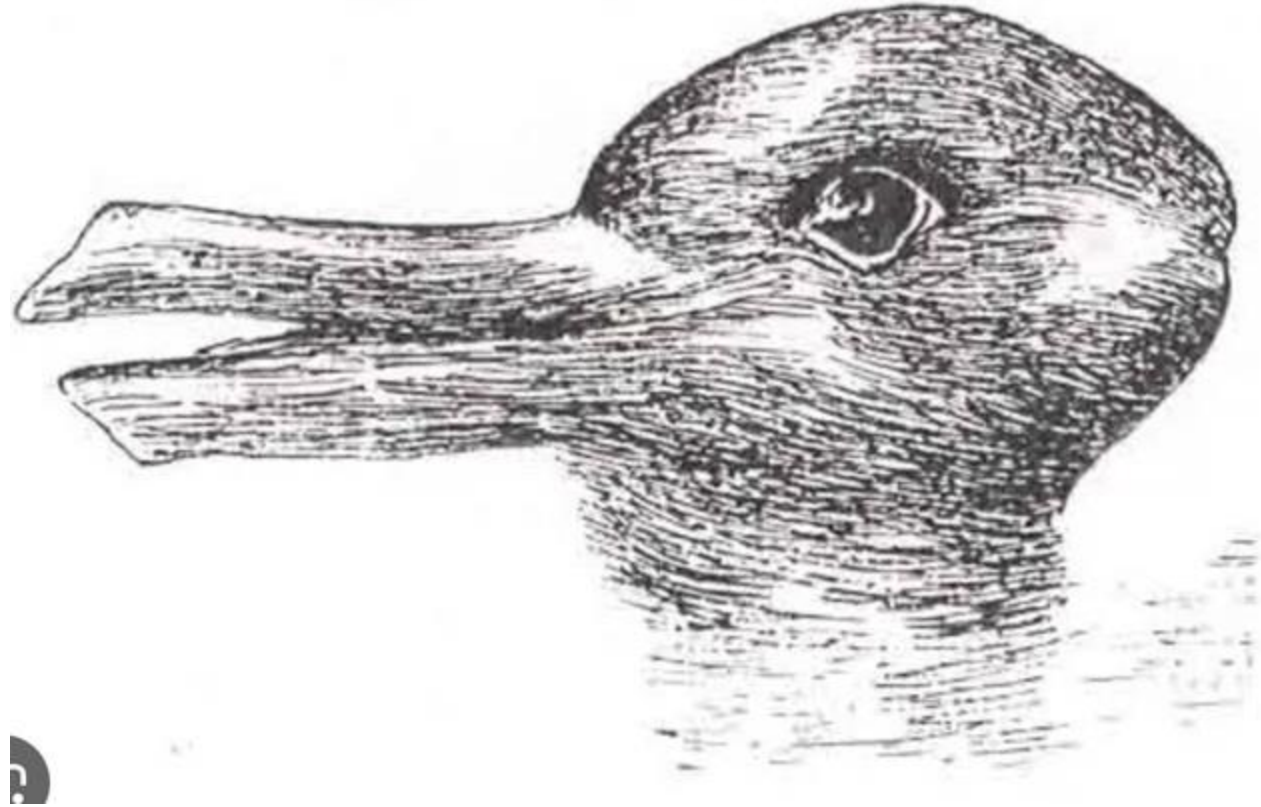
Language as Formal Logic



$$\begin{aligned} S &\rightarrow aS \mid bX \\ X &\rightarrow aX \mid bY \\ Y &\rightarrow aY \mid bZ \mid \Lambda \\ Z &\rightarrow aZ \mid \Lambda \end{aligned}$$



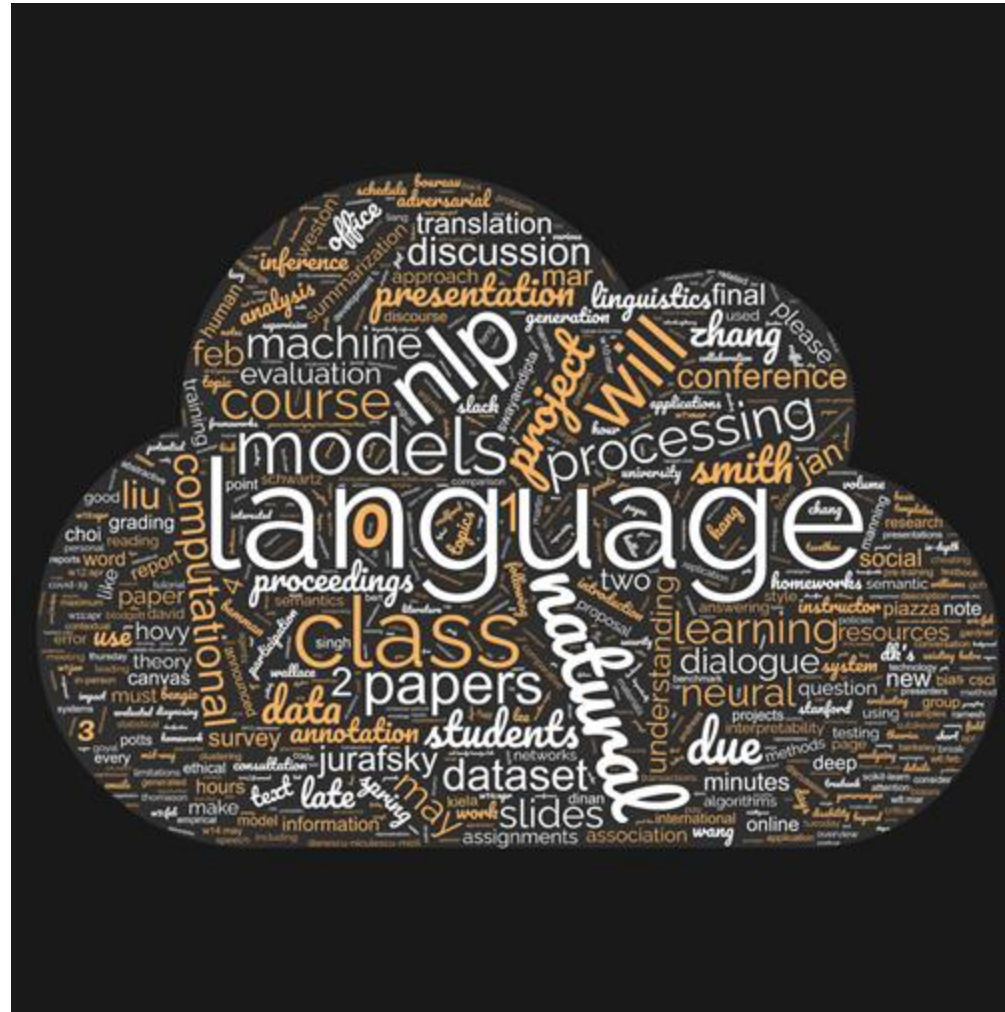
Language as Perspective and Social Interaction



NLP = Processing language
with computers



Processing as sorting and clouding



Word cloud generated with text on our class homepage using www.wordclouds.com

Processing as understanding sentiment

Reviews

Summary - Based on 1,668 reviews



What people are saying

ease of use	<div><div></div></div>	"Fun and easy to use".
value	<div><div></div></div>	"Great product at a great price".
battery	<div><div></div></div>	"use for email, skype, great battery life".
size	<div><div></div></div>	"This pad is light weight and very durable".
picture/video	<div><div></div></div>	"Crisp clear and fast".
design/style	<div><div></div></div>	"Fast and stylish tablet".
graphics	<div><div></div></div>	"The graphics are great".



Processing as assistant



Processing as question answering



- ❑ What year was Abraham Lincoln born?
- ❑ How many states were in the United States that year?
- ❑ How much Chinese silk was exported to England in the end of the 18th century?

It's alive: IBM's Watson supercomputer defeats humans in final jeopardy match, 2011

Processing as translation

Korean


↔



English

×



저번 시간에 내가 학생들에게 수업을 drop하라고 했는데, 몇명이나 drop했을지 너무 궁금하다.

jeobeon sigan-e naega
hagsaengdeulhante sueob-eul
drophalago haessneunde,
myeochmyeong-ina drop
haess-eulji neomu gung-
geumhada.



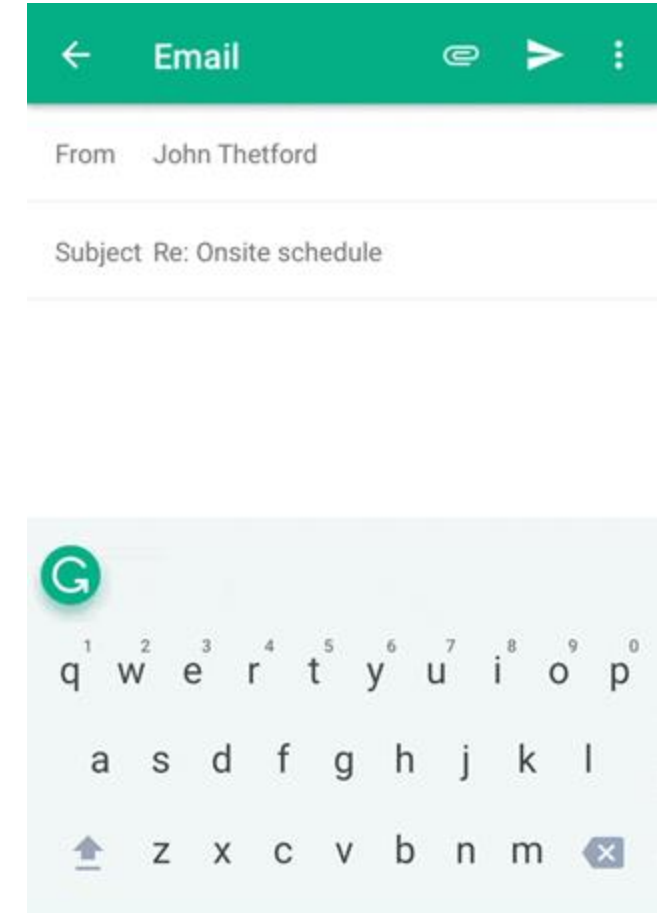
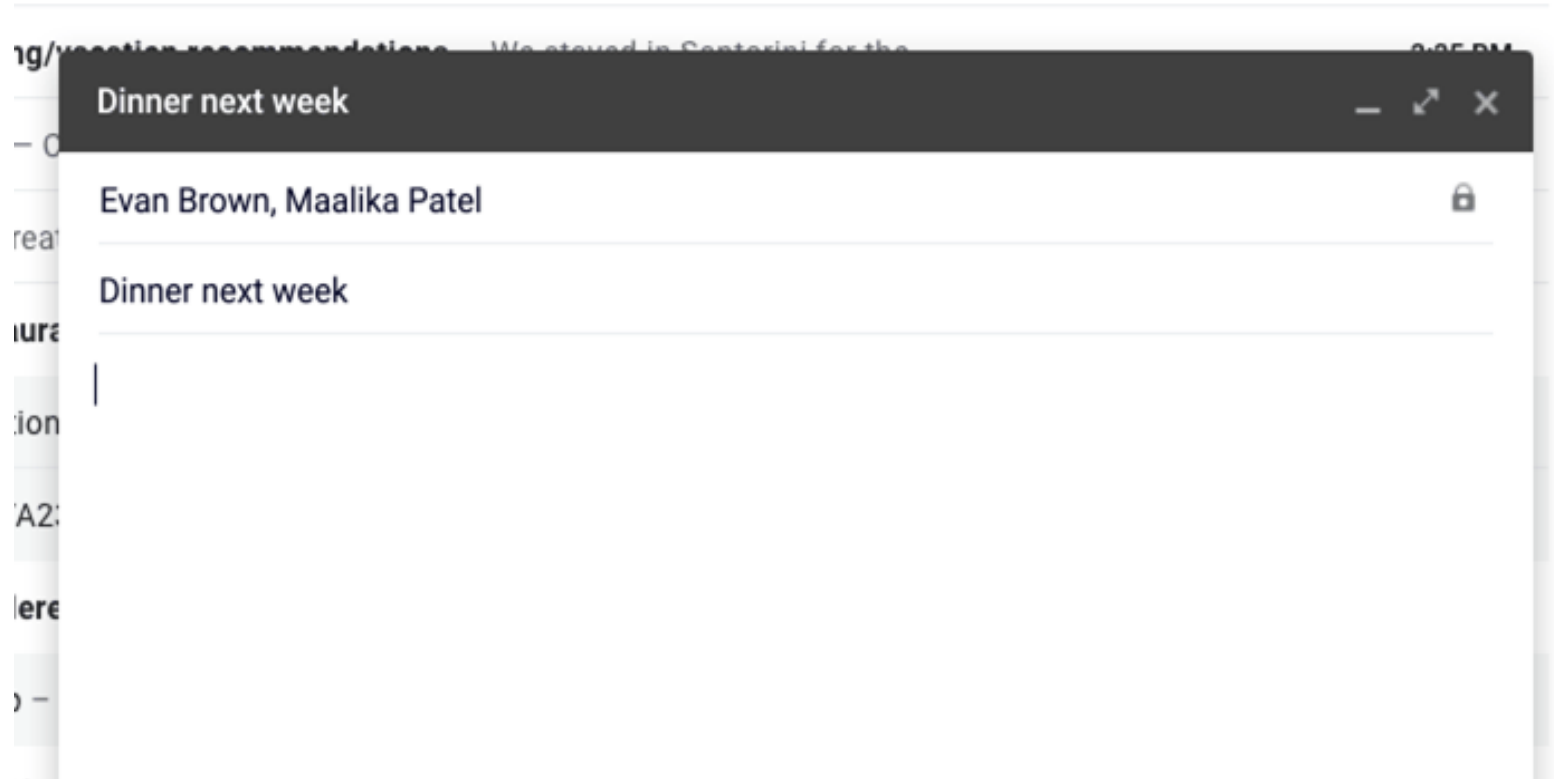
Last time I asked students to drop a class, I'm very curious about how many people dropped it.

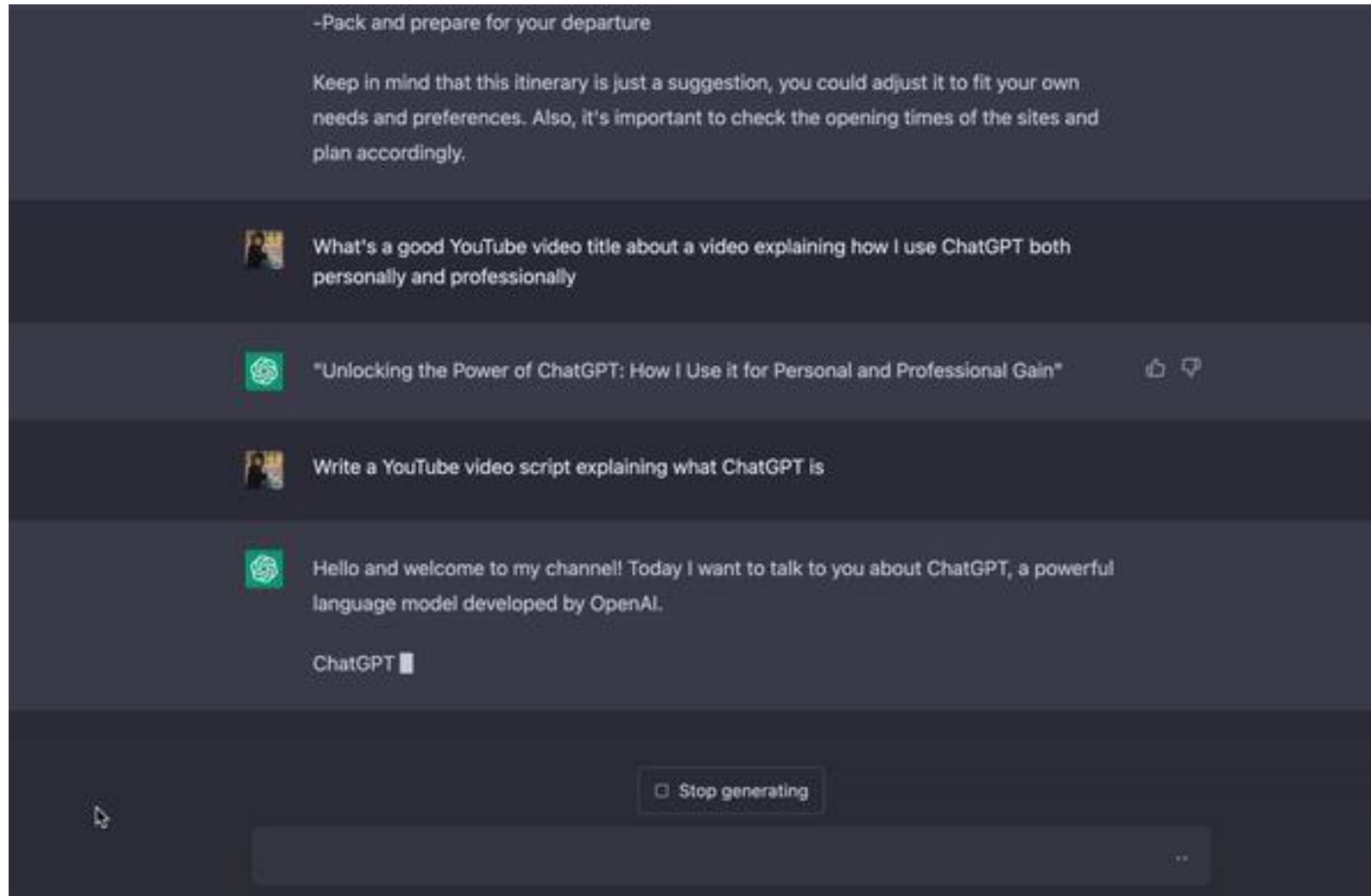
[Open in Google Translate](#) • [Feedback](#)



Processing as (email) assistant



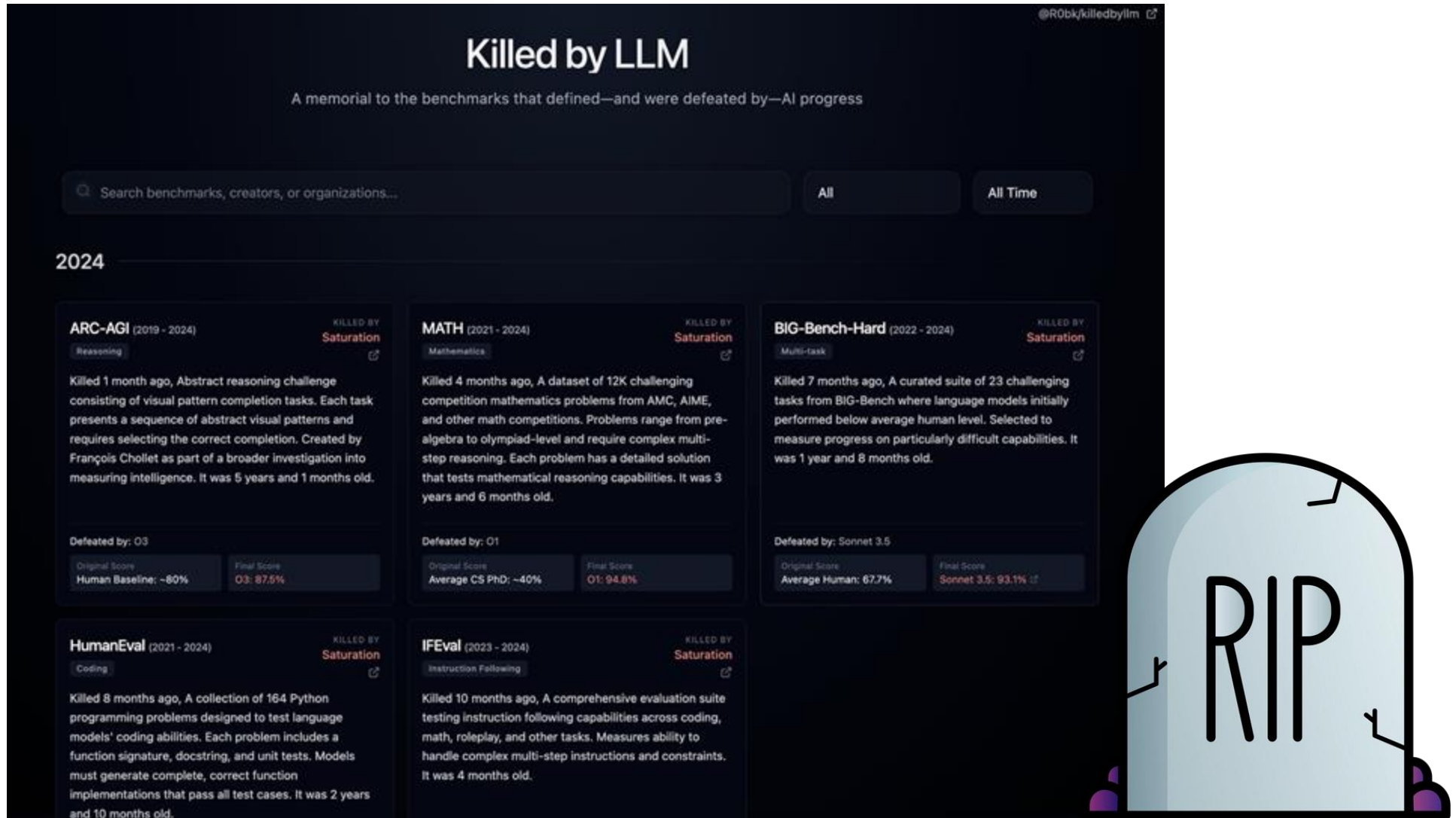
Processing as communication



Do LLM's solve every NLP
problem?



LLMs Keep Conquering New Benchmarks



Killed by LLM
A memorial to the benchmarks that defined—and were defeated by—AI progress

Search benchmarks, creators, or organizations... All All Time

2024

ARC-AGI (2019 - 2024)
Reasoning
Killed 1 month ago, Abstract reasoning challenge consisting of visual pattern completion tasks. Each task presents a sequence of abstract visual patterns and requires selecting the correct completion. Created by François Chollet as part of a broader investigation into measuring intelligence. It was 5 years and 1 months old.

Defeated by: O3
Original Score: Human Baseline: ~80%
Final Score: O3: 87.6%

MATH (2021 - 2024)
Mathematics
Killed 4 months ago, A dataset of 12K challenging competition mathematics problems from AMC, AIME, and other math competitions. Problems range from pre-algebra to olympiad-level and require complex multi-step reasoning. Each problem has a detailed solution that tests mathematical reasoning capabilities. It was 3 years and 6 months old.

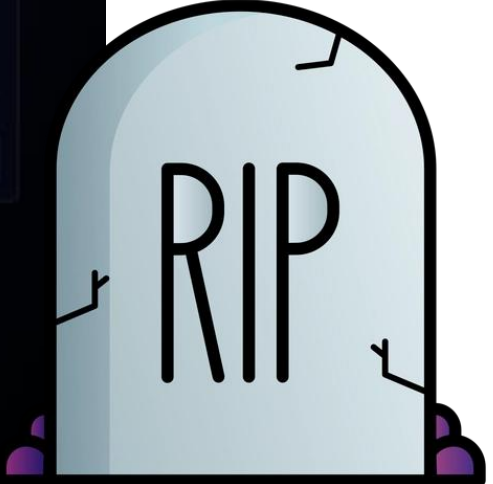
Defeated by: O1
Original Score: Average CS PhD: ~40%
Final Score: O1: 94.8%

BIG-Bench-Hard (2022 - 2024)
Multi-task
Killed 7 months ago, A curated suite of 23 challenging tasks from BIG-Bench where language models initially performed below average human level. Selected to measure progress on particularly difficult capabilities. It was 1 year and 8 months old.

Defeated by: Sonnet 3.5
Original Score: Average Human: 67.7%
Final Score: Sonnet 3.5: 93.1%

HumanEval (2021 - 2024)
Coding
Killed 8 months ago, A collection of 164 Python programming problems designed to test language models' coding abilities. Each problem includes a function signature, docstring, and unit tests. Models must generate complete, correct function implementations that pass all test cases. It was 2 years and 10 months old.

IFEval (2023 - 2024)
Instruction Following
Killed 10 months ago, A comprehensive evaluation suite testing instruction following capabilities across coding, math, roleplay, and other tasks. Measures ability to handle complex multi-step instructions and constraints. It was 4 months old.



How many r's in strawberry?

There are **2 R's** in "strawberry."



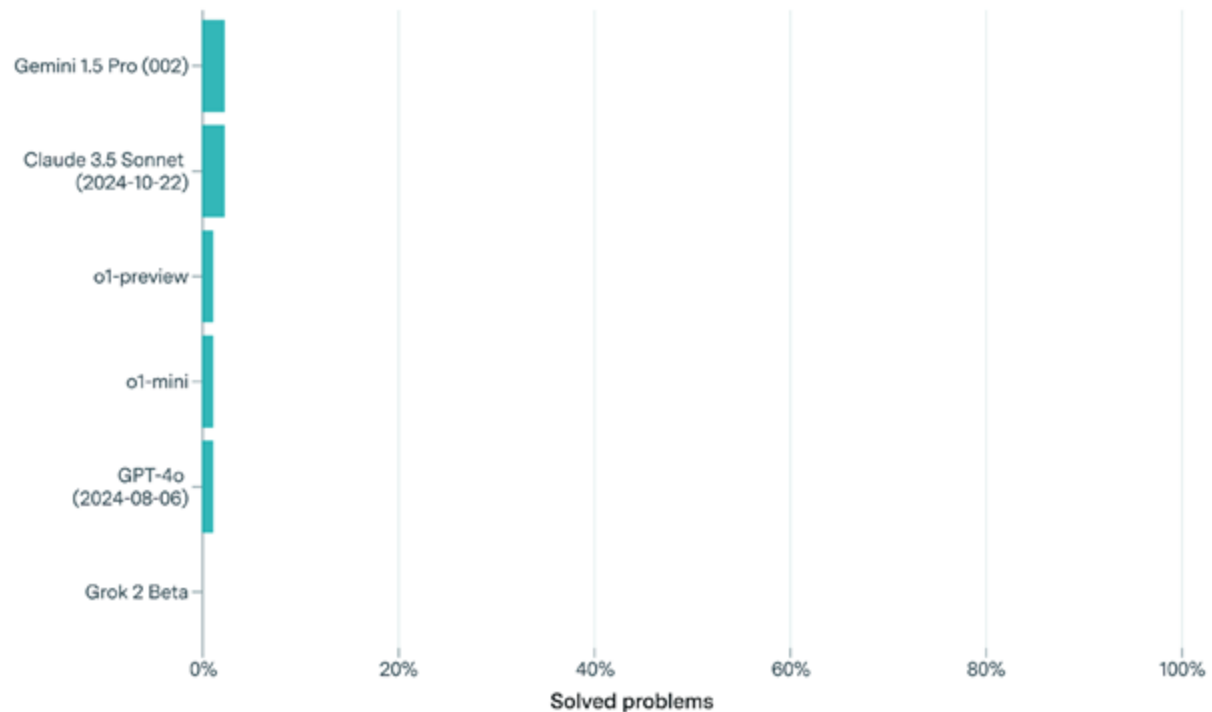


Table 3: **Performance comparison of various baselines on SWE-bench M.** The table shows results for different software development agent frameworks, including SWE-agent (with multimodal and JavaScript-specific variations) and a retrieval augmented generation (RAG) approach. Each system's success rate (% Resolved) and average cost (\$ Avg. Cost) per task are reported.

System	Model	% Resolved	\$ Avg. Cost
SWE-agent M	GPT-4o	12.2	2.94
	Claude 3.5 Sonnet	11.4	3.11
SWE-agent JS	GPT-4o	9.2	0.99
	Claude 3.5 Sonnet	12.0	3.11
SWE-agent Base	GPT-4o	12.0	2.07
	Claude 3.5 Sonnet	12.2	1.52
Agentless JS	GPT-4o	3.1	0.38
	Claude 3.5 Sonnet	6.2	0.42
RAG	GPT-4o	6.0	0.17
	Claude 3.5 Sonnet	5.0	0.15



Current models are unable to solve FrontierMath problems

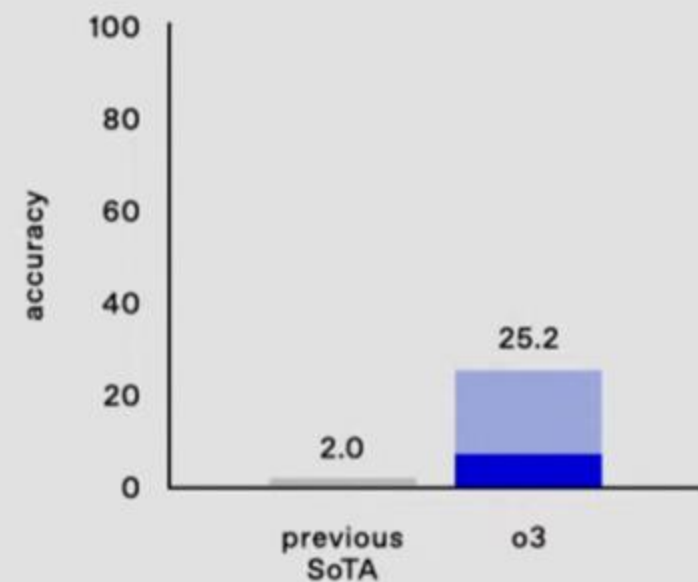


CC-BY

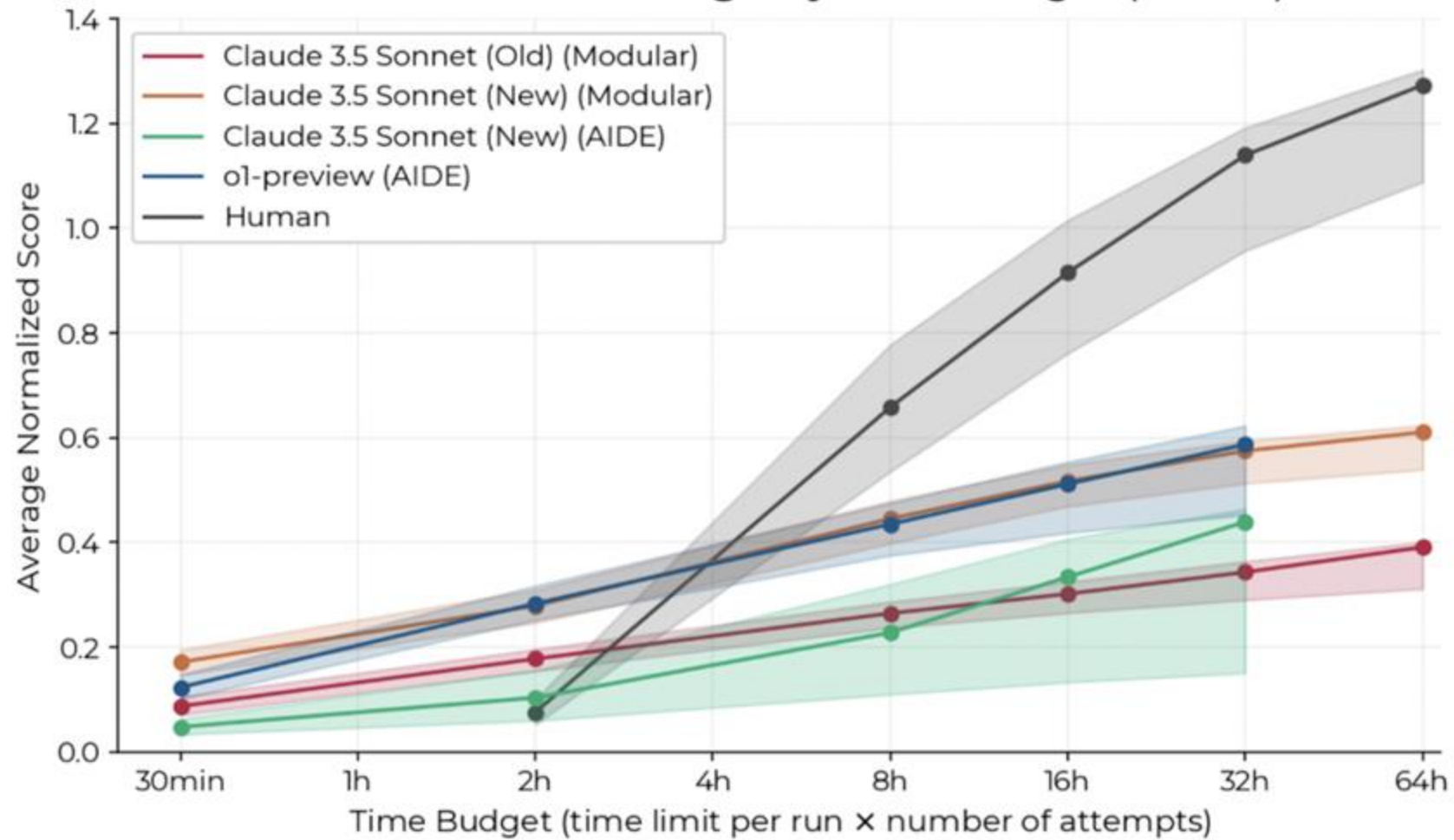
EPOCH AI

epochai.org

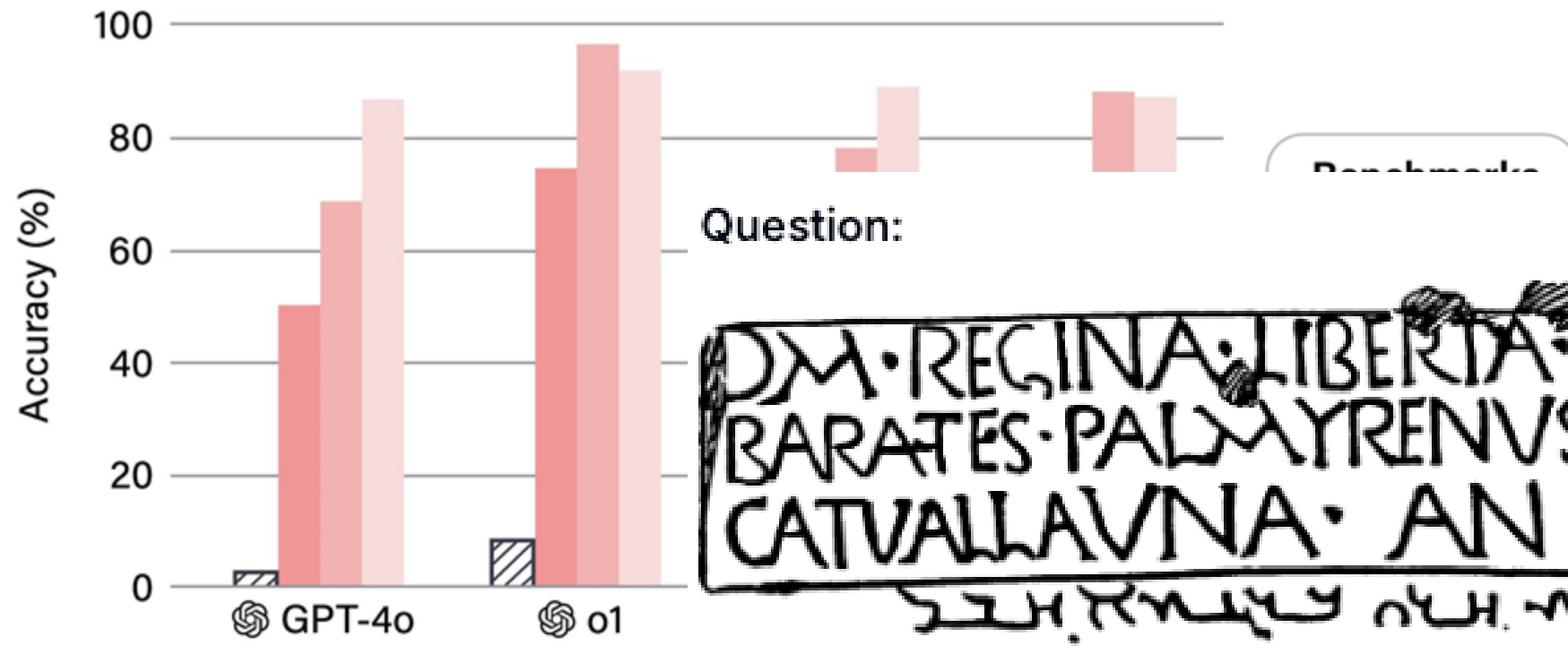
Research Math
(EpochAI Frontier Math)



Best Observed Score@k by Time Budget (95% CI)



Accuracy of LLMs Across Benchmarks

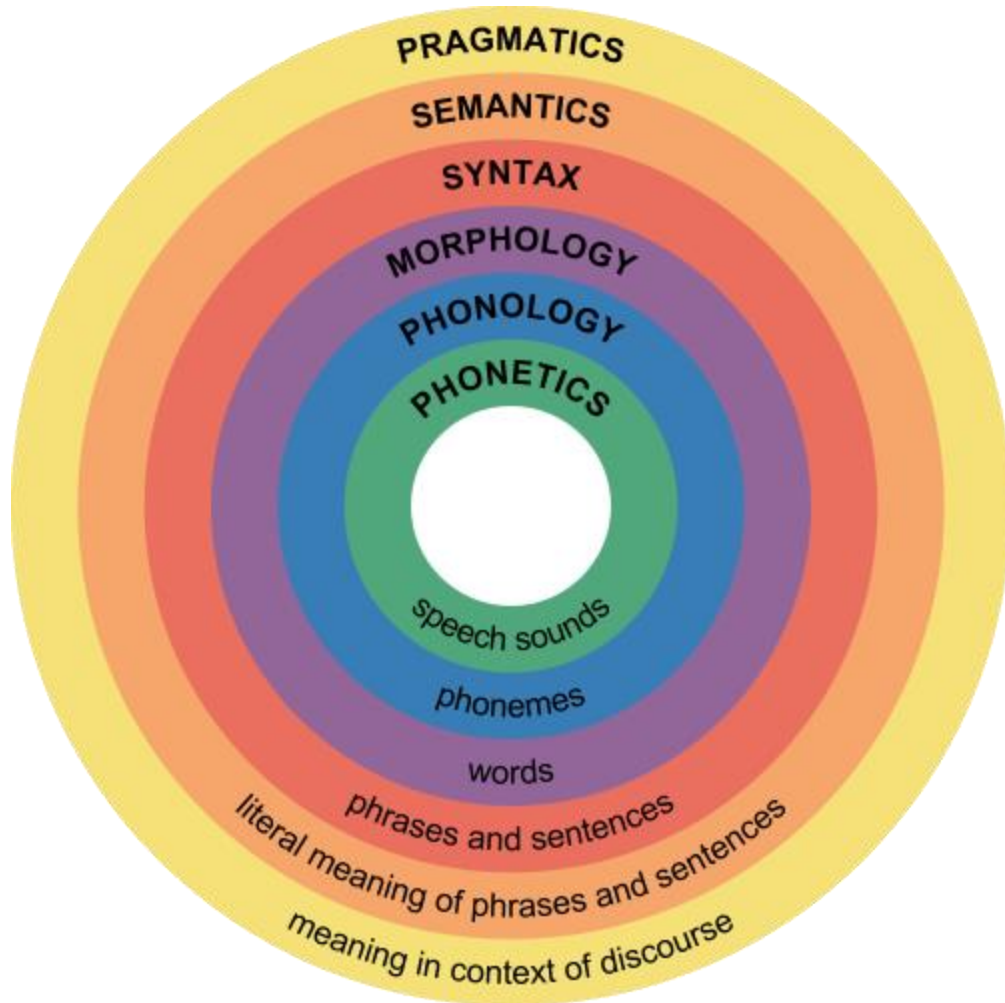


Mode Here is a representation of a Roman inscription, originally found on a tombstone. Provide a translation for the Palmyrene script. A transliteration of the text is provided: RGYN^o BT HRY BR ^eT^o HBL

What makes language so difficult to process?



Language consists of many levels of structure



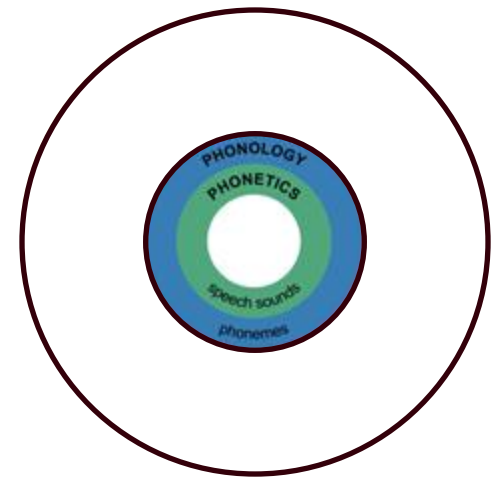
Humans fluently integrate all of these in generating and understanding language

This is a simple sentence



Phonology

□ Pronunciation modeling



SOUNDS

Th i a si e n

Example by Nathan Schneider

Words

- ❑ Tokenization
- ❑ Language modeling
- ❑ Spelling correction



WORDS

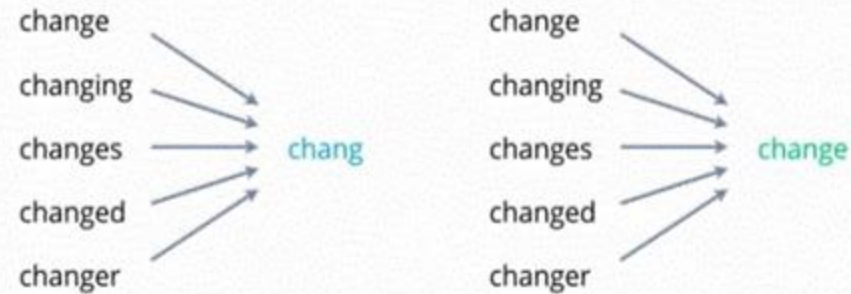
This is a simple sentence

Example by Nathan Schneider

Morphology

- ❑ Morphological analysis
- ❑ Tokenization
- ❑ Stemming / Lemmatization

Stemming vs Lemmatization



WORDS

This is a simple sentence

MORPHOLOGY

be
3sg
present

Read more about stemming and lemmatization
<https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html>

Example by Nathan Schneider



Parts of Speech (POS)

- Part-of-speech tagging



PART OF SPEECH

DT

VBZ

DT

JJ

NN

WORDS

This is a simple sentence

MORPHOLOGY

be
3sg
present

Example by Nathan Schneider

Parts of Speech (POS)

□ Part-of-speech tagging

PART OF SPEECH

WORDS

MORPHOLOGY

DT VBZ DT

This is a sir

be
3sg
present

CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential <i>there</i>
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP\$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	to
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VCN	Verb, past participle
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Wh-determiner
WP	Wh-pronoun
WP\$	Possessive wh-pronoun
WRB	Wh-adverb

Example by Nathan Schneider



Syntax

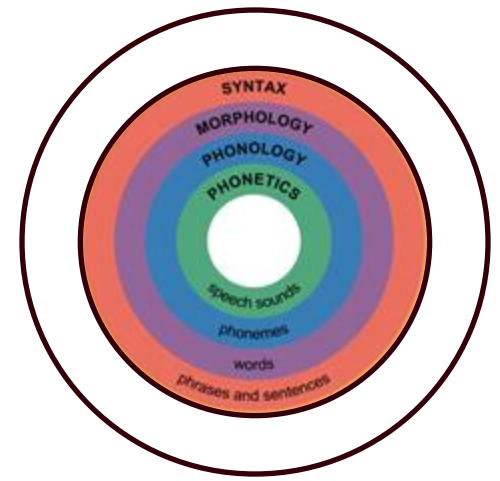
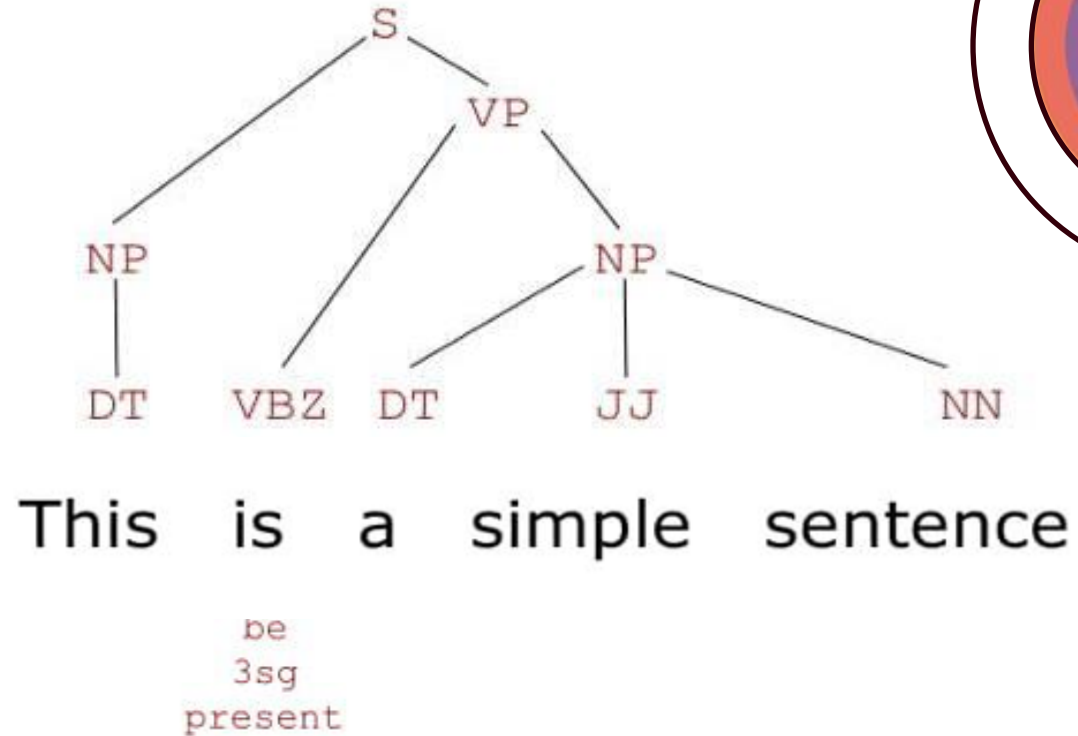
□ Syntax parsing

SYNTAX

PART OF SPEECH

WORDS

MORPHOLOGY

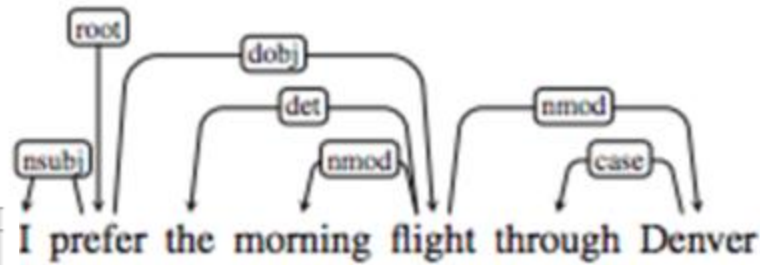
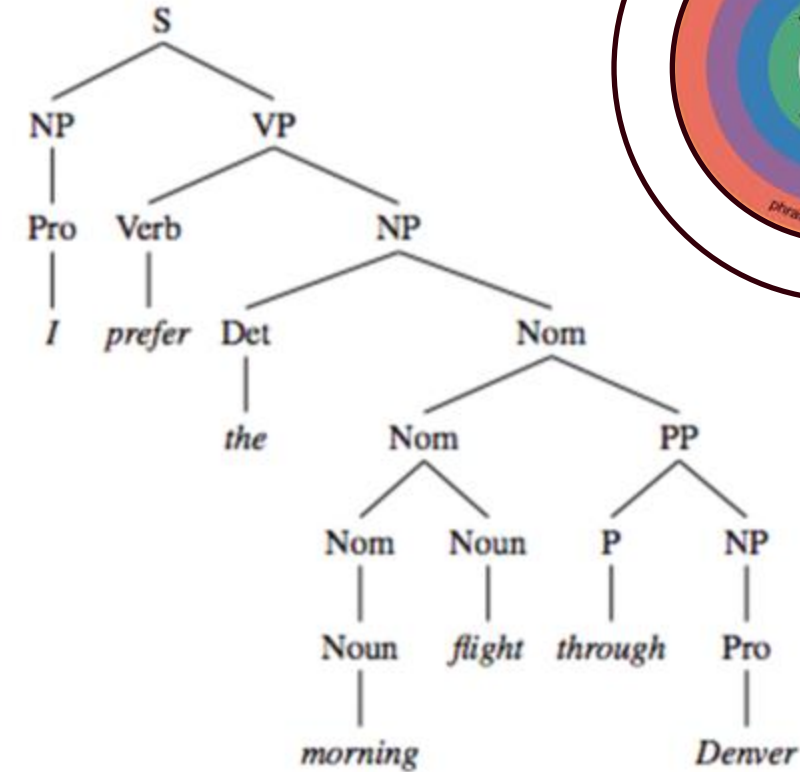
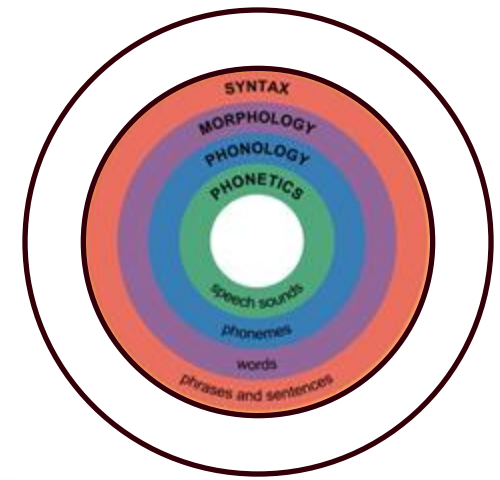


Example by Nathan Schneider

Syntax

□ Syntax parsing

- *Constituency Parsing*: break a sentence into sub-phrases
- *Dependency Parsing*: explore the dependencies between the words in a sentence



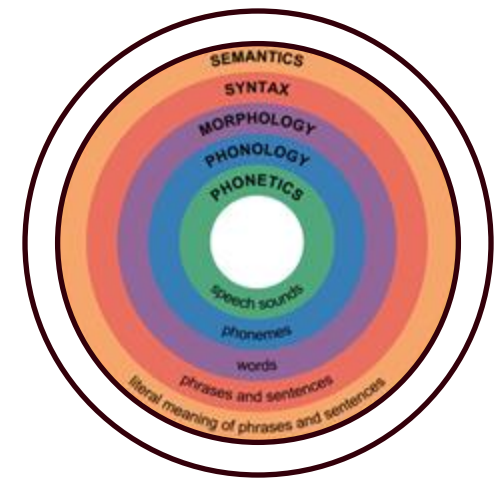
Clausal Argument Relations	Description
NSUBJ	Nominal subject
DOBJ	Direct object
IOBJ	Indirect object
CCOMP	Clausal complement
XCOMP	Open clausal complement
Nominal Modifier Relations	Description
NMOD	Nominal modifier
AMOD	Adjectival modifier
NUMMOD	Numeric modifier
APPOS	Appositional modifier
DET	Determiner
CASE	Prepositions, postpositions and other case markers
Other Notable Relations	Description
CONJ	Conjunct
CC	Coordinating conjunction

Figure 15.2 Selected dependency relations from the Universal Dependency set. (de Marneffe et al., 2014)

Example by Nathan Schneider



Semantics



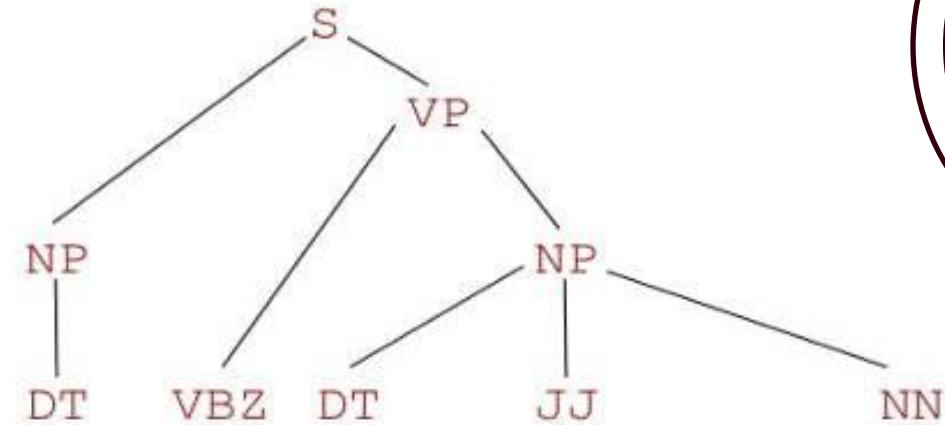
SYNTAX

PART OF SPEECH

WORDS

MORPHOLOGY

SEMANTICS



This is a simple sentence

be
3sg
present

SIMPLE1
having
few
parts

SENTENCE1
string of words
satisfying the
grammatical rules
of a language

- ☐ Named entity recognition
- ☐ Word sense disambiguation
- ☐ Semantic role labeling
- ☐ Frame semantics

Example by Nathan Schneider



Discourse (Pragmatics)

❑ Co-reference resolution

SYNTAX

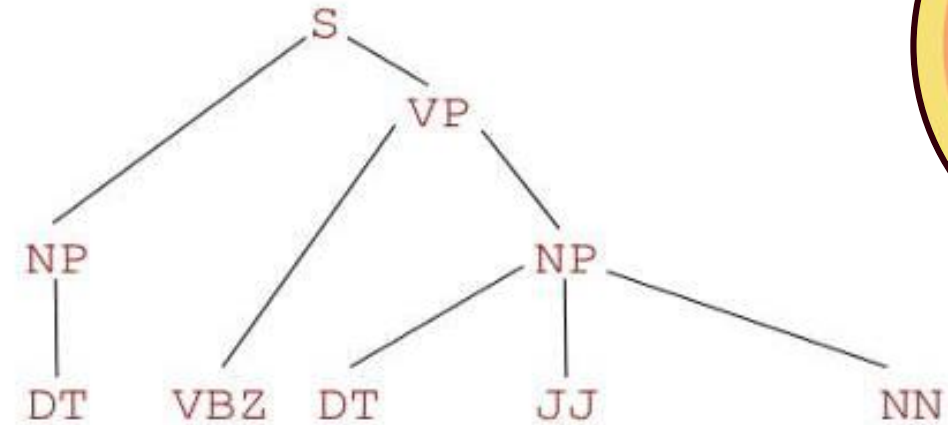
PART OF SPEECH

WORDS

MORPHOLOGY

SEMANTICS

DISCOURSE



This is a simple sentence

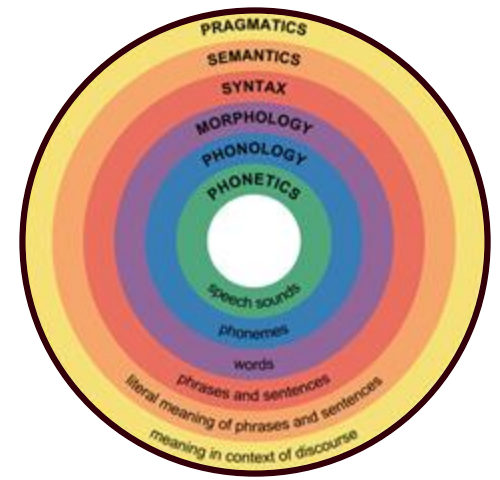
be
3sg
present

SIMPLE1
having
few
parts

SENTENCE1
string of words
satisfying the
grammatical rules
of a language

But it is an instructive one.

CONTRAST



Example by Nathan Schneider

What makes language difficult?

- ❑ Language is *ambiguous*
- ❑ Language needs to be *scaled*
- ❑ Language is *sparse*
- ❑ Language is *varying*
- ❑ Language is *implicit*
- ❑ Language is hard to *represent*



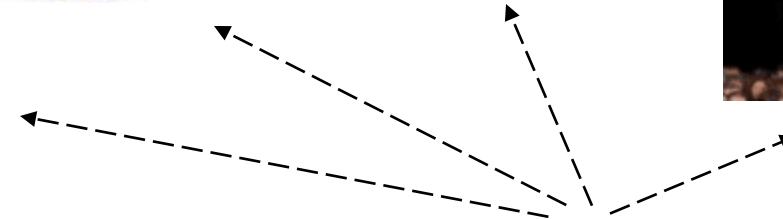
Ambiguity at multiple levels



Groucho Marx

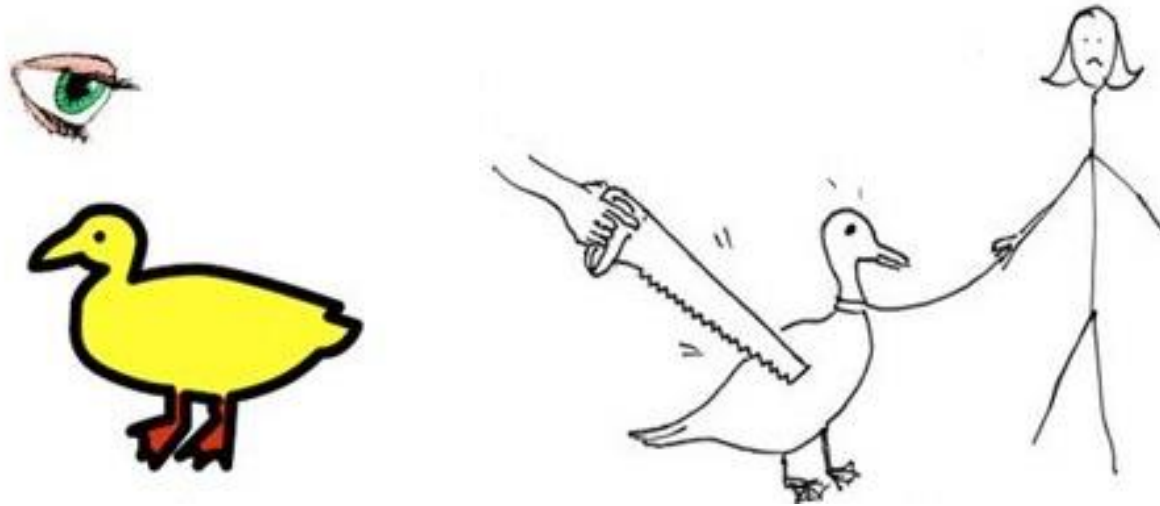
"One morning I shot an
elephant **in my pajamas**"

Ambiguity at multiple levels



"One morning I shot an
elephant **in my pajamas**"

"I saw her duck with a telescope"



- I used a telescope to observe a small web-footed broad-billed swimming bird belonging to a female person.
- I observed a small web-footed broad-billed swimming bird belonging to a female person. The bird had a telescope.
- I observed a female person move quickly downwards. The person had a telescope.
- I used a telescope to observe a female person move quickly downwards.
- I used a telescope to cut a small web-footed broad-billed swimming bird belonging to a female person.
- I used a telescope to observe heavy cotton fabric of plain weave belonging to a female person.
- I used a telescope to cut heavy cotton fabric of plain weave belonging to a female person.

Slide from Dhruv Batra and figure from Liang Huang



Scale: Applications x Languages



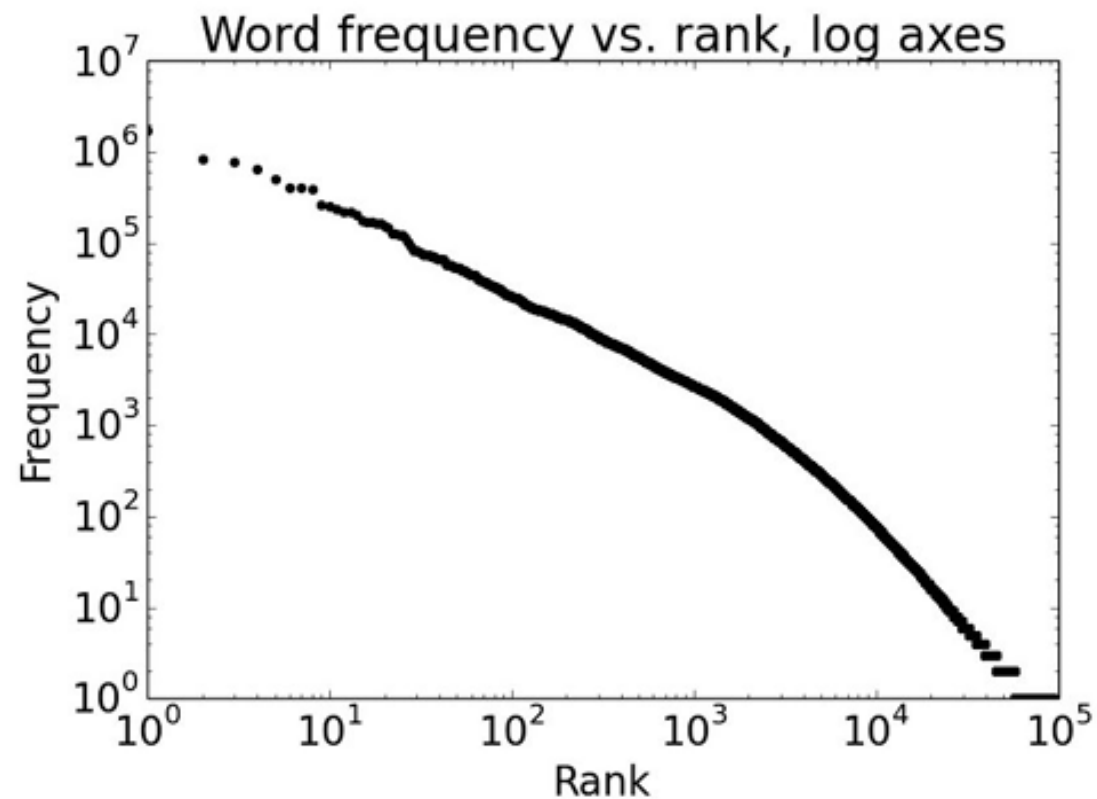
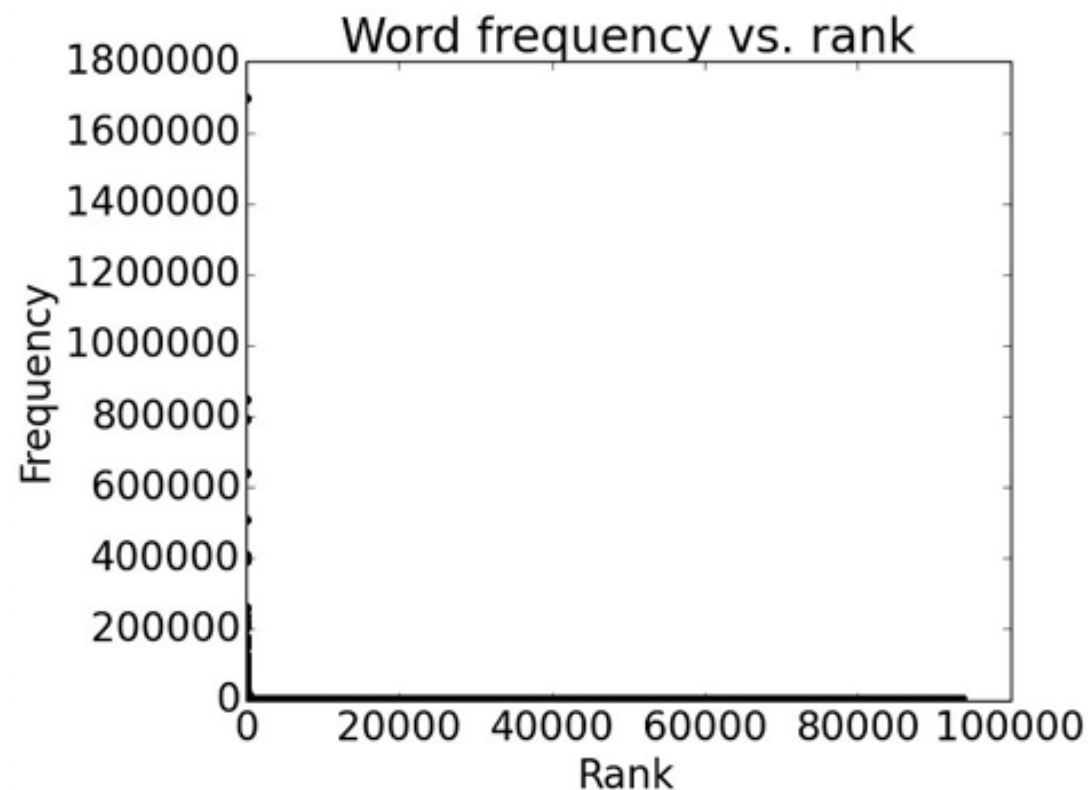
Sparsity

- ❑ Sparse data due to Zipf's Law
- ❑ Most frequent words in the English Europarl corpus (out of 24M word tokens)
- ❑ 36,231 occur only once
 - E.g., pseudo-rapporteur, lobby-ridden, perfunctorily, Lycketoft, UNCITRAL, policyfor, 145.95 ..

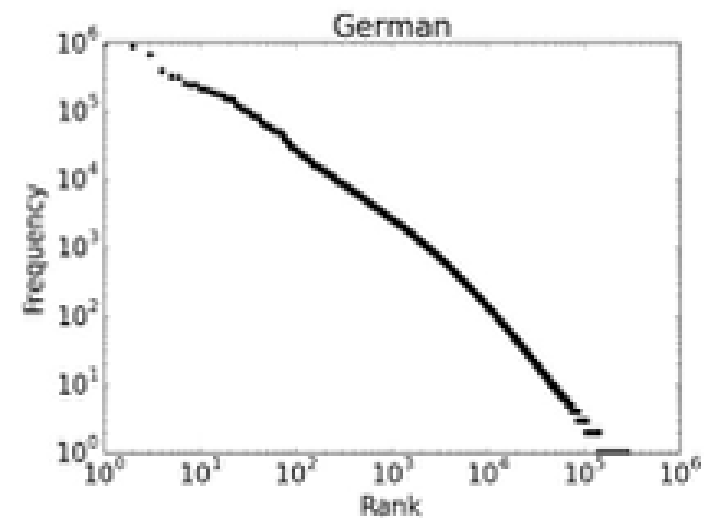
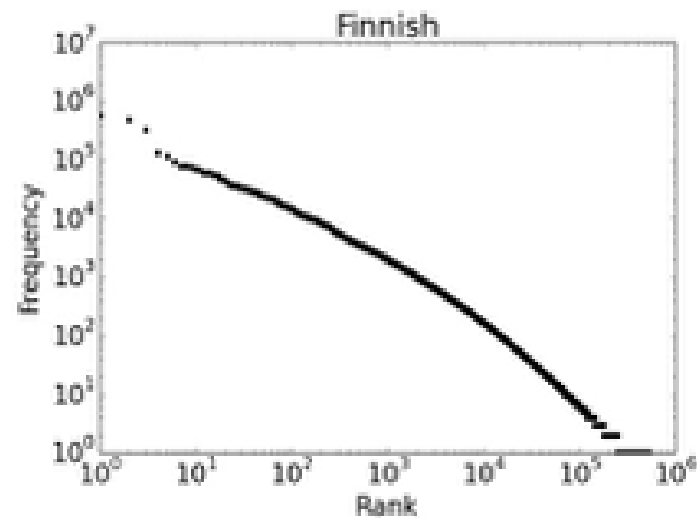
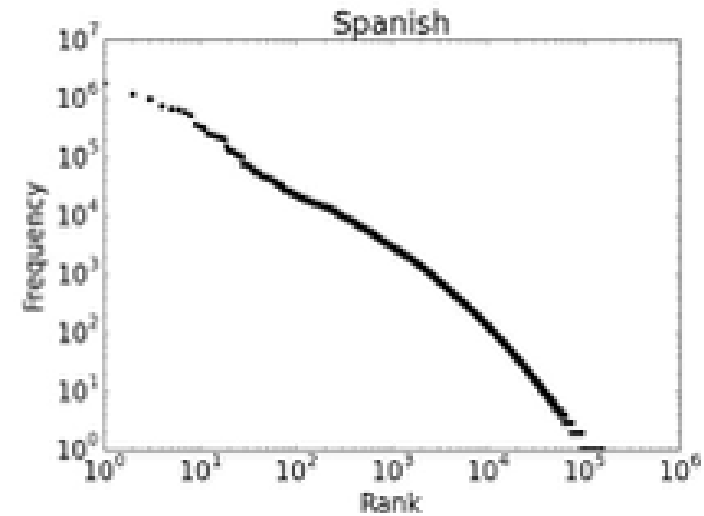
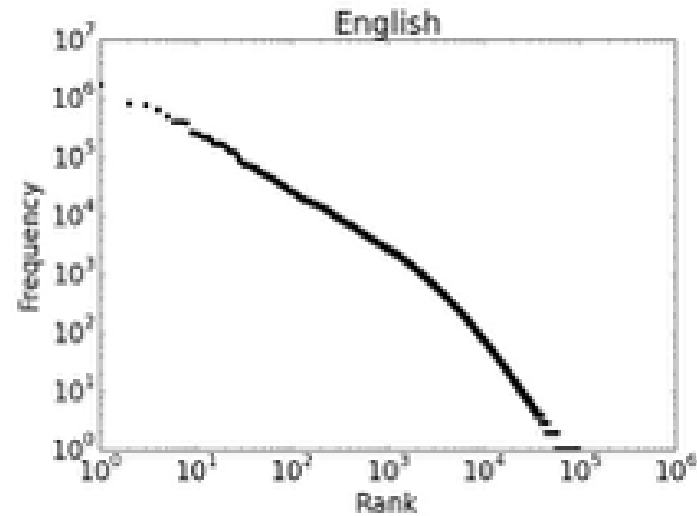
any word		nouns	
Frequency	Token	Frequency	Token
1,698,599	the	124,598	European
849,256	of	104,325	Mr
793,731	to	92,195	Commission
640,257	and	66,781	President
508,560	in	62,867	Parliament
407,638	that	57,804	Union
400,467	is	53,683	report
394,778	a	53,547	Council
263,040	I	45,842	States



Word Frequency Distribution

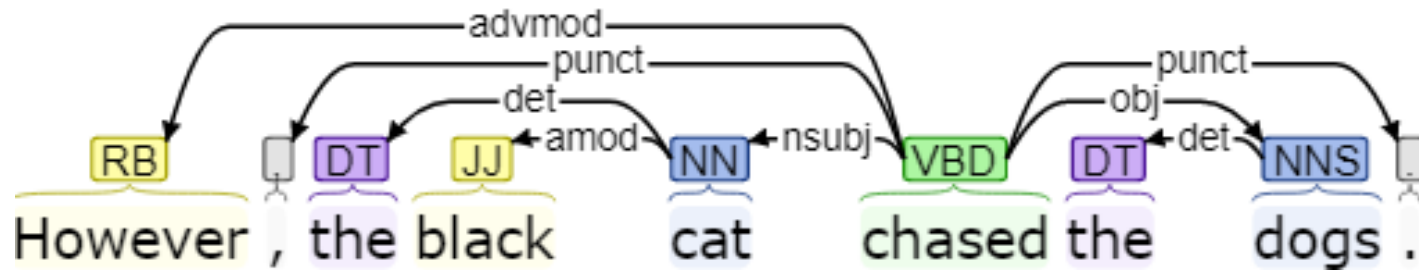


Zipf's Law



Variation over Domains

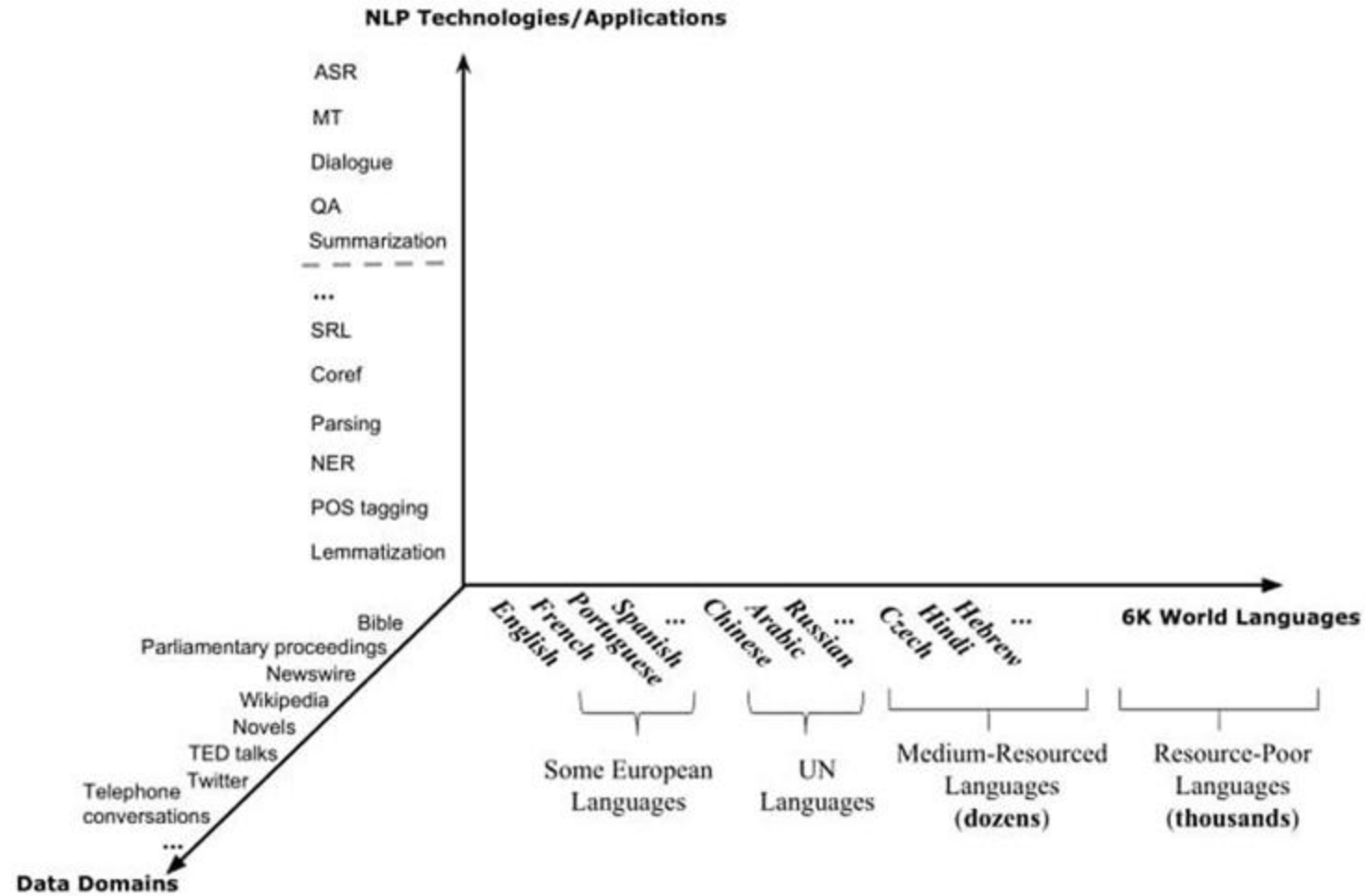
- ❑ Suppose you trained a part-of-speech tagger or parser on the Wall Street Journal



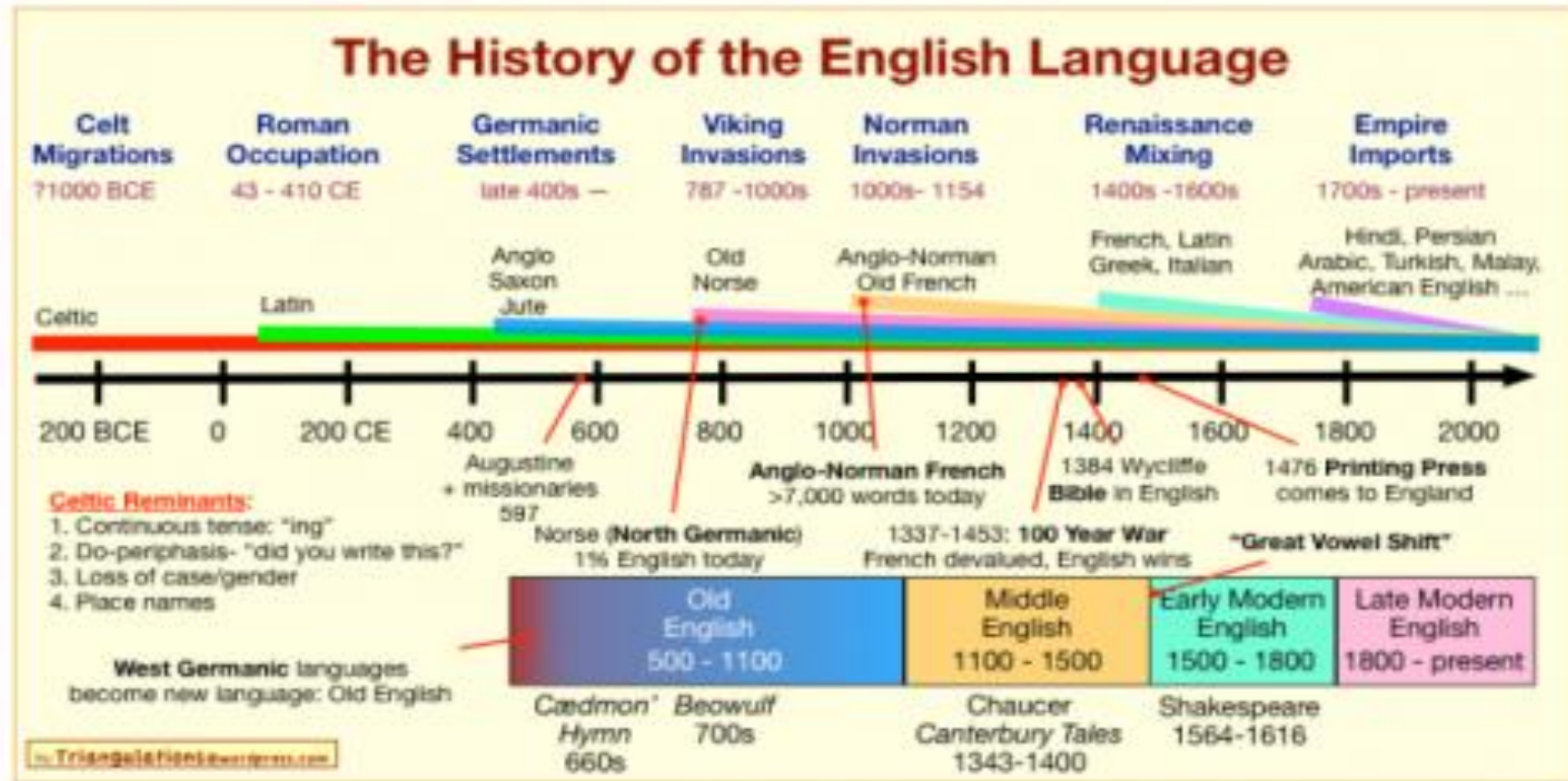
- ❑ What happens if you try to use the same tagger/parser for **social media text**?

@_rkpntnte hindi ko alam babe eh, absent ako
kanina I'm sick rn hahaha 😊👏

Application x Languages x Domains



Variation over Time



Variation over Time

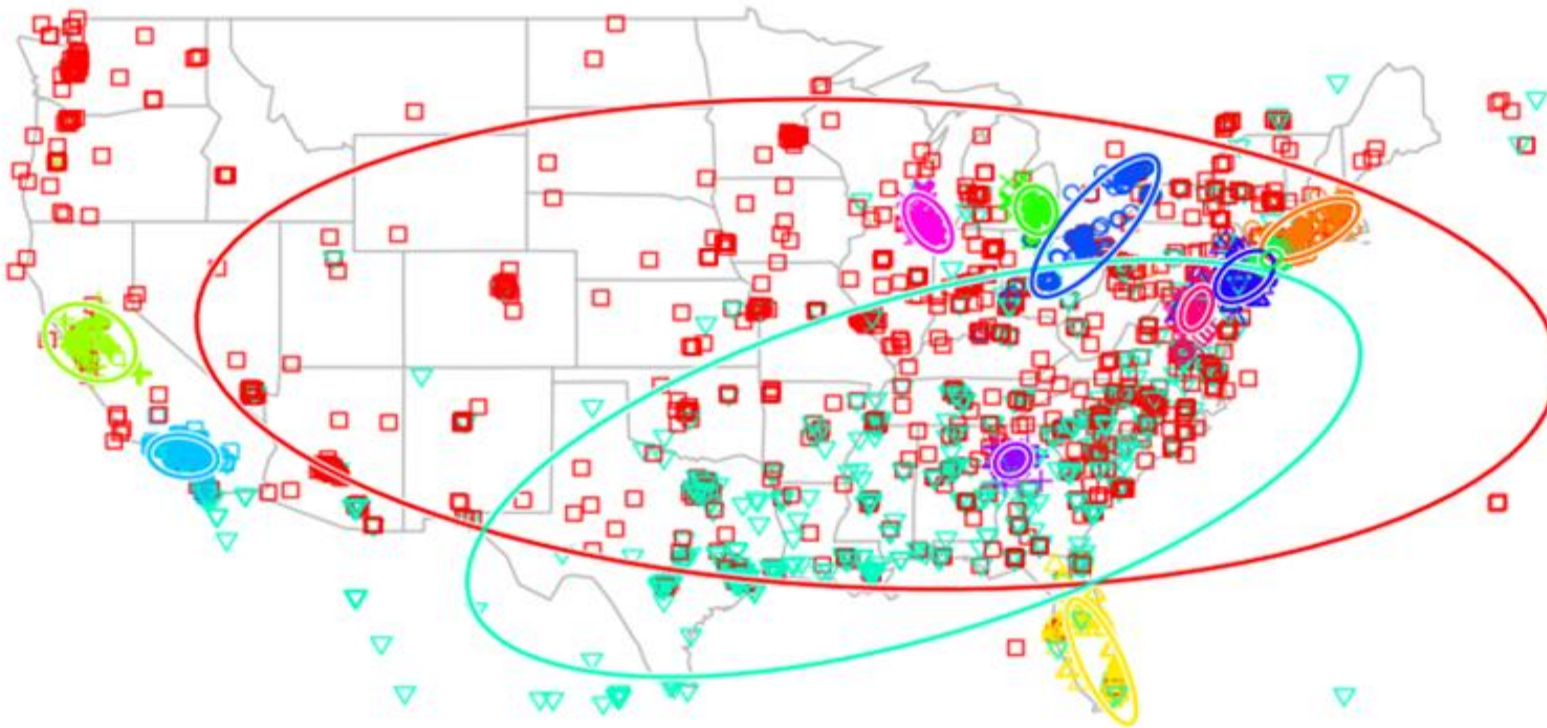


24 New Words Invented by Teenagers , NYT 20220331



<https://www.instagram.com/reel/C-NuNbutMD6/>

Variation over Location



A Latent Variable Model for Geographic Lexical Variation [Eisenstein et al., 2010]

British & American English



British	American
anticlockwise	counter
appetizer	starter
aubergine	eggplant
biscuit	cookie
boot	trunk
braces	suspenders
candyfloss	cotton candy
car park	parking lot
chemist	drugstore
chips	French fries
cot	crib
courgette	zucchini
crisps	chips
drawing pin	thumbtack
dressing gown	robe
dummy	pacifier
dustbin	garbage can
flannel	washcloth
flat	apartment
football	soccer
fringe	bangs
grill	broil

British	American
grill	broiler
hairslide	barrette
holiday	vacation
jumper	sweater
lift	elevator
mobile phone	cell phone
number plate	license plate
off-licence	liquor store
oven glove	oven mitt
parting	part
pavement	sidewalk
petrol	gas, gasoline
postbox	mailbox
postcode	zip code
public school	private school
pushchair	stroller
shopping trolley	shopping cart
skipping rope	jump rope
sledge	sled
state school	public school

www.englishgrammarhere.com



Beyond conventional meaning



Implicit meaning behind language and Pragmatics

□ Speech act [\[Austin 1962\]](#)

- "Could you please pass the salt to me?"

□ Implicature [\[Grice 1975\]](#)

- Alice: "Are you going to Paul's party?"
- Bob: "I have to work."

- labelling
- repeating
- answering
- **requesting (action)**
- requesting (answer)
- calling
- greeting
- protesting
- practicing



Unknown Representation

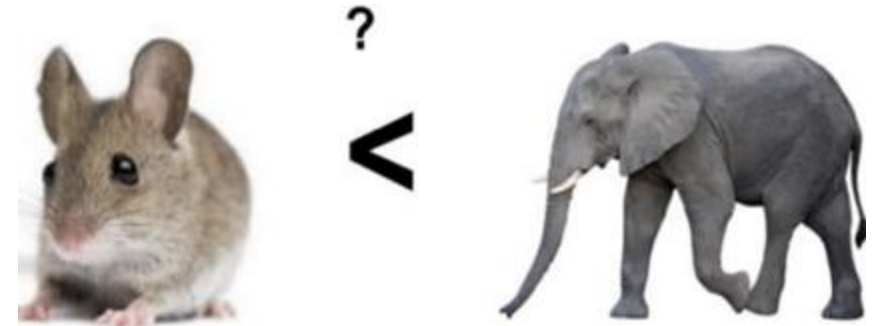
- ❑ We don't even know how to represent knowledge a human has/needs
- ❑ What is the meaning of word or sentence?
- ❑ How to model context or general knowledge?



"**Drink** this milk"



"Sunset is **beautiful**"



Elephants are **bigger than** mice?

Summary



- ❑ NLP is interdisciplinary
- ❑ Language consists of many levels of structure:
 - Phonology, syntax, semantics, discourse, pragmatics
- ❑ Processing language is difficult, due to
 - ambiguity, scales, sparsity, variation, implication, and representation
- ❑ Development of NLP models and representations grows rapidly
 - From rules to feature learning to RNNs to Transformers
- ❑ “Large” language models
 - Generalist AI or AGI via prompting and chat
 - Scaling law
 - Multimodal
 - Limitations? Future directions?



How to process language?



Methods

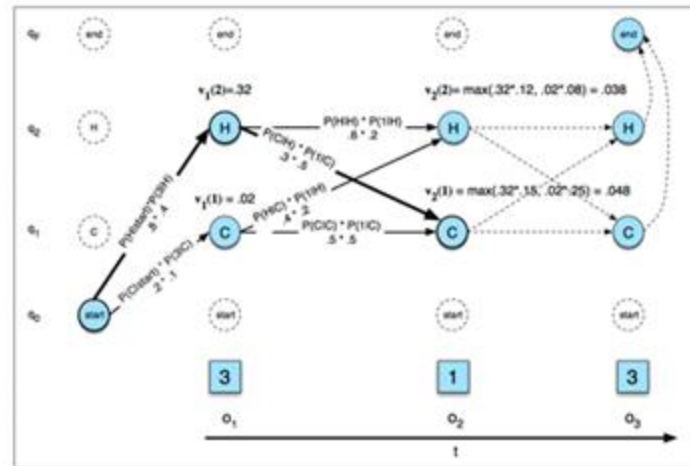
- ❑ Logic-based and rule-based NLP systems (~80s)
- ❑ Dynamic programming and Viterbi/CKY (~90s)
- ❑ Naïve Bayes, LogReg, HMM/CRF, SVM, N-gram LMs (~00s)

Some queries:

```
?- ancestor(mildred,mary).
yes % because parent(mildred,mary).

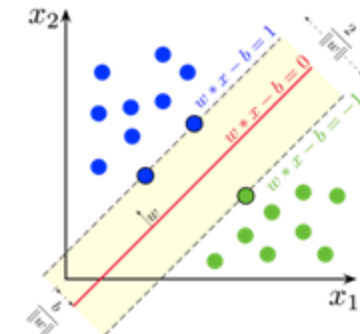
?- ancestor(irvin,nora).
yes % because
    % parent(irvin,ken) and
    % ancestor(ken,nora) because parent(ken,nora).

?- ancestor(chester,elizabeth).
yes % because
    % parent(chester,irvin)
    % and ancestor(irvin,elizabeth)
    % because parent(irvin,ken) and
    % ancestor(ken,elizabeth)
    % because parent(ken,elizabeth).
```



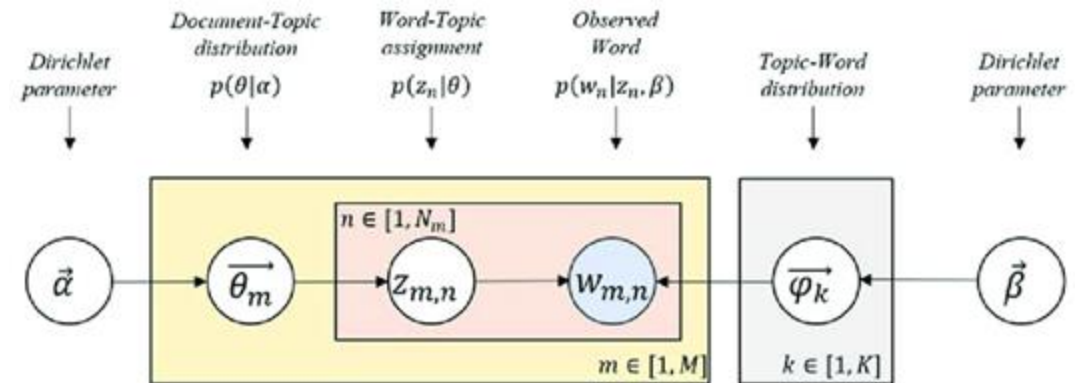
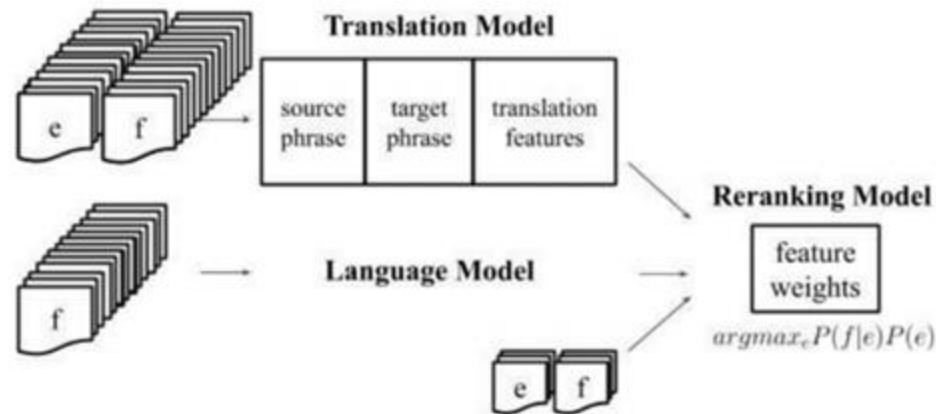
$$P(Y = y|X = x)$$

$$= \frac{P(Y = y)P(X = x|Y = y)}{\sum_y P(Y = y)P(X = x|Y = y)}$$



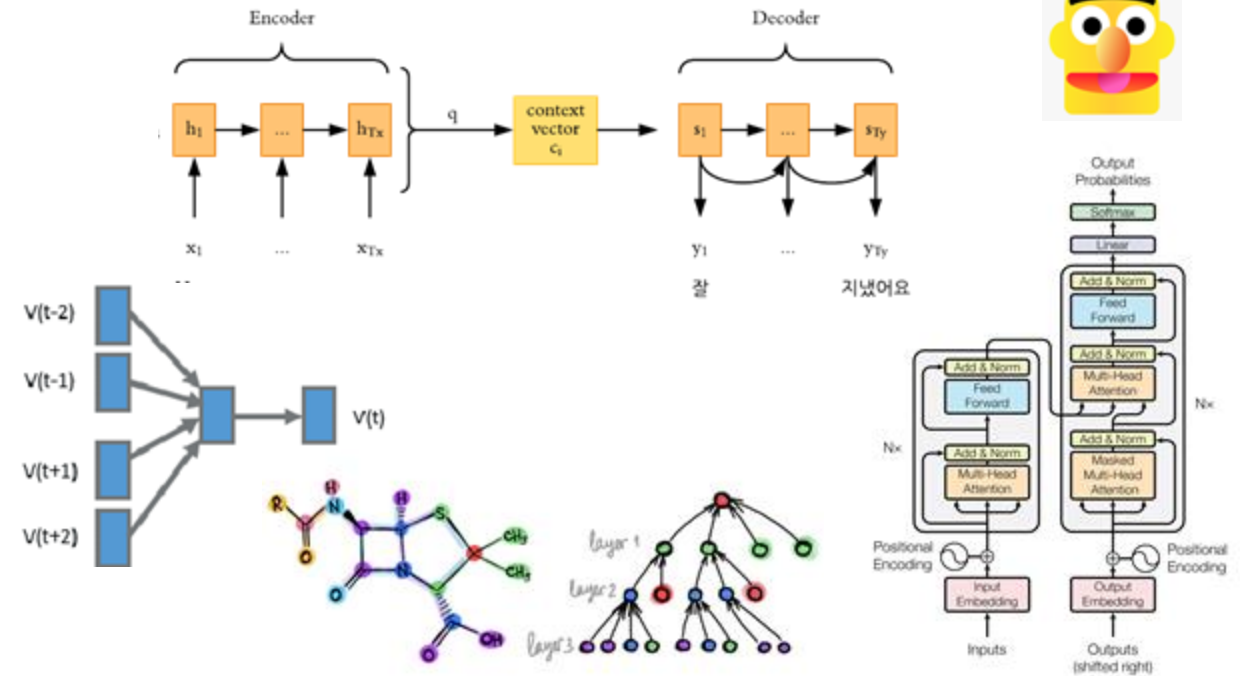
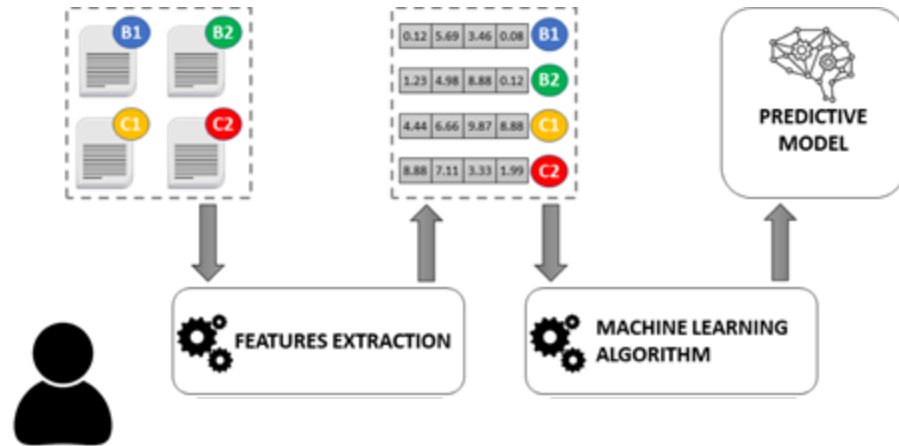
Methods

- ❑ Statistical NLP (~2005s)
- ❑ Latent variable models (Early ~2010s)
 - Specifying probabilistic structure between variables and inferring likely latent values



Representations

- ❑ Human-engineered features and SVMs (2005s ~ 2010s)
- ❑ Learned features/representations (2013s ~ 2018)



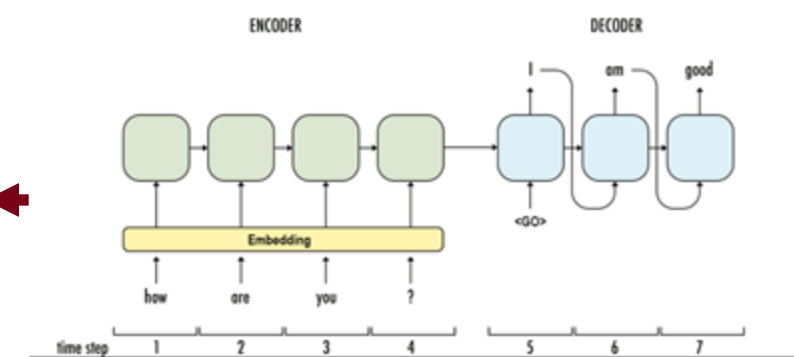
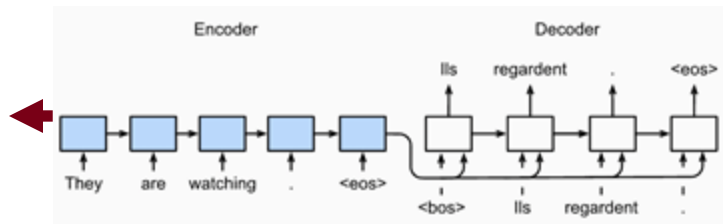
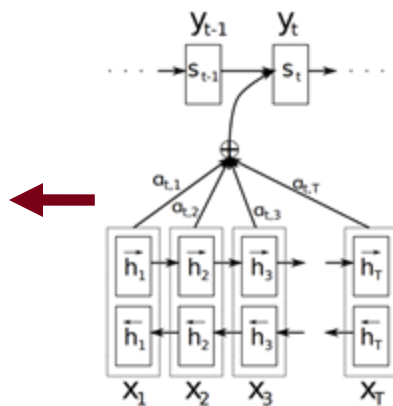
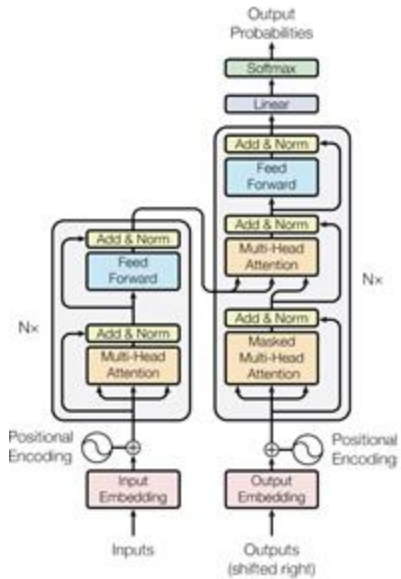
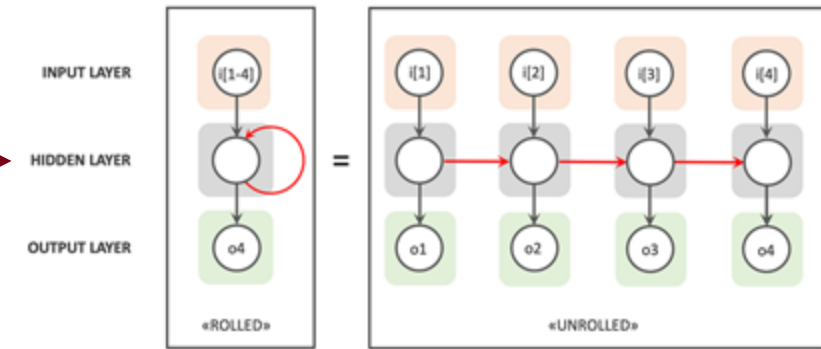
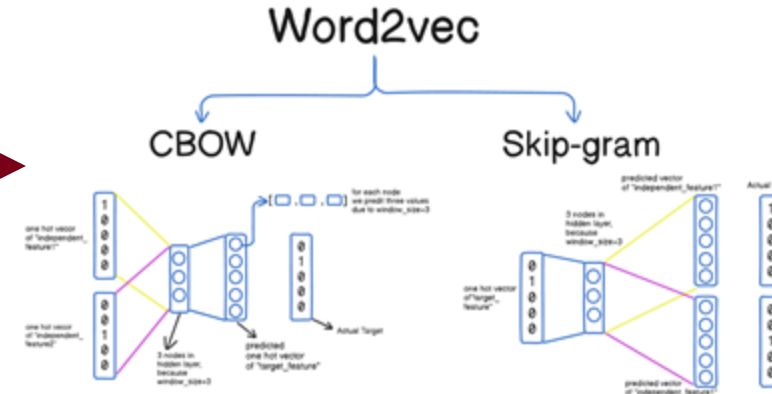
Representations (Developing Attention)

Term Frequency \times Inverse Document Frequency

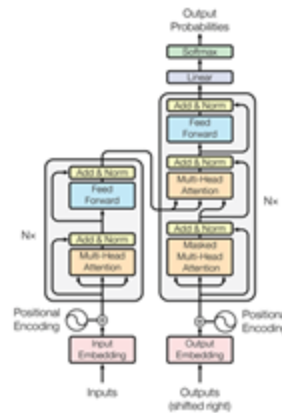
$$w_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

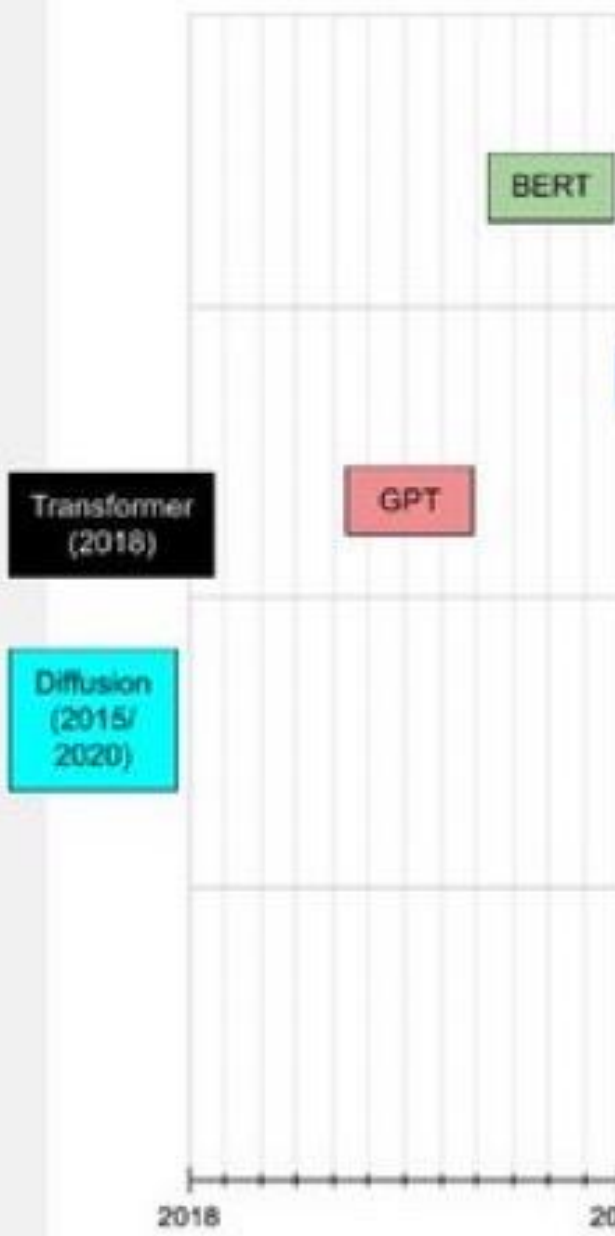
Text1: Basic Linux Commands for Data Science
Text2: Essential DVC Commands for Data Science

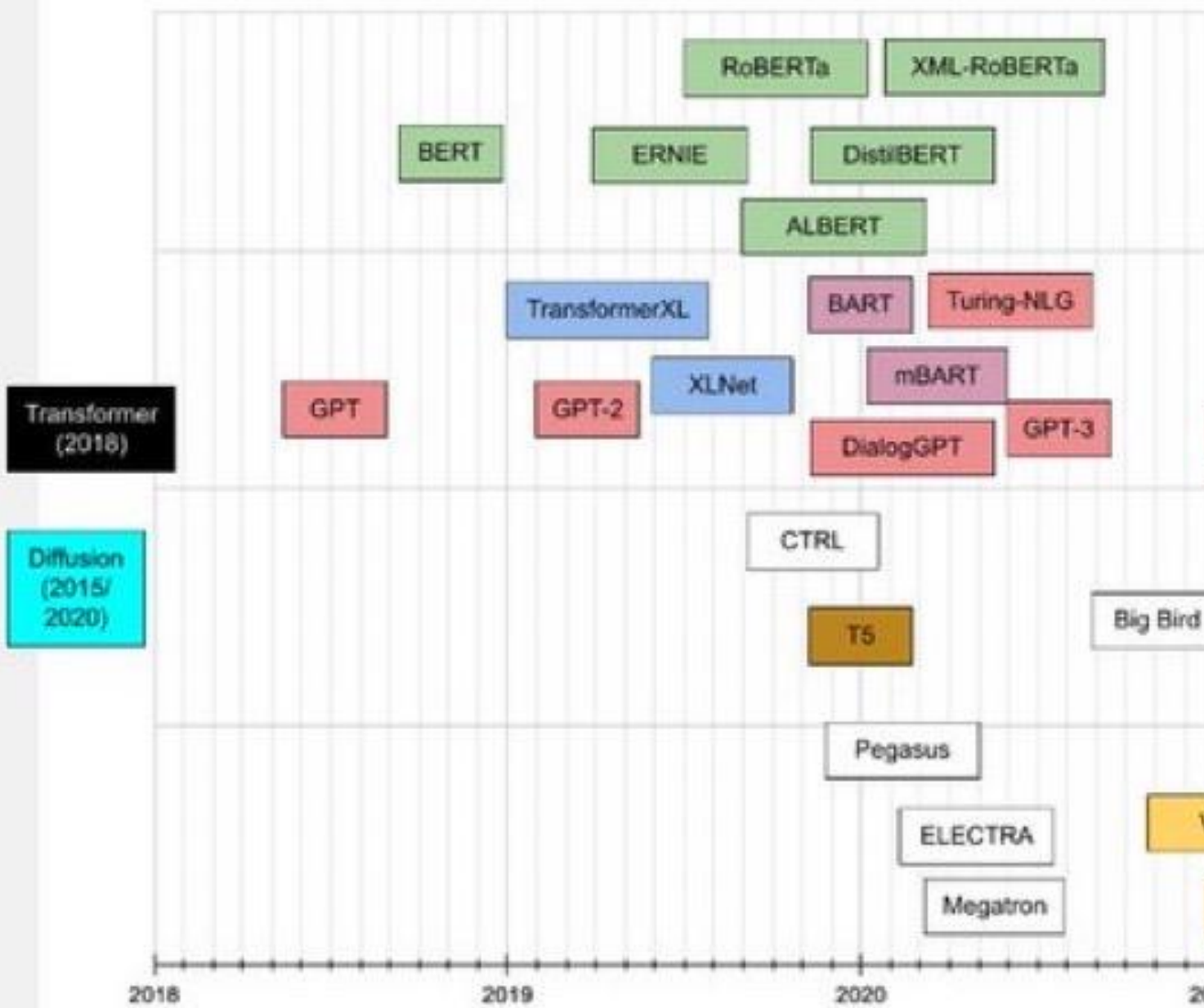
	basic	commands	data	etc	essential	for	linux	reference
Text 1	0.0	0.25	0.25	0.0	0.0	0.25	0.0	0.25
Text 2	0.0	0.25	0.25	0.0	0.0	0.25	0.0	0.25

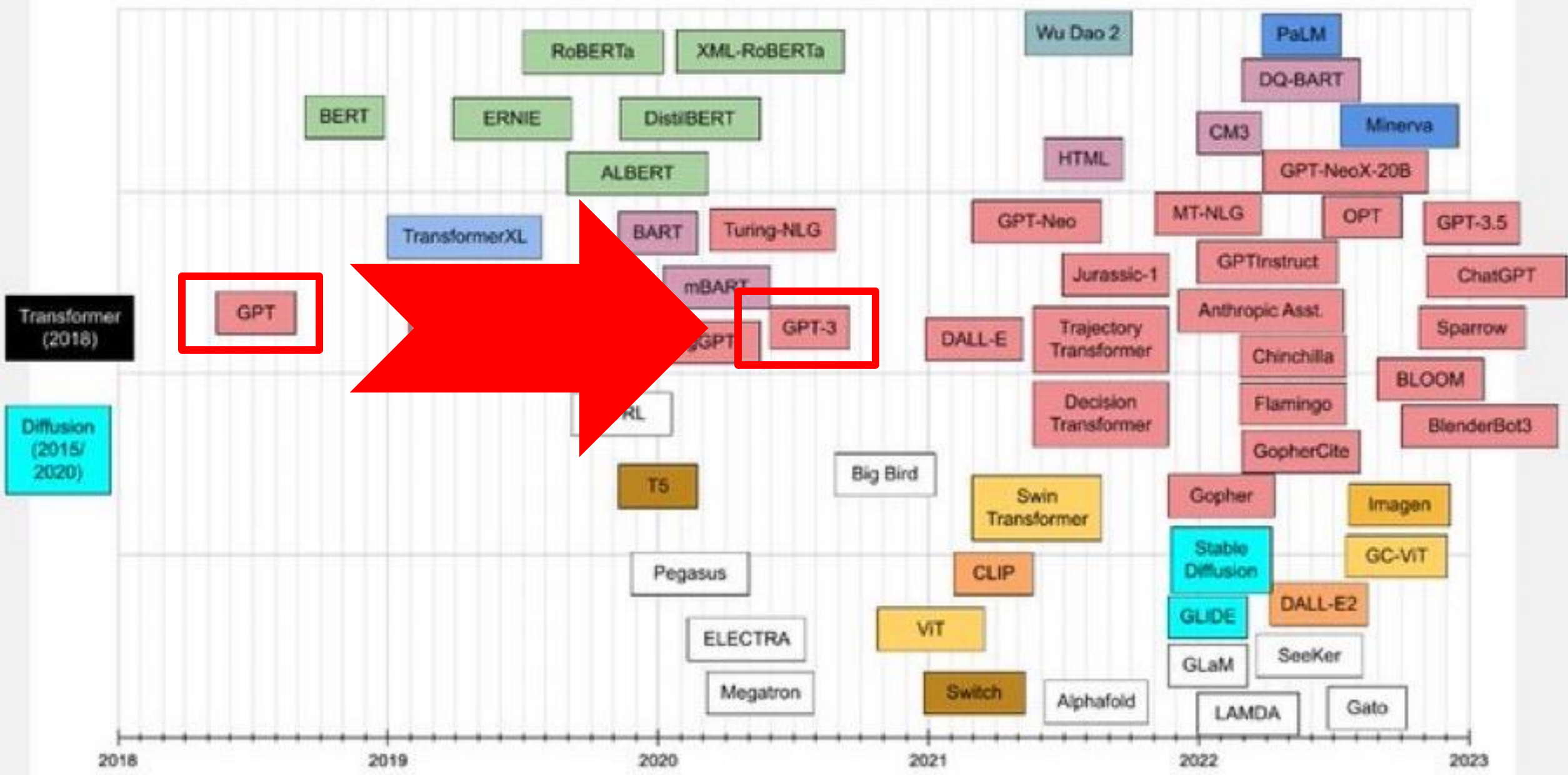


What happened in NLP over the last six years (2019-2025)?

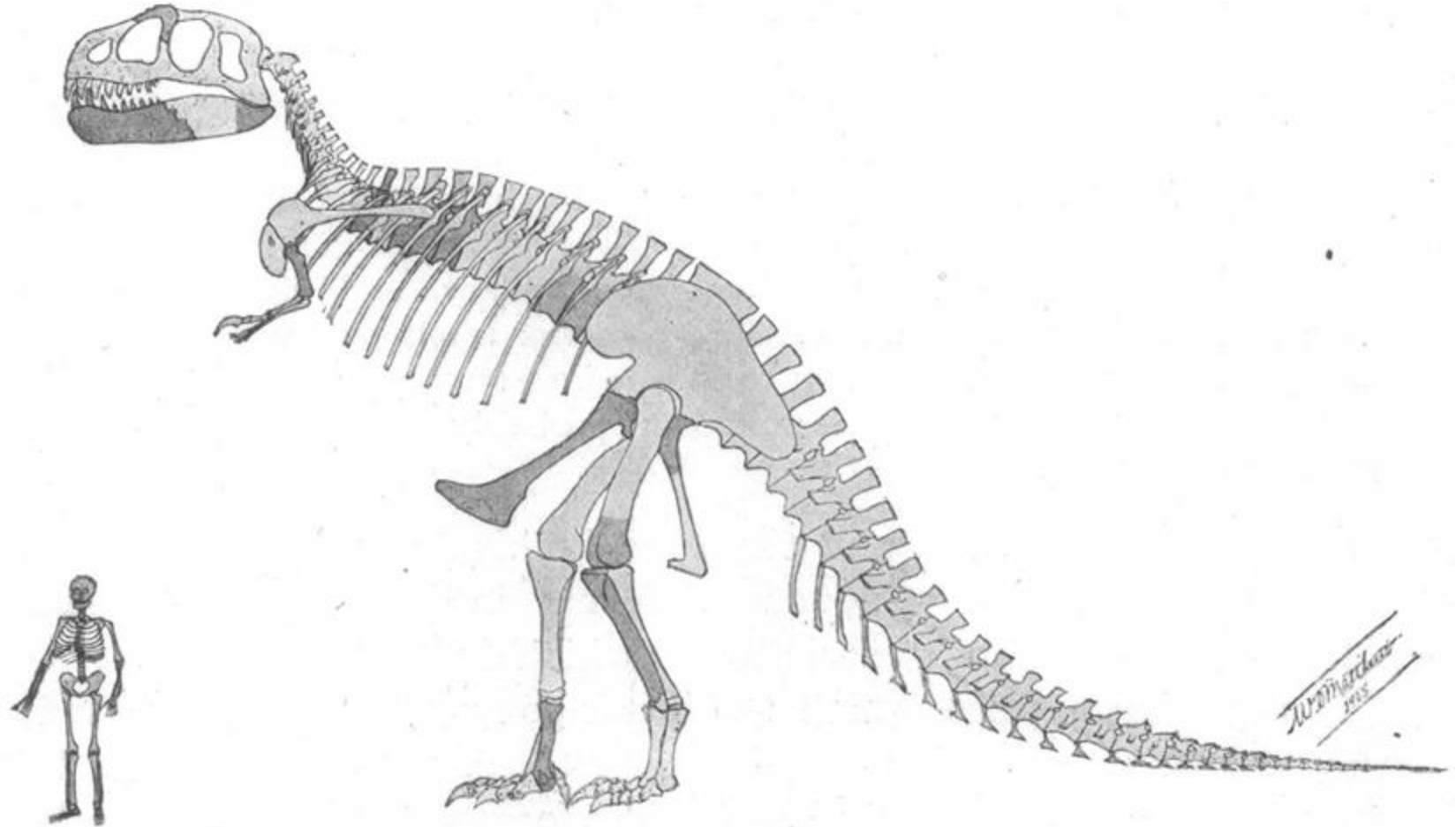








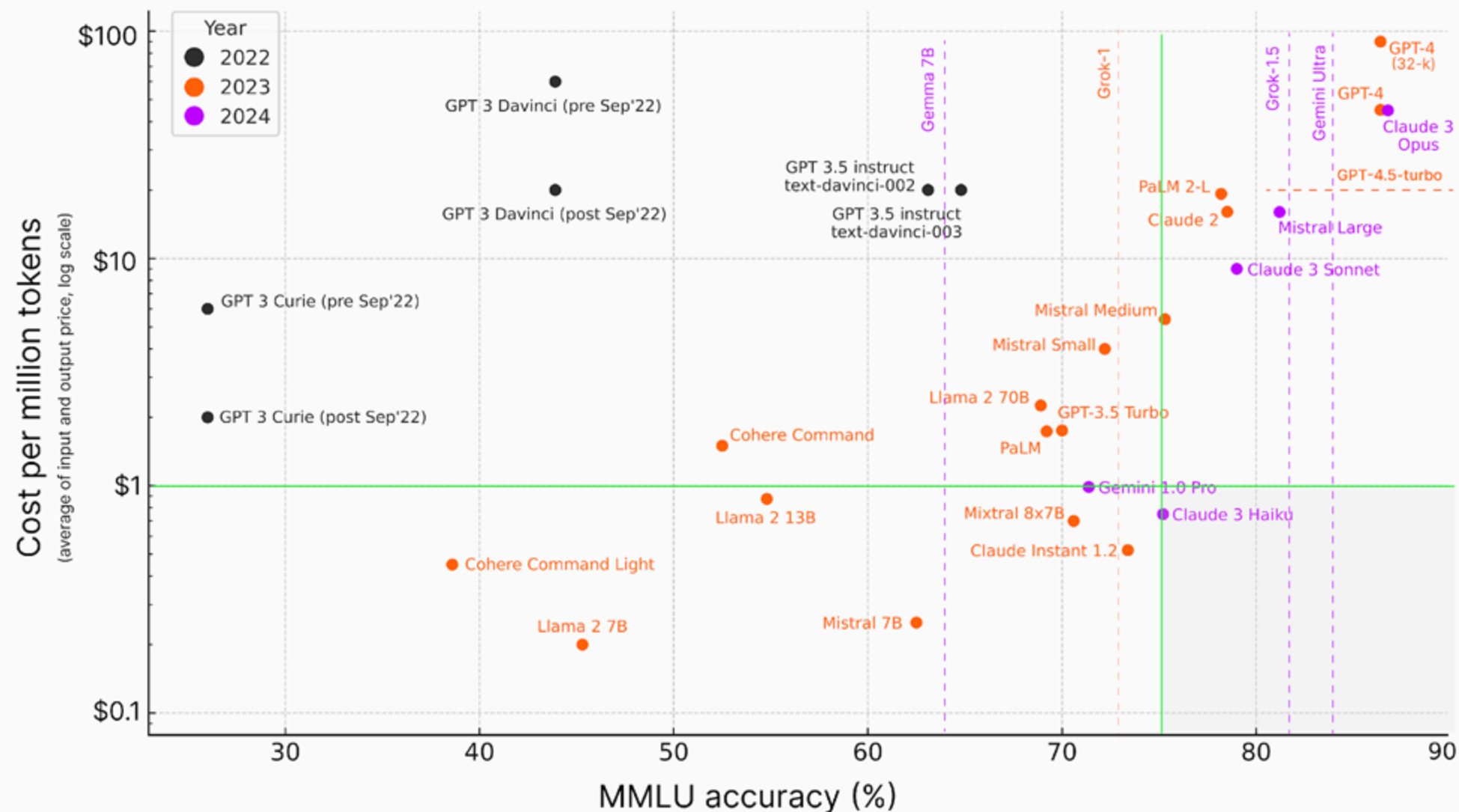
Scaling up!

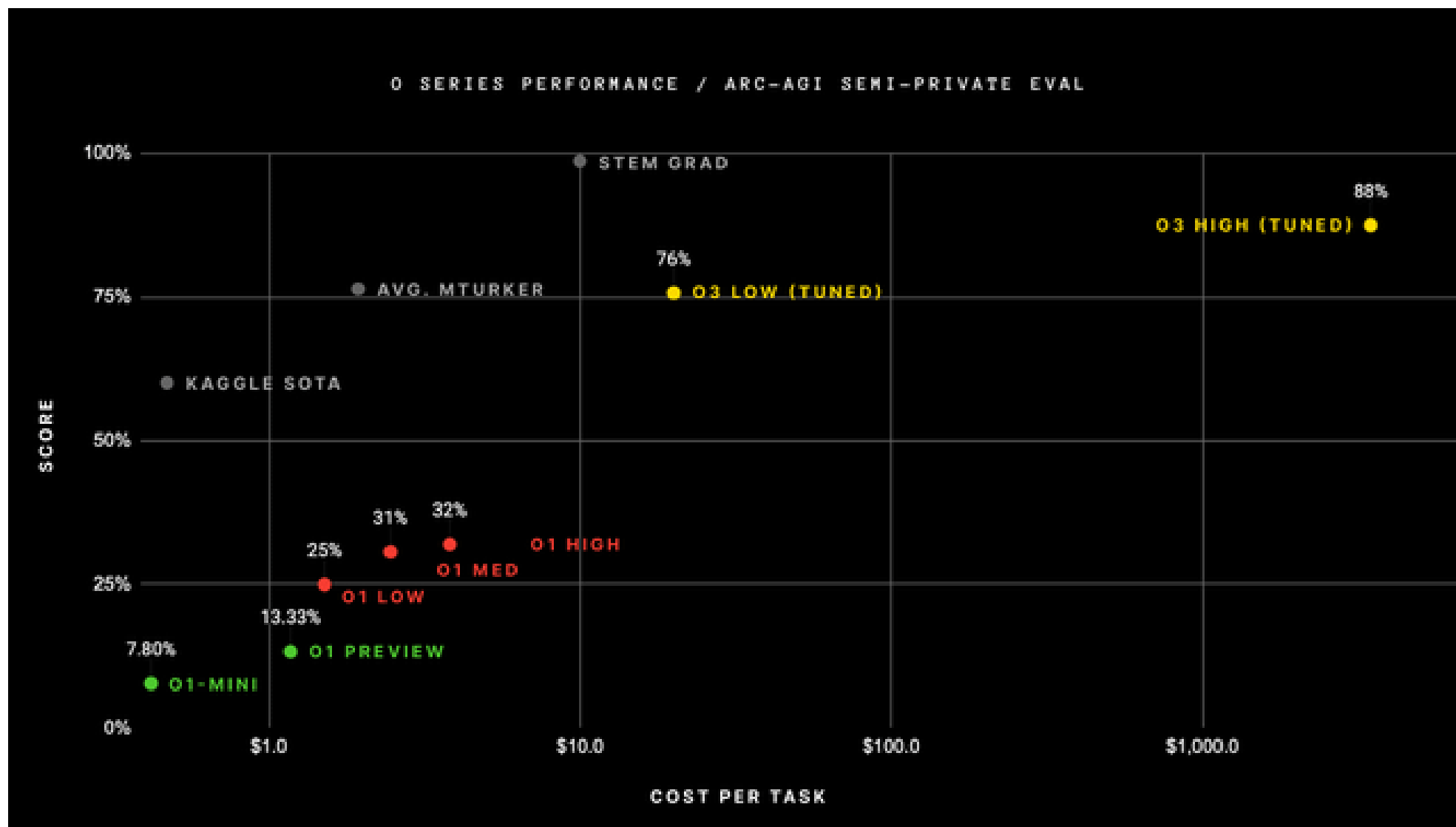


GPT-2
1.5B Parameters

GPT-3
175B Parameters

MMLU Performance vs. Cost Over Time (2022-2024)



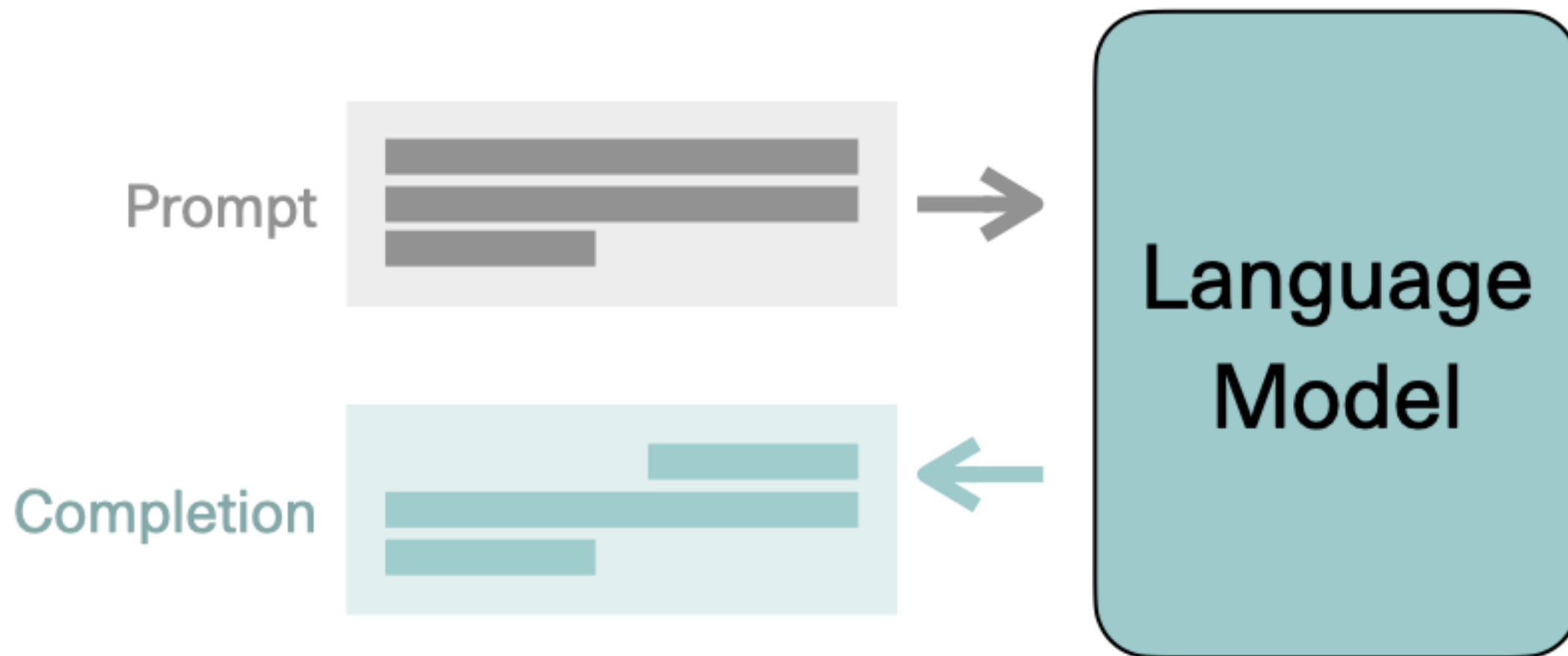


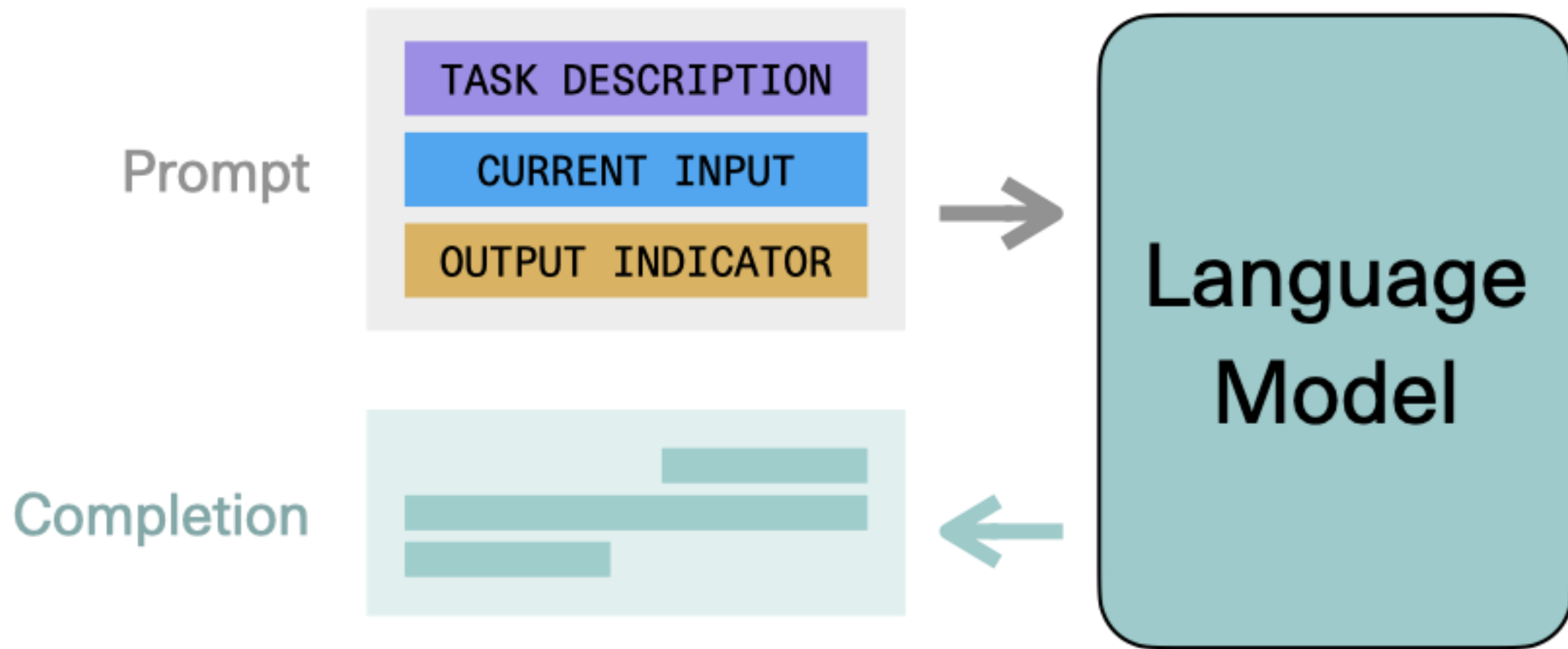
The Leading Players

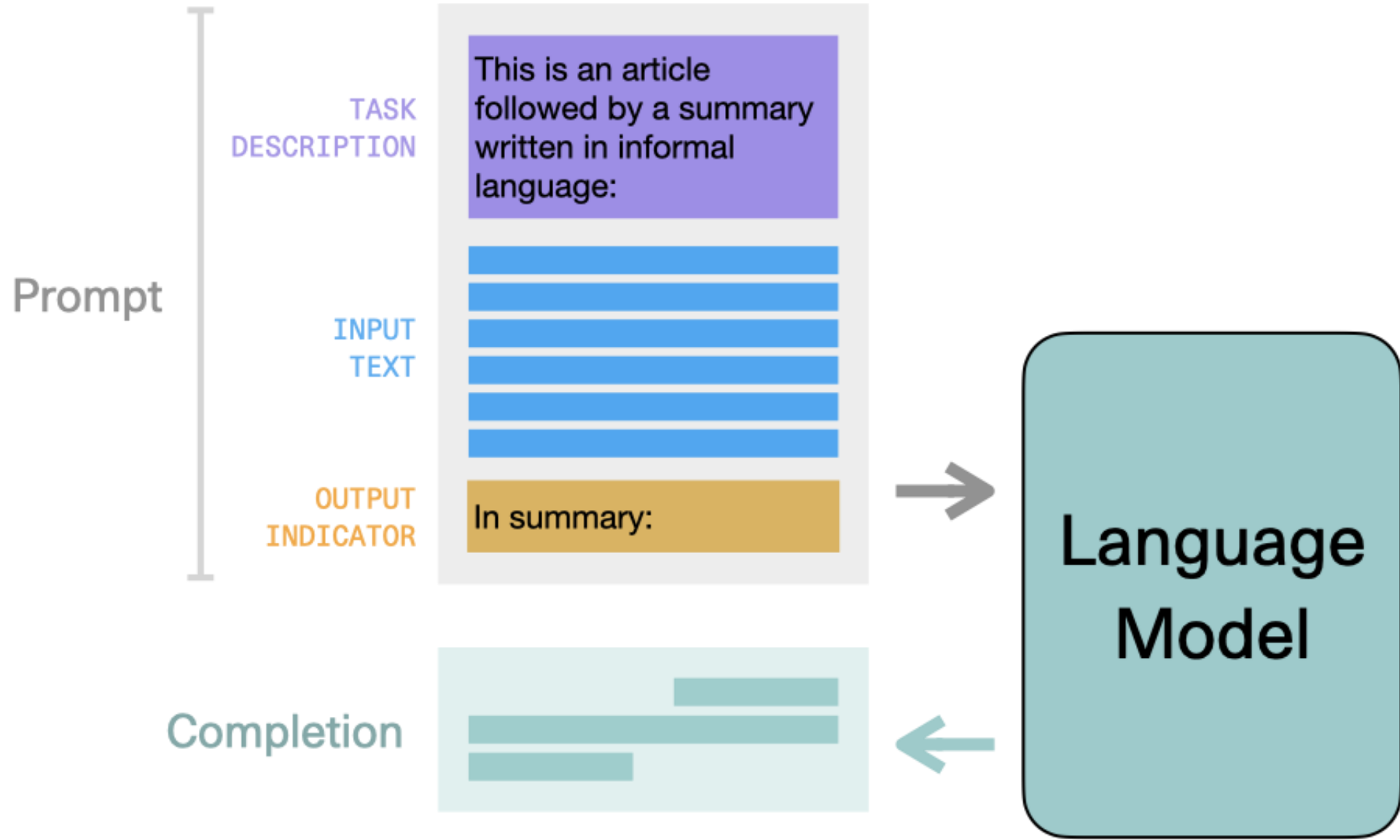


ANTHROPIC



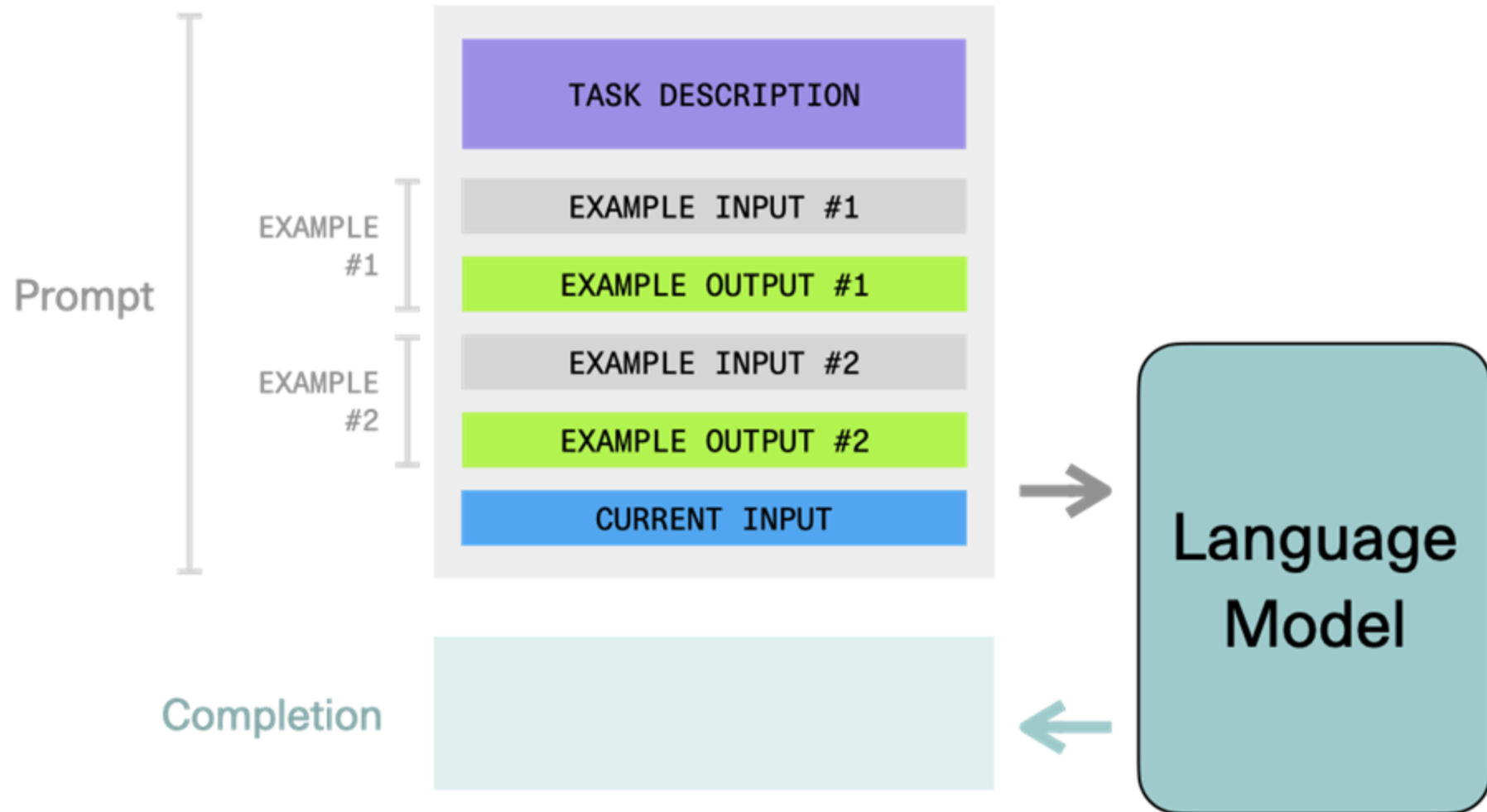






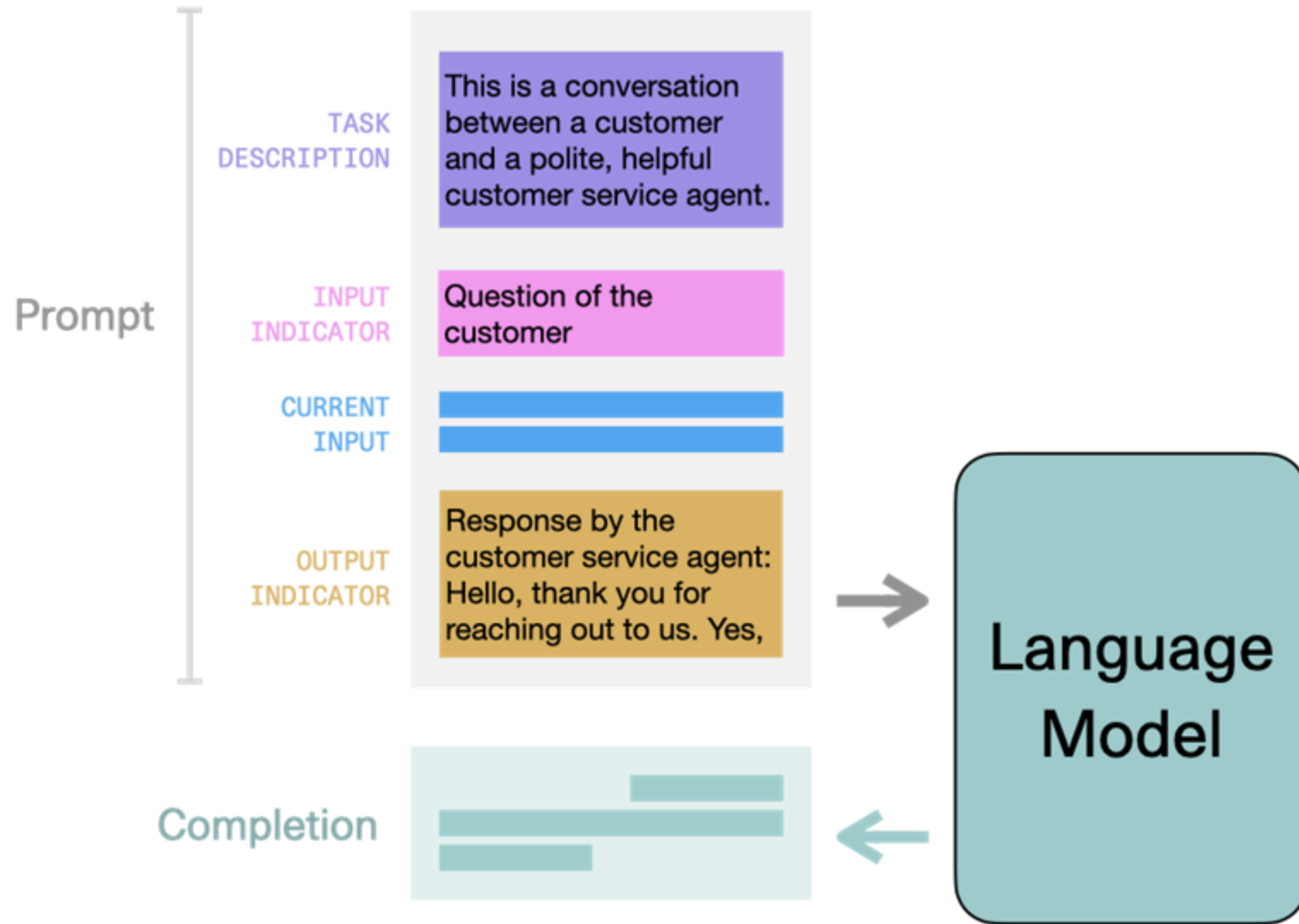
<https://docs.cohere.ai/prompt-engineering-wiki/>





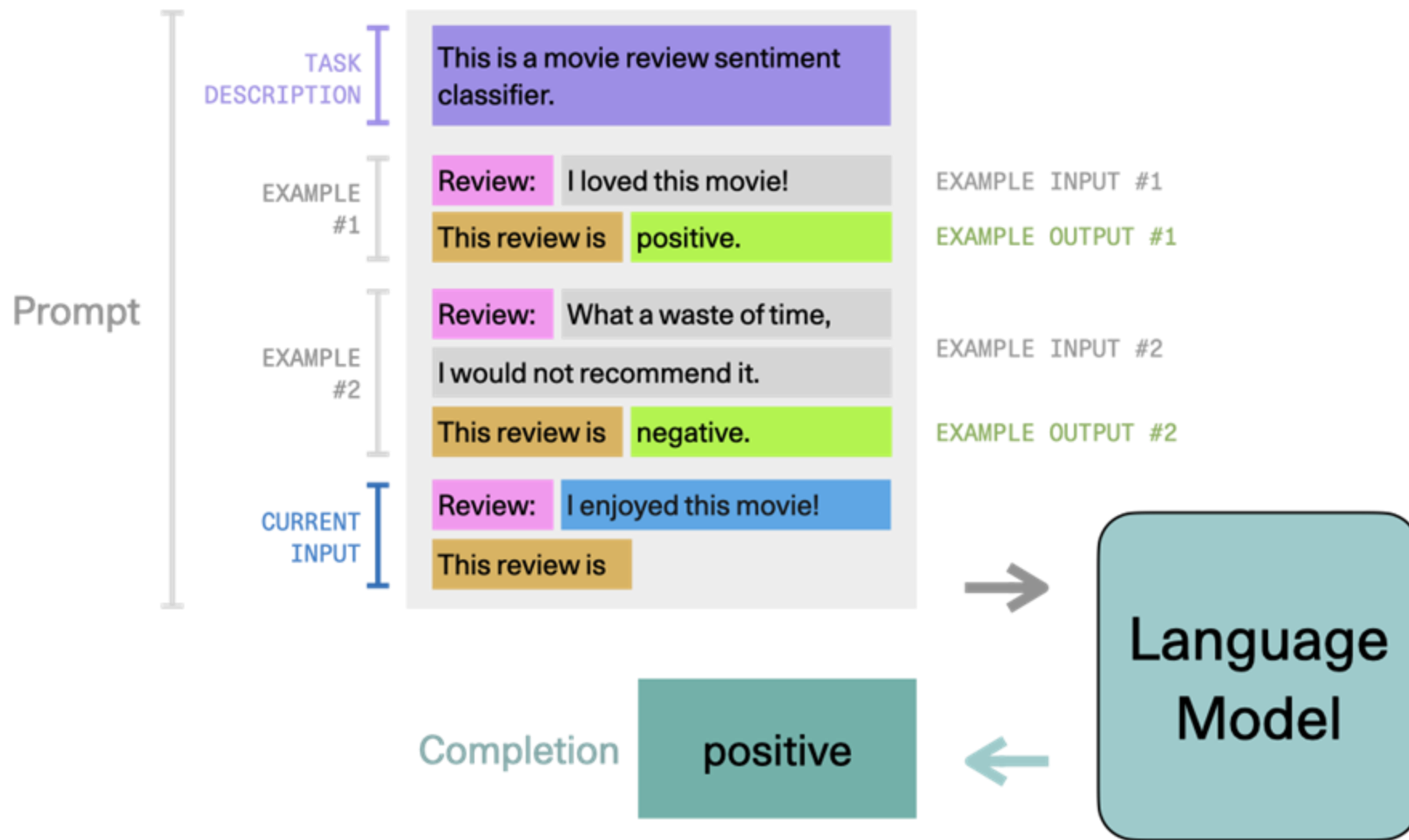
<https://docs.cohere.ai/prompt-engineering-wiki/>





<https://docs.cohere.ai/prompt-engineering-wiki/>





Sentence classification via Prompting

Input Temperature:0	Classify the sentences below as positive, negative, neutral: Sentence: I enjoyed this movie despite the gory violence. Classification: Positive — Sentence: It is beyond my comprehension how such a movie grossed over \$100 USD. Classification: Negative — Sentence: I can't say I hate it or love it. Classification: Neutral — Sentence: I endured the silly plot purely because of the excellent acting of the hero. Classification:
------------------------	---

<https://towardsdatascience.com/a-quiet-shift-in-the-nlp-ecosystem-84672b8ec7af>



Text Summarization via Prompting

Input Temperature:0	<p>Summarize this for a second-grade student:</p> <p>An atom is the smallest unit of ordinary matter that forms a chemical element.[1] Every solid, liquid, gas, and plasma is composed of neutral or ionized atoms. Atoms are extremely small, typically around 100 picometers across. They are so small that accurately predicting their behavior using classical physics—as if they were tennis balls, for example—is not possible due to quantum effects.</p>
------------------------	---

<https://towardsdatascience.com/a-quiet-shift-in-the-nlp-ecosystem-84672b8ec7af>



Relation Extraction via Prompting

Input	Identify drugs, diseases and genes as well as the relations between them.
Temperature:0	Sentence: Imatinib is used to treat cancer Entity1: Imatinib (drug) Entity2: cancer (disease) Relation: treat -- Sentence: Imatinib can cause abdominal pain Entity1: Imatinib (drug) Entity2: abdominal pain (disease) Relation: cause -- Sentence: EGFR is overexpressed in many forms of cancers Entity1: EGFR (gene) Entity2: cancers (disease) Relation: overexpressed -- Sentence: Dasatinib, nilotinib is used as a combination therapy for some cancers Entity1: Dasatinib (drug), nilotinib (drug) Entity2: cancers (disease) Relation: combination therapy -- Sentence: Her hypophysitis secondary to ipilimumab was well managed with supplemental hormones Entity1:

<https://towardsdatascience.com/a-quiet-shift-in-the-nlp-ecosystem-84672b8ec7af>



Email Generation via Prompting

Input Temperature:0	Generate full emails from simple commands. Here are some examples: Command: Thank John for his mother's day gift Email: John, Thank you so much for your thoughtful gift. I hope to see you soon - Mom. -- Command: Tell Sam to email the invoice Email:
------------------------	---

<https://towardsdatascience.com/a-quiet-shift-in-the-nlp-ecosystem-84672b8ec7af>



Code Generation via Prompting

Prompt

```
// Translate from C to Python
int add_one ( int x ){
    int m = 1;
    while ( x & m ) {
        x = x ^ m;
        m <<= 1;
    }
    x = x ^ m;
    return x; }
```

Model Response

<https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html>



Mathematical Reasoning via Prompting

Input Temperature:0	Calculate $4.5e1 + 1.5e2$
------------------------	---------------------------

 Jurassic-X (7.5B) →  Calculator

$4.5e1 + 1.5e2 = 195$

&frasl Explain answer

$x = (4.5e1 + 1.5e2)$

<https://towardsdatascience.com/a-quiet-shift-in-the-nlp-ecosystem-84672b8ec7af>



Chain-of-Thought Prompting

Few-shot CoT

Standard Prompting

Example 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?

Example Output


A: The answer is 11.

Prompt

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

Model Response

The answer is 50.



Standard prompting versus chain-of-thought prompting for an example grade-school math problem. Chain-of-thought prompting decomposes the prompt for a multi-step reasoning problem into intermediate steps (highlighted in yellow), similar to how a person would approach it.

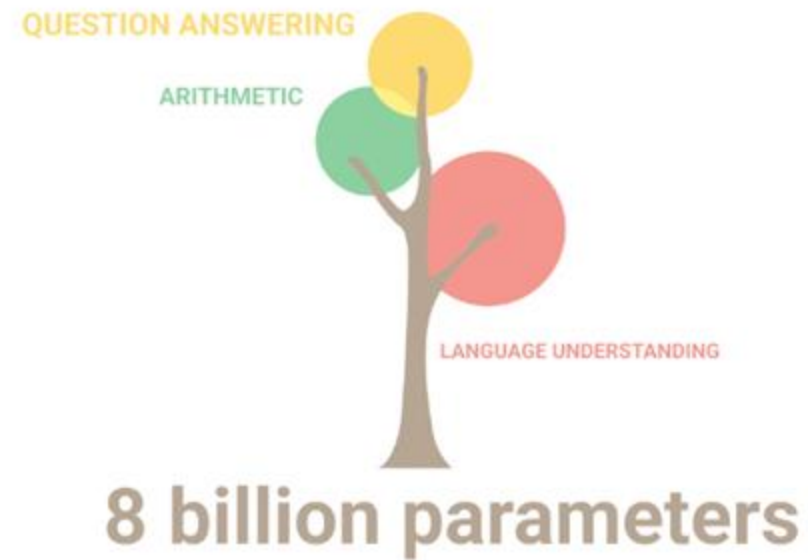
<https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html>



Law in language model

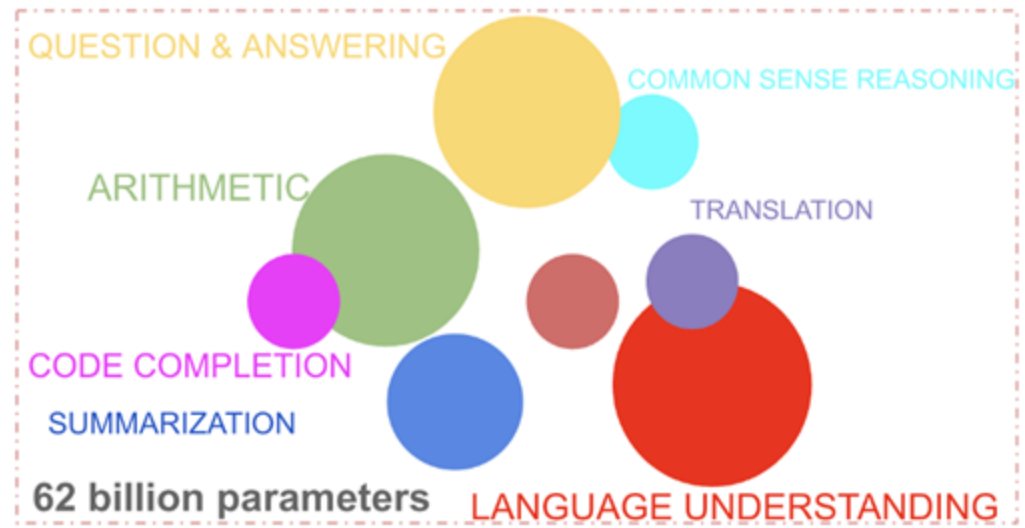
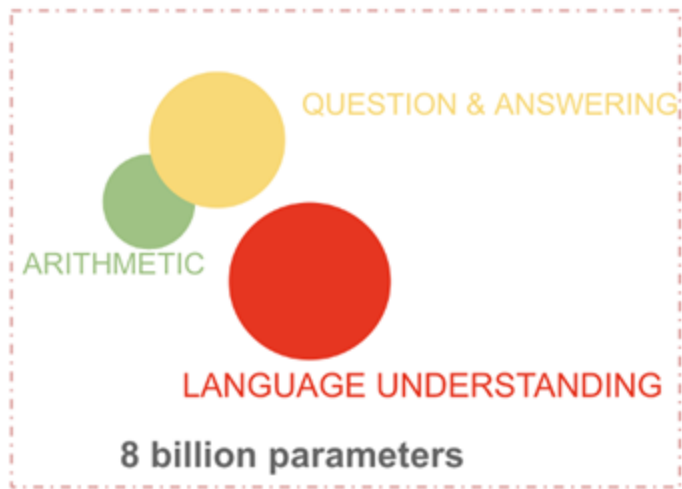
Figure 1: Exponential growth of number of parameters in DL models





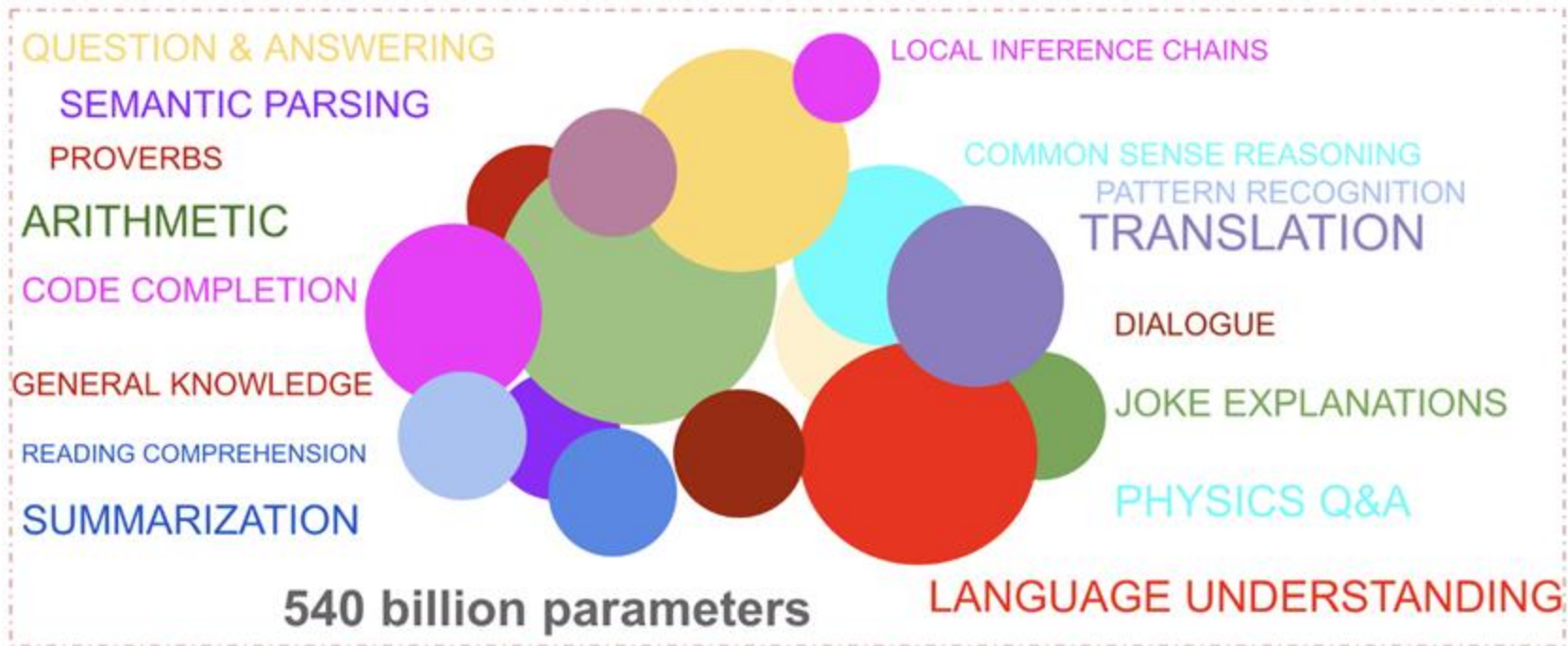
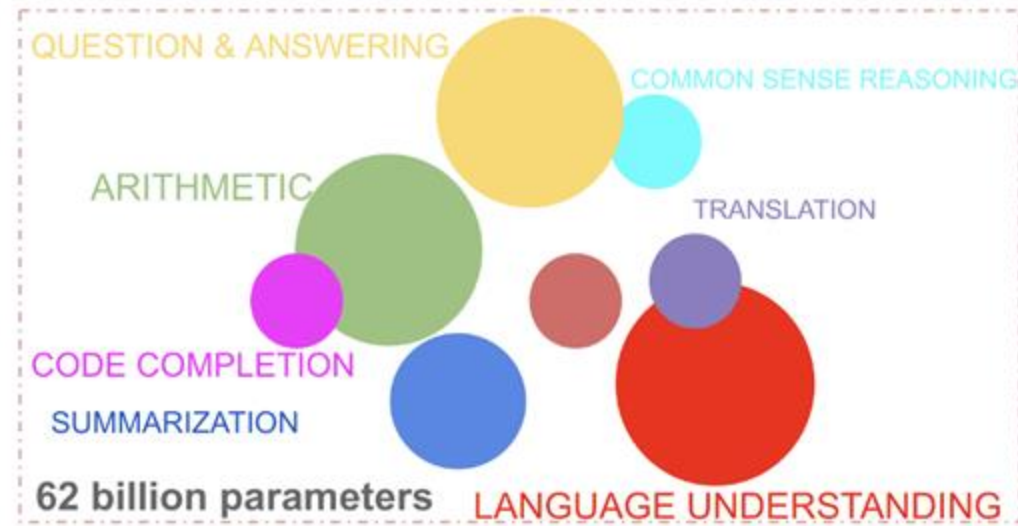
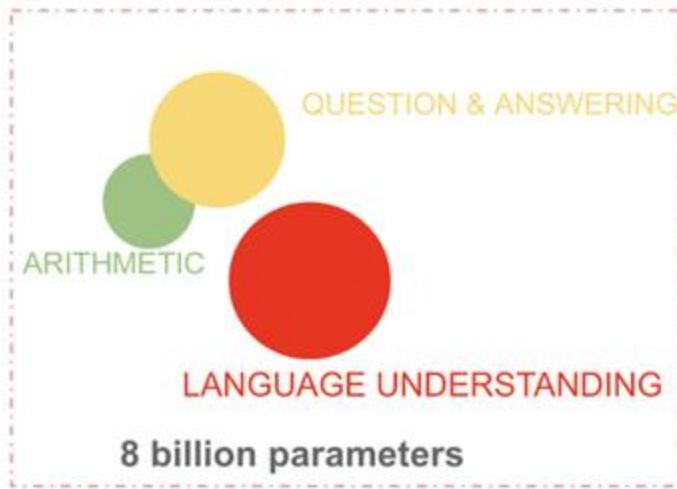
<https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html>





<https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html>

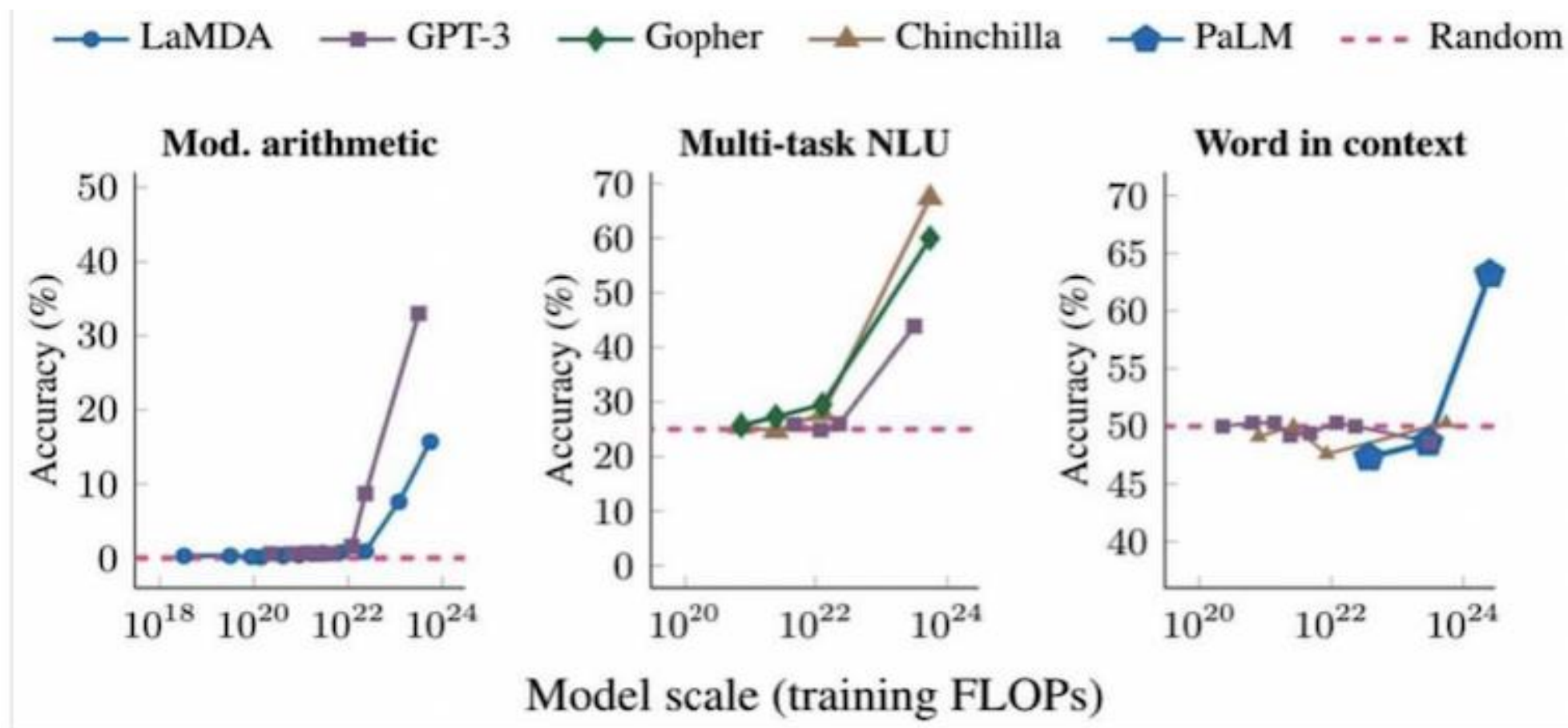




<https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html>

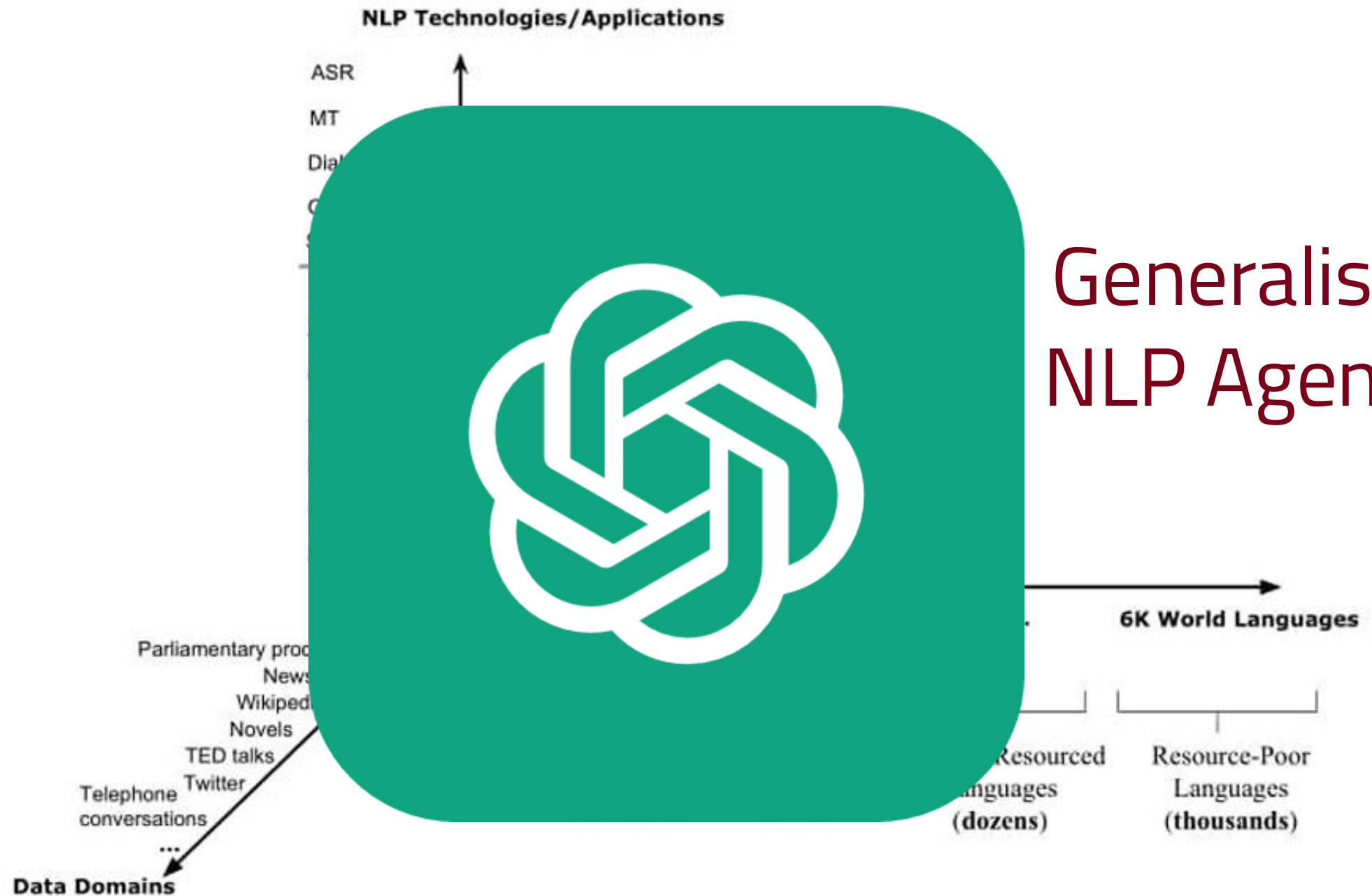


Emergent behavior from Scaling Law



Jeff Dean <https://ai.googleblog.com/2023/01/google-research-2022-beyond-language.html>





Generalist NLP Agent

Generalist AI Agent

PROMPT:

Email: "Dear Alan, how should I learn more about AI? Best, Jim"

Respond to this email diplomatically and comprehensively.



Dear Jim,

Thank you for your email. Learning about AI can be a fascinating and rewarding experience. There are many ways to do this, and the best approach will depend on your individual interests and goals. Here are a few suggestions to get you started:

- If you are new to AI, one of the best ways to learn more is to start with some of the foundational concepts and principles. This can help you gain a better understanding of how AI works, and what it is capable of. Some good resources for this include online courses, books, and tutorials.
- Another way to learn about AI is to explore some of the applications and

Try again



PROMPT:

Add an extra column. In the new column, insert an emoji symbolising the movie.



Here is a table showing ten slow-paced and relaxing movies set at night, with an added column for an emoji symbolizing the movie:

Movie Title	Director	Year Released	Emoji Symbol
Moonlight	Barry Jenkins	2016	🌙
Arrival	Denis Villeneuve	2016	👁️
The Tree of Life	Terrence Malick	2011	🌳
A Ghost Story	David Lowery	2017	👻
Her	Spike Jonze	2013	💻
The Theory of Everything	James Marsh	2014	🌌
Melancholia	Lars von Trier	2011	🌩️
The Road	John Hillcoat	2009	🚗
The Shawshank Redemption	Frank Darabont	1994	🏠
The Elephant Man	Try again	1980	🐘

<https://chat.openai.com/chat>



Generalist AI across different modalities



Jeff Dean <https://ai.googleblog.com/2023/01/google-research-2022-beyond-language.html>



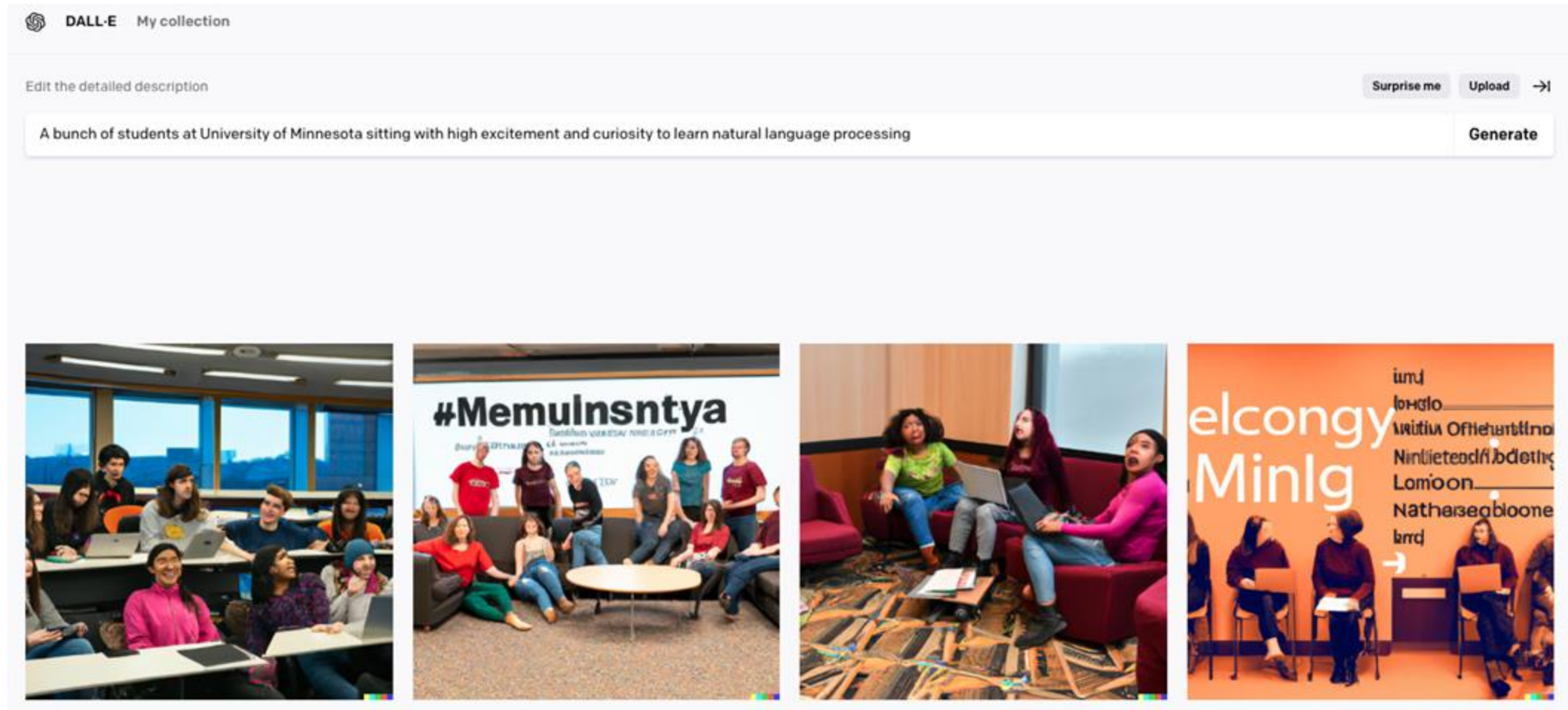
Scaling Law in Vision-Language Model



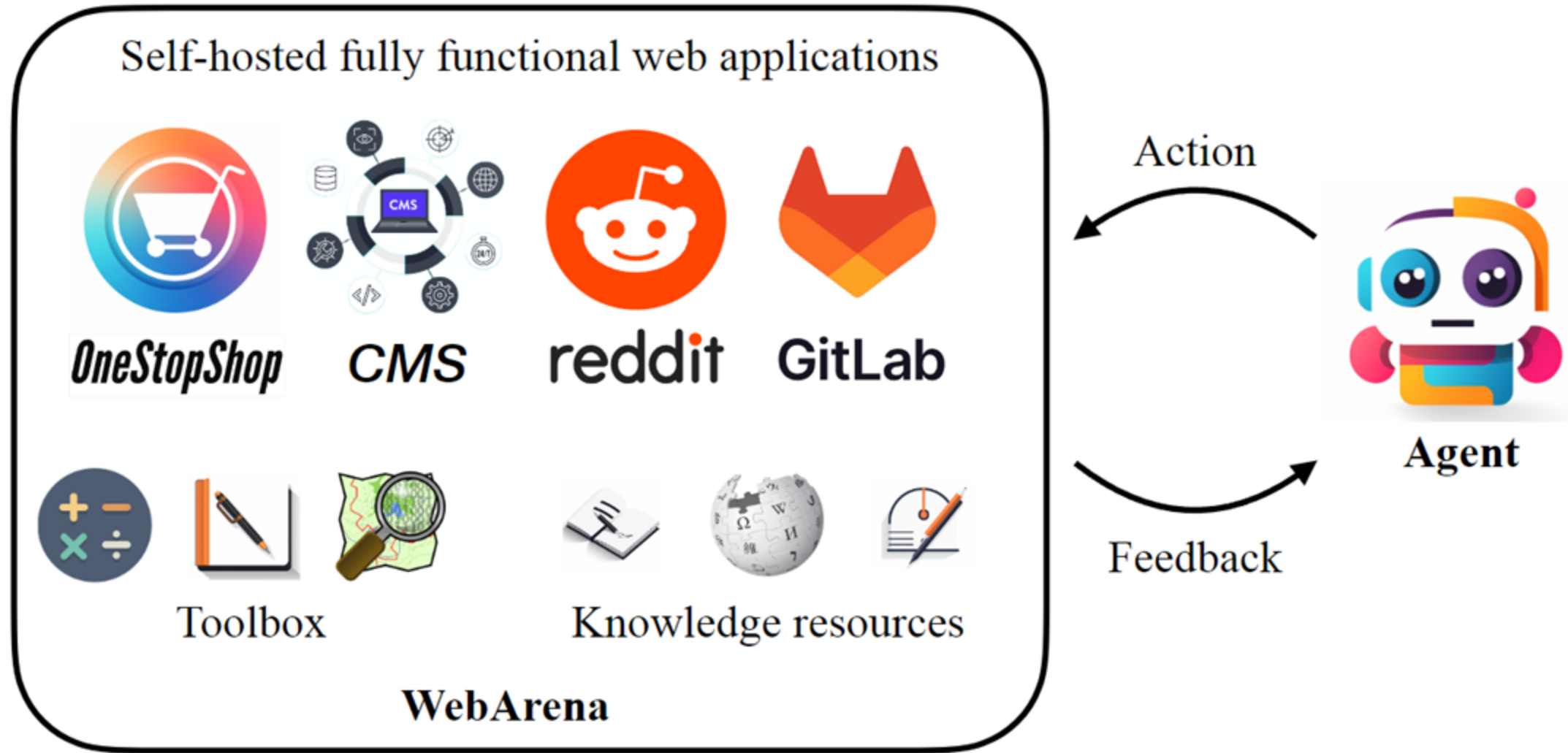
Figure 4. The generated image for the text *"A portrait photo of a kangaroo wearing an orange hoodie and blue sunglasses standing on the grass in front of the Sydney Opera House holding a sign on the chest that says Welcome Friends!"*. Note the model gets the text in the image "welcome friends" correct at 20B.

<https://towardsdatascience.com/a-quiet-shift-in-the-nlp-ecosystem-84672b8ec7af>

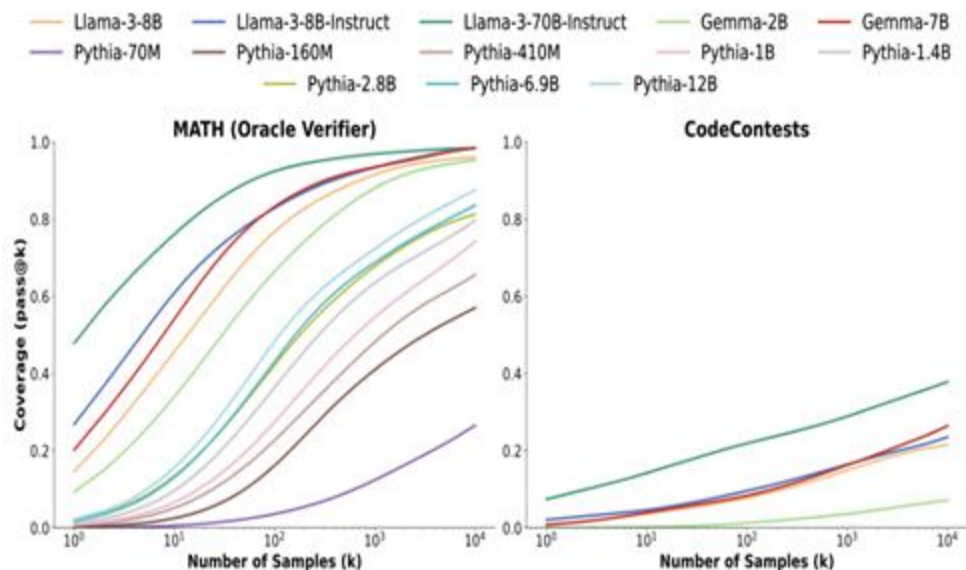
Beyond Language



Web Based Agent



Reasoning (Test-time compute/scaling)



Model	AIME 2024		MATH-500	GPQA Diamond	LiveCode Bench	CodeForces
	pass@1	cons@64	pass@1	pass@1	pass@1	rating
GPT-4o-0513	9.3	13.4	74.6	49.9	32.9	759
Claude-3.5-Sonnet-1022	16.0	26.7	78.3	65.0	38.9	717
OpenAI-o1-mini	63.6	80.0	90.0	60.0	53.8	1820
QwQ-32B-Preview	50.0	60.0	90.6	54.5	41.9	1316
DeepSeek-R1-Distill-Qwen-1.5B	28.9	52.7	83.9	33.8	16.9	954
DeepSeek-R1-Distill-Qwen-7B	55.5	83.3	92.8	49.1	37.6	1189
DeepSeek-R1-Distill-Qwen-14B	69.7	80.0	93.9	59.1	53.1	1481
DeepSeek-R1-Distill-Qwen-32B	72.6	83.3	94.3	62.1	57.2	1691
DeepSeek-R1-Distill-Llama-8B	50.4	80.0	89.1	49.0	39.6	1205
DeepSeek-R1-Distill-Llama-70B	70.0	86.7	94.5	65.2	57.5	1633

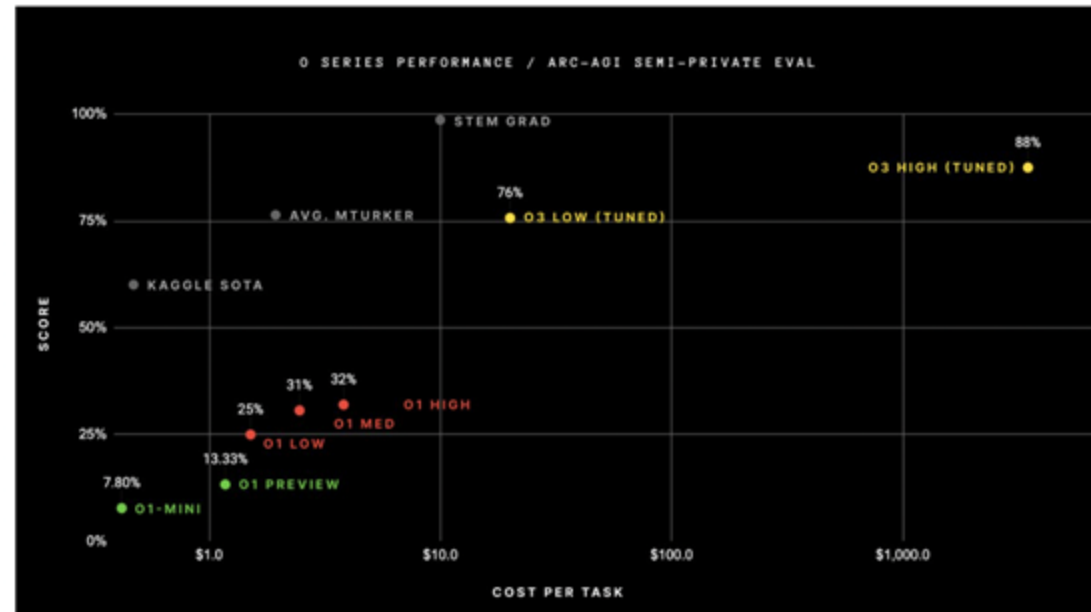
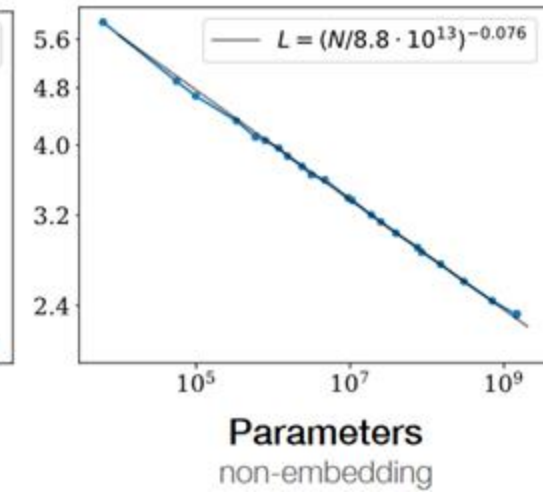
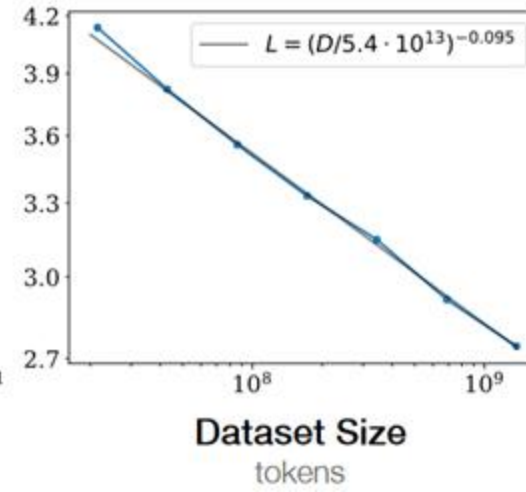
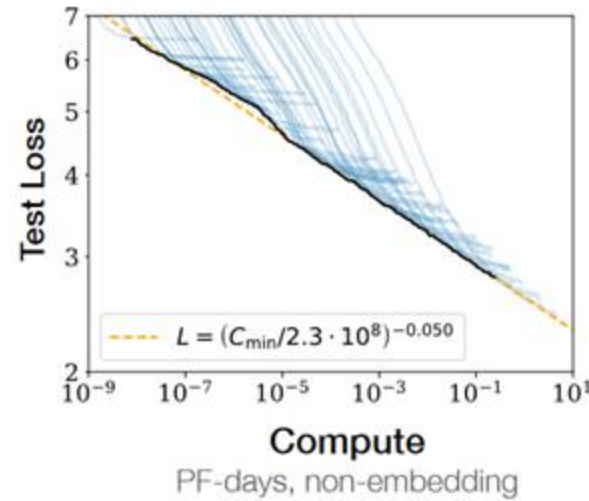
Table 5 | Comparison of DeepSeek-R1 distilled models and other comparable models on reasoning-related benchmarks.



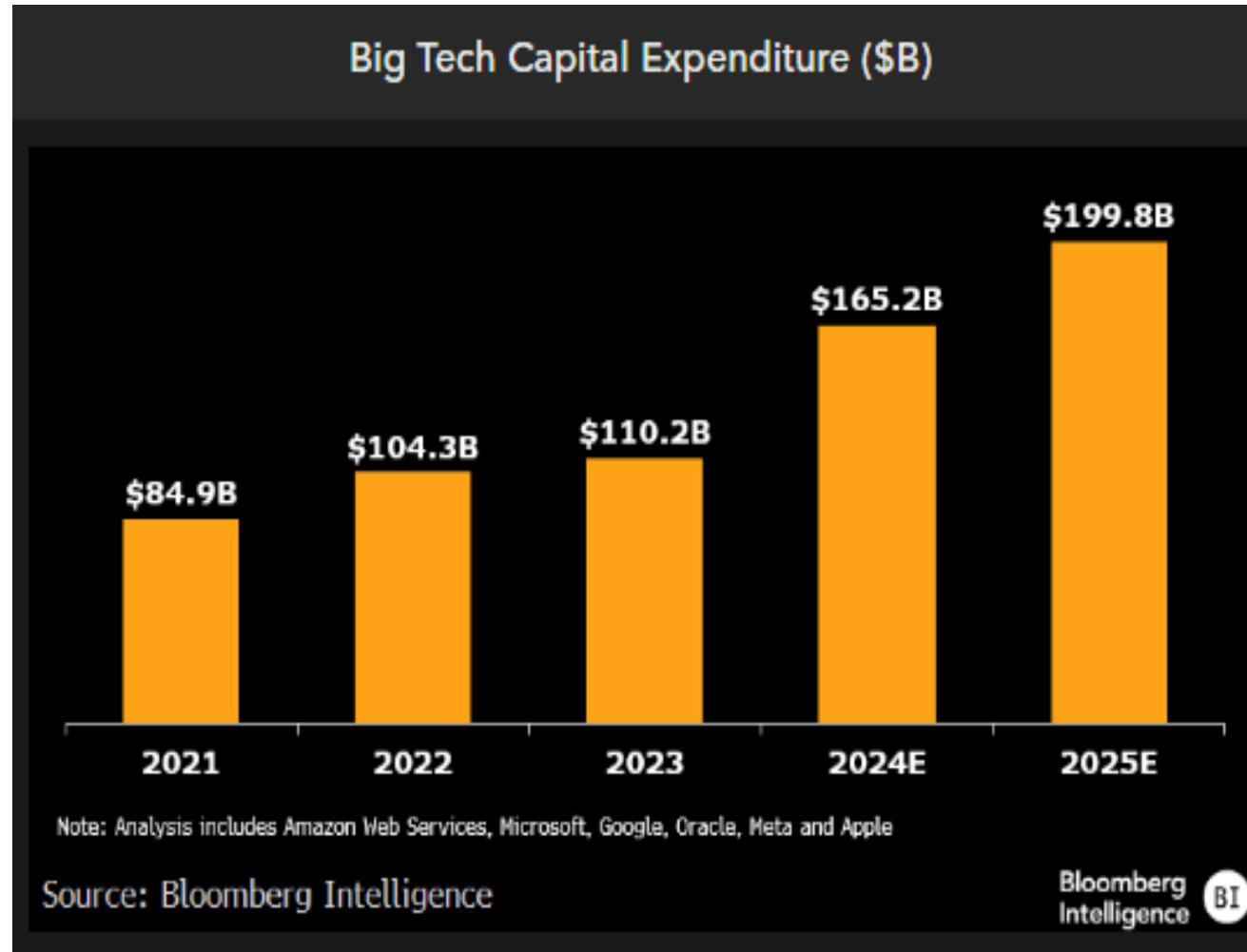
Limits and Impacts of LLMs to “Us”



Limits of scaling

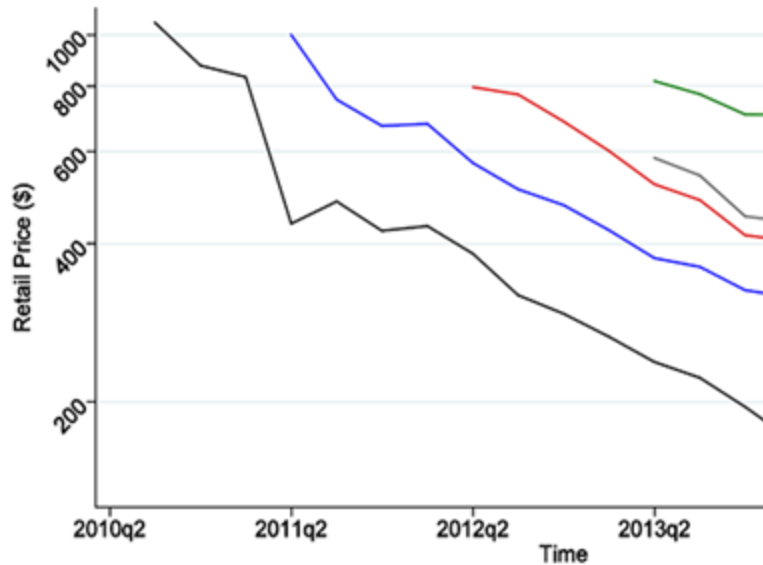


AI “arms race” by Big Tech



<https://www.bloomberg.com/professional/insights/technology/big-tech-2025-capex-may-hit-200-billion-as-gen-ai-demand-booms/>

What if They are Wrong?



BUSINESS INSIDER

Getty Images; Tyler Le/BI

DISCOURSE | TECH

AI hype is crashing into reality. Stay calm.



By [Hugh Langley](#)

Sep 4, 2025, 1:17 AM PT



Add us on

A market correction. A wake-up call. A great digestion. Call it what you want: AI is going through it.

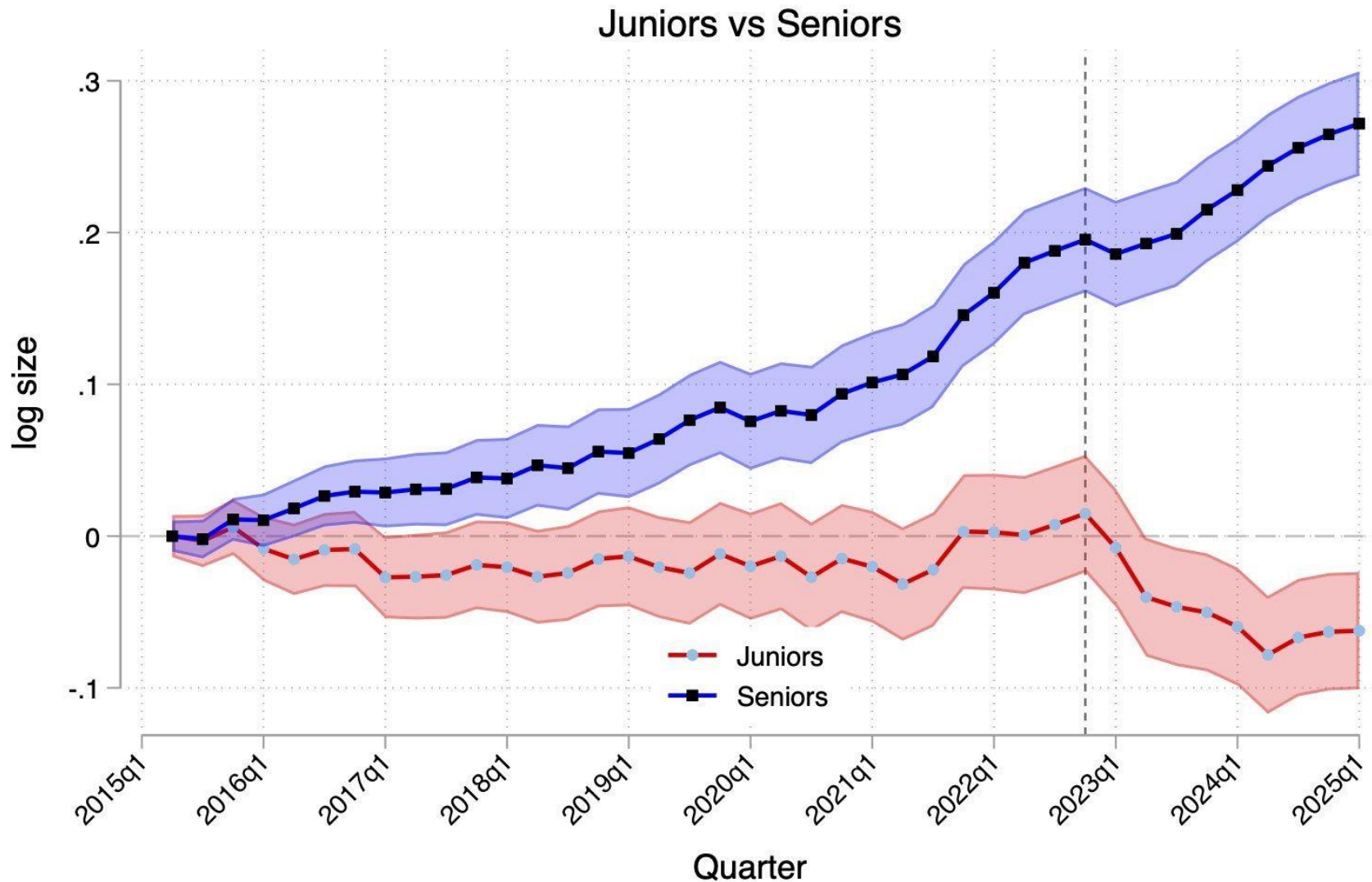
Two things appear to be happening in tandem. Businesses are starting to finally grasp what AI can — and importantly, can't

The Fever Dream of Imminent 'Superintelligence' Is Finally Breaking

Sept. 3, 2025

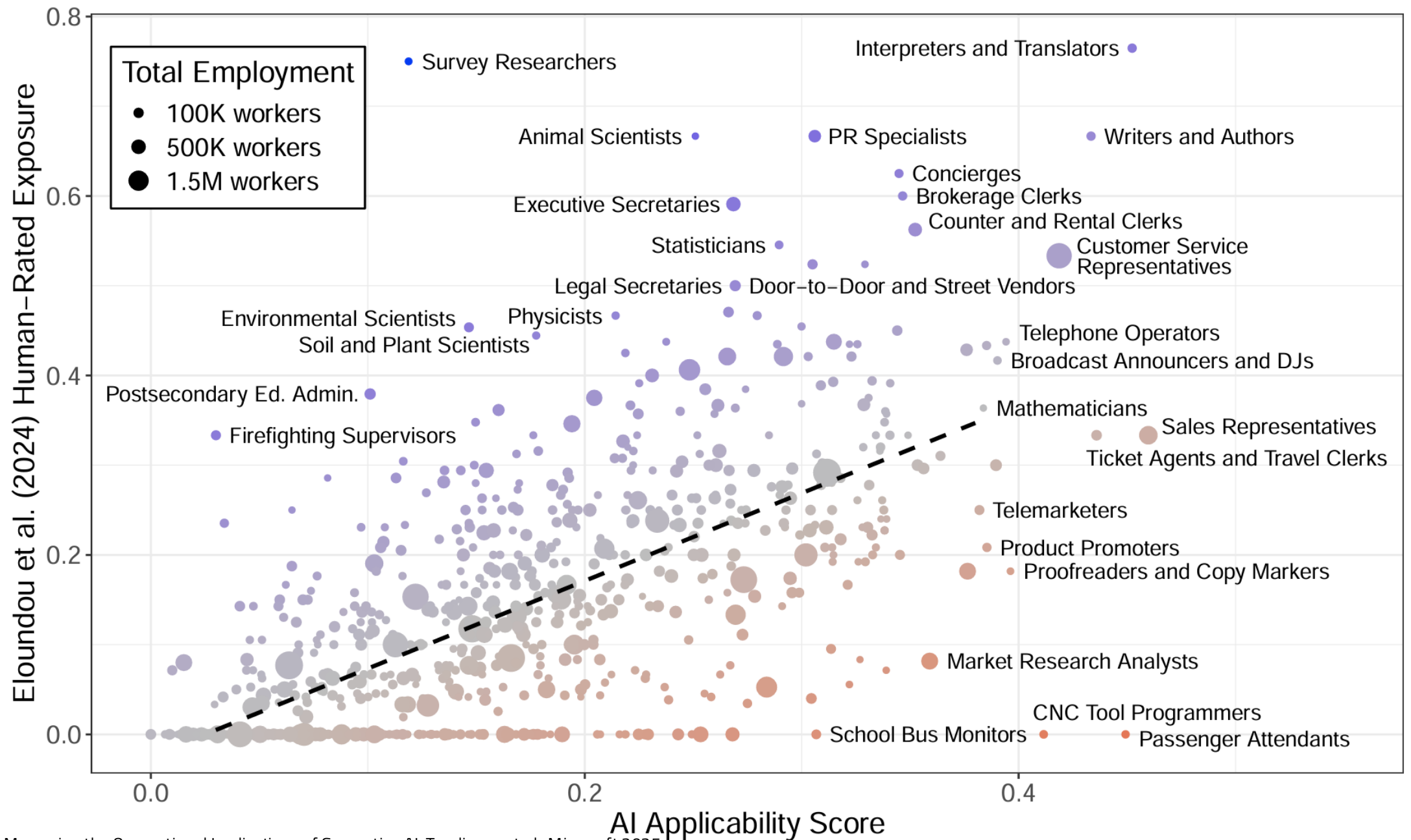


GenAI reduces the number of junior people hired



https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5425555





Working with AI: Measuring the Occupational Implications of Generative AI, Tomlinson et al., Microsoft 2025

Summary

- ❑ NLP is interdisciplinary
- ❑ Language consists of many levels of structure:
 - Phonology, syntax, semantics, discourse, pragmatics
- ❑ Processing language is difficult, due to
 - ambiguity, scales, sparsity, variation, implication, and representation
- ❑ Development of NLP models and representations grows rapidly
 - From rules to feature learning to RNNs to Transformers
- ❑ “Large” language models
 - Generalist AI or AGI via prompting and chat
 - Scaling law
 - Multimodal
 - Limitations? Future directions?

