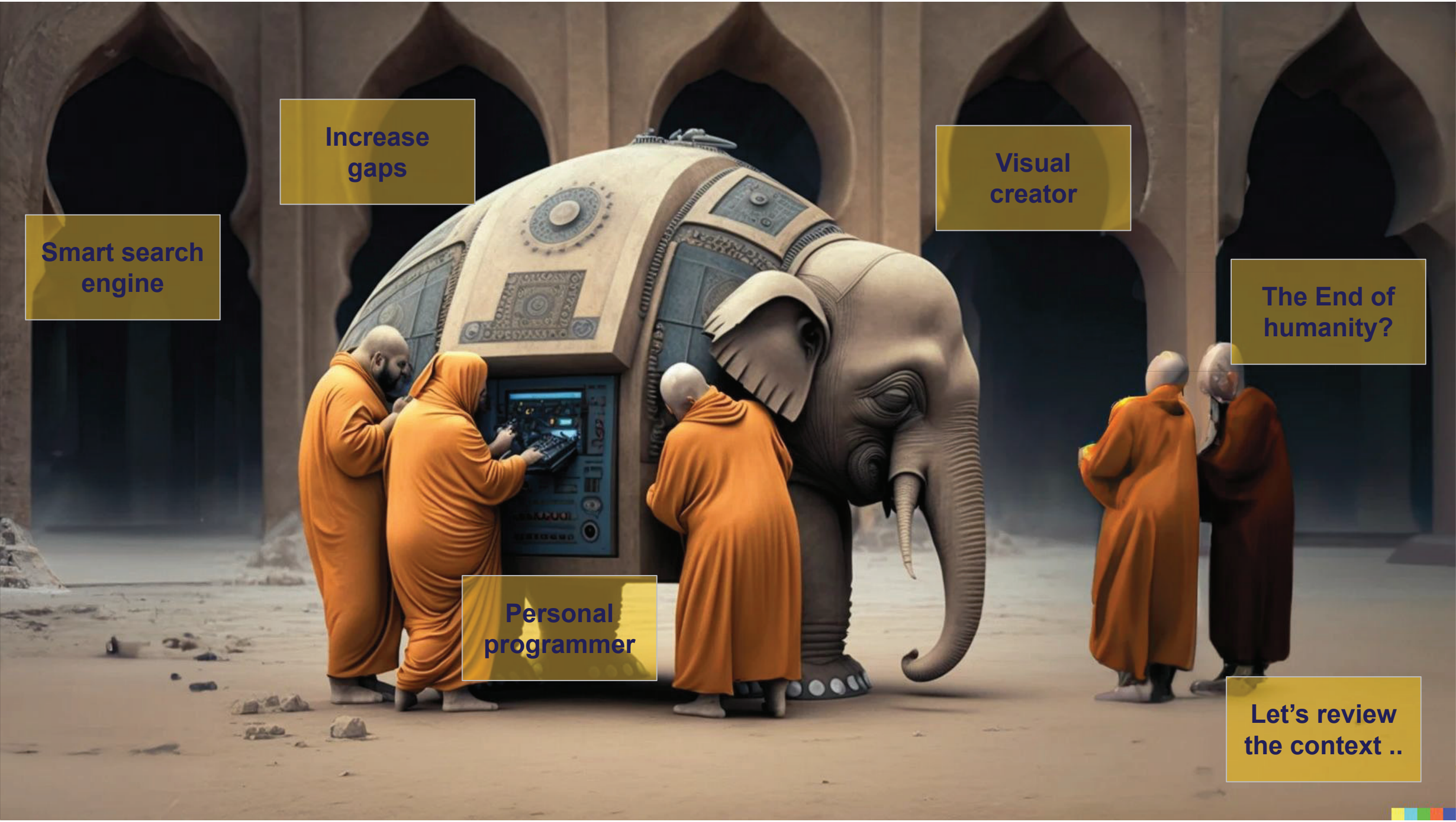
Today

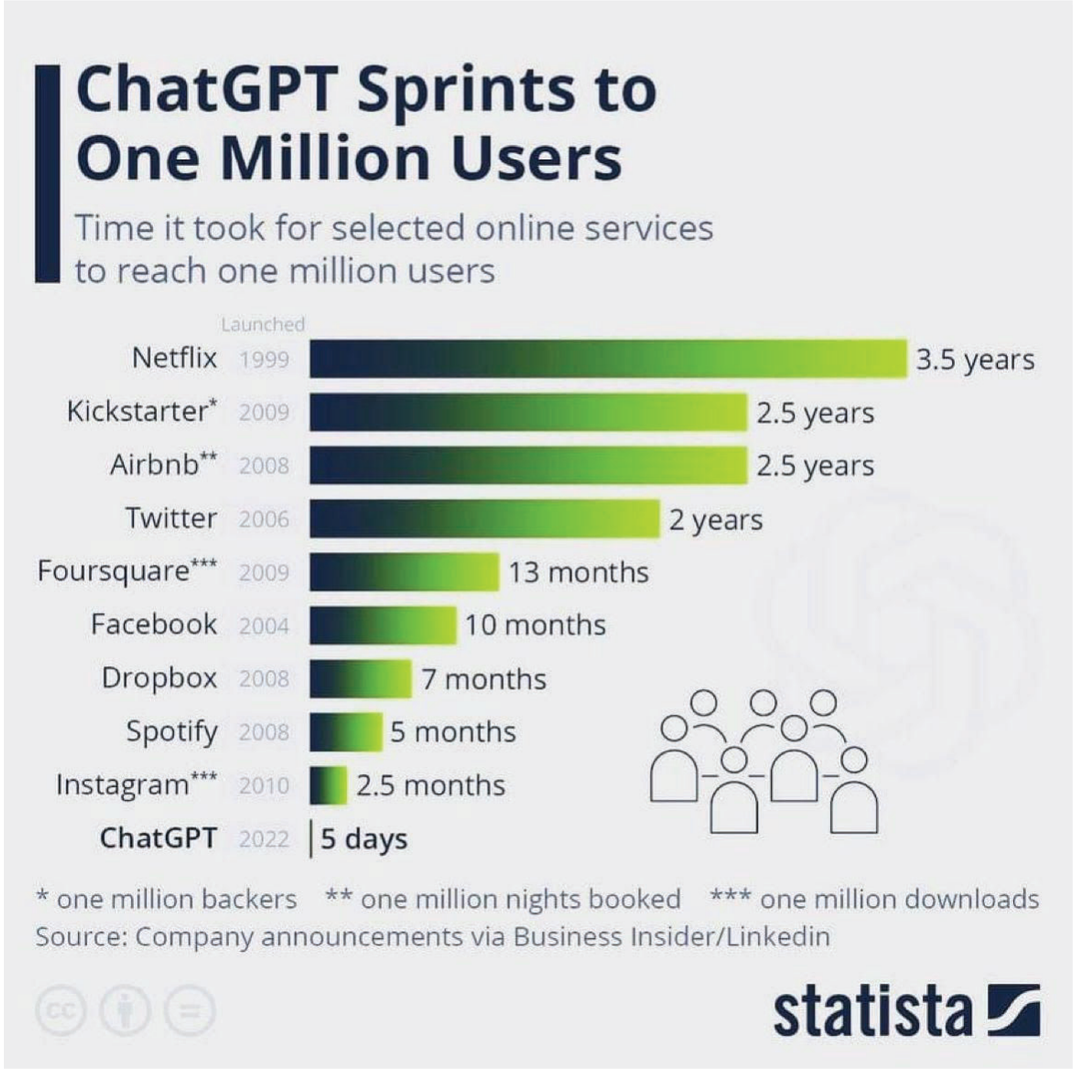
- (a) The Context: The AGI Elephant
  - The AGI Elephant 🐘
  - Context: Is AGI Just One More Tech?
  - Key AGI Concepts
  - How Big is This AGI Elephant?
- (b) Review: AGI Tools
- 10 min Break ☕
- (c) Learning: Tips That Teach Us
- (d) Open Q&A


Increase gaps
Visual creator
Smart search engine
The End of humanity?
Personal programmer
Let’s review the context ..

Rate of Adoption

What we learn:
People use (EZ use, like it)
People talk about it

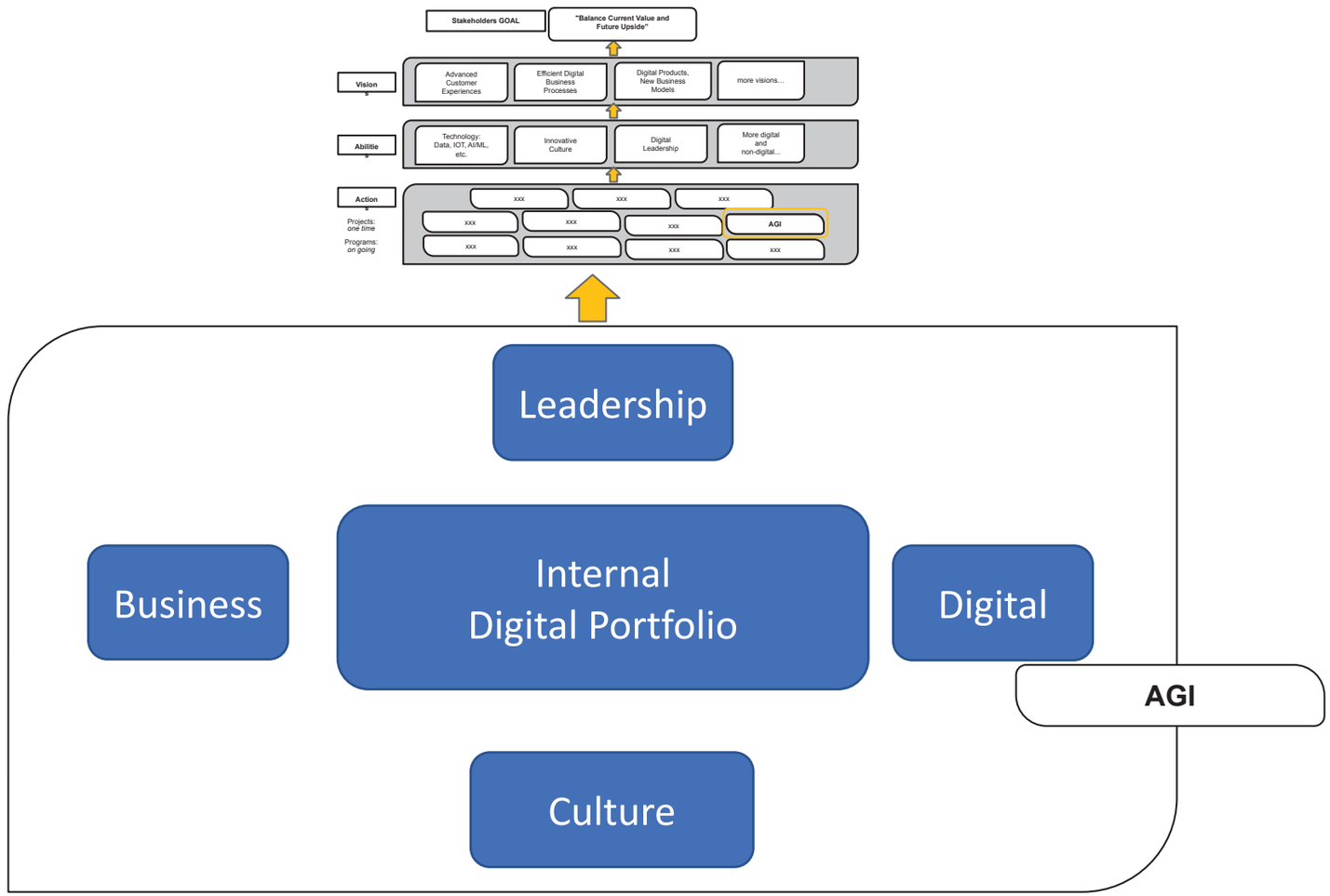
Much more:
Ability to scale (Microsoft)



Today

- (a) The Context: The AGI Elephant
  - The AGI Elephant 🐘
  - Context: Is AGI Just One More Tech?
  - Key AGI Concepts
  - How Big is This AGI Elephant?
- (b) Review: AGI Tools
- 10 min Break ☕
- (c) Learning: Tips That Teach Us
- (d) Open Q&A

Context: AGI as Any Other Transformational Technology



The Leadership Context

So, is AGI just one more Tech?

Or as big as: Book printing, Electricity, and the Internet?

or Even Bigger?

Today

- (a) The Context: The AGI Elephant
  - The AGI Elephant 🐘
  - Context: Is AGI Just One More Tech?
  - Key AGI Concepts
  - How Big is This AGI Elephant?
- (b) Review: AGI Tools
- 10 min Break ☕
- (c) Learning: Tips That Teach Us
- (d) Open Q&A

Key AGI Concepts

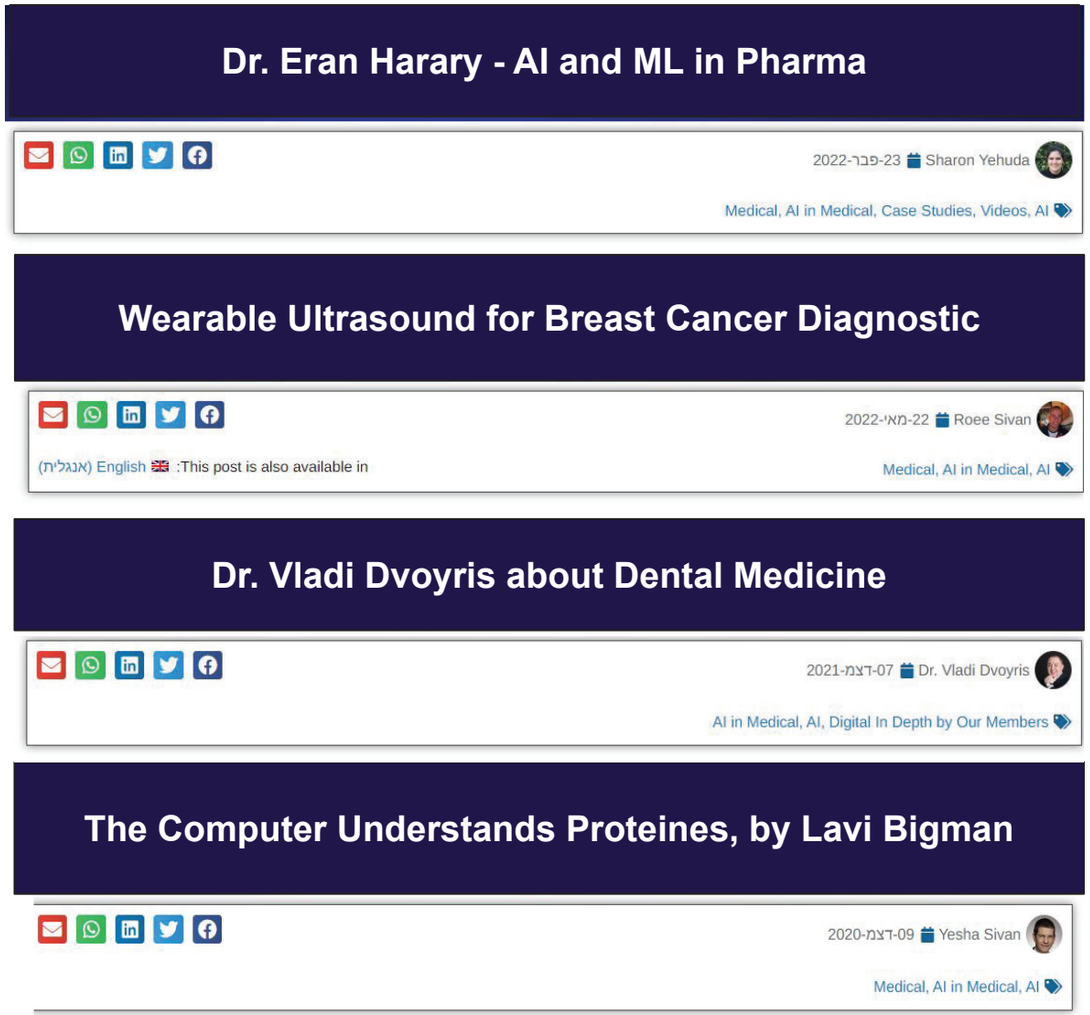
Baseline:

1. **Data Set** -- data that is used to train the model.
2. **Transformer** -- the “learning” tool that create the work corpus.
3. **Promt** -- what you ask the system to do.
4. **Chat Context** -- unique way to keep context (GPT instruct << key innovation)

Even more:

5. **“Connectionist” vs. “Symbolic”** (ala https://www.wolframalpha.com/)
6. **Alignment** -- what do we want the system to do -- is it aligned
7. **Guardrails** -- a rail that prevents people from falling off or being hit by something.
8. **Business Models** -- what will be the structure/nature of the market?

AI vs. AGI

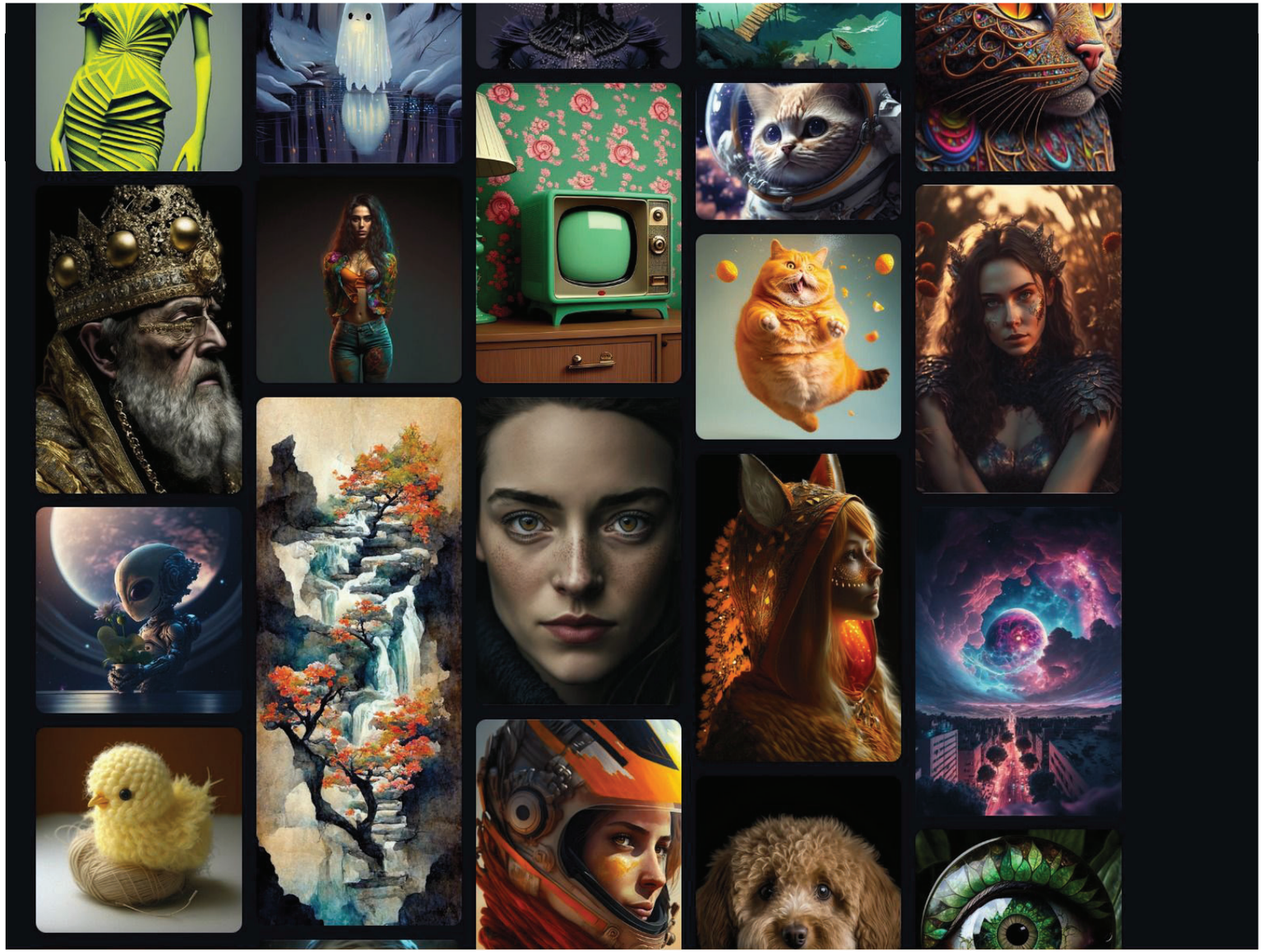


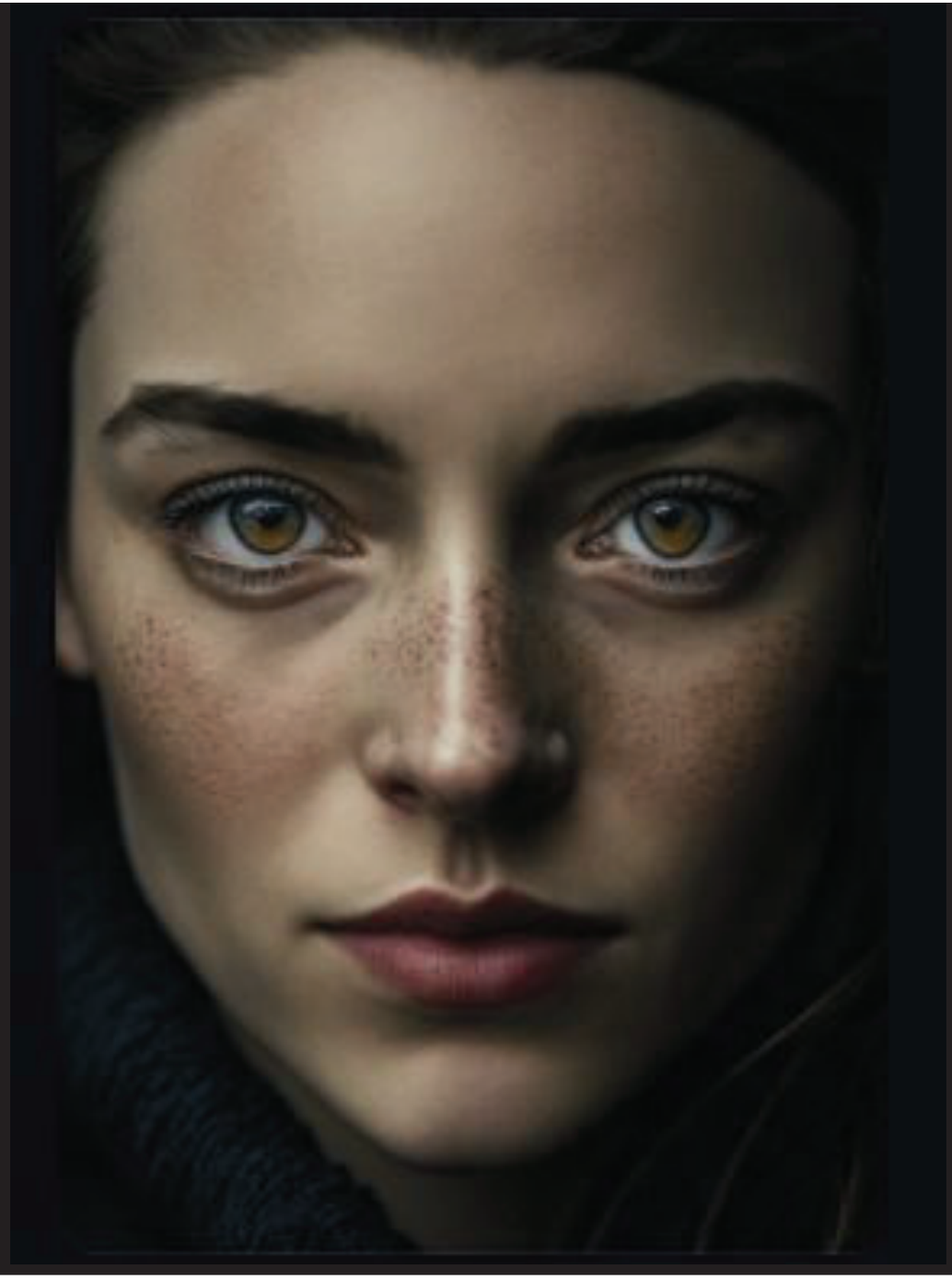
<>

AGI

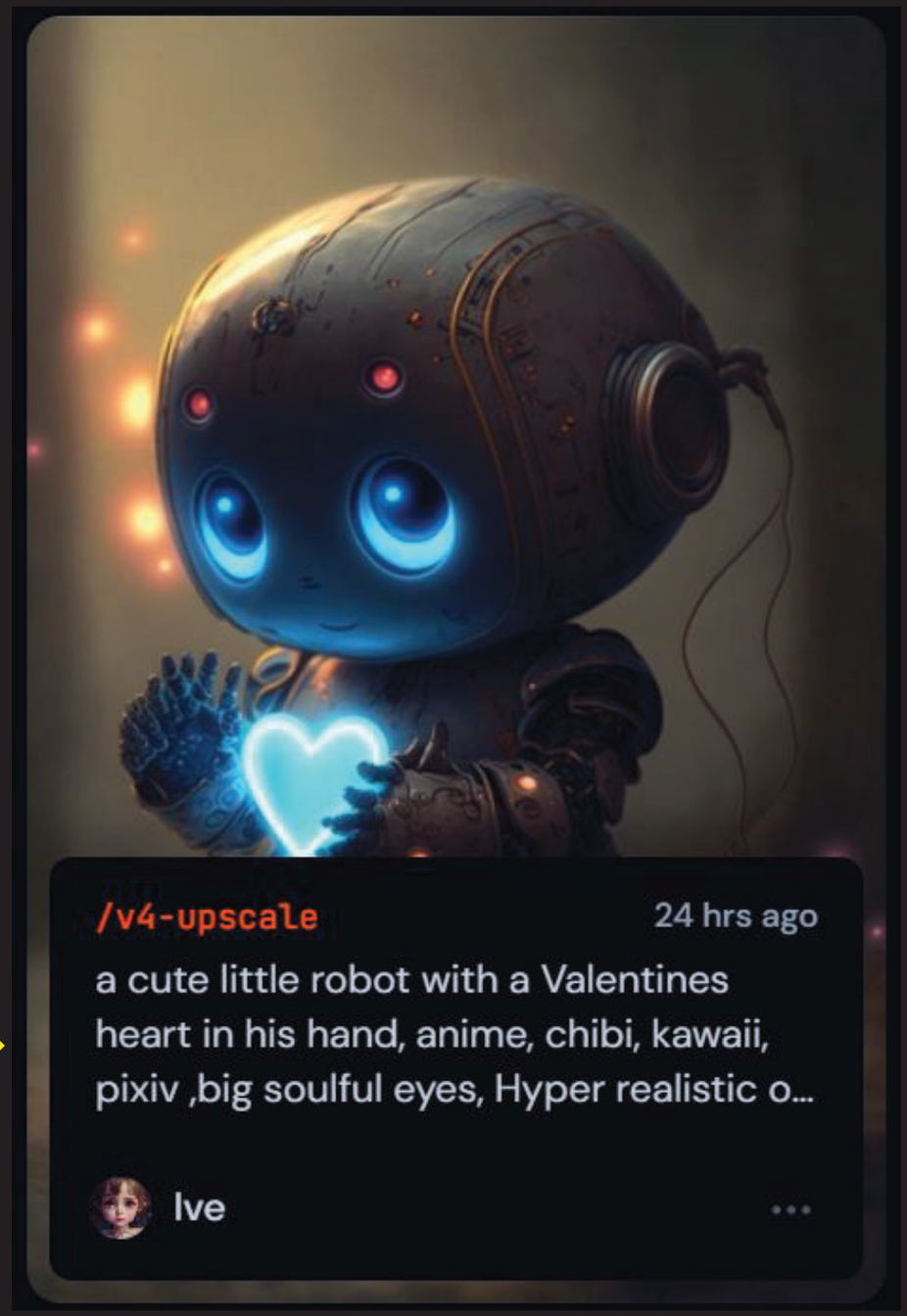
Artificial
General
Intelligence

AGI Visual

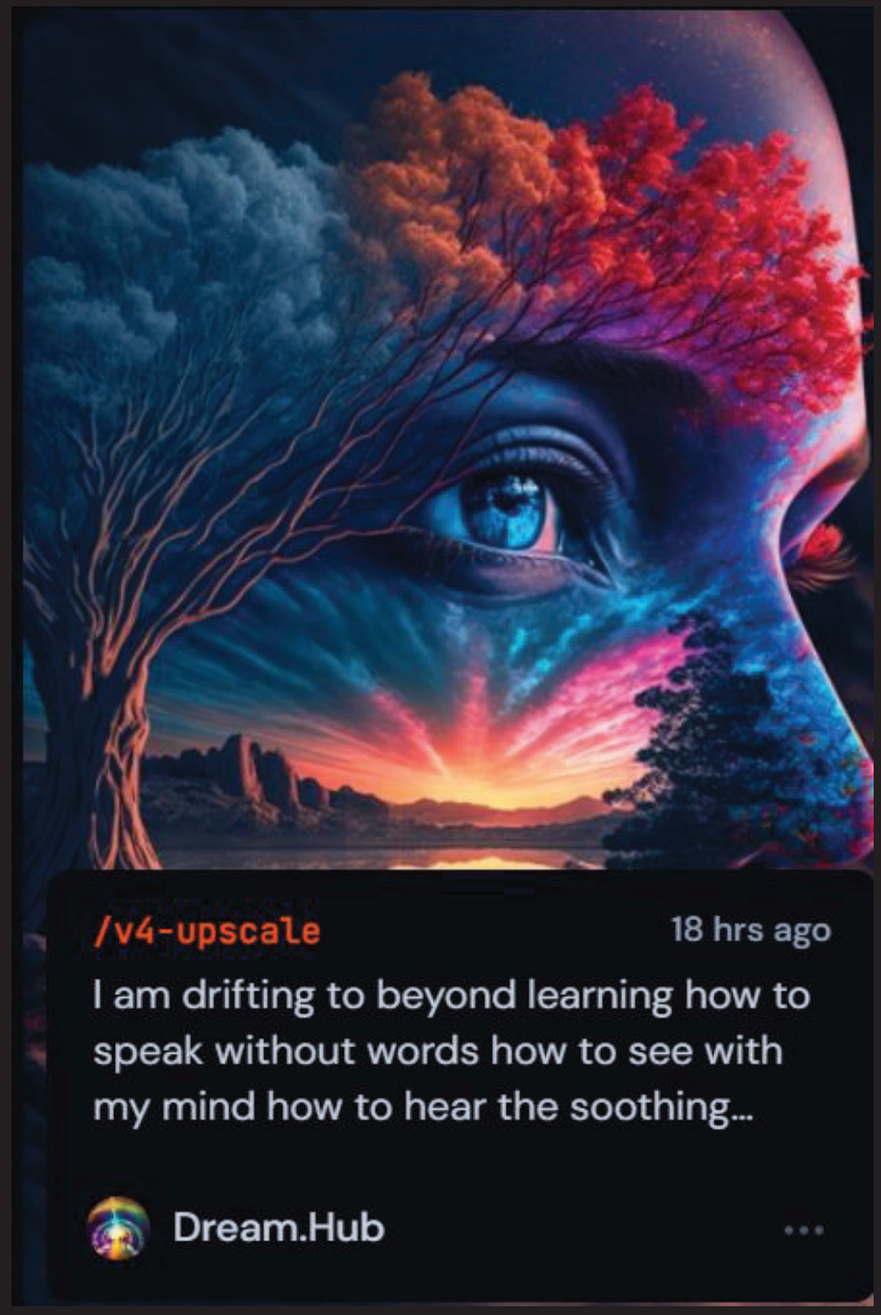




Concrete PROMPT



Artistic PROMPT



AGI for Robotics

OpenAI releases Point-E, an AI that generates 3D models

Converting the Point-E point clouds into meshes. ***Image Credits:*** *OpenAI*

AGI Combined

Text

Visual

Video

Voice

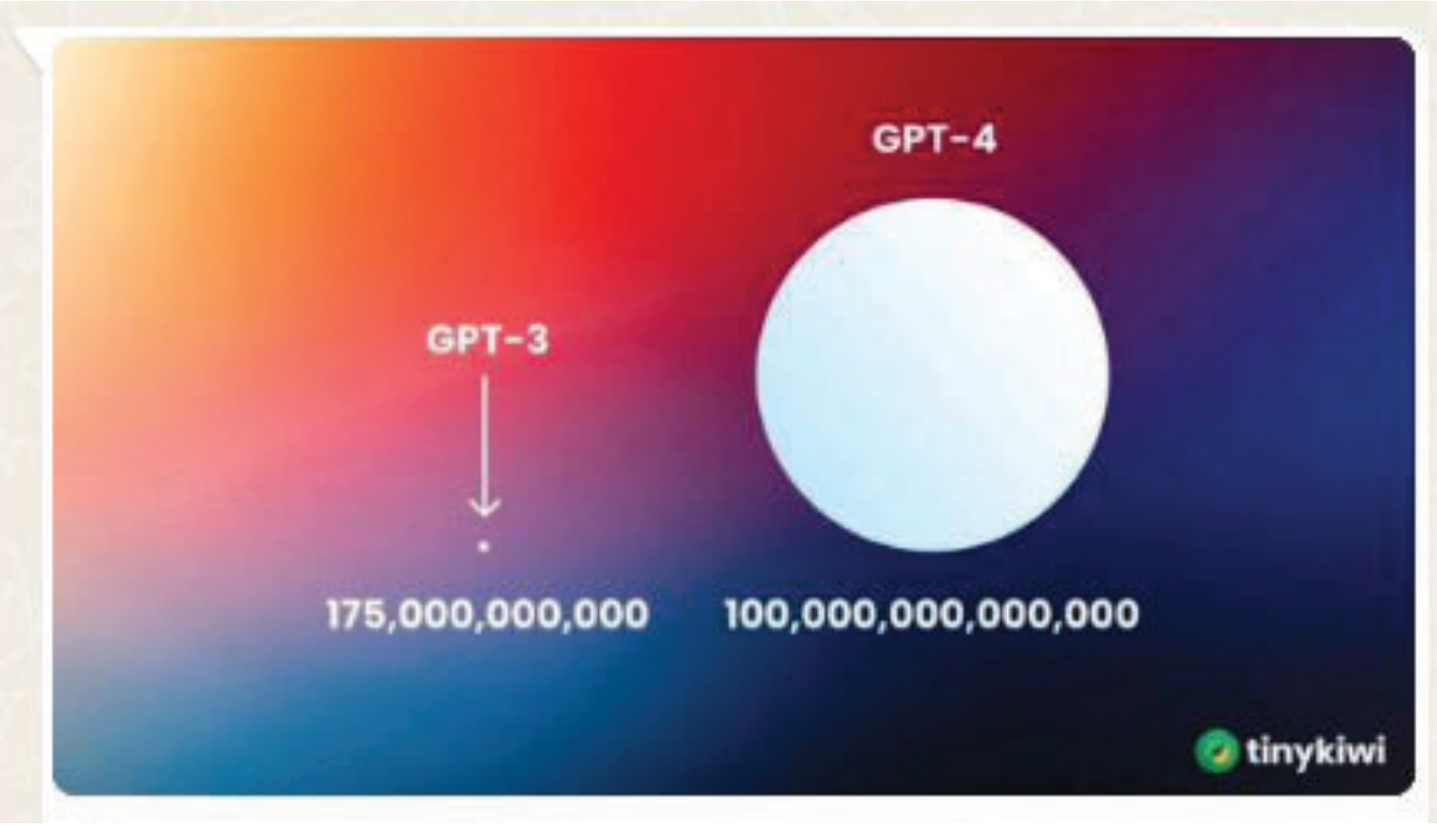
Movement

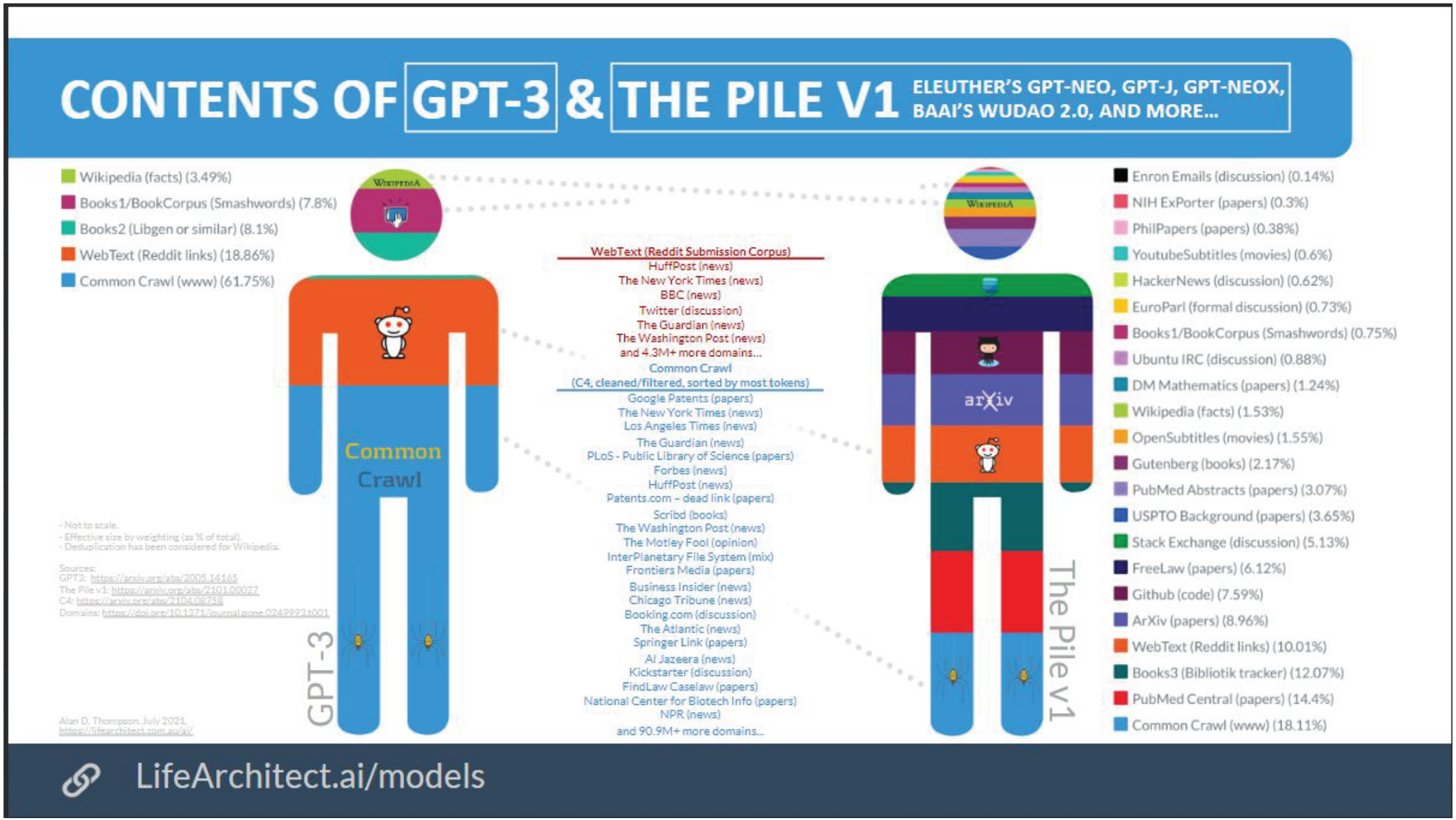
Any other types of data…

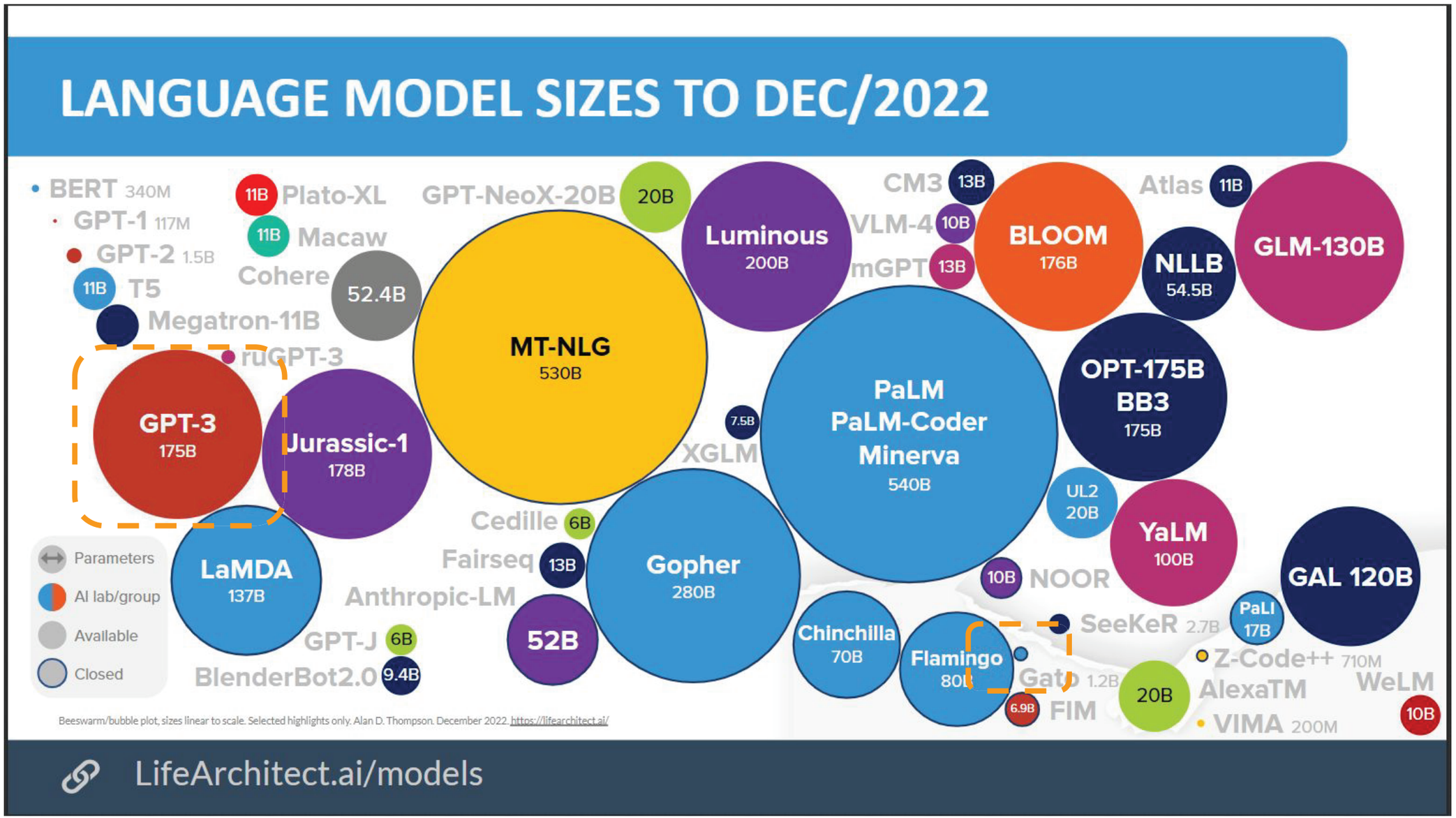
Today

- (a) The Context: The AGI Elephant
  - The AGI Elephant 🐘
  - Context: Is AGI Just One More Tech?
  - Key AGI Concepts
  - How Big is This AGI Elephant?
- (b) Review: AGI Tools
- 10 min Break ☕
- (c) Learning: Tips That Teach Us
- (d) Open Q&A

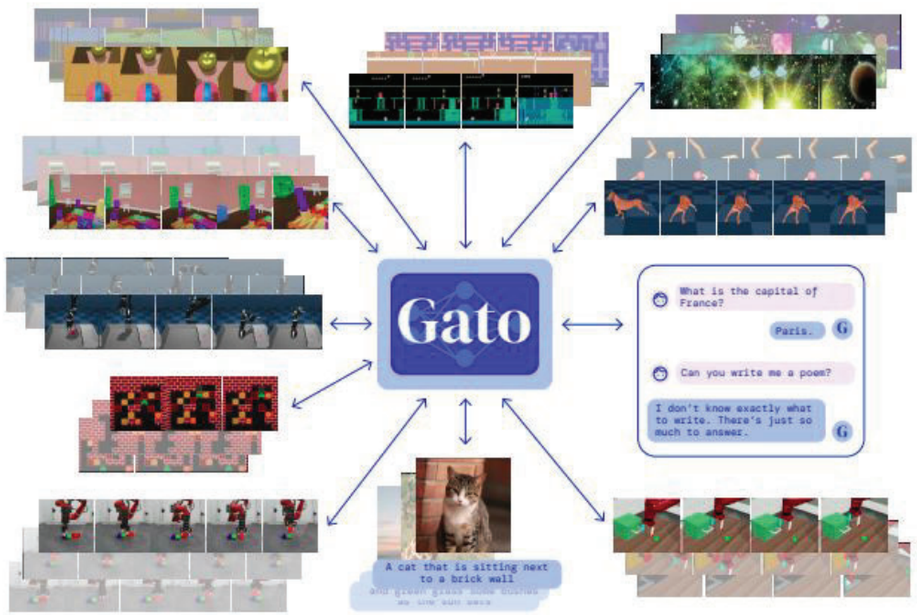
Machine Learning based on Corpus of text:




CONTENTS OF GPT-3 & THE PILE V1
ELEUTHER'S GPT-NEO, GPT-J, GPT-NEOX, BAAI'S WUDAO 2.0, AND MORE...
Wikipedia (facts) (3.49%)
Books1/BookCorpus (Smashwords) (7.8%)
Books2 (Libgen or similar) (8.1%)
WebText (Reddit links) (18.86%)
Common Crawl (www) (61.75%)
Common Crawl
GPT-3
- Not to scale.
- Effective size by weighting (as % of total).
- Deduplication has been considered for Wikipedia.
Sources:
GPT3: https://arxiv.org/abs/2005.14165
The Pile v1: https://arxiv.org/abs/2101.00027
C4: https://arxiv.org/abs/2104.08758
Domains: https://doi.org/10.1371/journal.pone.0249993.t001
Alan D. Thompson. July 2021.
https://lifearchitect.com.au/ai/
WebText (Reddit Submission Corpus)
HuffPost (news)
The New York Times (news)
BBC (news)
Twitter (discussion)
The Guardian (news)
The Washington Post (news)
and 4.3M+ more domains...
Common Crawl
(C4, cleaned/filtered, sorted by most tokens)
Google Patents (papers)
The New York Times (news)
Los Angeles Times (news)
The Guardian (news)
PLoS - Public Library of Science (papers)
Forbes (news)
HuffPost (news)
Patents.com – dead link (papers)
Scribd (books)
The Washington Post (news)
The Motley Fool (opinion)
InterPlanetary File System (mix)
Frontiers Media (papers)
Business Insider (news)
Chicago Tribune (news)
Booking.com (discussion)
The Atlantic (news)
Springer Link (papers)
Al Jazeera (news)
Kickstarter (discussion)
FindLaw Caselaw (papers)
National Center for Biotech Info (papers)
NPR (news)
and 90.9M+ more domains...
arXiv
The Pile v1
Enron Emails (discussion) (0.14%)
NIH ExPorter (papers) (0.3%)
PhilPapers (papers) (0.38%)
YoutubeSubtitles (movies) (0.6%)
HackerNews (discussion) (0.62%)
EuroParl (formal discussion) (0.73%)
Books1/BookCorpus (Smashwords) (0.75%)
Ubuntu IRC (discussion) (0.88%)
DM Mathematics (papers) (1.24%)
Wikipedia (facts) (1.53%)
OpenSubtitles (movies) (1.55%)
Gutenberg (books) (2.17%)
PubMed Abstracts (papers) (3.07%)
USPTO Background (papers) (3.65%)
Stack Exchange (discussion) (5.13%)
FreeLaw (papers) (6.12%)
Github (code) (7.59%)
ArXiv (papers) (8.96%)
WebText (Reddit links) (10.01%)
Books3 (Bibliotik tracker) (12.07%)
PubMed Central (papers) (14.4%)
Common Crawl (www) (18.11%)
LifeArchitect.ai/models


LANGUAGE MODEL SIZES TO DEC/2022
BERT 340M
GPT-1 117M
GPT-2 1.5B
11B T5
Megatron-11B
ruGPT-3
11B Plato-XL
11B Macaw
Cohere
52.4B
GPT-NeoX-20B
20B
MT-NLG
530B
Luminous
200B
CM3 13B
VLM-4 10B
mGPT 13B
BLOOM
176B
Atlas 11B
NLLB
54.5B
GLM-130B
OPT-175B
BB3
175B
GPT-3
175B
Jurassic-1
178B
7.5B
XGLM
PaLM
PaLM-Coder
Minerva
540B
UL2
20B
YaLM
100B
GAL 120B
Cedille 6B
Fairseq 13B
Anthropic-LM
52B
Gopher
280B
10B NOOR
SeeKeR 2.7B
PaLI
17B
Z-Code++ 710M
AlexaTM
WeLM
10B
LaMDA
137B
GPT-J 6B
BlenderBot2.0 9.4B
Chinchilla
70B
Flamingo
80B
Gato 1.2B
6.9B
FIM
20B
VIMA 200M
Parameters
AI lab/group
Available
Closed
Beeswarm/bubble plot, sizes linear to scale. Selected highlights only. Alan D. Thompson. December 2022. https://lifearchitect.ai/
LifeArchitect.ai/models

Note: Gato by Google Deepmind



GATO – A New Generalist Artificial Intelligence Agent

Into the World of GATO

Datasets

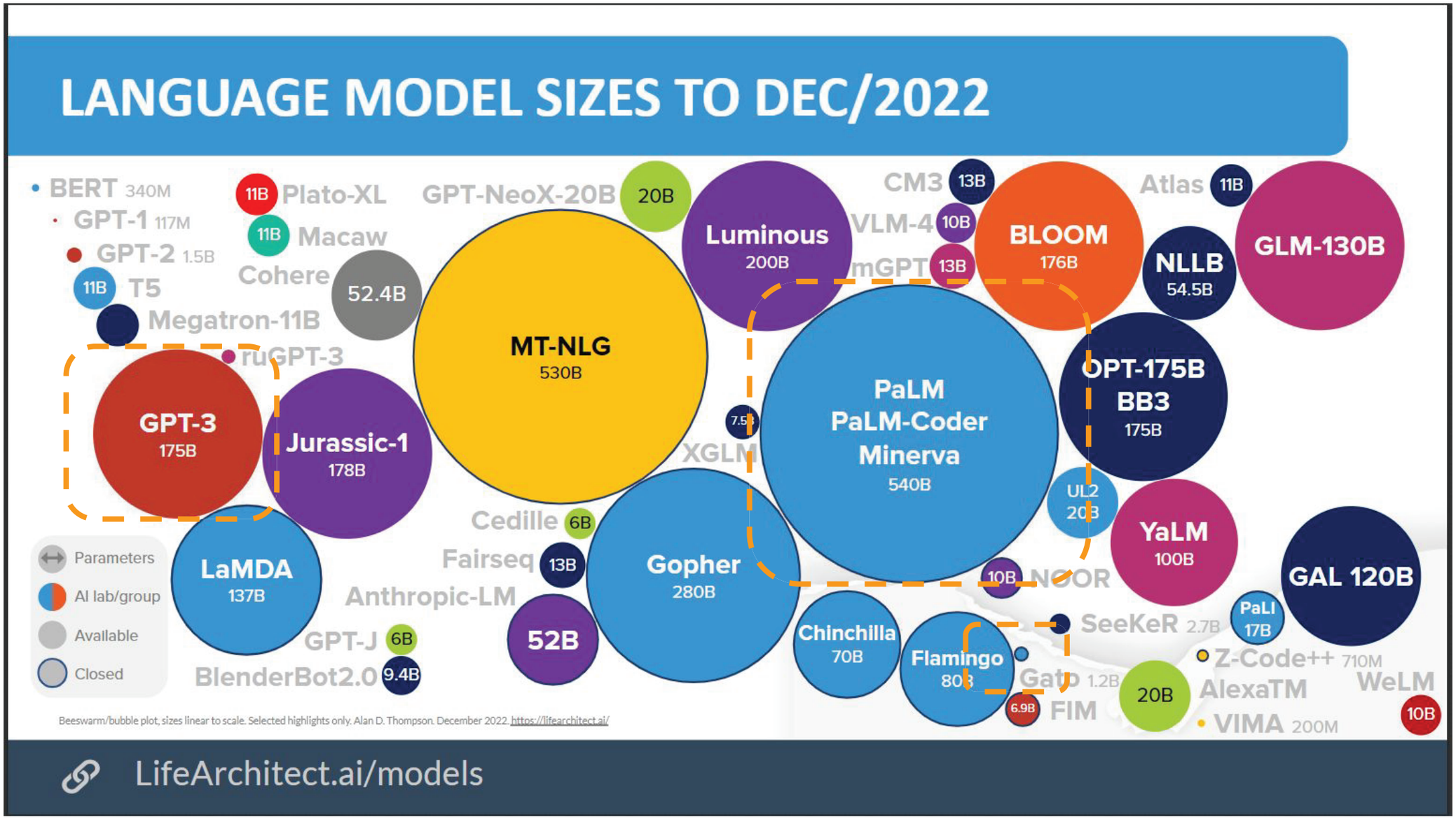
Gato is trained using a variety of datasets, including agent experience in both simulated and real-world settings, as well as natural language and image datasets. The below tables describes the datasets used for GATO training.

Control environment	Tasks	Episodes	Approx. Tokens	Sample Weight
DM Lab	254	16.4M	194B	9.35%
ALE Atari	51	63.4K	1.26B	9.5%
ALE Atari Extended	28	28.4K	565M	10.0%
Sokoban	1	27.2K	298M	1.33%
BabyAI	46	4.61M	22.8B	9.06%
DM Control Suite	30	395K	22.5B	4.62%
DM Control Suite Pixels	28	485K	35.5B	7.07%
DM Control Suite Random Small	26	10.6M	313B	3.04%
DM Control Suite Random Large	26	26.1M	791B	3.04%
Meta-World	45	94.6K	3.39B	8.96%
Procgen Benchmark	16	1.6M	4.46B	5.34%
RGB Stacking simulator	1	387K	24.4B	1.33%
RGB Stacking real robot	1	15.7K	980M	1.33%
Modular RL	38	843K	69.6B	8.23%
DM Manipulation Playground	4	286K	6.58B	1.68%
Playroom	1	829K	118B	1.33%
Total	596	63M	1.5T	85.3%

Vision / language dataset	Sample Weight
MassiveText	6.7%
M3W	4%
ALIGN	0.67%
MS-COCO Captions	0.67%
Conceptual Captions	0.67%
LTIP	0.67%
OKVQA	0.67%
VQAV2	0.67%
Total	14.7%

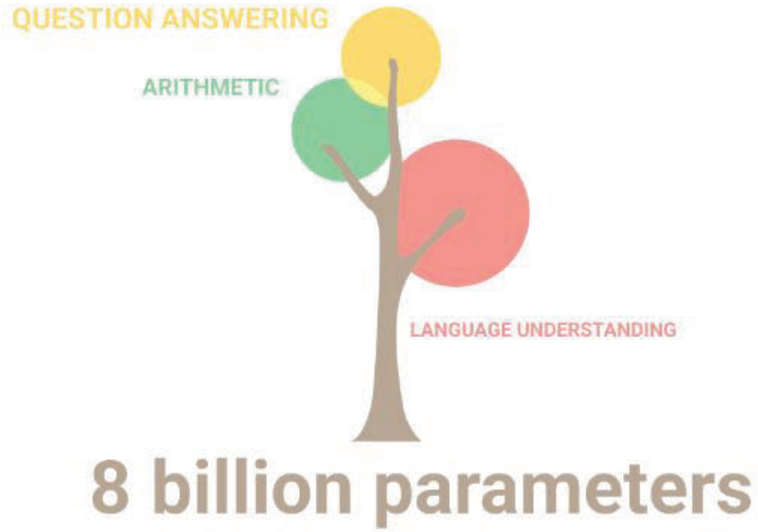
https://arxiv.org/pdf/2205.06175.pdf

The data contained in the final dataset used to train the GATO model is widely spread in different domains, those are


LANGUAGE MODEL SIZES TO DEC/2022
BERT 340M
GPT-1 117M
GPT-2 1.5B
11B T5
Megatron-11B
11B Plato-XL
11B Macaw
Cohere
52.4B
ruGPT-3
GPT-3
175B
Jurassic-1
178B
LaMDA
137B
GPT-J 6B
BlenderBot2.0 9.4B
MT-NLG
530B
GPT-NeoX-20B
20B
Luminous
200B
Cedille 6B
Fairseq
13B
Anthropic-LM
52B
Gopher
280B
CM3 13B
VLM-4 10B
mGPT 13B
BLOOM
176B
PaLM
PaLM-Coder
Minerva
540B
XGLM
7.5B
Atlas 11B
NLLB
54.5B
GLM-130B
OPT-175B
BB3
175B
UL2
20B
YaLM
100B
10B NOOR
GAL 120B
PaLI
17B
SeeKeR 2.7B
Z-Code++ 710M
AlexaTM
WeLM
10B
VIMA 200M
Chinchilla
70B
Flamingo
80B
6.9B
20B
Parameters
AI lab/group
Available
Closed
Beeswarm/bubble plot, sizes linear to scale. Selected highlights only. Alan D. Thompson. December 2022. https://lifearchitect.ai/
LifeArchitect.ai/models

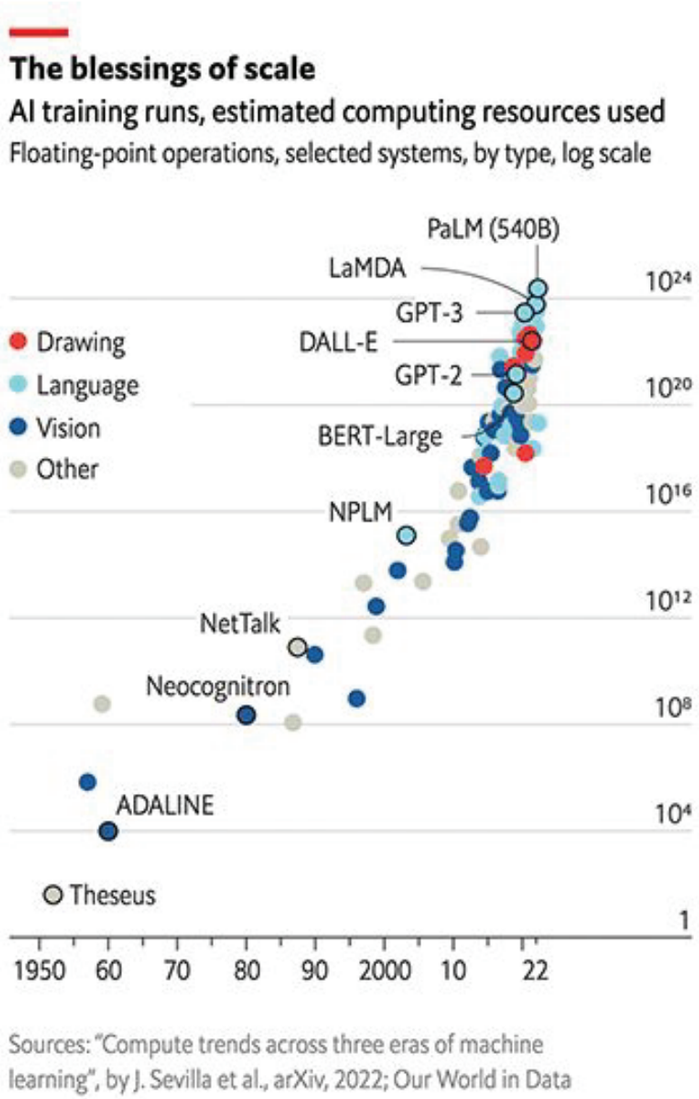
https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html

For example PaLM by google:



Source: Google Deep Mind

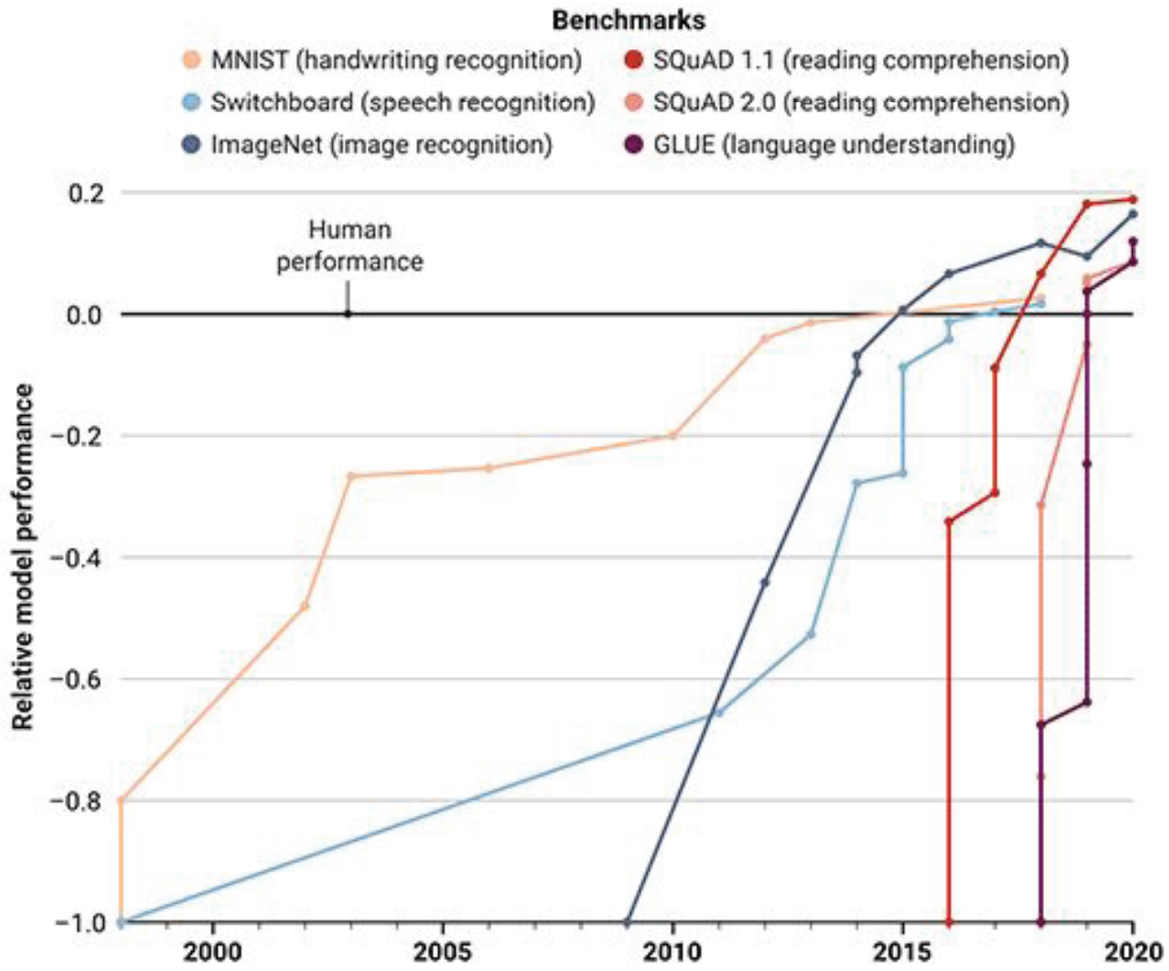
Scale:



Source: Sequoia Dec, 2022

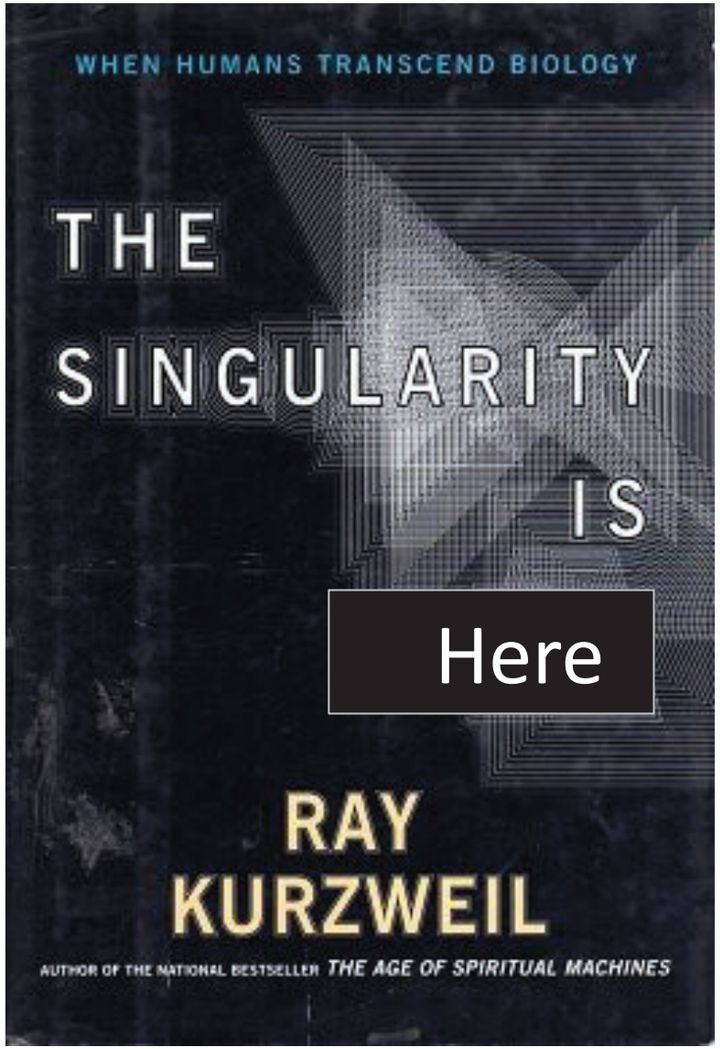
Machines Are Fast Learns

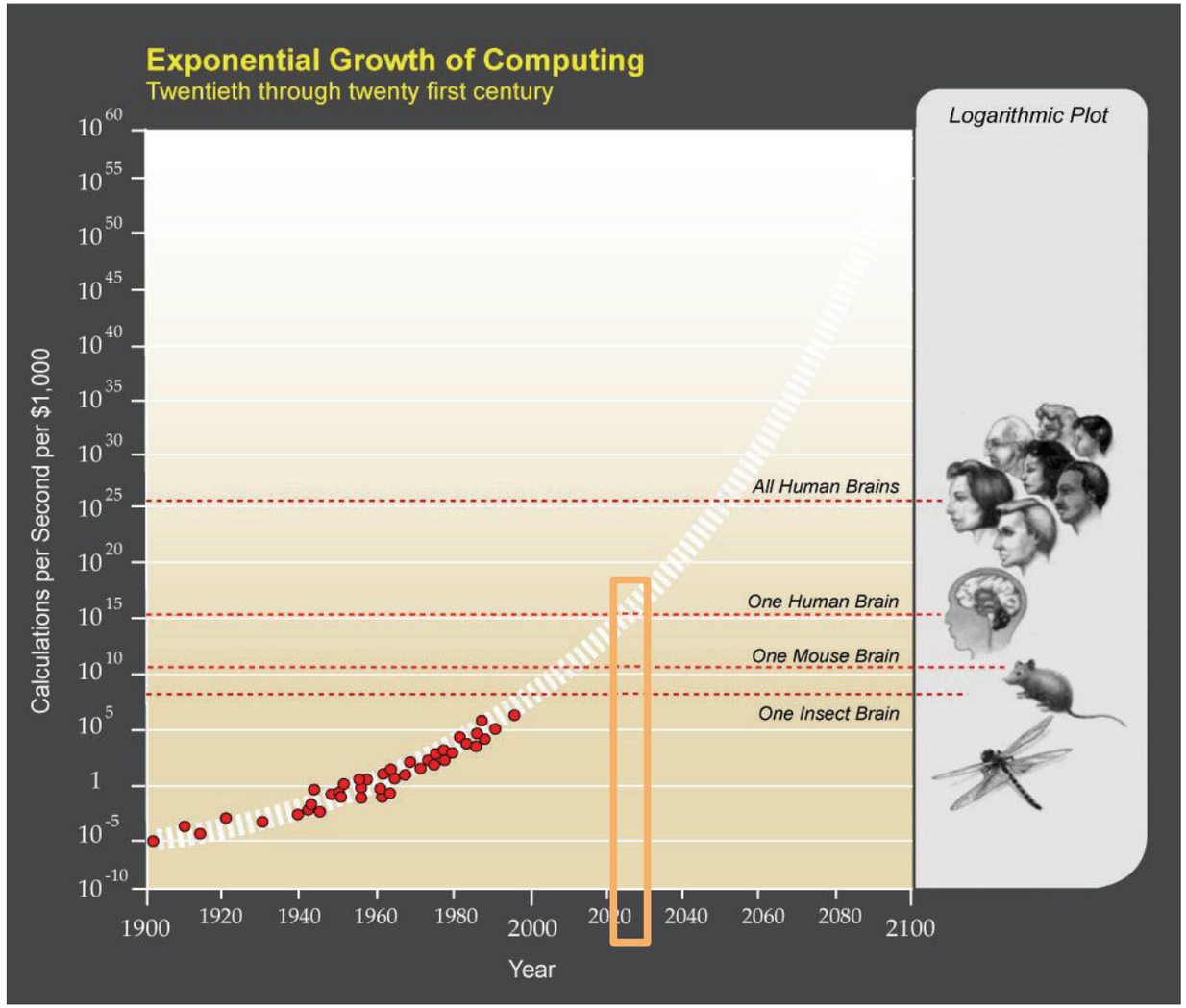
The speed at which artificial intelligence models master benchmarks and surpass human baselines is accelerating. But they often fall short in the real world.



(GRAPHIC) K. FRANKLIN/*SCIENCE*; (DATA) D. KIELA *ET AL.*, DYNABENCH: RETHINKING BENCHMARKING IN NLP, DOI:10.48550/ARXIV.2104.14337

Source: Sequoia Dec, 2022





2005 -- By Courtesy of Ray Kurzweil and Kurzweil Technologies, Inc. - en:PPTExponentialGrowthof_Computing.jpg, CC BY 1.0, https://commons.wikimedia.org/w/index.php?curid=3324354