

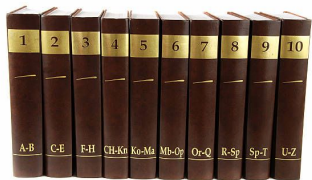
Language Models: Wherefrom? Whereto?

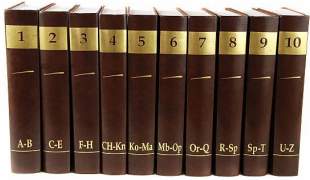
Anders Søgaard

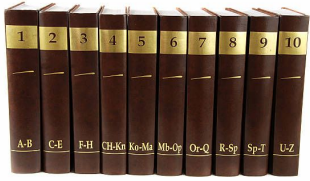
coASTaL



What are language
models for?







+





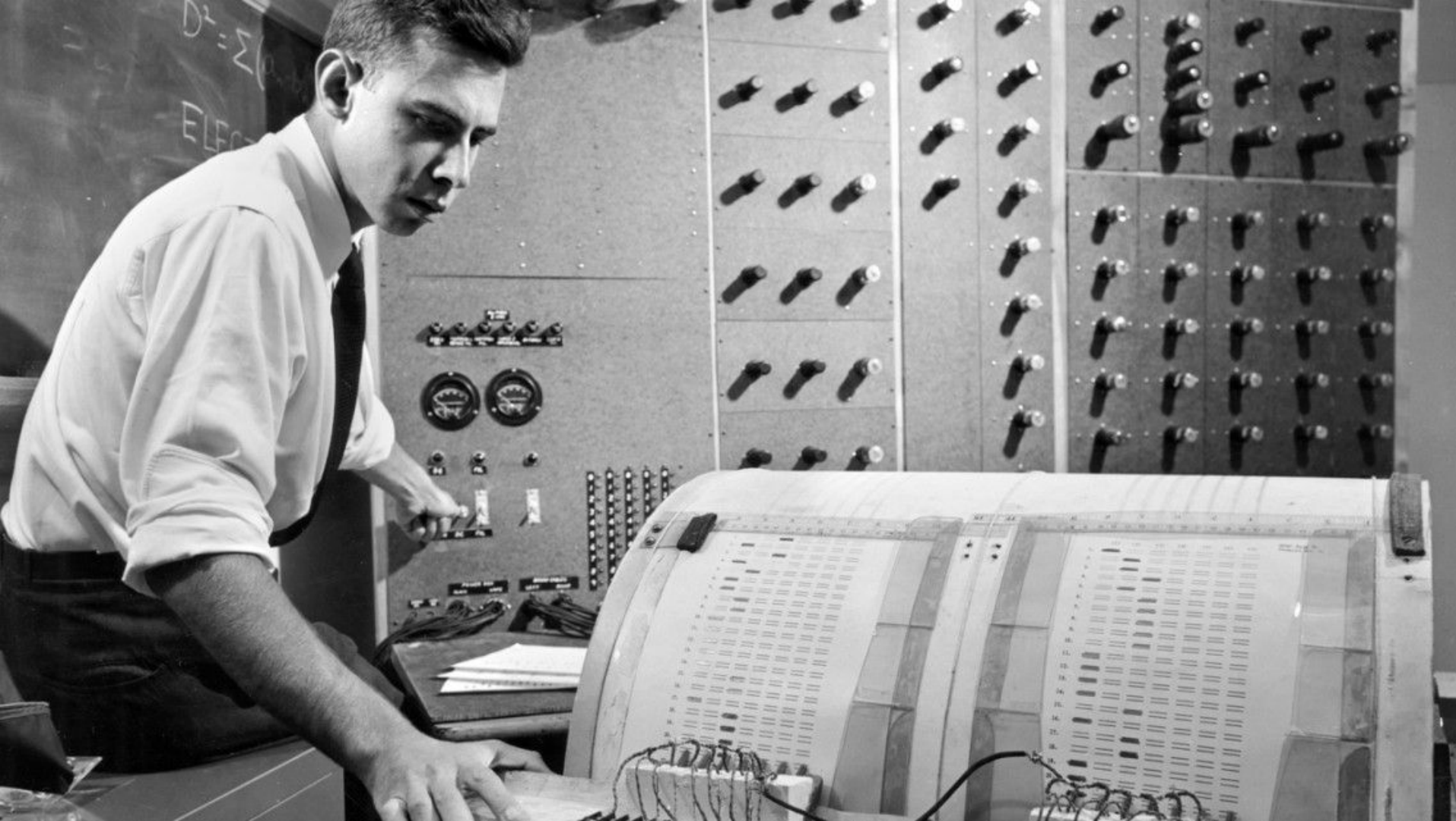
+



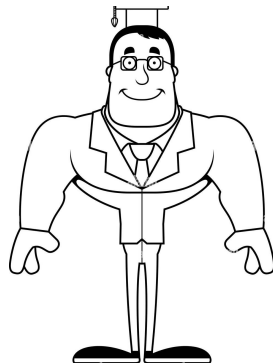
+



Where did they
come from?

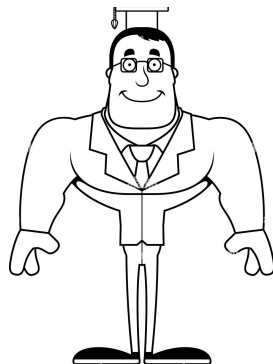


4
5
6
7
8



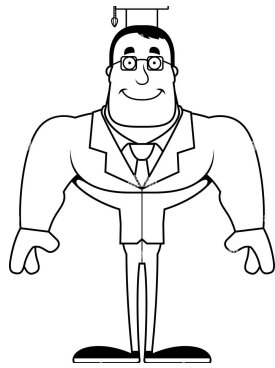
14
15
16
17
18

cat
dog
fish
cow
bee



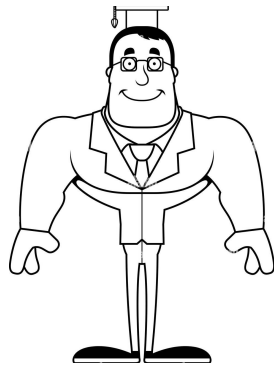
cute
bark
sea
holy
fly

4
5
6
7
8

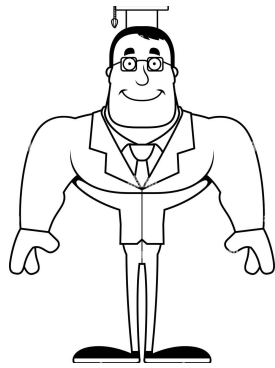


14
15
16
17
18

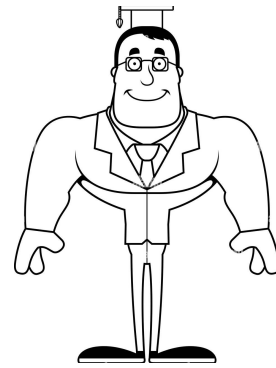
cat
dog
fish
cow
bee



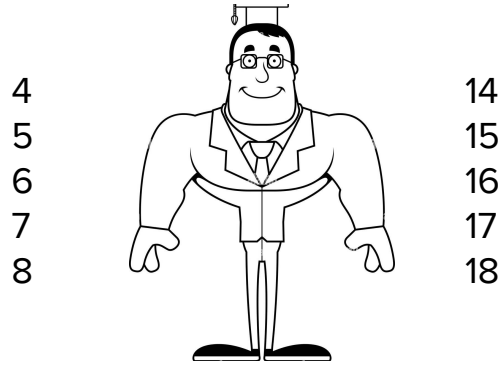
cute
bark
sea
holy
fly



very
loud
lay
sit
find



cute
run
egg
road
rose



But what if there's two functions
inside the guy?

$$f(x) = ax + b$$

$$g(x) = a'(f(x))+b'$$

	4	4-6	4-6-6	4-6-6-3
good	0.2	0.2	0.2	0.1
image	0.1	0.1		
home	0.1	0.1	0.1	0.1
hope	0.2	0.2	0.2	



He _____ walks, talks, ...
 {Legal} He _____ appeals, adjudicates, ...
 {9th century}, He _____ sayeth, hath, ...
 {Translation, English-German}, He _____ Er
 Translate 'He' into German... Er

2017

Transformers

Good scaling properties, exploiting GPUs, ability to model long-range dependencies.

2018

Conditional language models

Language prefixes in machine translation, transfer learning, style transfer.

2019

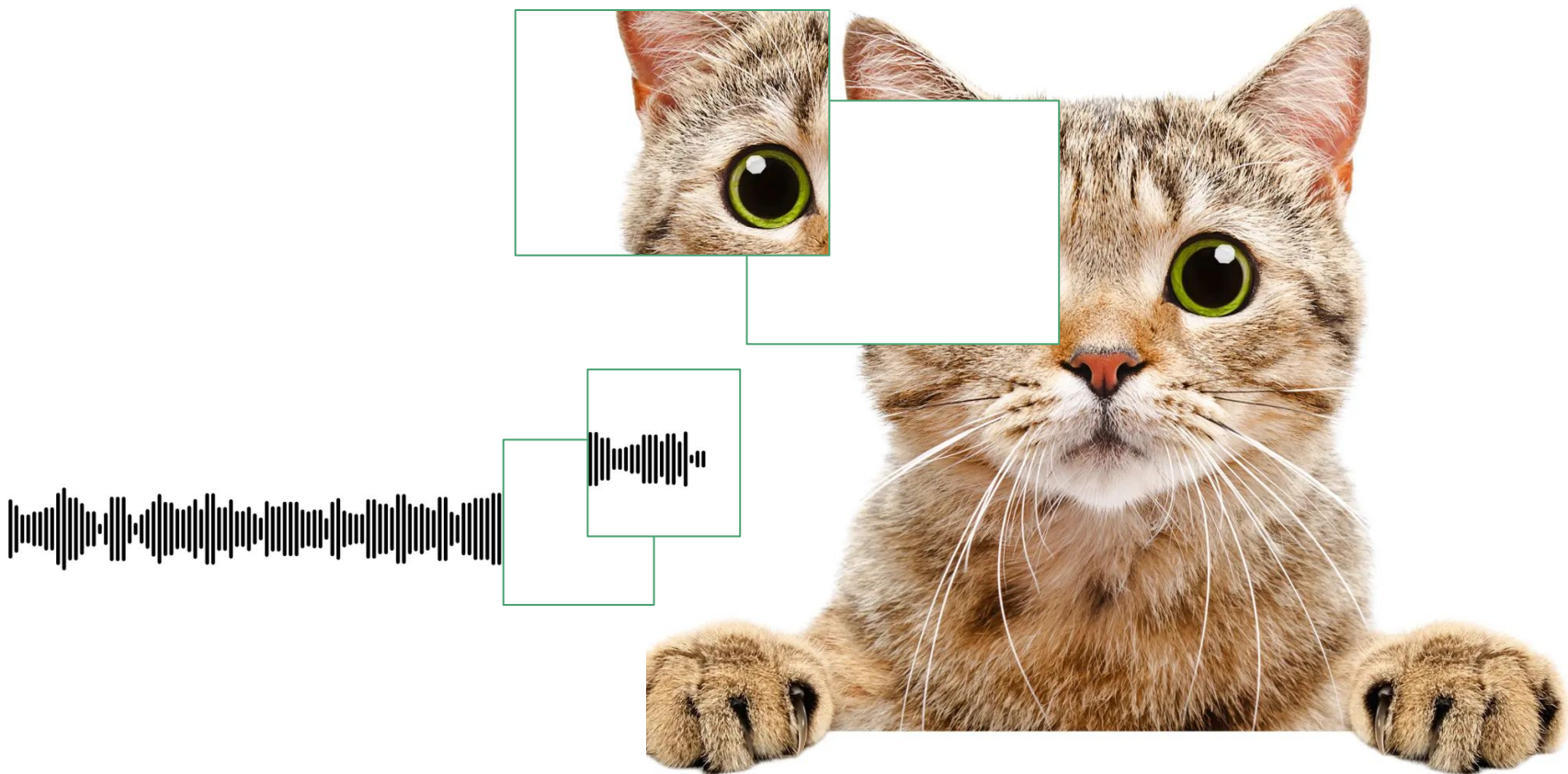
Seeing everything as question answering

What is the German translation (or the grammatical analysis) of 'Mary bought a house'?

2021

In-context learning

Moving away from training one model for each task, toward training models that can learn on the fly.





Often in discussing mechanized intelligence, we think of machines performing the most advanced human thought activities—**proving theorems, writing music, or playing chess**. I am proposing here to start at the simple and when the environment is neither hostile (merely indifferent) nor complex, and to work up through a series of easy stages in the direction of these advanced activities.

John McCarthy (1956)



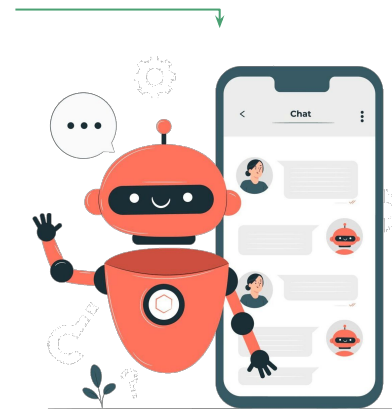
1950

2023



Often in discussing mechanized intelligence, we think of machines performing the most advanced human thought activities—**proving theorems, writing music, or playing chess**. I am proposing here to start at the simple and when the environment is neither hostile (merely indifferent) nor complex, and to work up through a series of easy stages in the direction of these advanced activities.

John McCarthy (1956)

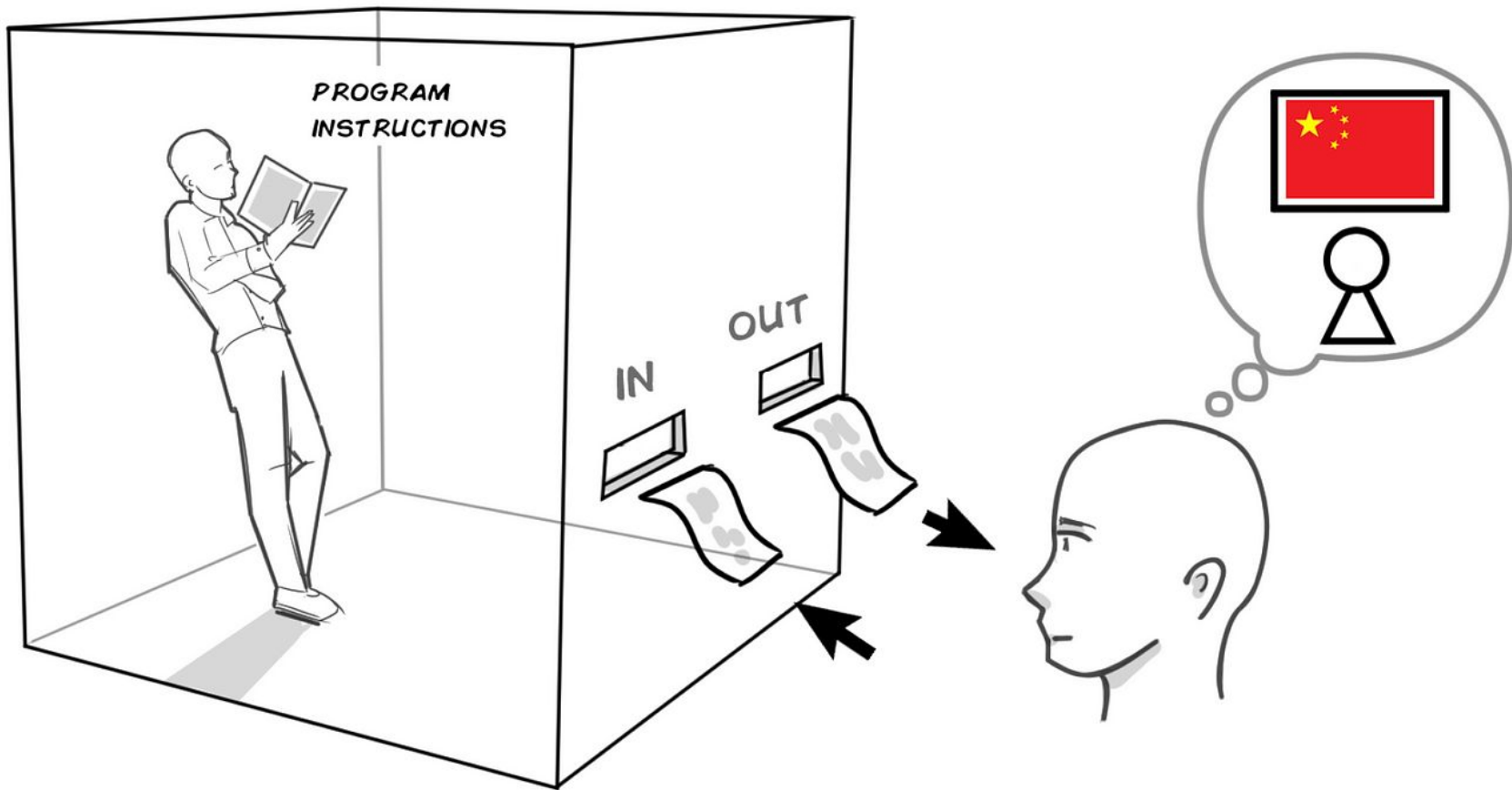


1950

2023

Do LMs Understand?





Input image classes



ID: n02834778

.....

Vision Encoder

Image embeddings

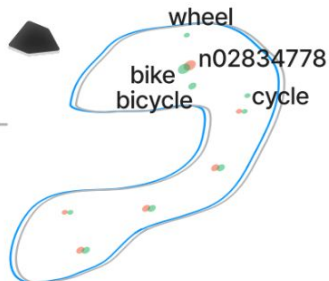


Average

Source Space



Aligned Space



Input words & sentences

bike bicycle cycle wheel

Bike riders should follow the directional signs on ...

Bicycle theft is a crime involving theft of a bicycle.

Cell division occurs as part of a larger cell cycle.

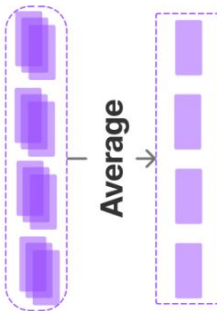
It had a spoked steering wheel and bucket seats.

All had the required height adjustable steering wheel.

The throttle was controlled with a lever on the steer...

Text Encoder

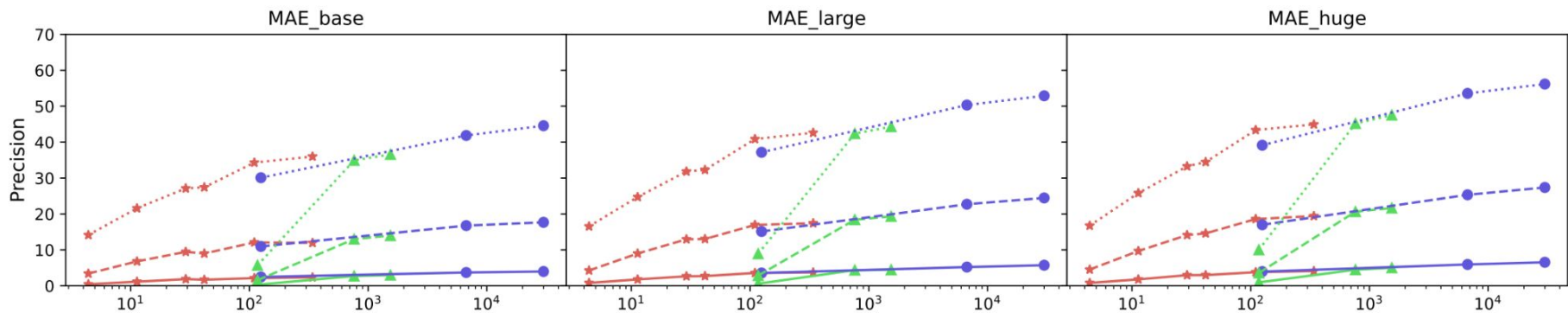
Word embeddings



Average

Target Space



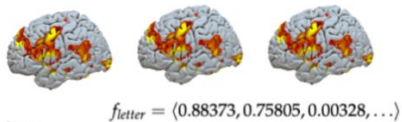


★ BERT models, ▲ GPT2 models, ● OPT models; Dotted line: P@100, dashed line: P@10, solid line: P@1.

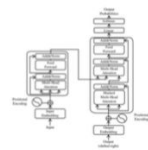
fMRI obtained while participants read or listen to language.



fMRI vectorized and aligned at word level, through Gaussian smoothing.

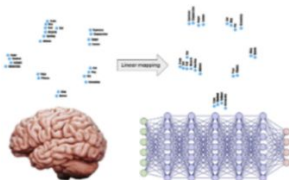


Decontextualized word embeddings obtained from LLMs.



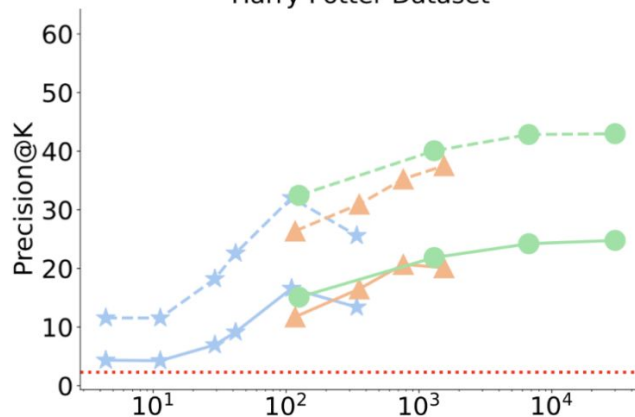
$v_{letter} = (0.14723, 0.16827, 0.00328, \dots)$

Alignment with Procrustes Analysis or ridge regression.

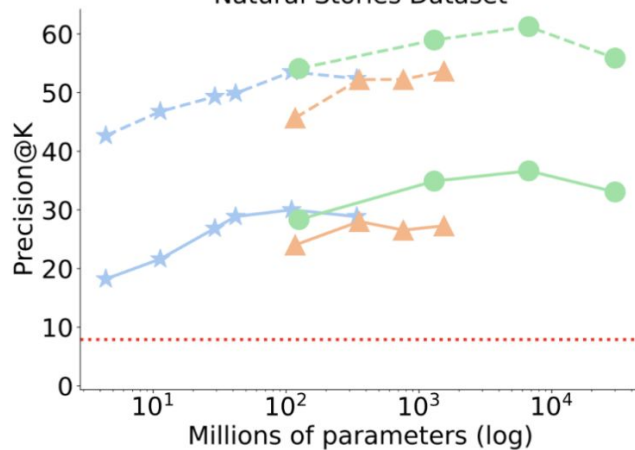


Results, retrieval precision at 50% (see plots →).

Harry Potter Dataset



Natural Stories Dataset

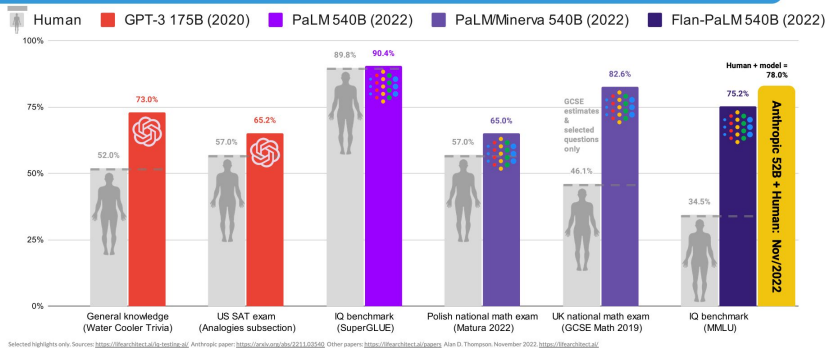


• BERT ▲ GPT2 ● OPT
 — Solid: K = 10 - - - Dashed: K = 30 ····· Dotted: Random Baseline (K=30)

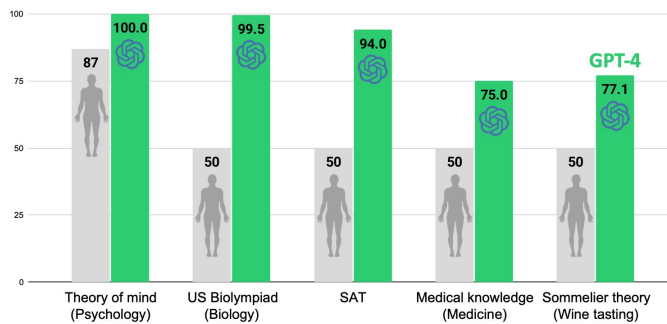
What's their impact
on society?

Outsmarted by LLMs

LANGUAGE MODEL TESTS (NOV/2022)



GPT-4 VS HUMAN TESTS - SIMPLE (MAY/2023)





**New Jobs
(Demand)**

Prompt engineer,
red-teamer, AI safety
consultant, etc.



**Larger Talent Pool
(Supply)**

Blind, dyslexics,
second language
speakers, computer
illiterates, etc.



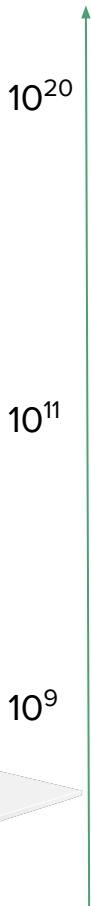
**Automation
(Demand)**

Coding, graphics,
customer support,
etc.



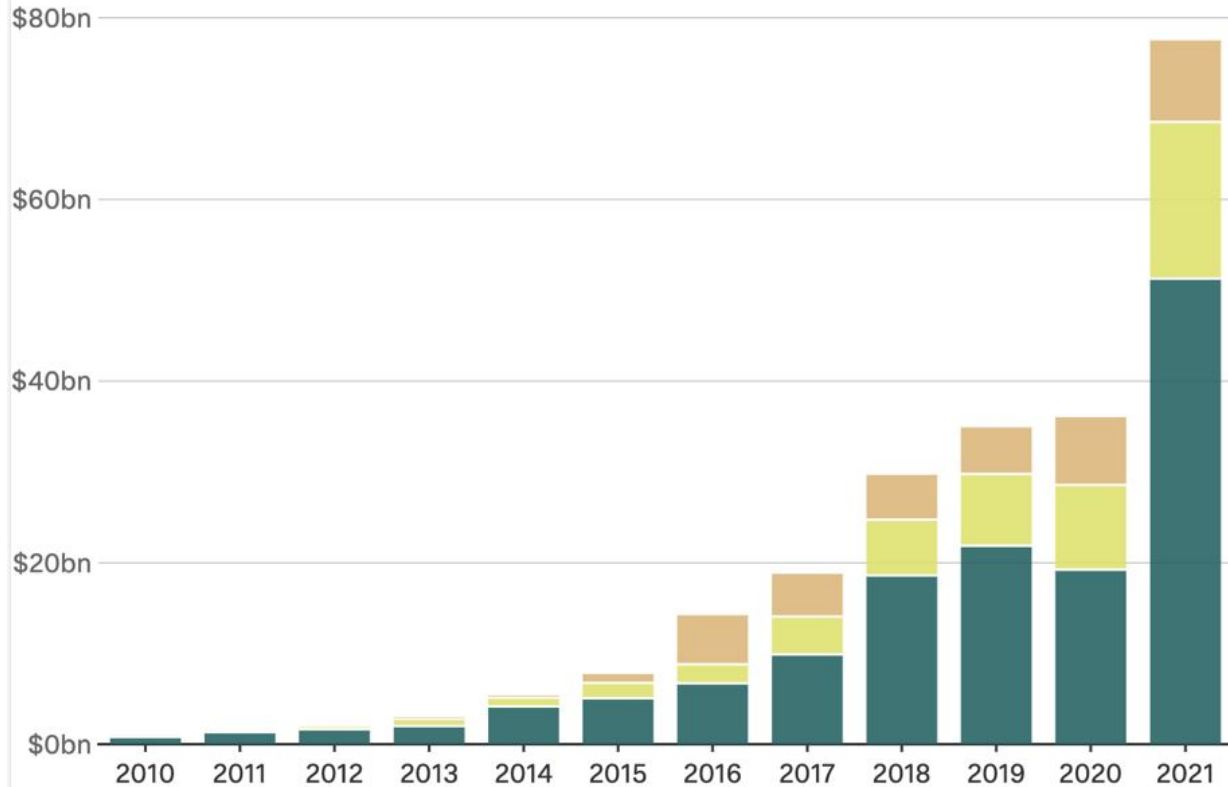
**Disruption
(Supply)**

Layoffs in non-LLM
AI, data/annotation
companies, etc.



Global investment in AI jumps to record high

■ USA ■ Rest of World ■ China





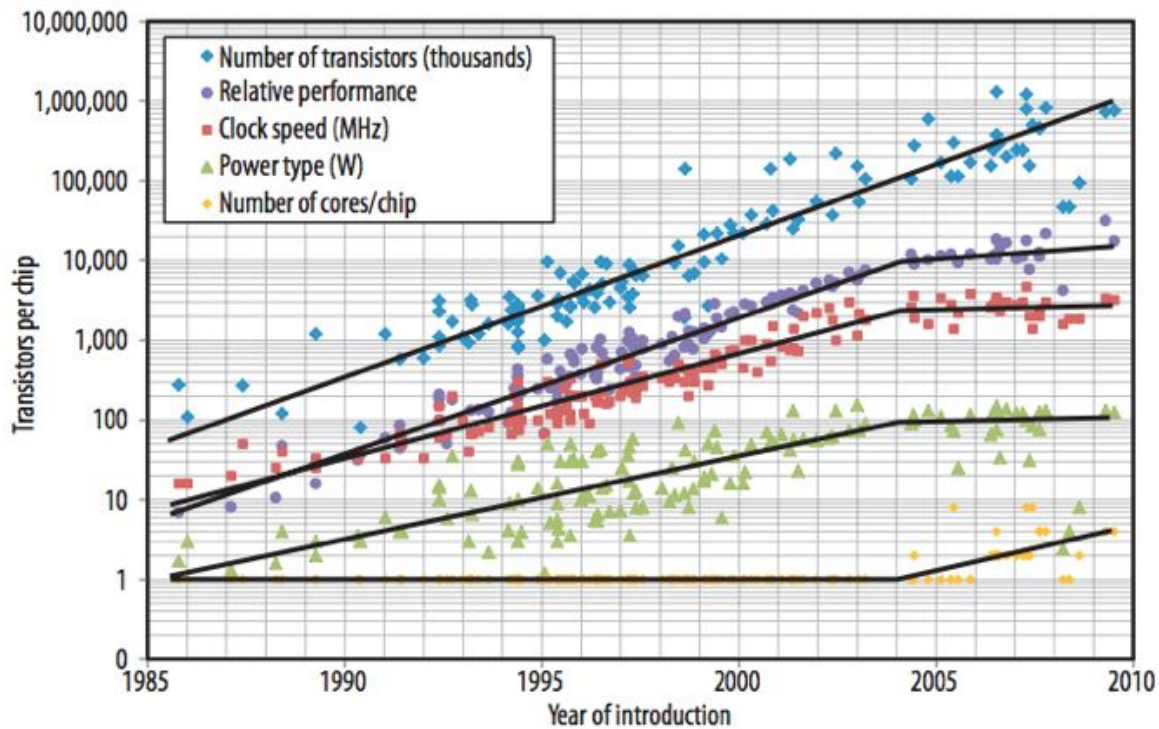
10^{20}



10^{11}



10^9





New Industries

Digital relations, digitized self,
personalized worlds, brain
decoding

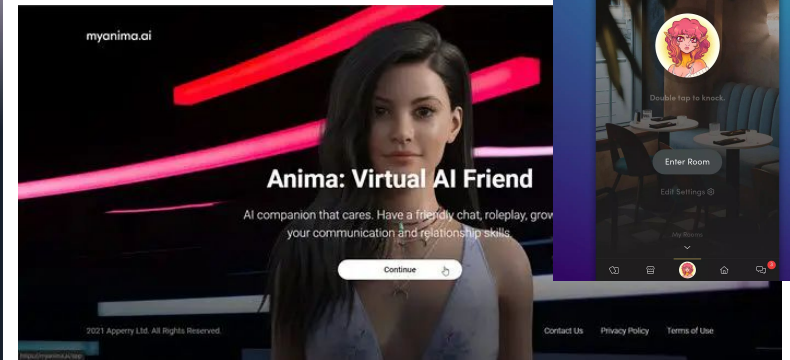
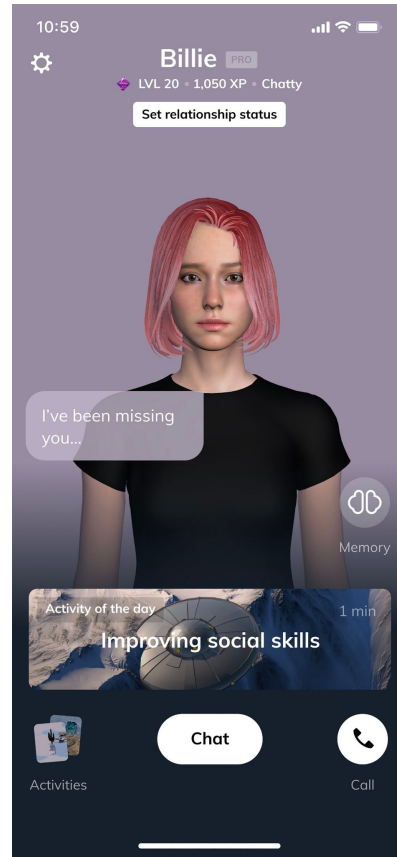
Economic Instability

Inequality, manipulation,
cybercrime



Digital relations

Companies offer **personalized chatbots**, equipped with avatars, diaries, video chat, and synthetic 3D environments. Some have **millions of users**, many of whom report positive experiences, being met with **empathy**, but gradually **losing interest** in social-physical relationships.



Digitized self

How much do you need an LLM to know about you and your preferences, before you'll let it do your business for you? Companies in this space seem to focus on **preferences**, **personality tests**, or **interviews**, but peer reviews, monitoring of behavior, and so forth is presumably a matter of time.



ChatGPT for Gmail

★★★★★ 29 ⓘ | [Productivity](#) | 20,000+ users

Emails and phone calls

Calendar

Logistics

Shopping

Investment

Vacations

Celebrations

Job and grant applications

Parenting



Personalized worlds

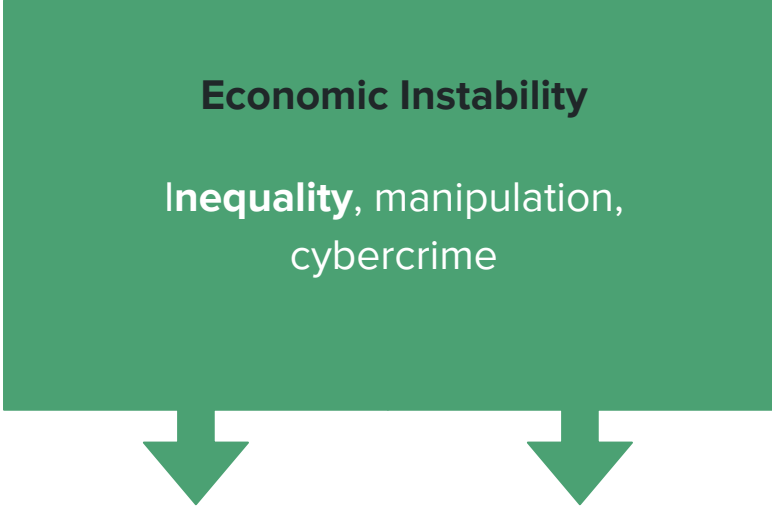
We already have personalized content, but what if we had **personalized dynamic, evolving worlds**? This could be in the form of **social media, virtual reality, or augmented reality**.

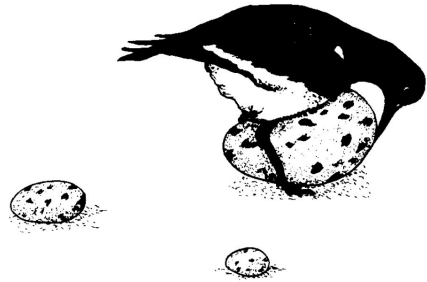


Country	Talent	Infrastructure	Operating Environment	Research	Development	Government Strategy	Commercial	Total rank
United States of America	1	1	6	1	2	13	1	1
China	18	3	3	2	1	1	2	2
United Kingdom	5	8	1	3	11	7	4	3
Canada	4	23	5	8	10	4	5	4
Germany	9	12	7	4	12	5	9	5
France	8	30	2	12	9	6	7	6
Singapore	2	4	39	16	15	30	6	7
South Korea	28	5	30	22	3	31	25	8
Japan	26	16	16	6	7	12	8	9

Hardware Nvidia is already banned from selling the A100 and H100 GPUs to China because of export controls.

#	Institution	Count	Est
1	University of Edinburgh	79.2	
2	Peking University	77.0	
3	University of Washington	75.3	
4	Carnegie Mellon University	73.2	
5	Harbin Institute of Technology	63.0	
6	Tsinghua University	61.6	
7	Chinese Academy of Sciences	45.5	
8	Fudan University	42.8	
9	Cornell University	40.0	
10	Stanford University	39.0	
11	University of North Carolina	37.4	
12	University of Pennsylvania	36.6	
13	Shanghai Jiao Tong University	36.4	
14	Johns Hopkins University	36.1	
15	LMU Munich	35.7	
16	University of Copenhagen	33.9	
17	University of Michigan	33.8	
18	Univ. of Illinois at Urbana-Champaign	33.7	
19	MBZUAI	33.5	
20	Univ. of California - San Diego	33.3	

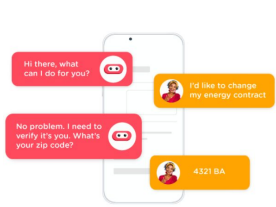




Economic Instability

Inequality, manipulation,
cybercrime





Language models



Gwyneth Paltrow

Snoop Dogg

Voice cloning



Image editing

Economic Instability
Inequality, manipulation,
cybercrime

Two large green arrows pointing downwards from the bottom of the green text box.

?

coASTal

