

Tree of Thoughts

(a) Input-Output Prompting

AI News

Samuel Albanie



Scaling Speech Tech to 1000+ Languages

MEGABYTE: Predicting Million-byte Sequences with Multiscale Transformers

Lili Yu^{*1} Dániel Simig^{*1} Colin Flaherty^{*2} Armen Aghajanyan¹ Luke Zettlemoyer¹ Mike Lewis¹

Abstract

Autoregressive transformers are spectacular models for short sequences but scale poorly to long sequences such as high-resolution images, podcasts, code, or books. We propose MEGABYTE, a multiscale decoder architecture that enables end-to-end differentiable modeling of sequences of over one million bytes. MEGABYTE segments sequences into patches and uses a *local* sub-model within patches and a *global* model between patches. This enables sub-quadratic self-attention, much larger feedforward layers for the same compute, and improved parallelism during decoding—unlocking better performance at reduced cost for both training and generation. Extensive experiments show that MEGABYTE allows byte-level models to perform competitively with subword models on long context language modeling, achieve state-of-the-art density estimation on ImageNet, and model audio from raw files. Together, these results establish

MegaByte

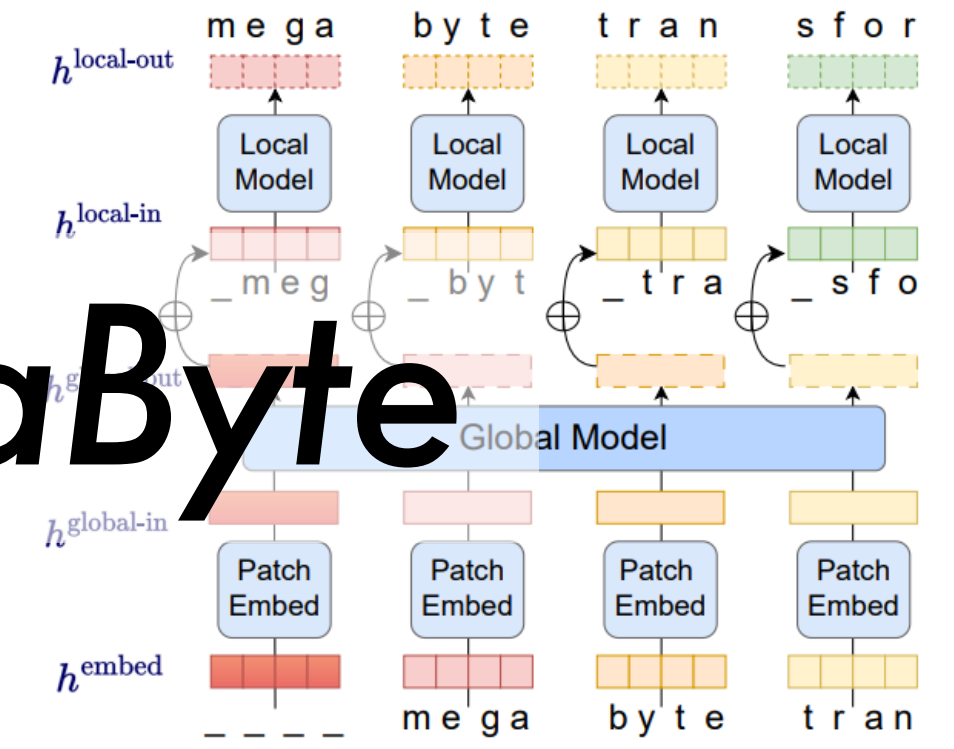


Figure 1. Overview of MEGABYTE with patch size $P = 4$. A small *local* model autoregressively predicts each patch byte-by-byte, using the output of a larger *global* model to condition on

v2 [cs.LG] 19 May 2023

LABOUR LONG TERM
Home About Us Supporters Briefings FAQs Blog Contact

Great British Cloud

Great British Cloud and BritGPT: the UK's AI Industrial Strategy Must Play to Our Strengths

AI is a crucial technology of our time. The UK Government should lay out a plan for founding, and over the next parliament ramping up investment in, two new publicly owned companies: Great British Cloud to £1-10 billion, and BritGPT to up to £1 billion

20 May 2023 | 14 min read

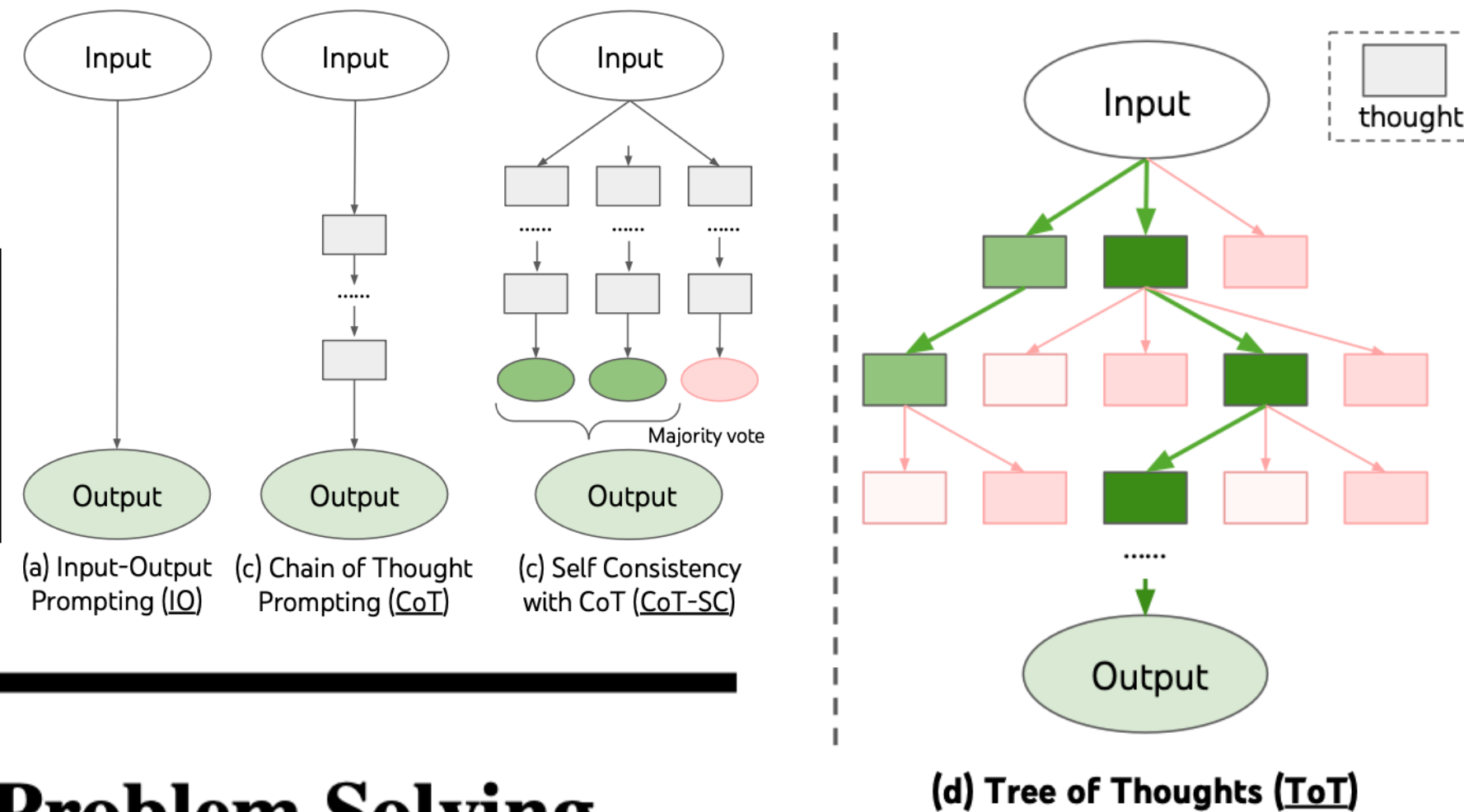
Download PDF

Executive Summary

AI is a crucial technology of our time, a transformative general-purpose technology, and will be crucial to our economy and security. The UK Government needs to act to strengthen our economy and security, reduce our dependence and vulnerability, and set us up for success. However, the AI supply chain is mind-bogglingly complex, specialised, and expensive. We have to choose where to invest. On AI industrial strategy, we must be careful with our investments and play to our strengths.

Tree of Thoughts

"generalises over the popular 'Chain of Thought' approach to **prompting**"



Tree of Thoughts: Deliberate Problem Solving with Large Language Models

17th May 2023

Method	Success
IO prompt	7.3%
CoT prompt	4.0%
CoT-SC (k=100)	9.0%
ToT (ours) (b=1)	45%
ToT (ours) (b=5)	74%
IO + Refine (k=10)	27%
IO (best of 100)	33%
CoT (best of 100)	49%

Table 2: Game of 24 Results.

Shunyu Yao
Princeton University

Dian Yu
Google DeepMind

Jeffrey Zhao
Google DeepMind

Izhak Shafran
Google DeepMind

Thomas L. Griffiths
Princeton University

Yuan Cao
Google DeepMind

Karthik Narasimhan
Princeton University

Abstract

Language models are increasingly being deployed for general problem solving across a wide range of tasks, but are still confined to token-level, left-to-right decision-making processes during inference. This means they can fall short in tasks that require exploration, strategic lookahead, or where initial decisions play a pivotal role. To surmount these challenges, we introduce a new framework for language model inference, “Tree of Thoughts” (ToT), which generalizes over the popular “Chain of Thought” approach to prompting language models, and enables exploration over coherent units of text (“thoughts”) that serve as intermediate steps toward problem solving. ToT allows LMs to perform deliberate decision making

The screenshot shows the GitHub repository page for 'ysymyth/tree-of-thought-llm'. The repository is public and has 169 watchers, 26 forks, and 650 stars. It contains a commit history with files like 'data', 'logs', 'prompts', 'scripts', 'tasks', '.gitignore', 'models.py', 'readme.md', and 'run.py'. The 'readme.md' file is expanded, showing the title 'Tree of Thoughts (ToT)' and a link to the paper: 'Code for paper Tree of Thoughts: Deliberate Problem Solving with Large Language Models. Also check its tweet thread in 1min.'

Scaling Speech Tech to 1000+ Languages

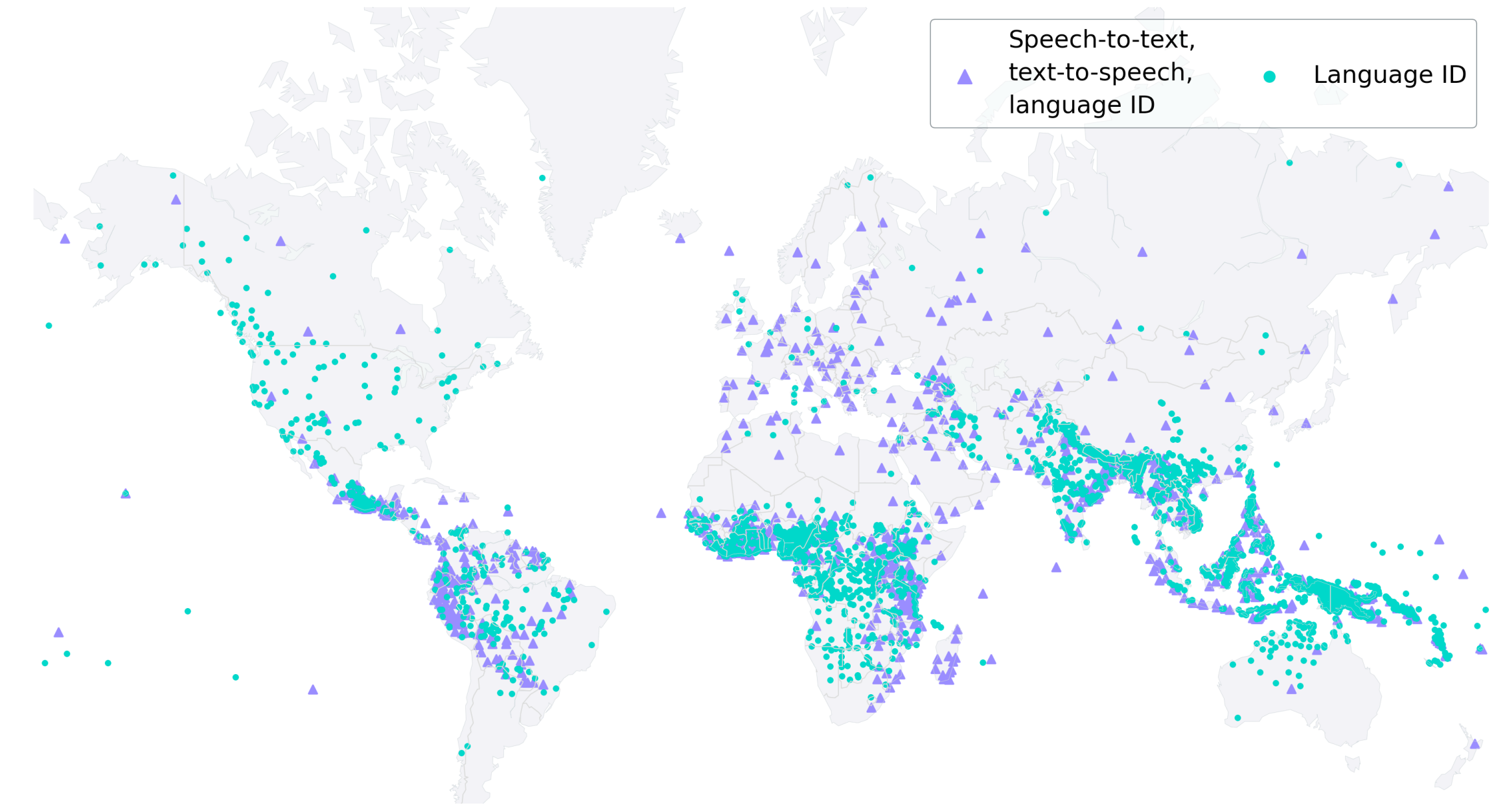
22nd May 2023

"wav2vec 2.0 models covering 1,406 languages"

Uses "readings of publicly available religious texts"

"ASR & synthesis models for 1,107 languages"

"language identification for 4,017 languages"



Scaling Speech Technology to 1,000+ Languages

Vineel Pratap[◇] Andros Tjandra[◇] Bowen Shi[◇]

Paden Tomasello Arun Babu Sayani Kundu* Ali Elkahky[†]

Zhaoheng Ni Apoorv Vyas Maryam Fazel-Zarandi Alexei Baevski[‡]

Yossi Adi[§] Xiaohui Zhang Wei-Ning Hsu Alexis Conneau[¶]

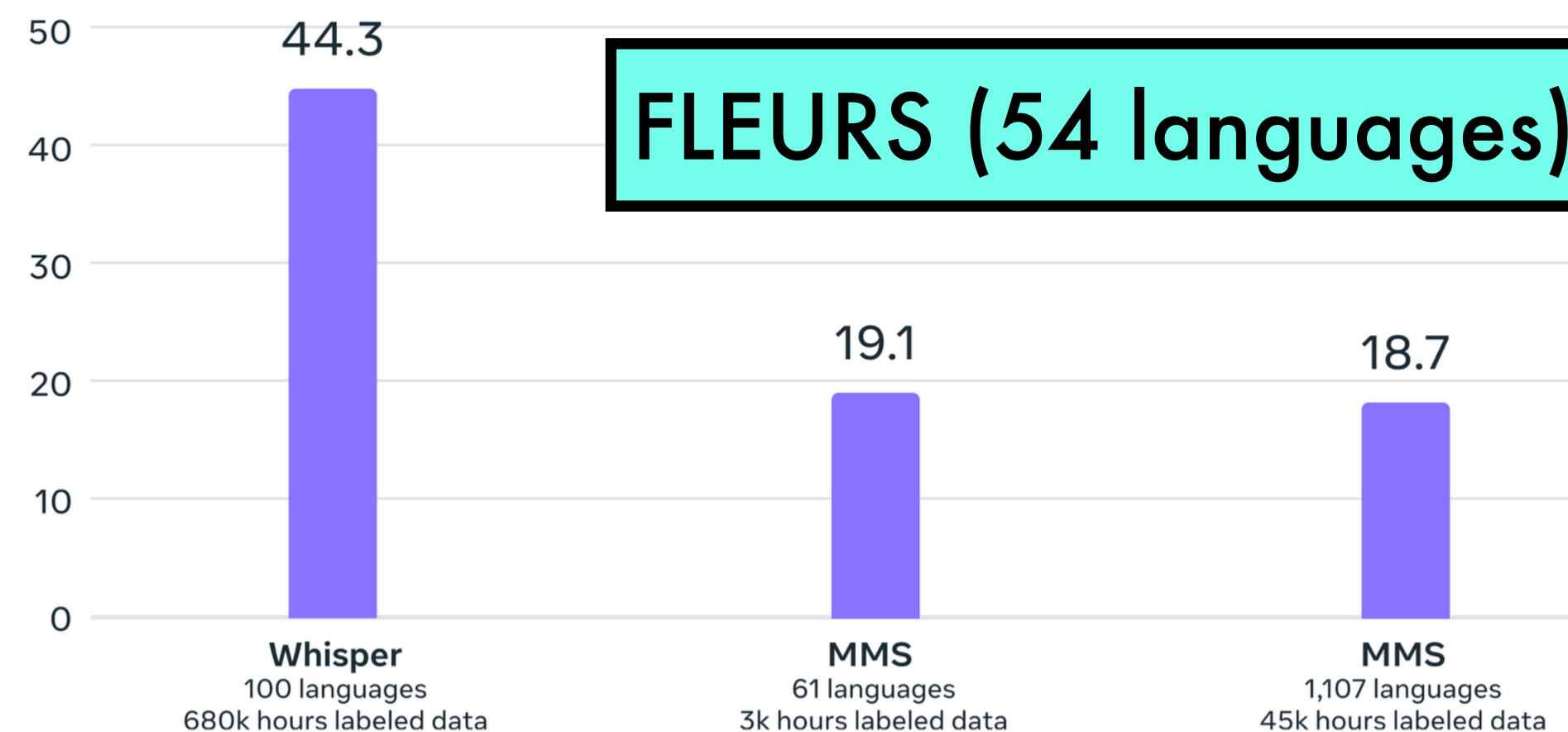
Michael Auli[◇]

Meta AI [◇] core team

Abstract

Expanding the language coverage of speech technology has the potential to improve access to information for many more people. However, current speech technology is restricted to about one hundred languages which is a small fraction of the over 7,000 languages spoken around the world. The Massively Multilingual Speech (MMS) project increases the number of supported languages by 10-40x, depending on the task. The main ingredients are a new dataset based on readings of publicly available religious texts and effectively leveraging self-supervised learning. We built pre-trained wav2vec 2.0 models covering 1,406 languages, a single multilingual automatic speech recognition model for 1,107 languages, speech synthesis models for the same number of languages, as well as a language identification model for 4,017 languages. Experiments show that our multilingual speech recognition model more than halves the word error rate compared to the state-of-the-art monolingual model on 54 languages.

Word error rates



A screenshot of a GitHub repository page for 'fairseq/examples/mms'. The page shows a commit history table with columns for Name, Last commit message, and Last commit date. Below the table is a README file titled 'MMS: Scaling Speech Technology to 1000+ languages'. The README text is partially visible, starting with 'The Massively Multilingual Speech (MMS) project expands speech technology from about 100 languages to over 1,000 by building a single multilingual speech recognition model supporting...'

VisionLLM

18th May 2023

"...a unified perspective for vision and language tasks by treating images as a **foreign language**"

VisionLLM: Large Language Model is also an Open-Ended Decoder for Vision-Centric Tasks

Wenhai Wang^{*1}, Zhe Chen^{*2,1}, Xiaokang Chen^{*3,1}, Jiannan Wu^{*4,1}, Xizhou Zhu^{5,1},
Gang Zeng³, Ping Luo^{4,1}, Tong Lu², Jie Zhou⁶, Yu Qiao¹, Jifeng Dai^{†6,1}
¹OpenGVLab, Shanghai AI Laboratory ²Nanjing University ³Peking University
⁴The University of HongKong ⁵SenseTime Research ⁶Tsinghua University

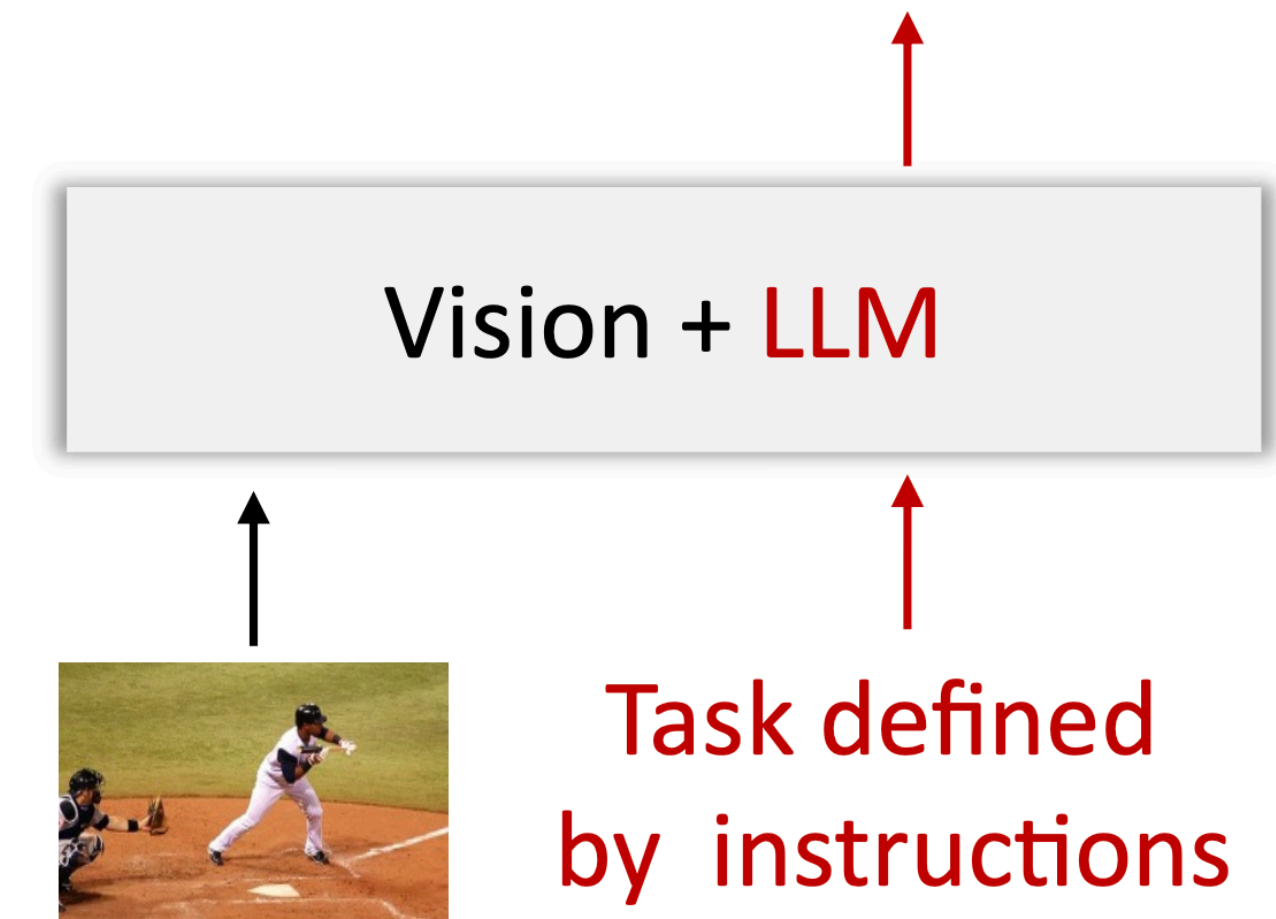
Code: <https://github.com/OpenGVLab/VisionLLM>

Demo: <https://github.com/OpenGVLab/InternGPT>

Abstract

Large language models (LLMs) have notably accelerated progress towards artificial general intelligence (AGI), with their impressive zero-shot capacity for user-tailored tasks, endowing them with immense potential across a range of applications. However, in the field of computer vision, despite the availability of numerous powerful vision foundation models (VFMs), they are still restricted to tasks in a pre-defined form, struggling to match the open-ended task capabilities of LLMs. In this work, we present an LLM-based framework for vision-centric tasks, termed VisionLLM. This framework provides a unified perspective for vision and language

Desired output:
<c1> <p1> <p3> ...



"with a **generalist LLM**... our model can achieve **60% mAP on COCO**"

ToolkenGPT

"In context learning...only allows for a few shots of demonstrations"

19th May 2023

"LLMs with external tools... a promising approach to solving complex problems"

ToolkenGPT "represents each tool as a token ('toolken')"

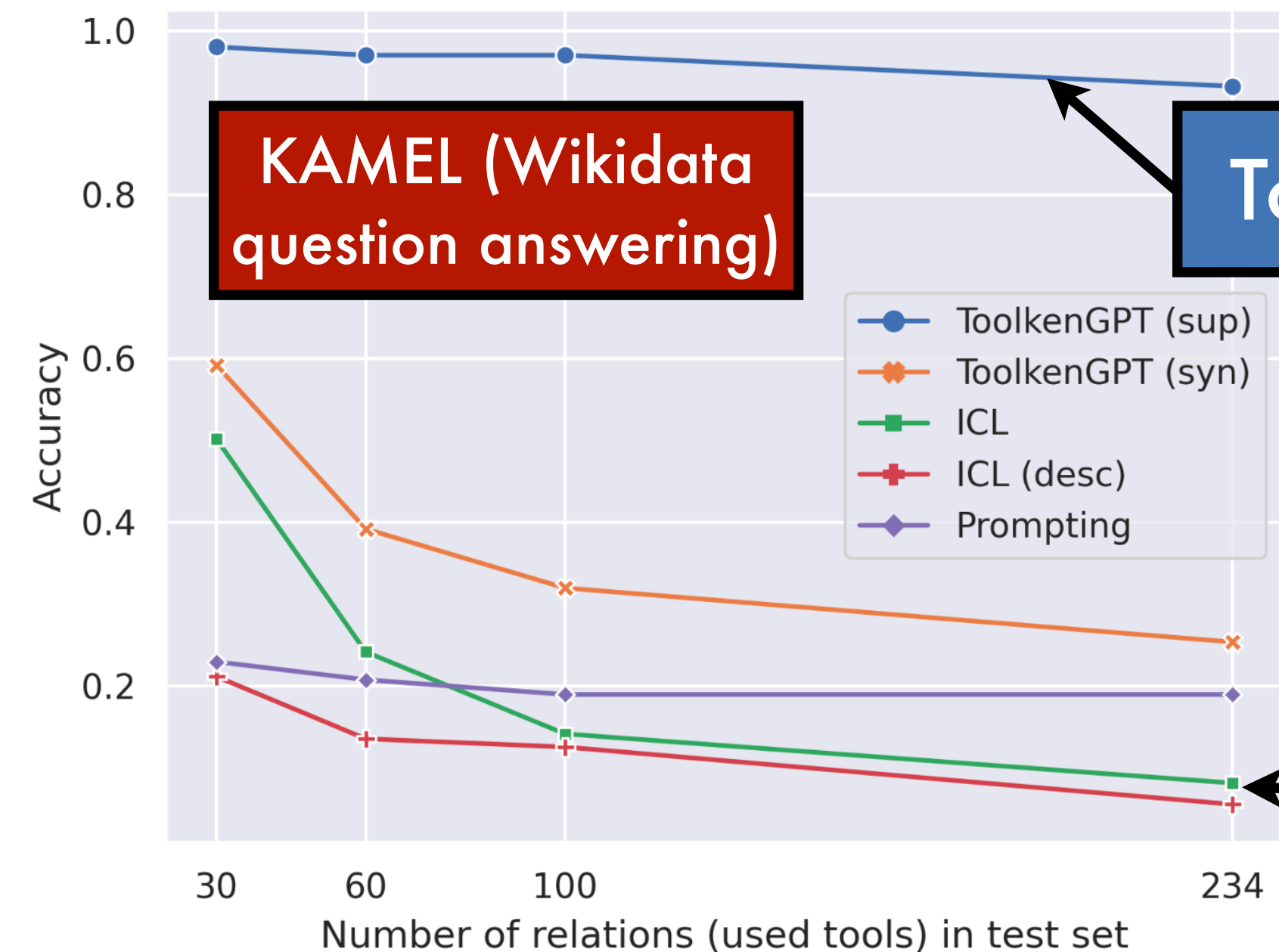
"...allows extensive demonstration data for learning the embeddings (toolkens)"

ToolkenGPT: Augmenting Frozen Language Models with Massive Tools via Tool Embeddings

Shibo Hao¹, Tianyang Liu¹, Zhen Wang^{1,2}, Zhiting Hu¹
¹UC San Diego, ²Mohamed bin Zayed University of Artificial Intelligence
{s5hao, ti1040, zhw085, zhh019}@ucsd.edu

Abstract

Augmenting large language models (LLMs) with external tools has emerged as a promising approach to solving complex problems. However, traditional methods, which finetune LLMs with tool demonstration data, can be both costly and restricted to a predefined set of tools. Recent in-context learning paradigm issues, but the limited context length only allows for a few shots of leading to suboptimal understandings of the tools. Moreover, numerous tools to choose from, in-context learning could complete. In this paper, we propose an alternative approach, ToolkenGPT, with the benefits of both sides. Our approach represents each tool as a token and learns an embedding for it, enabling tool calls in the same way as generating a regular word token. Once a toolken is triggered, the LLM is prompted to



KAMEL (Wikidata question answering)

ToolkenGPT

ICL

"when toolken is predicted, LLM switches into a special mode...and injects the outputs back into the generation"

MEGABYTE

19th May 2023

"Multi-scale decoder architecture"

"Segments sequences into patches:"

"Local submodel within patches"

"Global model between patches"

"Sub-quadratic self-attention"

"Per-patch feedforward layers"

"Parallelism in Decoding"

"Tokenization-free autoregressive sequence modelling at scale"

MEGABYTE: Predicting Million-byte Sequences with Multiscale Transformers

Lili Yu^{*1} Dániel Simig^{*1} Colin Flaherty^{*2} Armen Aghajanyan¹ Luke Zettlemoyer¹ Mike Lewis¹

Abstract

Autoregressive transformers are spectacular models for short sequences but scale poorly to long sequences such as high-resolution images, podcasts, code, or books. We propose MEGABYTE, a multi-scale decoder architecture that enables end-to-end differentiable modeling of sequences of over one million bytes. MEGABYTE segments sequences into patches and uses a *local* submodel within patches and a *global* model between patches. This enables sub-quadratic self-attention, much larger feedforward layers for the same compute, and improved parallelism during decoding—unlocking better performance at reduced cost for both training and generation. Extensive experiments show that MEGABYTE allows byte-level models to perform competitively with subword models on long context language modeling, achieve state-of-the-art density estimation on ImageNet, and model audio from raw files. Together, these results estab-

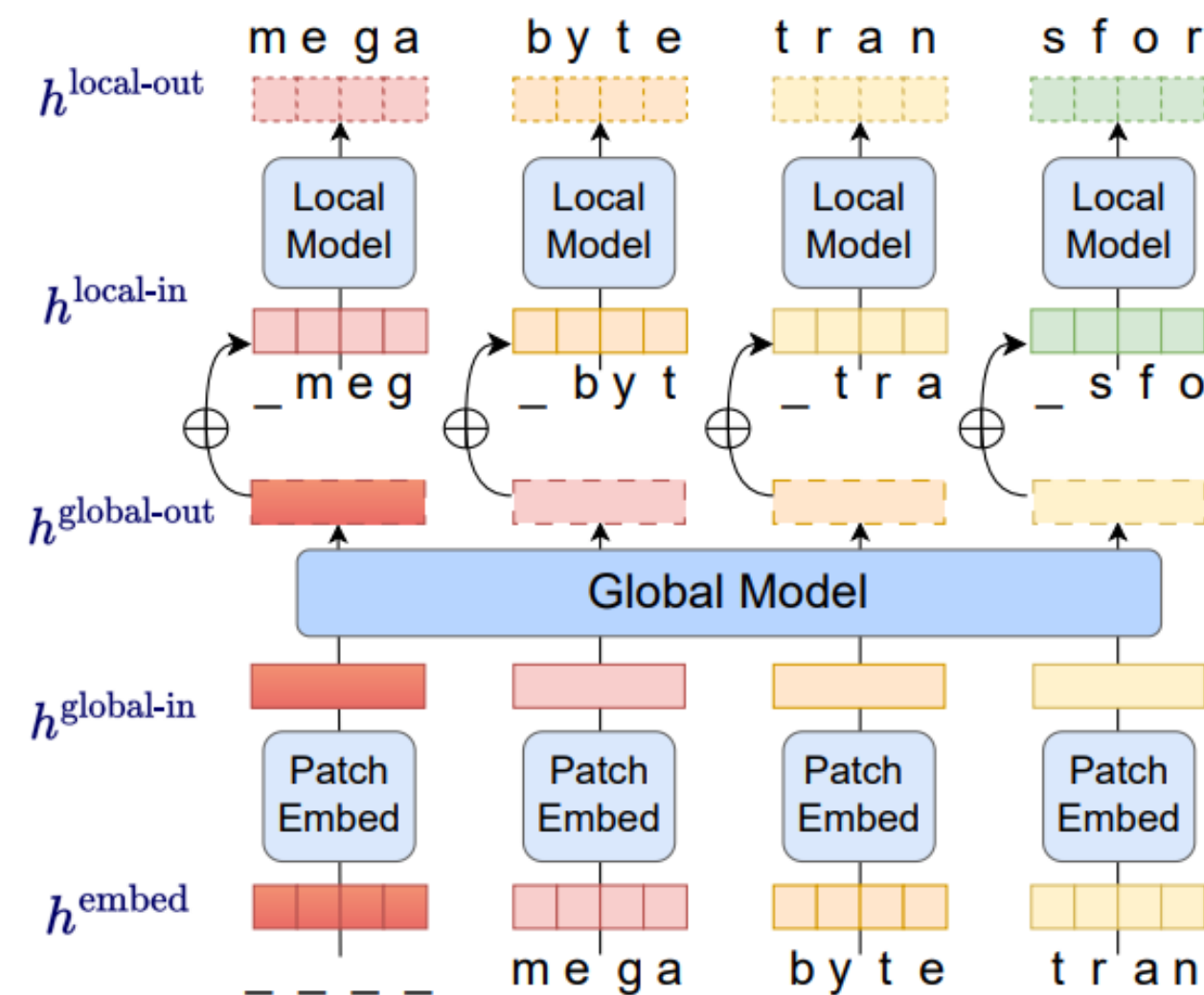
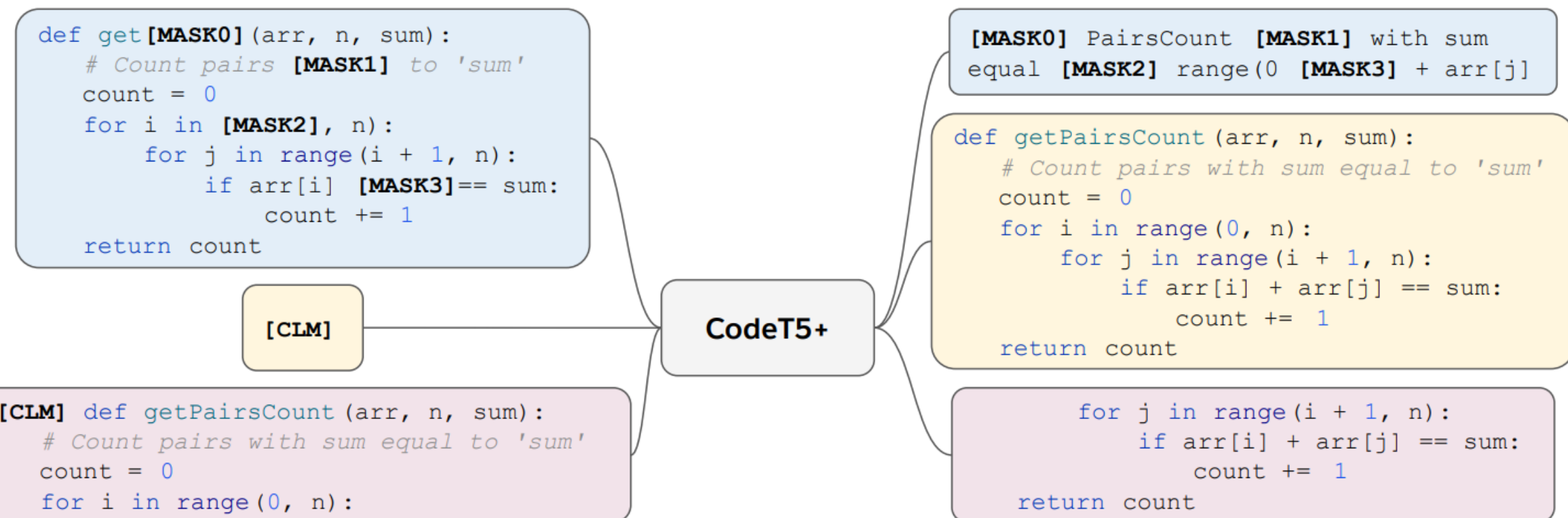


Figure 1. Overview of MEGABYTE with patch size $P = 4$. A small *local* model autoregressively predicts each patch byte-by-byte, using the output of a larger *global* model to condition on

CodeT5+

20th May 2023



span denoising

Decoder-only causal LM

Seq2Seq causal LM

CodeT5+: Open Code Large Language Models for Code Understanding and Generation

Yue Wang*, Hung Le*, Akhilesh Deepak Gotmare, Nghi D.Q. Bui, Junnan Li, Steven C.H. Hoi
Salesforce AI Research
<https://github.com/salesforce/CodeT5/tree/main/CodeT5+>

Abstract

Large language models (LLMs) pretrained on vast source code have achieved prominent progress in code intelligence. However, existing code LLMs have two main limitations in terms of architecture and pretraining tasks. First, they often adopt a specific architecture (encoder-only or decoder-only) or rely on a unified encoder-decoder network for different downstream tasks. The former paradigm is limited by inflexibility in applications while in the latter, the model is treated as a single system for all tasks, leading to suboptimal performance on a subset of tasks. Secondly, they often employ a limited set of pretraining objectives which might not be relevant to some downstream tasks and hence result in substantial performance degrade. To address these limitations, we propose “CodeT5+”, a family of encoder-decoder LLMs for code in which component modules can be flexibly combined to suit a wide range of downstream code tasks. Such flexibility is enabled by our proposed mixture of pretraining objectives to mitigate the pretrain finetune

Model	Pass@1 on HumanEval
LLaMA 7B	10.5
LaMDA 137B	14.0
InCoder 6B	15.2
GPT-NeoX 20B	15.4
CodeT5+ 770M	15.5
LLaMA 13B	15.8
PaLM 62B	15.9
AlphaCode 1.1B	17.1
LLaMA 33B	21.7
Replit 3B	21.9
CodeGeeX 13B	22.9
LLaMA 65B	23.7
PaLM 540B	26.2
CodeGen-mono 16B	29.3
CodeT5+ 16B	30.9
code-cushman-001	33.5
StarCoder 15B	33.6
InstructCodeT5+ 16B	35.0

salesforce / CodeT5 Public

Code Issues 14 Pull requests 3 Actions Projects Security Insights

main CodeT5 / CodeT5+ Go to file Add file

yuewang-cuhk update readme 430d7e3 · 3 days ago History

Name	Last commit message	Last commit date
..		
humaneval	add humaneval evaluation	3 days ago
README.md	update readme	3 days ago
codet5p_architecture.png	reorganize the repo	last week
codet5p_overview.png	update readme	last week

CodeT5+

Official research release for the CodeT5+ models (220M , 770M , 2B , 6B 16B) for a wide range of Code Understanding and Generation tasks. Find out more via our [blog post](#).

Title: [CodeT5+: Open Code Large Language Models for Code Understanding and Generation](#)

Authors: [Yue Wang*](#), [Hung Le*](#), [Akhilesh Deepak Gotmare](#), [Nghi D.Q. Bui](#), [Junnan Li](#), [Steven C.H. Hoi](#) (* indicates equal contribution)

2 [cs.CL] 20 May 2023

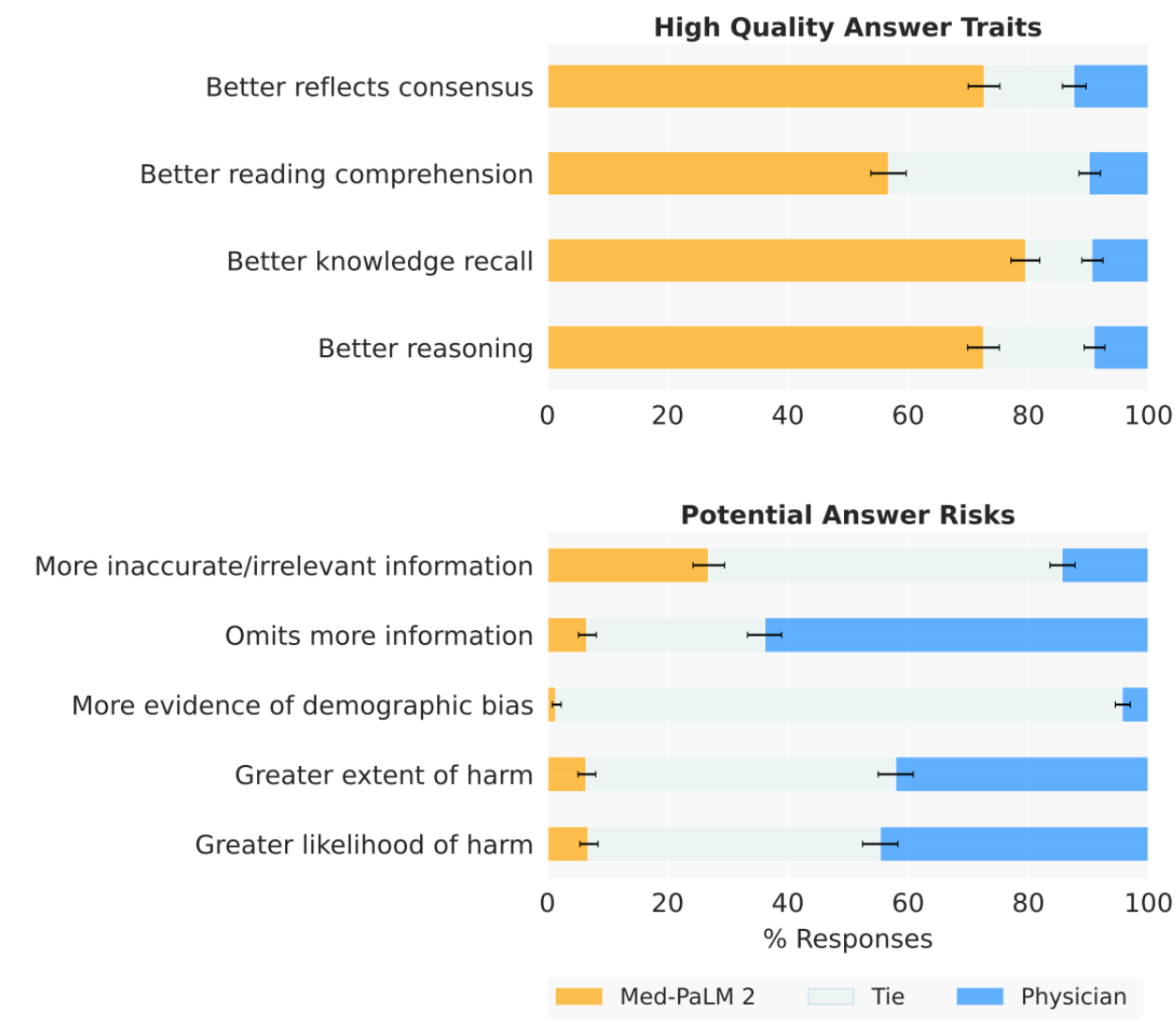
Med-PaLM 2

16th May 2023

PaLM 2

+ medical domain finetuning

+ *ensemble refinement* prompting



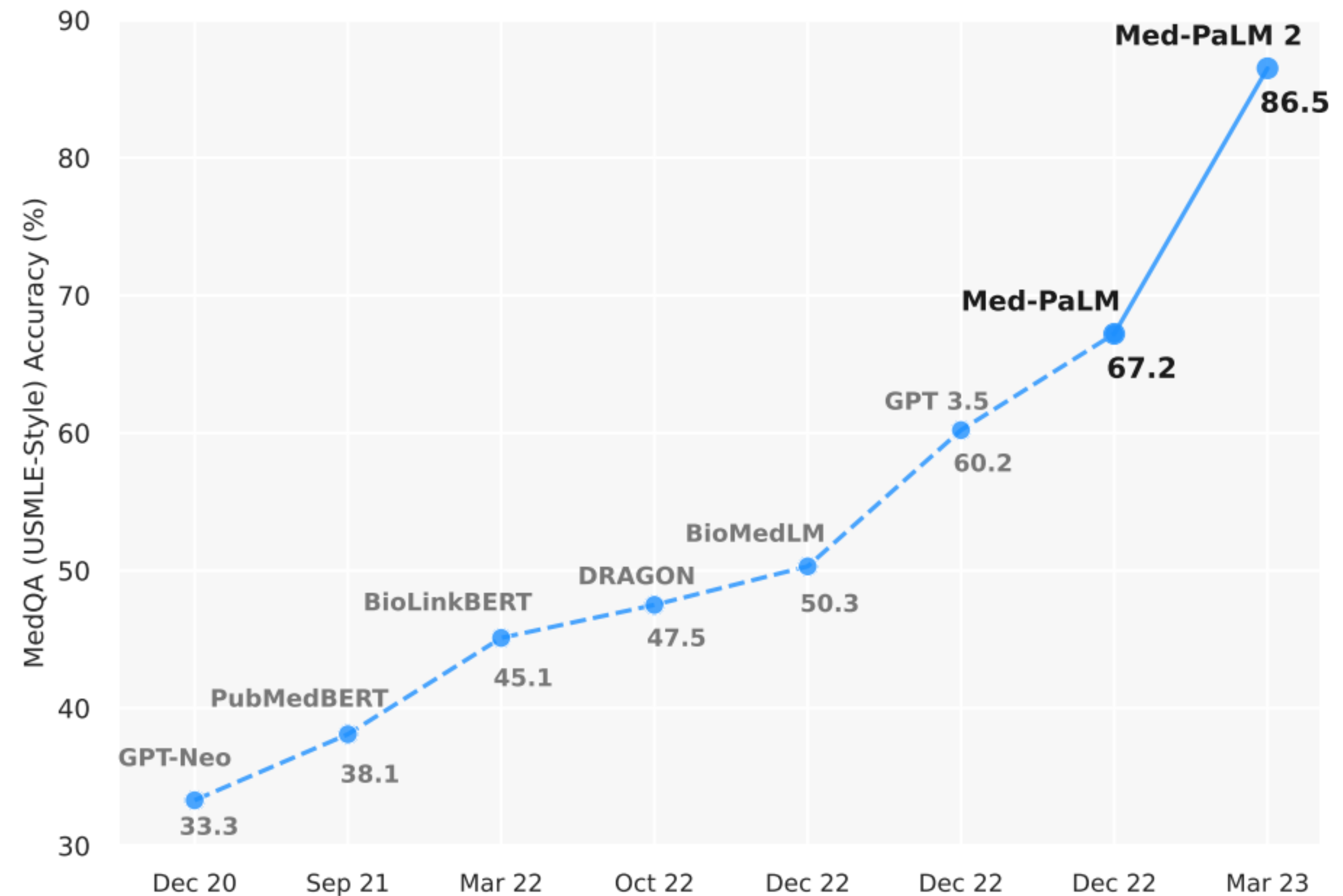
Towards Expert-Level Medical Question Answering with Large Language Models

Karan Singhal^{*,1}, Tao Tu^{*,1}, Juraj Gottweis^{*,1}, Rory Sayres^{*,1}, Ellery Wulczyn¹, Le Hou¹, Kevin Clark¹, Stephen Pfohl¹, Heather Cole-Lewis¹, Darlene Neal¹, Mike Schaekermann¹, Amy Wang¹, Mohamed Amin¹, Sami Lachgar¹, Philip Mansfield¹, Sushant Prakash¹, Bradley Green¹, Ewa Dominowska¹, Blaise Aguera y Arcas¹, Nenad Tomasev², Yun Liu¹, Renee Wong¹, Christopher Semturs¹, S. Sara Mahdavi¹, Joelle Barral¹, Dale Webster¹, Greg S. Corrado¹, Yossi Matias¹, Shekoofeh Azizi^{†,1}, Alan Karthikesalingam^{†,1} and Vivek Natarajan^{†,1}

¹Google Research, ²DeepMind,

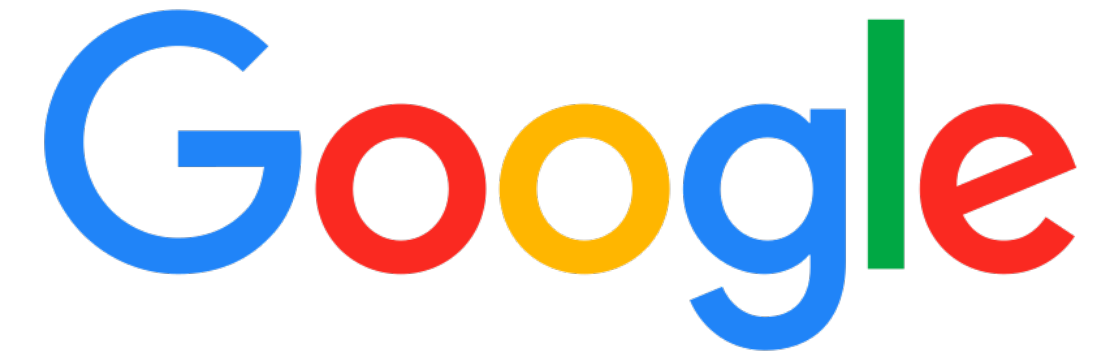
Recent artificial intelligence (AI) systems have reached milestones in “grand challenges” ranging from Go to protein-folding. The capability to retrieve medical knowledge, reason over it, and answer medical questions comparably to physicians has long been viewed as one such grand challenge.

Large language models (LLMs) have catalyzed significant progress in medical question answering; Med-PaLM was the first model to exceed a “passing” score in US Medical Licensing Examination (USMLE)



[cs.CL] 16 May 2023

Getting ViT in Shape



22nd May 2023

"small vision models can perform on par with larger ones with the same compute...."

"if we optimize their **shape**"

Getting ViT in Shape: Scaling Laws for Compute-Optimal Model Design

"SoViT-400m/14 achieves 90.3%
finetuning acc on ILSRCV2012"

Ibrahim Alabdulmohsin*, Xiaohua Zhai*, Alexander Kolesnikov, Lucas Beyer*

Google DeepMind
Zürich, Switzerland

{ibomohsin,xzhai,akolesnikov,lbeyer}@google.com

"MLP dim. is scaled faster than depth"

depth is scaled faster than width

Abstract

Scaling laws have been recently employed to derive compute-optimal model size (number of parameters) for a given compute duration. We advance and refine such methods to infer compute-optimal *model shapes*, such as width and depth, and successfully implement this in vision transformers. Our shape-optimized vision transformer, SoViT, achieves results competitive with models that exceed twice its size, despite being pre-trained with an equivalent amount of compute. For example,

"According to..." prompting

22nd May 2023

"LLMs struggle with hallucination"

"according-to prompting: directing LLMs to ground responses against previously observed text"

Prompt "Based on evidence from Wikipedia..."

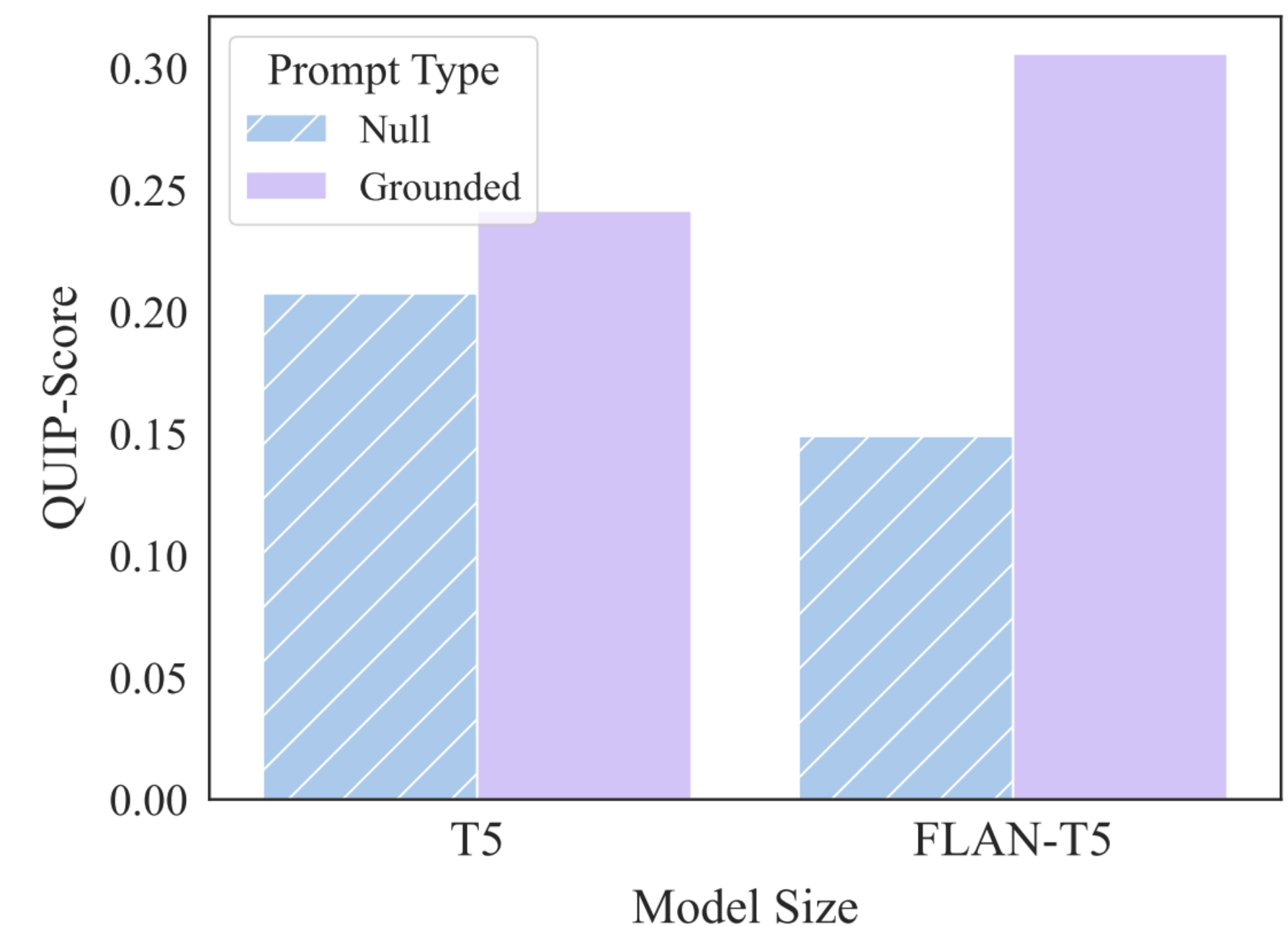
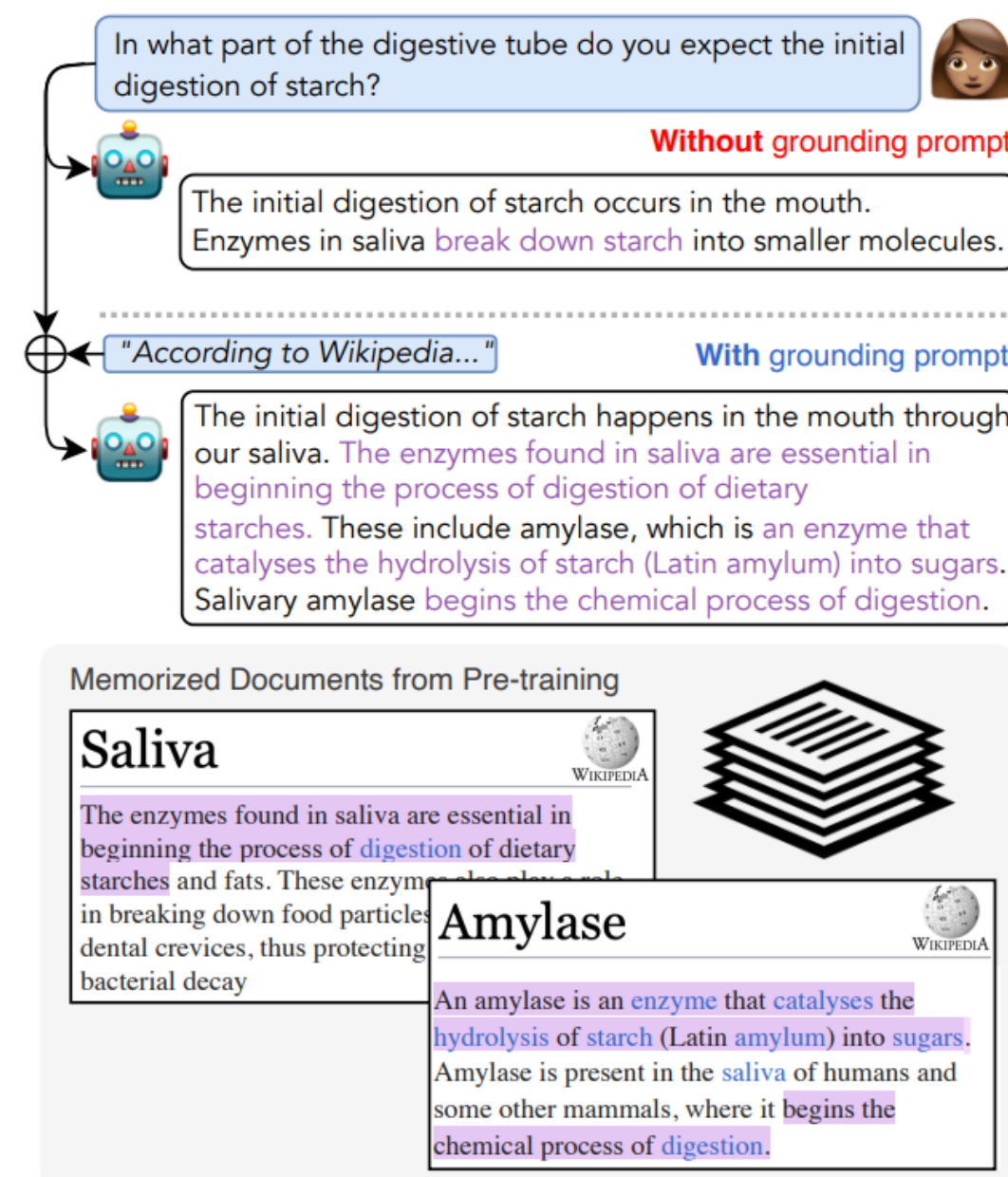
"According to ..."

Prompting Language Models Improves Quoting from Pre-Training Data

Orion Weller*, Marc Marone*, Nathaniel Weir,
Dawn Lawrie, Daniel Khashabi, Benjamin Van Durme
Johns Hopkins University
{owell2,mmarone1}@jhu.edu

Abstract

Large Language Models (LLMs) may hallucinate and generate fake information, despite pre-training on factual data. Inspired by the journalistic device of "according to sources", we propose according-to prompting: directing LLMs to ground responses against previously observed text. To quantify this grounding, we propose a novel evaluation metric (QUIP-Score) that measures the extent to which model-produced answers are directly found in underlying text corpora. We illustrate with experiments on Wikipedia that these prompts improve grounding under our metrics, with the additional benefit of often improving end-task performance. Furthermore, prompts that ask the model to decrease grounding (or to ground to other corpora) decrease grounding, indicating the ability of language models to increase or decrease grounded generations on request.



Complex Claim Verification

19th May 2023

"fully automated pipeline to **check real-world claims** by retrieving raw evidence from the web"

Complex Claim Verification with Evidence Retrieved in the Wild

Jifan Chen Grace Kim Aniruddh Sriram Greg Durrett Eunsol Choi

Department of Computer Science
The University of Texas at Austin
jfchen@cs.utexas.edu

Abstract

Evidence retrieval is a core part of automatic fact-checking. Prior work makes simplifying assumptions in retrieval that depart from real-world use cases: either no access to evidence, access to evidence curated by a human fact-checker, or access to evidence available long after the claim has been made. In this work, we present the first fully automated pipeline to check real-world claims by retrieving raw evidence from the web. We restrict our retriever to only search documents available prior to the claim's making, modeling the realistic scenario where an emerging claim needs to be checked. Our pipeline includes five components: claim decomposition, raw document retrieval, fine-grained evidence re-

Claim: James Quintero stated on October 10, 2016: "When San Francisco banned plastic grocery bags, you saw the number of instances of people going to the ER ... spike."

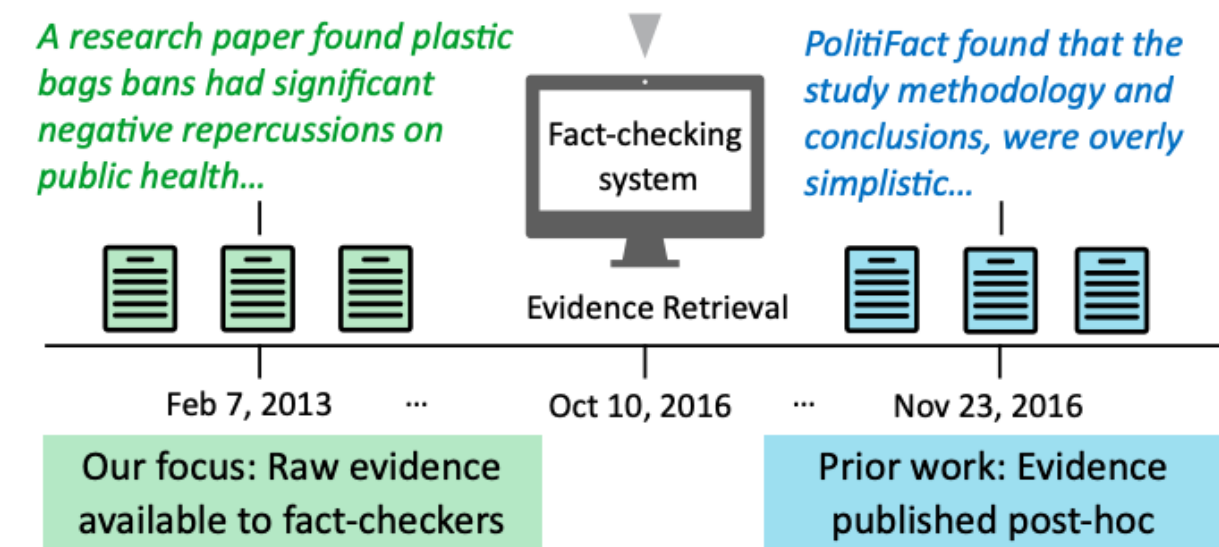


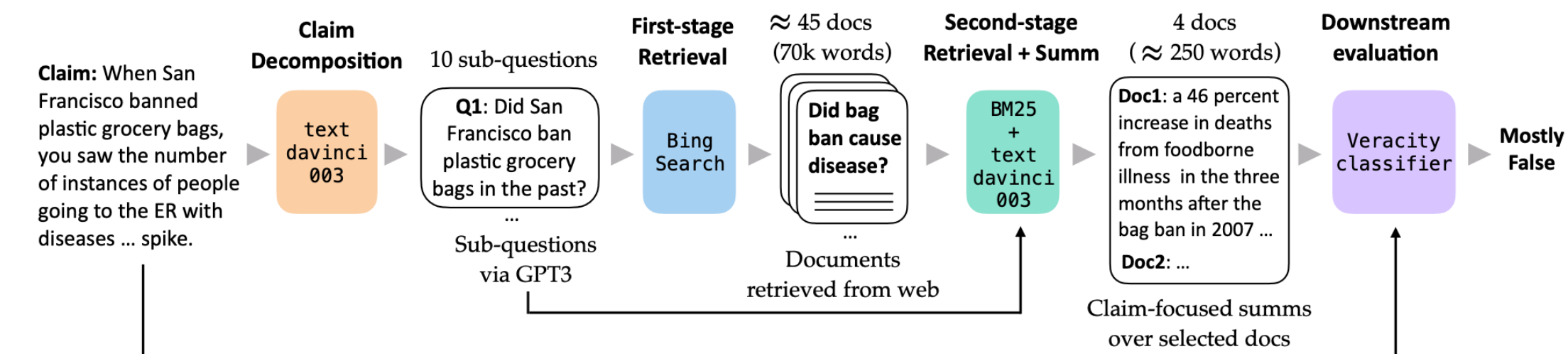
Figure 1: Our fact-check setting addresses realistic claims using evidence retrieved prior to when the claim was made.

work that does tackle real-world claims either relies on access to a document set which contains

Claim: James Quintero stated on October 10, 2016 in a panel discussion at SXSW Eco: When San Francisco banned plastic grocery bags, "you saw the number of instances of people going to the ER with things like salmonella and other related illnesses" spike.

text-davinci-003

- Q1: Did San Francisco ban plastic grocery bags in the past?
- Q2: Did the number of people going to the ER for salmonella and other related illnesses increase after the plastic bag ban in San Francisco?
- ...
- Q10: Was the plastic bag ban directly caused the increase of the number of people going to the ER for salmonella and other related illnesses in San Francisco?



LM vs LM

22nd May 2023

"key idea is that an incorrect claim is likely to result in **inconsistency**..."

"multi-turn interaction between:"

(1) "LM that generated the claim"

(2) "LM (acting as an examiner)"

LM vs LM: Detecting Factual Errors via Cross Examination

Roi Cohen¹ May Hamri¹ Mor Geva² Amir Globerson^{1,3}

¹Blavatnik School of Computer Science, Tel Aviv University

²Google DeepMind ³Google Research

{roi1, mayhamri}@mail.tau.ac.il, pipek@google.com, gamir@tauex.tau.ac.il

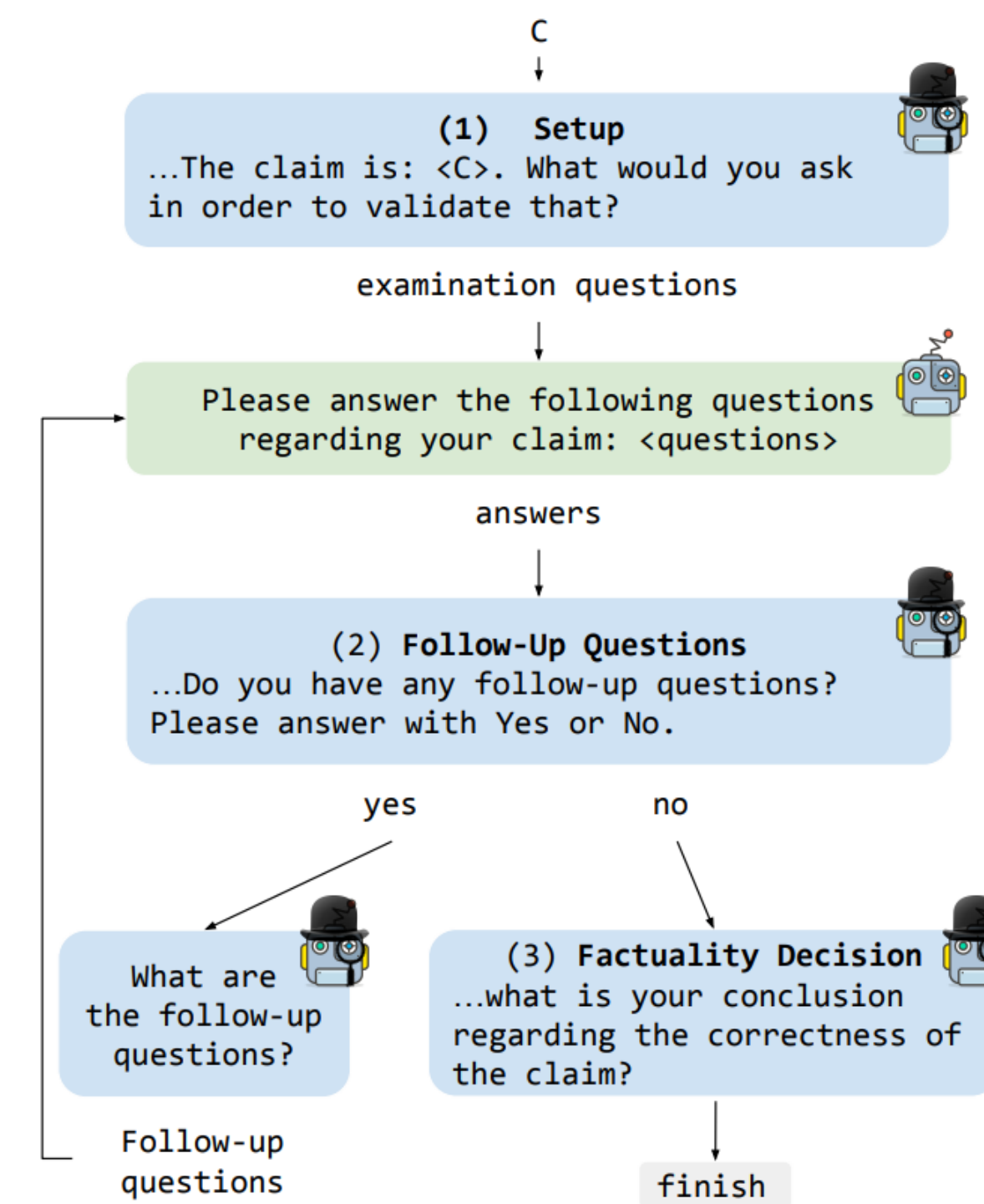
Abstract

A prominent weakness of modern language models (LMs) is their tendency to generate factually incorrect text, which hinders their usability. A natural question is whether such factual errors can be detected automatically. Inspired by truth-seeking mechanisms in law, we propose a factuality evaluation framework for LMs that is based on cross-examination. Our key idea is that an incorrect claim is likely to result in inconsistency with other claims that the model generates. To discover such inconsistencies, we facilitate a multi-turn interaction between the LM that generated the claim and

The Greek god of marriage is Hera.

1. Who is the Greek god of marriage?
2. What is the name of the wife of Zeus in Greek mythology?
3. Is Hera associated with marriage in any way?
4. Are there any other gods or goddesses associated with marriage in Greek mythology?

1. **The Greek god of marriage is Hymenaeus.**
2. The name of the wife of Zeus in Greek mythology is Hera.
3. Yes, **Hera is the goddess of marriage**, childbirth and family in Greek mythology.
4. Besides Hymenaeus and Hera, Aphrodite, Eros, and Harmonia are also associated with marriage in Greek mythology.



	LAMA		
	P	R	F1
AYS	82.3	25.2	38.6
IDK	49.1	52.4	50.7
LMvLM	85.1	70.7	76.7
LMvLM (Majority)	86.6	75.8	80.8
- Follow-up	83.8	68.1	75.1

AI News

ANTHROPIC

Company Announcements

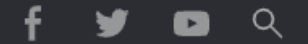
Anthropic Raises \$450 Million in Series C Funding to Scale Reliable AI Products

May 23, 2023 · 2 min read

engadget

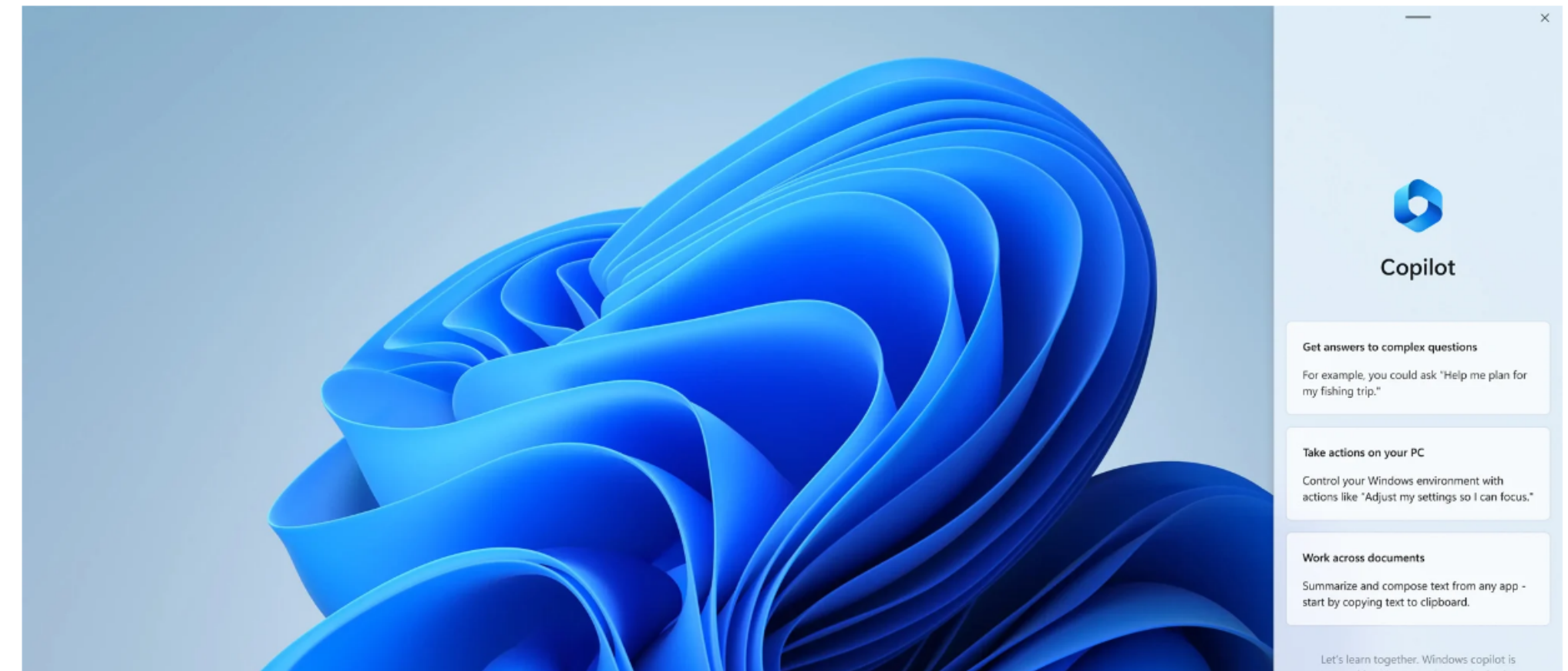
Login

Reviews Buying Guides Gaming Gear Entertainment Tomorrow Deals News Video Podcasts



Microsoft puts AI in the heart of Windows 11 with Windows Copilot

A chat with Panos Panay about the company's next big AI leap.



"led by **Spark Capital** with participation from Google, Salesforce Ventures, Sound Ventures, Zoom Ventures"

AI News

LABOUR LONG TERM

Home About Us Supporters Briefings FAQs Blog Contact

Great British Cloud and BritGPT: the UK's AI Industrial Strategy Must Play to Our Strengths

AI is a crucial technology of our time. The UK Government should lay out a plan for founding, and over the next parliament ramping up investment in, two new publicly owned companies: Great British Cloud to £1-10 billion, and BritGPT to up to £1 billion

20 May 2023 | 14 min read

Download PDF ↓

Executive Summary

AI is a crucial technology of our time, a transformative general-purpose technology, and will be crucial to our economy and security. The UK Government needs to act to strengthen our economy and security, reduce our dependence and vulnerability, and set us up for success. However, the AI supply chain is mind-bogglingly complex, specialised, and expensive. We have to choose where to invest. On AI industrial strategy, we must be careful with our investments and play to our strengths.

"prioritise investments in public cloud and public foundation models ...not try to be at the forefront of chip production"

Next Steps for Neeva

SRIDHAR RAMASWAMY AND VIVEK RAGHUNATHAN ON 05/20/23

We started Neeva with the mission to take search back to its users. Having worked on search and search ads for over a decade, we sincerely believed that there was space for a model of search that put user and not advertiser interests first—a private, ads-free experience.

"we will be shutting down neeva.com and our consumer search product..."

"...acquiring users has been really hard"

Technomancers.ai

e/acc



Delos Prime | May 13, 2023

EU AI Act To Target US Open Source Software

In a bold stroke, the EU's amended AI Act would ban American companies such as OpenAI, Amazon, Google, and IBM from providing API access to generative AI.

"seems to mean is that you can open source traditional machine learning models but **not generative AI.**"

"This is a deeply corrupt piece of legislation"

Published on May 16, 2023 In Endless Origins

No, European Draft AI Act Does Not Target Open-Source Softwares

The act have introduced a range of bans on what the European Parliament refers to as "intrusive and discriminatory uses of AI systems."

By Lokesh Choudhary



"the draft AI act does **not target** the open-source ecosystem"

AI News

FEATURED

Hardware

MTIA v1: Meta's first-generation AI inference accelerator

May 18, 2023 · 8 min read



FEATURED

Research

Pursuing groundbreaking scale and accelerating research using Meta's Research SuperCluster

May 18, 2023 · 8 min read



16K NVIDIA A100s

AI News

AI

Together raises \$20M to build open source generative AI models

Kyle Wiggers @kyle_l_wiggers / 10:36 PM GMT+1 • May 15, 2023

Comment

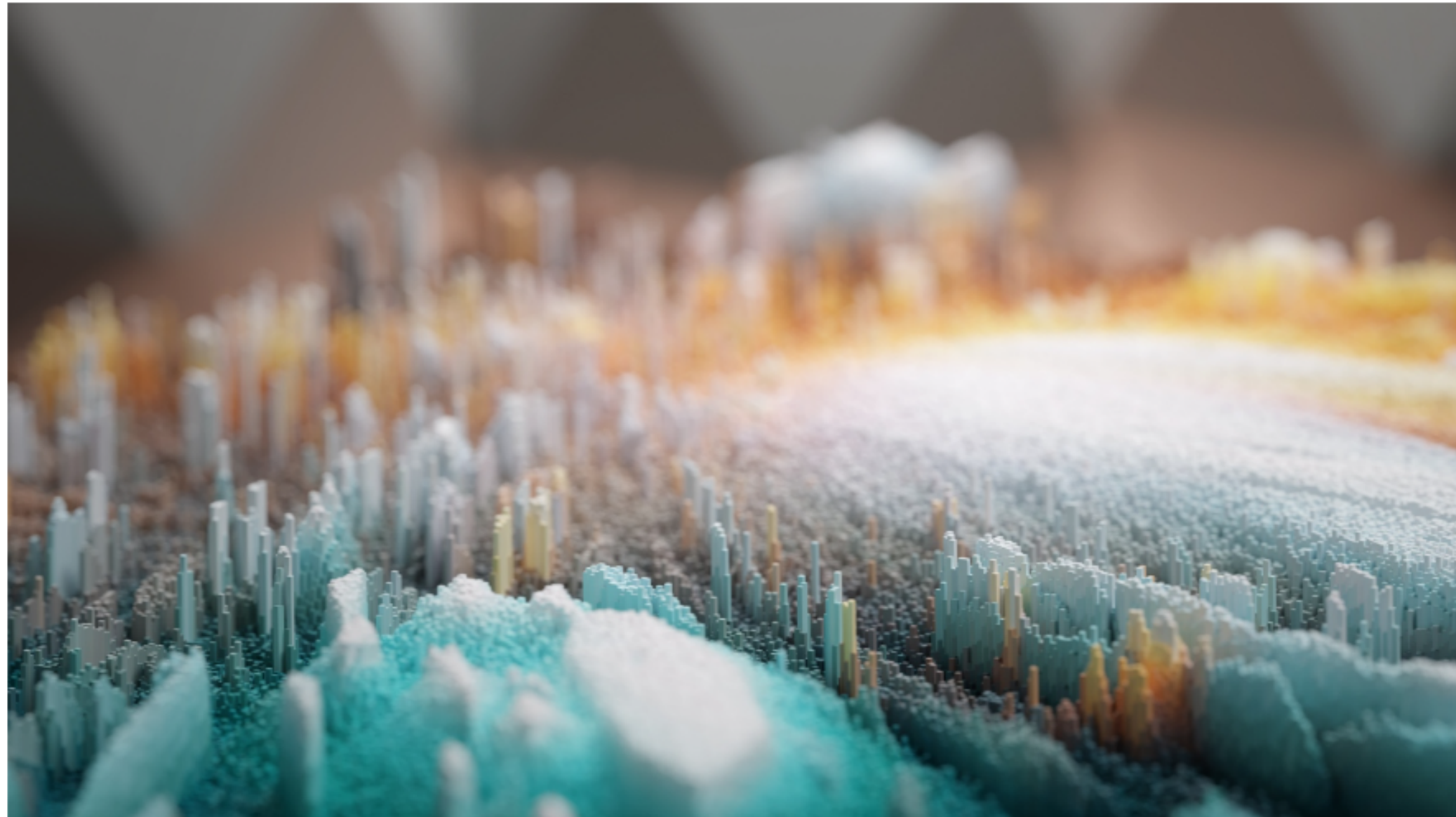


Image Credits: koto_feja / Getty Images

Generative AI — AI that can write essays, create artwork and music, and more — continues to attract outside investor attention. According to one source, generative AI startups raised \$1.7 billion in Q1 2023, with an additional \$10.68 billion worth of deals [announced](#) in the quarter but not yet completed.

THE WALL STREET JOURNAL.

English Edition | Print Edition | Video | Audio | Latest Headlines | More

Home World U.S. Politics Economy Business Tech Markets Opinion Books & Arts Real Estate Life & Work Style Sports

◆ WSJ NEWS EXCLUSIVE | TECH

Apple Restricts Employee Use of ChatGPT, Joining Other Companies Wary of Leaks

The iPhone maker is concerned workers could release confidential data as it develops its own similar technology

By [Aaron Tilley](#) [Follow](#) and [Miles Kruppa](#) [Follow](#)

Updated May 18, 2023 7:35 pm ET

Share

Resize

Listen (2 min)



"Development of superhuman machine intelligence"
Sam Altman, CEO of ChatGPT creator OpenAI, touted the benefits of AI and acknowledged potential downsides of the technology during a Senate subcommittee hearing. Photo: Patrick Semansky/Associated Press



AI News

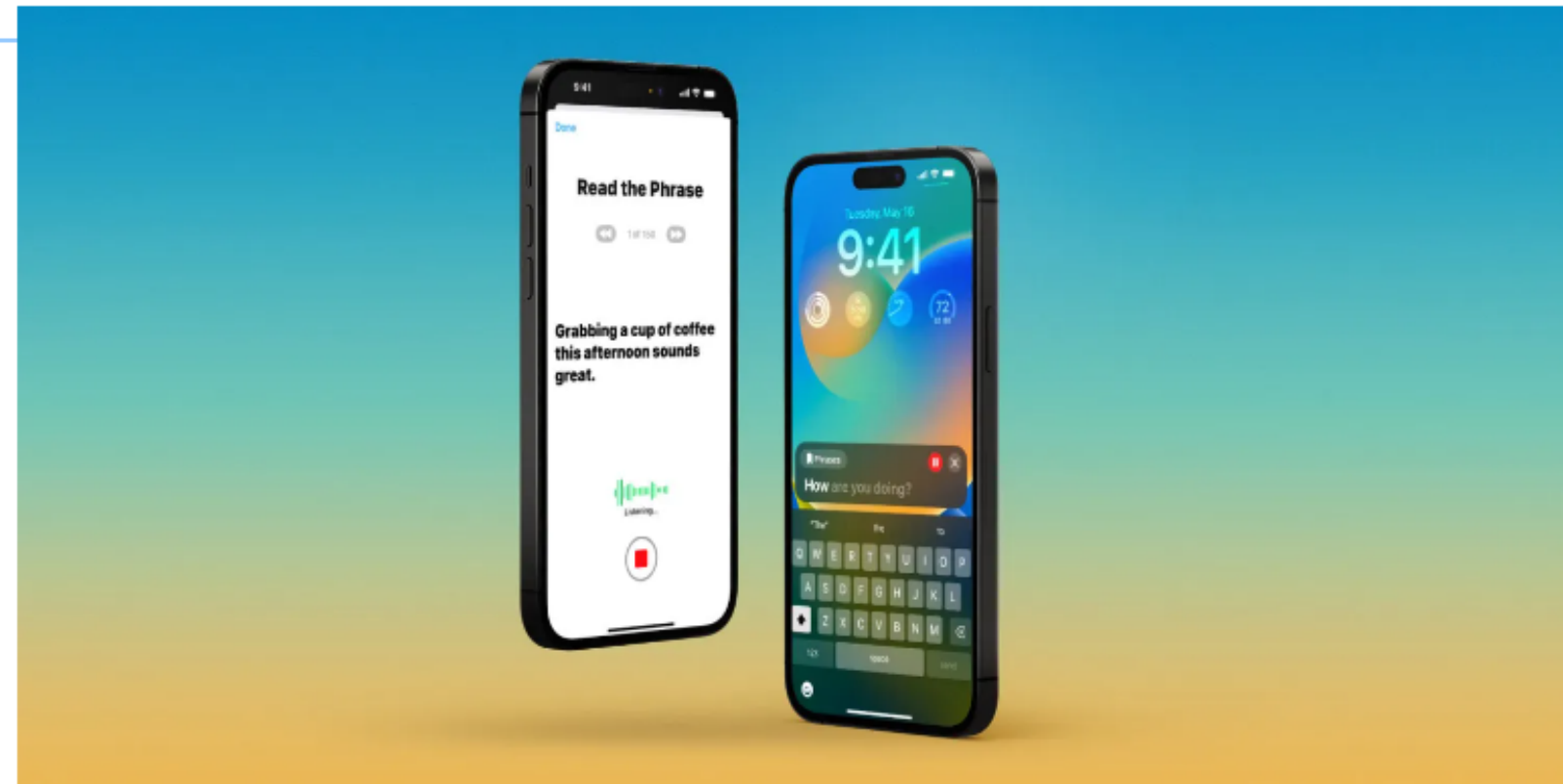
9T05Mac ▾

Exclusives Store iPhone ▾ Mac ▾ Watch ▾ iPad ▾ Music ▾ TV ▾ Guides ▾

ACCESSIBILITY IOS 17

Apple's new 'Personal Voice' feature can create a voice that sounds like you or a loved one in just 15 minutes

 Chance Miller | May 16 2023 - 6:07 am PT |  39 Comments



As part of its [preview of iOS 17 accessibility updates](#) coming this year, Apple has announced a pair of new features called Live Speech and Personal Voice. Live Speech allows users to type what they want to say and have it be spoken out.

Personal Voice, on the other hand, is a way for people who are at risk of losing their ability to speak to create and save a voice that sounds like them. Apple says it's designed for people at risk of losing their ability to speak, such as those with a recent diagnosis of ALS.

Here's how Apple describes the new Live Speech feature coming later this year:



News ▾ Ukraine war Features Economy Opinion Video

News | Technology

Five key takeaways from OpenAI's CEO Sam Altman's Senate hearing

ChatGPT creator and OpenAI CEO Sam Altman urges lawmakers to regulate artificial intelligence during a hearing.



OpenAI CEO Sam Altman testifies before a Senate Judiciary Privacy, Technology and the Law Subcommittee hearing on Capitol Hill in Washington, DC, on May 16. [Elizabeth Frantz/Reuters]

17 May 2023

Sam Altman, the chief executive of ChatGPT's OpenAI, testified before members of a Senate subcommittee on Tuesday about the need to regulate the increasingly powerful artificial intelligence technology being created inside his company and others like Google and Microsoft.

AI

Gary Marcus is happy to help regulate AI for US government: 'I'm interested'

Connie Loizos @cookie / 6:35 AM GMT+1 • May 19, 2023

 Comment



 Image Credits: Jack Gruber / USA TODAY

On Tuesday of this week, neuroscientist, founder and author Gary Marcus, sat between OpenAI CEO Sam Altman and Christina Montgomery, who is IBM's chief privacy trust officer, as all three testified before the Senate Judiciary Committee for

Governance of superintelligence

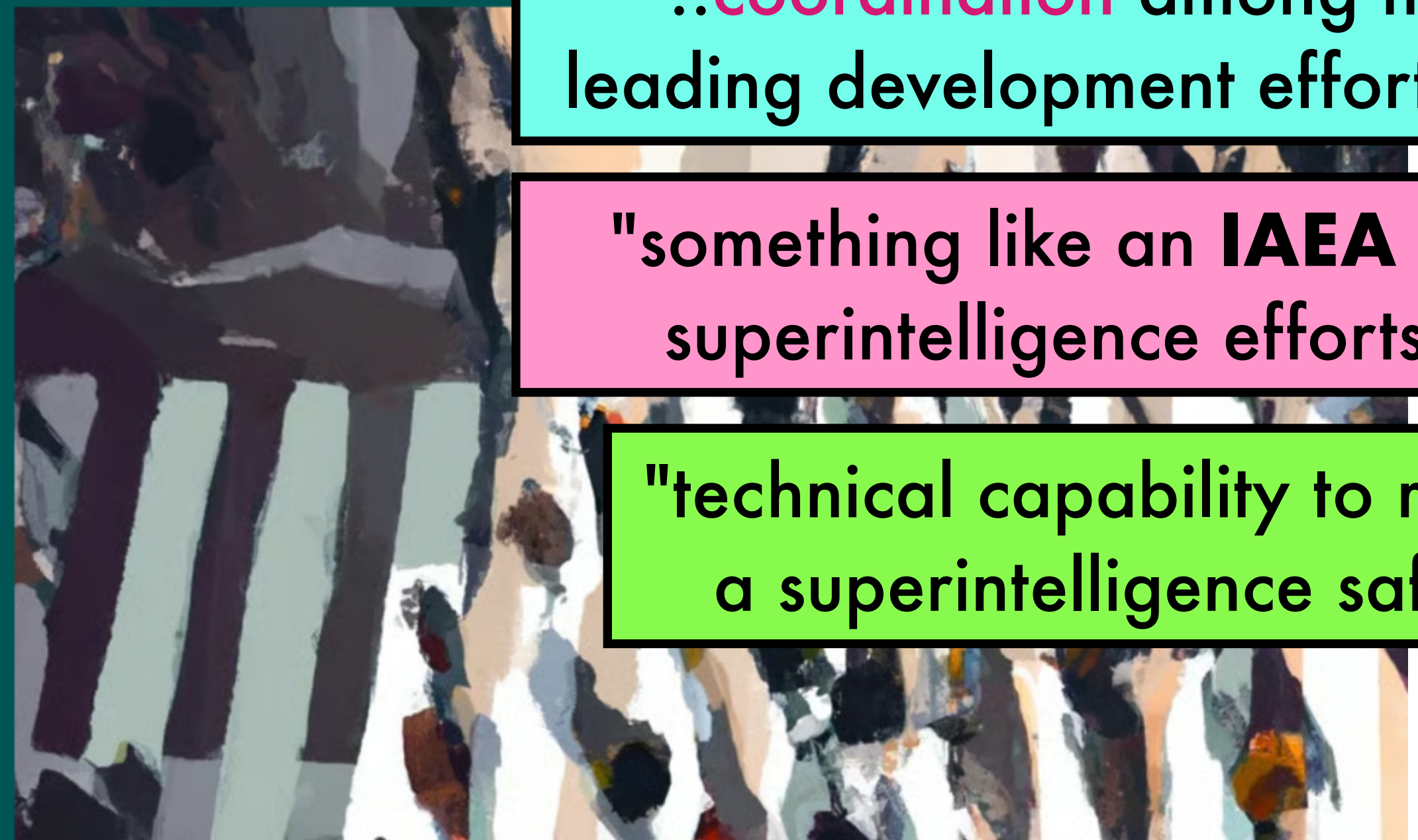
Now is a good time to start thinking about the governance of superintelligence—future AI systems dramatically more capable than even AGI.

"Given the possibility of existential risk, we can't just be **reactive**."

"..**coordination** among the leading development efforts..."

"something like an **IAEA** for superintelligence efforts"

"technical capability to make a superintelligence safe"



"A potentially **rogue AI** is an autonomous AI system that could behave in ways that would be catastrophically harmful to a large fraction of humans..."

"We should definitely avoid designing **survival instincts** into AI systems."

Yoshua Bengio

Home Profile Research Publications Students Media Presentations News English



How Rogue AIs may Arise

Published **22 May 2023** by [yoshuabengio](#)

The rise of powerful AI dialogue systems in recent months has precipitated debates about AI risks of all kinds, which hopefully will yield an acceleration of governance and regulatory frameworks. Although there is a general consensus around the need to regulate AI to protect the public from harm due to discrimination and biases as well as disinformation, there are profound disagreements among AI scientists regarding the potential for dangerous loss of control of powerful AI systems, also known as **existential risk from AI**, that may arise when an AI system can autonomously act in the world (without humans in the loop to check that these actions are acceptable) in ways that could potentially be catastrophically harmful. Some view these risks as a distraction for the more concrete risks and harms that are already occurring or are on the horizon. Indeed, there is a lot of uncertainty and lack of clarity as to how such catastrophes could happen. In this blog post we start a set of formal definitions, hypotheses and resulting claims about AI systems which could harm humanity and then discuss the possible conditions under which such catastrophes could arise,



Recognized worldwide as one of the leading experts in artificial intelligence, **Yoshua Bengio** is most known for his pioneering work in deep learning, earning him the Turing Award, "the Nobel Prize of Computing," with Geoffrey Hinton and Yann LeCun.

He is a Full Professor at Université de Montréal, and the Founder and Scientific Director of Mila – Quebec AI Institute. He co-directs the CIFAR Learning in Machines & Brains program as Senior Fellow and acts as Scientific Director of IVADO.

In 2019, he was awarded the prestigious Killam Prize and in 2022, became the computer scientist with the highest h-index in the world. He is a Fellow of both the Royal Society of London and Canada, Knight of the Legion of Honor of France and Officer of the Order of Canada.

Concerned about the social impact of AI and the objective that

Get 12 weeks for ~~\$29.99~~ \$6

THE
NEW YORKER

Newsletter Sign In 

[MORE ON A.I.](#) The potential and perils of a rapidly evolving technology. [Read More »](#)

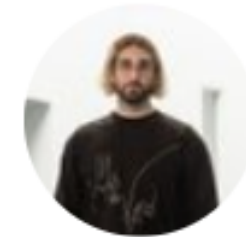
ANNALS OF ARTIFICIAL INTELLIGENCE

CAN WE STOP RUNAWAY A.I.?

Technologists warn about the dangers of the so-called singularity. But can anything actually be done to prevent it?

By **Matthew Hutson**

May 16, 2023



Aidan Gomez  

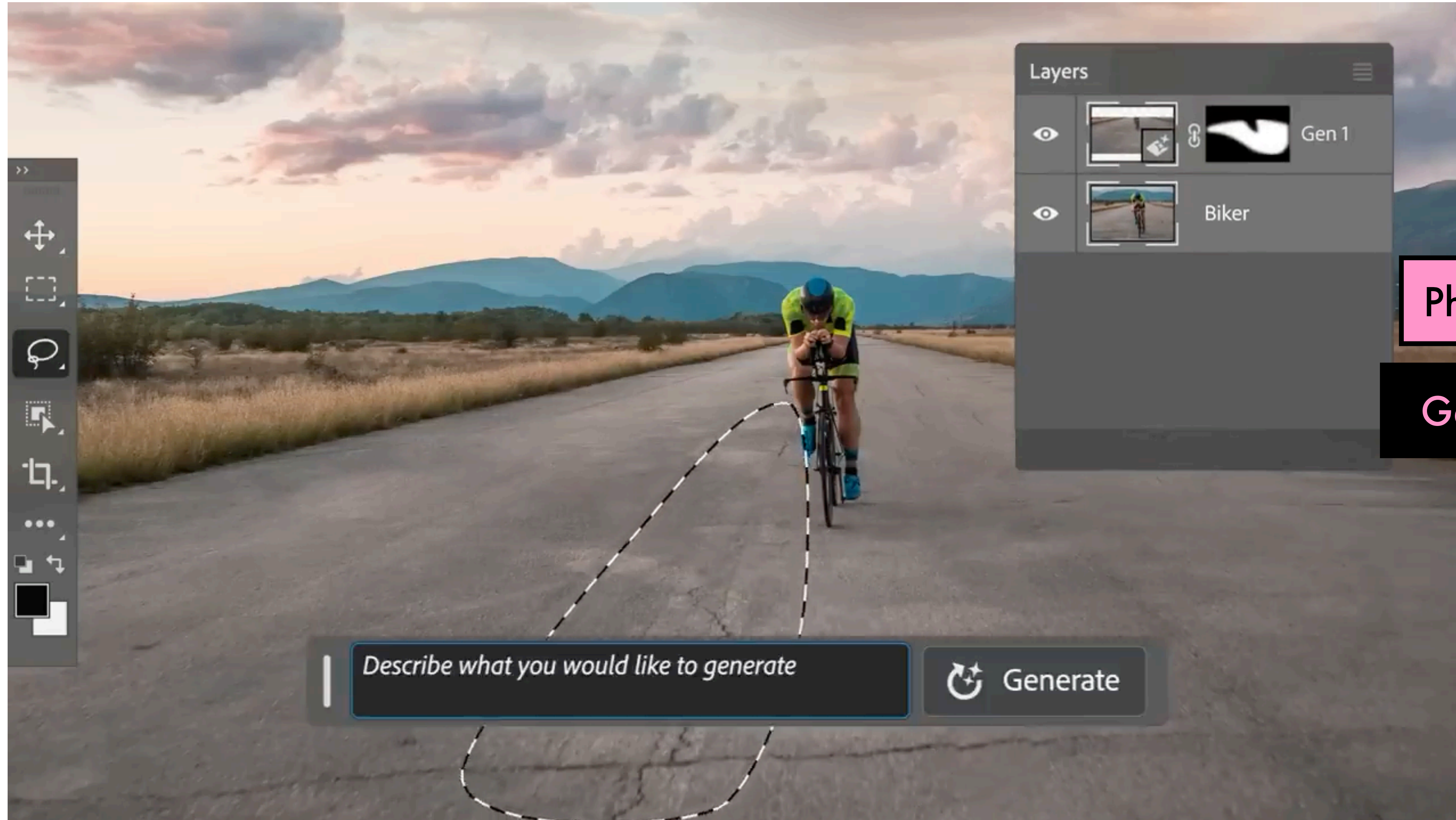
@aidangomezzz

it's pretty extraordinary that we have a literal doomsday cult steering the narrative on tech regulation. Not a figurative doomsday cult, a literal "omnipotent entity is coming, it will annihilate us all" doomsday cult.

3:56 PM · May 17, 2023 · **141.3K** Views

136 Retweets **22** Quotes **949** Likes **59** Bookmarks

Adobe Generative Fill



Photoshop & Firefly

Generative fill tool

Pandas AI

Pandas AI

Adds generative AI capabilities to Pandas library

PandasAI 🐼

Release **v0.2.14**  **ci passing**  **cd failing**  PandasAI **downloads/month 23k** License **MIT**  **Open in Colab**

Pandas AI is a Python library that adds generative artificial intelligence capabilities to Pandas, the popular data analysis and manipulation tool. It is designed to be used in conjunction with Pandas, and is not a replacement for it.



VoxCeleb Challenge

ABOUT TIMELINE TRACKS DATA METRICS REGISTRATION VOXCELEB



VoxSRC 23

Challenge for speaker recognition "in the wild"

Welcome to the 2023 VoxCeleb Speaker Recognition Challenge! The goal of this challenge is to probe how well current methods can recognize speakers from speech obtained 'in the wild'. The data is obtained from YouTube videos of celebrity interviews, as well as news shows, talk shows, and debates - consisting of audio from both professionally edited videos as well as more casual conversational audio in which background noise, laughter, and other artefacts are observed in a range of recording environments.

The workshop will be held in conjunction with Interspeech 2023.

Timeline

May 20th Development set for verification tracks released.

May 31st Development set for diarisation tracks released.

June 1st Test set released and evaluation server open.

Machine Learning Street Talk



AI Podcast

In-depth technical
discussions with researchers,
engineers & scientists

No Priors Podcast



Interview podcast

Discussions with AI
engineers, researcher &
founders

[NO-PRIORS.COM](https://no-priors.com)

Google Codey

DEVELOPERS

AI-powered coding, free of charge with Colab

17th May 2023

May 17, 2023
2 min read

Google Colab will soon introduce AI coding features using Google's most advanced family of code models, Codey.



Chris Perry
Group Product Manager,
Colab

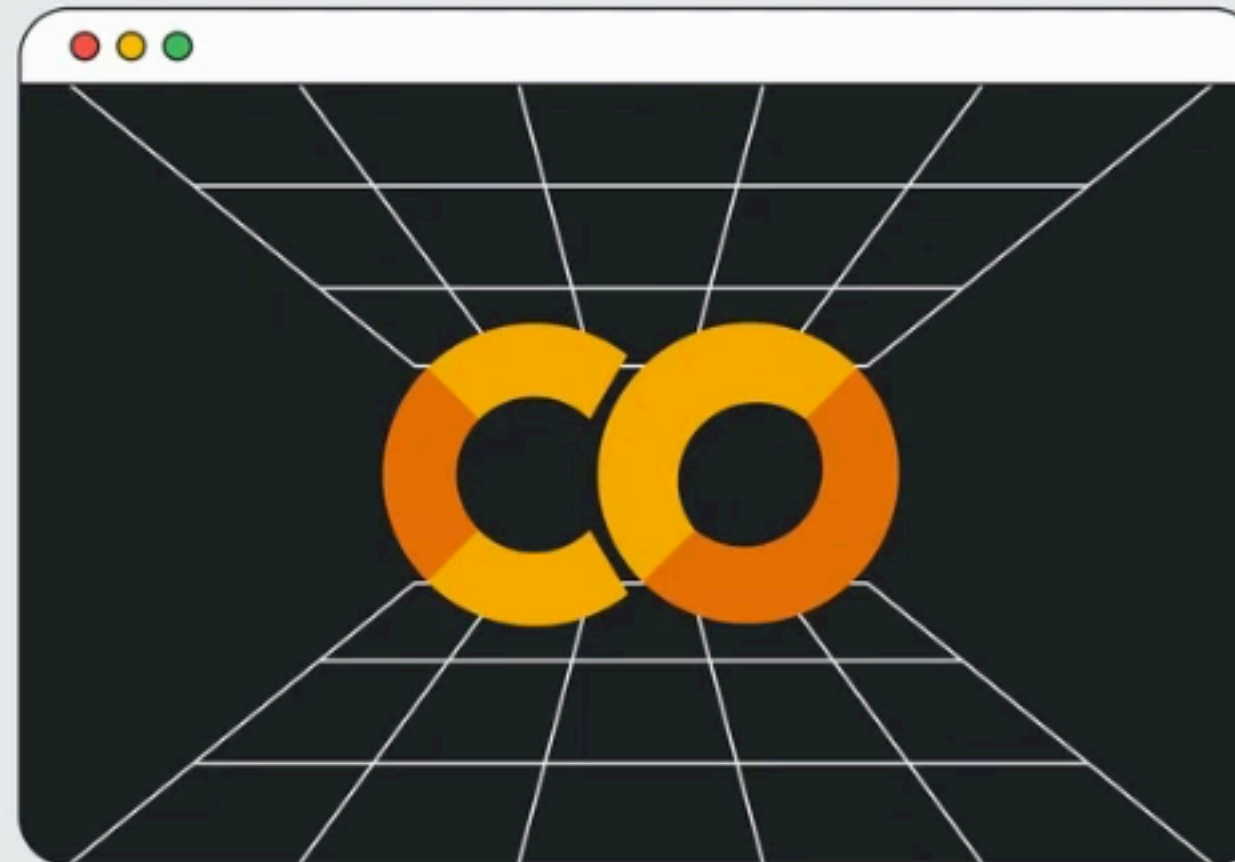


Shrestha Basu Mallick
Senior Product
Manager, Google Labs

Share

AI-powered Colab

Code generation from text,
autocompletion, chat



Since 2017, Google Colab has been the easiest way to start programming in Python. Over 7 million people, including students, already use Colab to access these powerful computing resources, free of charge, without having to install or manage any software. It's a great tool for machine learning, data analysis, and education — and now it's getting even better with advances in AI.

Today, we're announcing that Colab will soon add AI coding features like code completions, natural language to code generation and even a code-assisting chatbot. Colab will use Codey, a family of code

22nd May 2023

Why are so many giants of AI getting GPTs so badly wrong?

Introduction



Fergal Reid · Follow

9 min read · 1 day ago

Some big names in AI (e.g. Yann LeCun, Rodney Brooks, Noam Chomsky) are **seriously underestimating** the capabilities of large language models.



31



"GPT-4 clearly seems to have a world model"

Some big names in AI (e.g. Yann LeCun, Rodney Brooks, Noam Chomsky) are seriously underestimating the capabilities of large language models.

As evidence, I present specific examples I've generated using GPT-4, increasing in sophistication.

The examples take time to read, and its OK to skim them; but I think specific examples are the best way to counter some very sweeping claims.

The underestimating I'm seeing doesn't reassure me about the AI Safety

"The underestimating I'm seeing doesn't reassure me about the AI Safety debate at all"

"...if so many experts are getting it so wrong today, this isn't encouraging."

12th May 2023

RSS DS+AI Section Newsletter

The Alan Turing Institute has failed to develop modern AI in the UK

MARTIN GOODSON

12 MAY 2023

♥ 23

💬 11

🔄 1

📌 s

"at best irrelevant to the development of modern AI in the UK"

"the Turing has been completely blindsided by recent breakthroughs"

The UK's flagship institute for artificial intelligence, the Alan Turing Institute, has been at best irrelevant to the development of modern AI in the UK. Along with the AI council, which advises the government on AI, the Turing has been completely blindsided by recent breakthroughs in artificial intelligence based on large language models (LLMs).

The institute's Annual reports for the last four years do not refer to LLMs at all. There is no record of its website or Director mentioning them until a few months ago. It's as if the most important development in the history of AI has completely passed it by.

"radical change is needed"

Opinion **EU defence**

European governments need to start taking defence innovation seriously

Technologies that will protect democracy should not be outsourced

NATHAN BENAICH

+ Add to myFT



Drone footage released by the Ukrainian Army shows an attack on a Russian tank column in Kyiv in March last year © Armed Forces of Ukraine/Reuters

24th May 2023

"Europe's absent track record in defence innovation is striking"

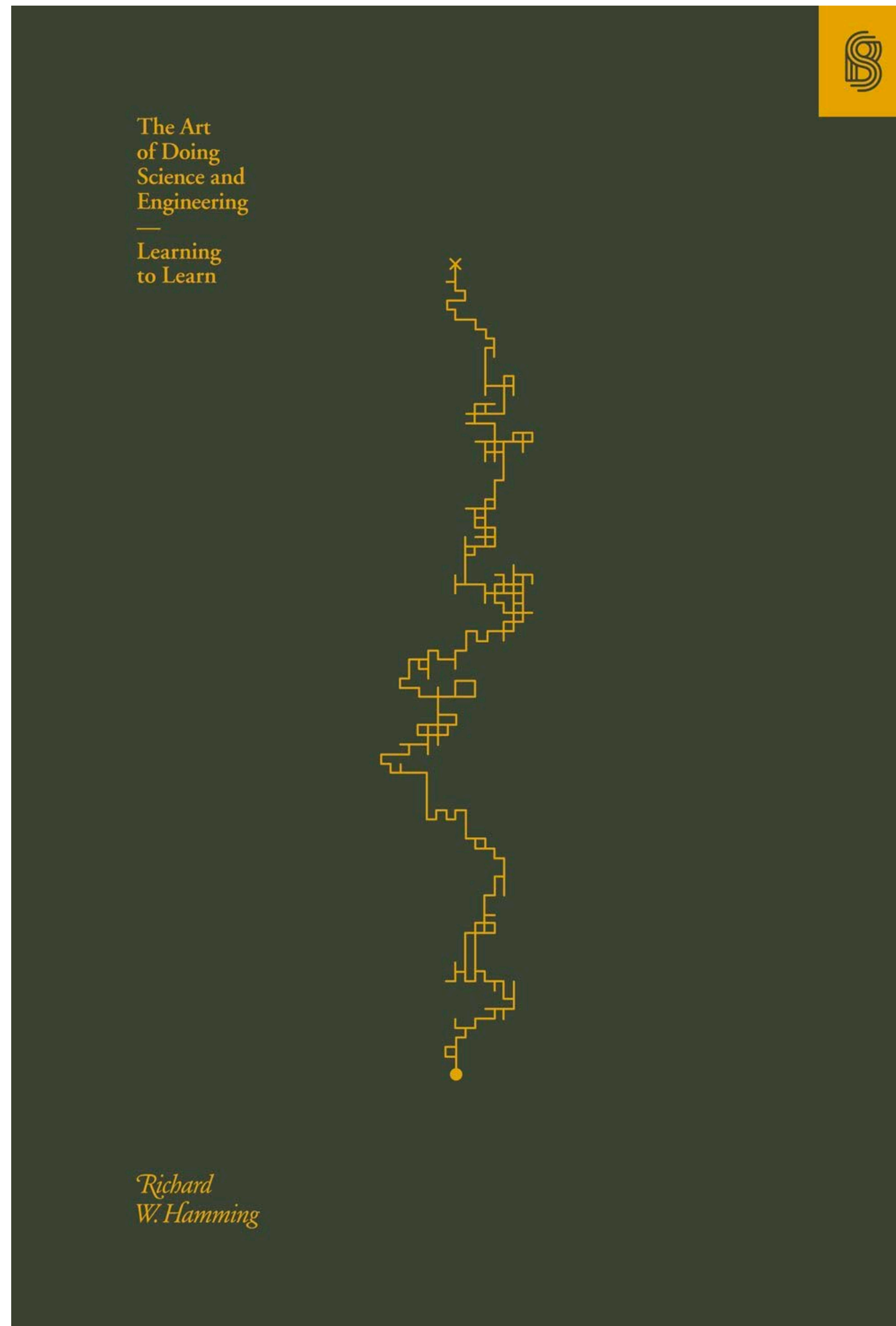
"Democracy won't defend itself with the next grocery-delivery app."

"lack of political courage"

"reluctance to make choices"

"a culture of penny pinching"

Samuel's Book Recommendation



Unsolicited book recommendation

"The Art of Doing Science and Engineering:
Learning to Learn"

Richard Hamming (1997)

What is it? Many insights gathered from a long career as a leading engineer and scientist

Filtir - fact-checking AI claims

