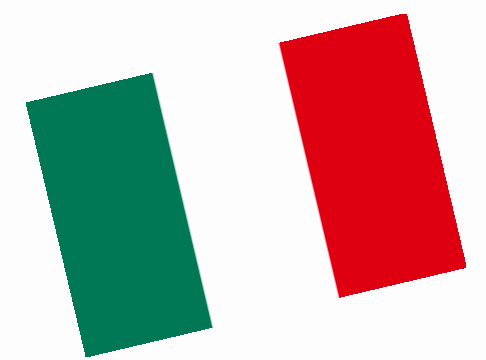
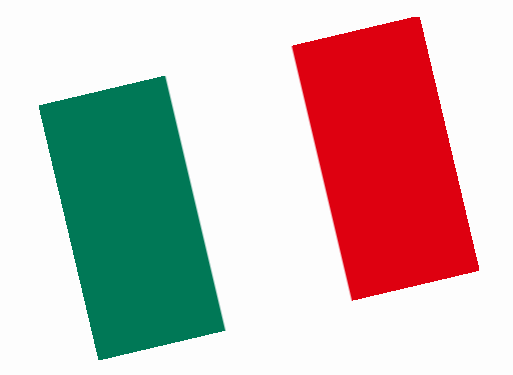


The Unintended Consequences of Censoring Digital Technology – Evidence from Italy's ChatGPT Ban*

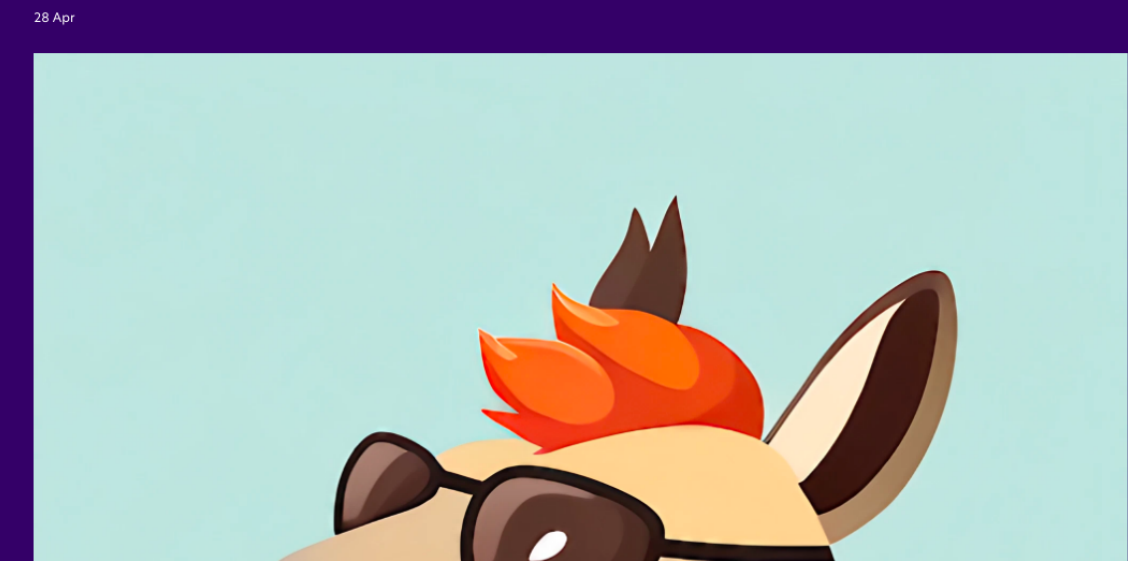


David Kreitmeir†
Paul A. Raschky‡

April 20, 2023



Stability AI releases StableVicuna, the AI World's First Open Source RLHF LLM Chatbot



'The Godfather of A.I.' Leaves Google and Warns of Danger Ahead

For half a century, Geoffrey Hinton nurtured the technology at the heart of chatbots like ChatGPT. Now he worries it will cause serious harm.



Abstract
We analyse the effect of the ban of ChatGPT, a generative pre-trained transformer chatbot, on individual productivity of GitHub users in Italy. We use data on the hourly coding output of over 8,000 professional developers to analyse the impact of the ban on individual productivity. Our results show a significant decrease in productivity following the announcement of the ban.

DIGICHINA ABOUT LATEST WORK EVENTS

Home » Work » **Translation: Measures for th**

Research

JAMA Internal Medicine | [Original Investigation](#)

Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum

John W. Ayers, PhD, MA; Adam Poliak, PhD; Mark Dredze, PhD; Eric C. Leas, PhD, MPH; Zechariah Zhu, BS; Jessica B. Kelley, MSN; Dennis J. Faix, MD; Aaron M. Goodman, MD; Christopher A. Longhurst, MD, MS; Michael Hogarth, MD; Davey M. Smith, MD, MAS

IMPORTANCE The rapid expansion of virtual health care has caused a surge in patient messages concomitant with more work and burnout among health care professionals. Artificial intelligence (AI) assistants could potentially aid in creating answers to patient questions by drafting responses that could be reviewed by clinicians.

OBJECTIVE To evaluate the ability of an AI chatbot assistant (ChatGPT), released in November 2022, to provide quality and empathetic responses to patient questions.

DESIGN, SETTING, AND PARTICIPANTS In this cross-sectional study, a public and nonidentifiable database of questions from a public social media forum (Reddit's r/AskDocs) was used to randomly draw 195 exchanges from October 2022 where a verified physician responded to a public question. Chatbot responses were generated by entering the original question into a fresh session (without prior questions having been asked in the session) on December 22 and 23, 2022. The original question along with anonymized and randomly ordered physician and chatbot responses were evaluated in triplicate by a team of licensed health care

- + Invited Commentary
- + Related article
- + Supplemental content

AI News

Apr 2023

ABSTRACT

Large language models are transforming research on machine learning while galvanizing public debates. Understanding not

Samuel
@Samuel
The latest con
A biological l
Available via
2:36 PM
46 Retv

ber Binz, Zeynep
Cybernetics

Banning ChatGPT

18th April 2023



The Unintended Consequences of Censoring Digital Technology – Evidence from Italy's ChatGPT Ban*

David Kreitmeir[†]
Paul A. Raschky[‡]

April 20, 2023

Abstract

We analyse the effects of the ban of ChatGPT, a generative pre-trained transformer chatbot, on individual productivity. We first compile data on the hourly coding output of over 8,000 professional GitHub users in Italy and other European countries to analyse the impact of the ban on individual productivity. Combining the high-frequency data with the sudden announcement of the ban in a difference-in-differences framework, we find that the output of Italian developers decreased by

Reduction in **GitHub releases** of $\approx 50\%$
in the 2 days after ban

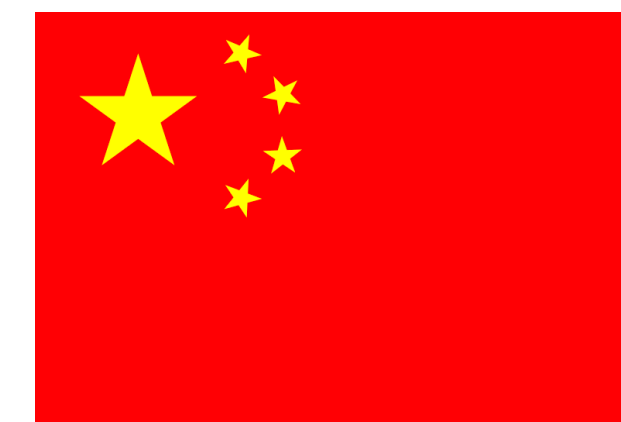
VPN searches increased by $\approx 50\%$

Tor bridge use increased by $\approx 9.4\%$

"While these measures may be well-intended, they are often **ineffective** and can lead to short-term disruptions in output."

China Release Draft For Comment

18th April 2023



DIGICHINA

ABOUT LATEST WORK EVENTS



Home » Work »

Translation: Measures for the Management of Generative Artificial Intelligence Services (Draft for Comment) – April 2023

Novel rules about training data and accuracy of generated media circulated for comment

Published
April 12, 2023

By:
Seaton Huang
Helen Toner
Zac Haluza
Rogier Creemers
Graham Webster



THIS IMAGE WAS GENERATED BY STABLE DIFFUSION IN RESPONSE TO A PROMPT: "Chinese officials announce a draft policy to govern the use of generative artificial intelligence algorithms, pixar-style 3d render"

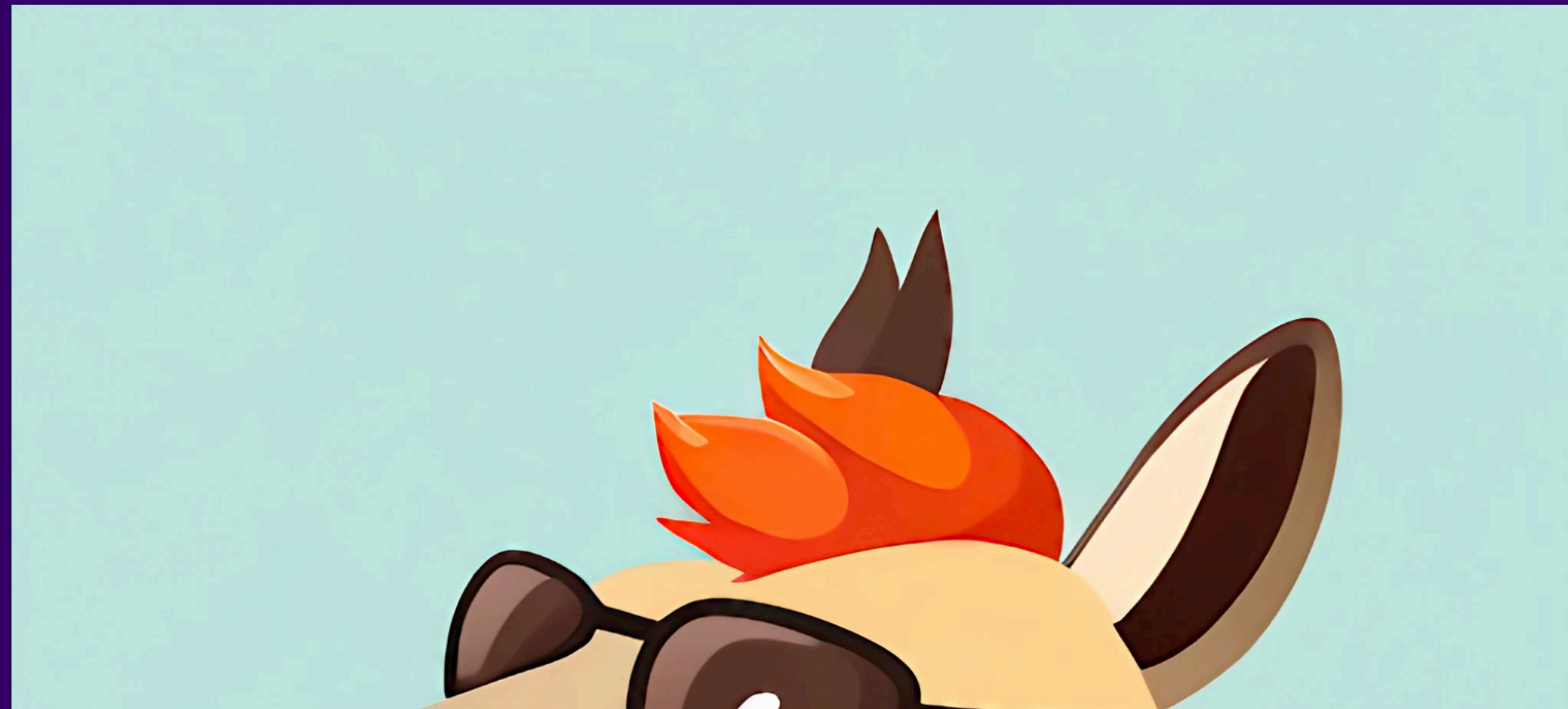
HELEN TONER

Director of Strategy and Foundational Research Grants, Center for Security and Emerging Technology, Georgetown University;

A second important new element is that in a single sentence, Article 5 appears to do away with a conundrum that has been puzzling European Union policy makers for months. It specifies that companies providing access to generative AI via “programmable interfaces”—aka APIs like those released by OpenAI and Google—are responsible for all content produced. (Disclosure: I am a member of the board of directors of OpenAI.) This stance is appealingly simple, but seems likely to run into practical hurdles. While the original AI developer can and should be responsible for some types of problems, this approach would hold them liable for everything, including issues arising from choices the downstream client company makes about app design or how to restrict user behavior. It will be interesting to see if this provision remains intact throughout the comment and review process.

Stability AI releases StableVicuna, the AI World's First Open Source RLHF LLM Chatbot

28 Apr



2

StableVicuna by CarperAI

[CarperAI/stable-vicuna-13b-delta](#)

 Duplicate Space


Duplicate the Space to skip the queue and run in a private space

Chatbot

Is "AI news" by Samuel Albanie an absolutely stupendous source of up to date information about the AI ecosystem?

The reliability of information found on Samuel Albanie's "AI news" can vary depending on the quality of the sources he utilizes. It is always important to keep an open mind and conduct additional research to ensure the information provided is accurate.

Send a message

 Send

Stop

Clear History

StableVicuna by CarperAI

CarperAI/stable-vicuna-13b-delta

Duplicate Space

Duplicate the Space to skip the queue and run in a private space

Chatbot

Consider the identity:
 $9 * 4 + 6 * 6 = 72$
Can you modify exactly one integer (and not more than that!) on the left hand side of the equation so the right hand side becomes 99?
- Let's think step by step, write down a plan, and then write down your solution as: "The solution is: A * B + C * D".

Send

Stop

Clear History

Stability AI releases DeepFloyd IF, a powerful text-to-image model that can smartly integrate text into images

28 Apr

Today Stability AI, together with its multimodal AI research lab DeepFloyd, announced the research release of DeepFloyd IF, a powerful text-to-image cascaded pixel diffusion model.

DeepFloyd IF is a state-of-the-art text-to-image model released on a non-commercial, research-permissible license that provides an opportunity for research labs to examine and experiment with advanced text-to-image generation approaches. In line with other Stability AI models, Stability AI intends to release a DeepFloyd IF model fully open source at a future date.

Description and Features

- **Deep text prompt understanding:**

The generation pipeline utilizes the large language model [T5-XXL-1.1](#) as a text encoder. A significant amount of text-image cross-attention layers also provides better prompt and image alliance.

- **Application of text description into images:**

Incorporating the intelligence of the T5 model, DeepFloyd IF generates coherent and clear text alongside objects of different properties appearing in various spatial relations. Until now, these use cases have been challenging for most text-to-image models.

- **A high degree of photorealism:**

This property is reflected by the impressive zero-shot FID score of 6.66 on the [COCO dataset](#) (FID is a main metric used to evaluate the performance of text-to-image models; the lower the score, the better).

- **Aspect ratio shift:**

The ability to generate images with a non-standard aspect ratio, vertical or horizontal, as well as the standard square aspect.

- **Zero-shot image-to-image translations:**

Image modification is conducted by (1) resizing the original image to 64 pixels, (2) adding noise through forward diffusion, and (3) using backward diffusion with a new

DeepFloyd IF

The DeepFloyd IF model has been initially released as a non-commercial research-only model. Please make sure you read and abide to the [LICENSE](#) before using it.

🔄 Duplicate Space to skip the queue

A sign by the river in Grantchester Meadows says " Carpe diem - swim here!"

Generate

Enter a negative prompt



Pick your favorite generation to upscale.

Upscale

Advanced options

☰ Examples

high quality dslr photo, a photo product of a lemon inspired by natural and organic materials, wooden accents, intricately decorated with glowing vines of led lights, inspired by baroque luxury

paper quilling, extremely detailed, paper quilling of a nordic mountain landscape, 8k rendering

letters made of candy on a plate that says "diet"

a photo of a violet baseball cap with yellow text: "deep floyd". 50mm lens, photo realism, cine lens. violet baseball cap says "deep floyd". reflections, render. yellow stitch text "deep floyd"

Studying GPT-3.5 With Psychiatry

21st April 2023

Inducing anxiety in large language models in exploration and bias

Julian Coda-Forno^{1,2,+}, Kristin Witte^{1,+}, Akshay K. Jagadish^{1,+}, Marcel Binz¹, Zeynep Akata², and Eric Schulz^{1,*}

¹MPRG Computational Principles of Intelligence, Max Planck Institute for Biological Cybernetics

²University of Tübingen

*eric.schulz@tue.mpg.de

+these authors contributed equally to this work

ABSTRACT

Large language models are transforming research on machine learning while galvanizing public debates not only when these models work well and succeed but also why they fail and misbehave is of great societal relevance. To turn the lens of computational psychiatry, a framework used to computationally describe and modify aberrant outputs produced by these models. We focus on the Generative Pre-Trained Transformer 3.5 and subject it to tasks commonly studied in psychiatry. Our results show that GPT-3.5 responds robustly to a common anxiety questionnaire, producing higher anxiety scores than human subjects. Moreover, GPT-3.5's responses can be predictably changed by using emotion-inducing prompts. Emotion-induction not only influences GPT-3.5's behavior in a cognitive task measuring exploration but also influences its behavior in a previously-established task measuring biases such as racism. GPT-3.5 shows a strong increase in biases when prompted with anxiety-inducing text. Thus, it is likely that communication to large language models has a strong influence on their behavior in applied settings. These results progress our understanding of prompt engineering and demonstrate the usefulness of methods taken from computational psychiatry for studying the capable algorithms to which we increasingly delegate authority and autonomy.

"GPT-3.5 responds robustly to a common anxiety questionnaire, producing higher anxiety scores than human subjects."

"GPT-3.5 shows a strong increase in biases when prompted for anxiety-inducing text."

"In some sense, psychotherapy is just a form of prompt-engineering for humans."

Can insights from psychiatry help steer AI?

Scaling Transformer to 1M tokens and beyond with RMT

19th April 2023

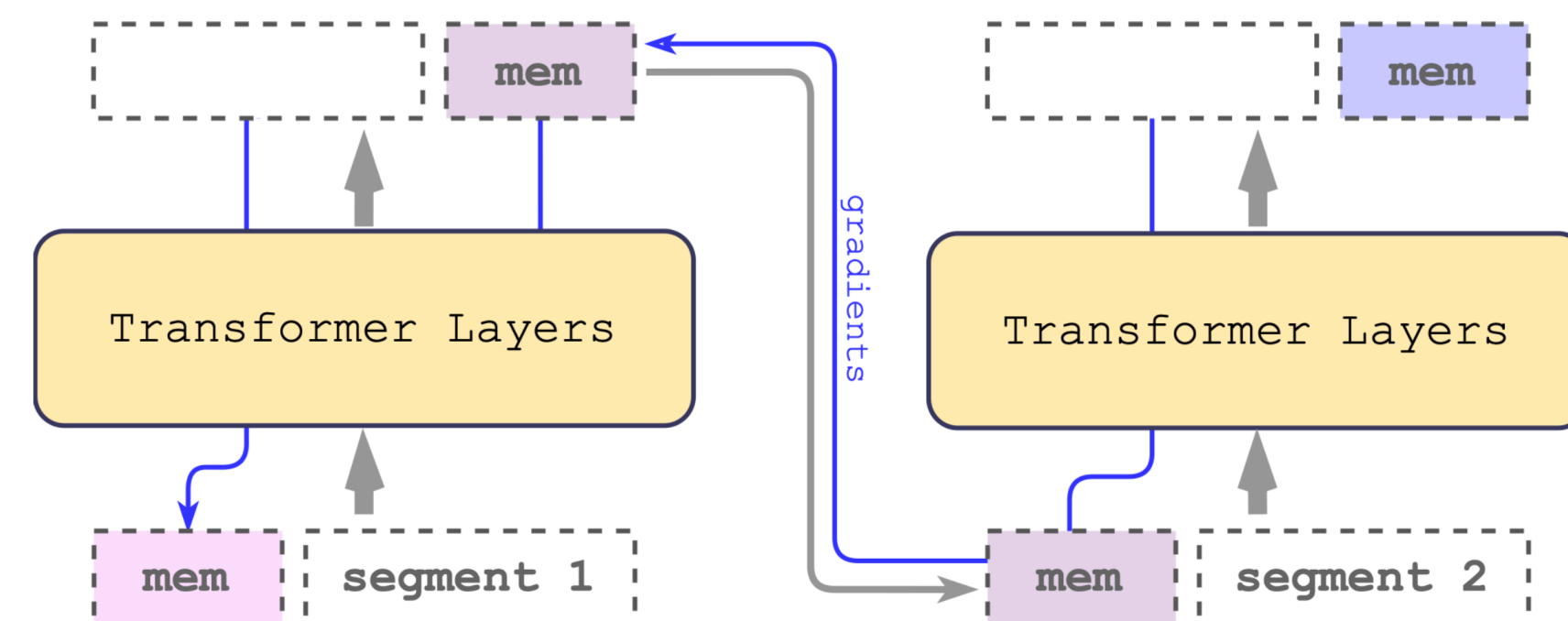
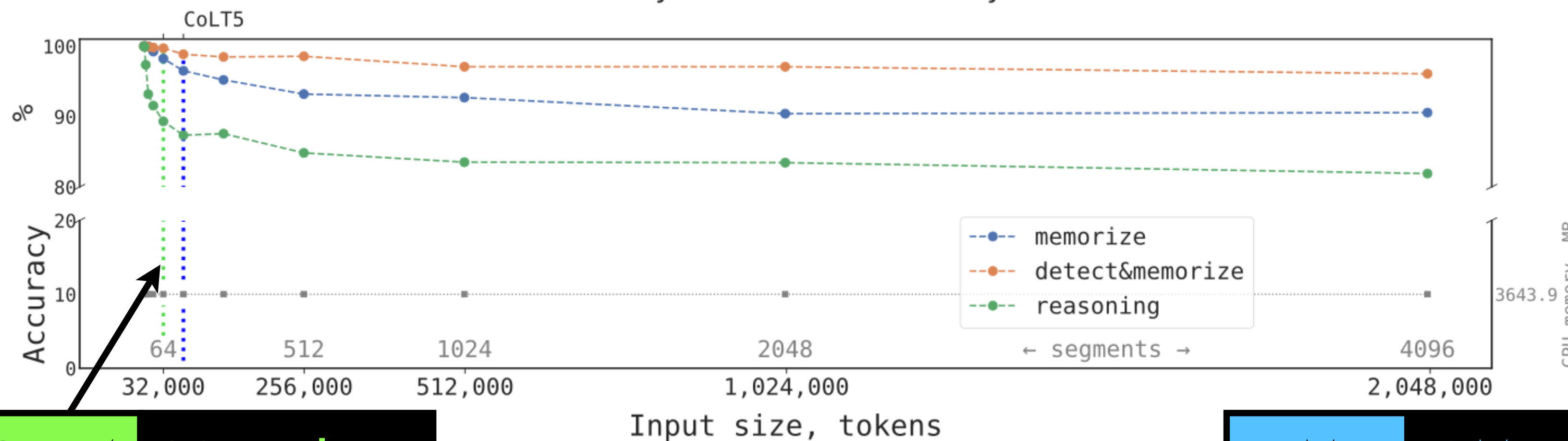
Aydar Bulatov¹
bulatov@deeppavlov.ai

Yuri Kuratov^{1,2}
kuratov@airi.net

Mikhail S. Burtsev^{1,3}
mbur@lims.ac.uk

¹DeepPavlov
²Artificial Intelligence Research Institute (AIRI)
³London Institute for Mathematical Sciences

Memory retrieval accuracy



Recurrent Memory Mechanism

GPT-4 32K tokens

RMT 2M tokens

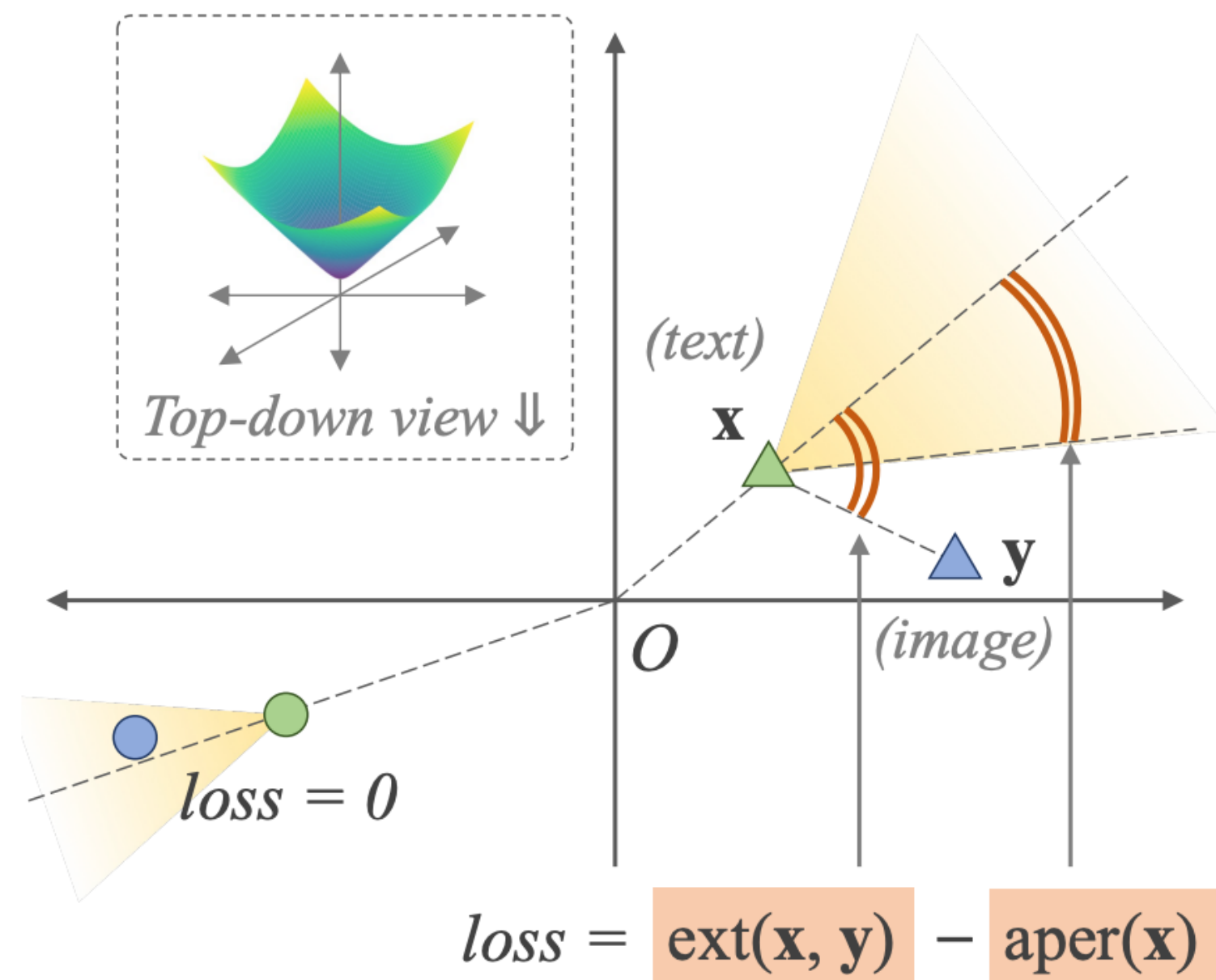
Hyperbolic Image-Text Representations

8th April 2023

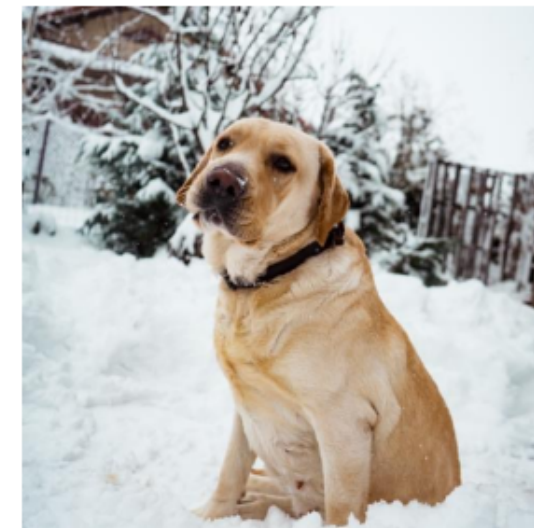
Karan Desai¹ Maximilian Nickel² Tanmay Rajpurohit³ Justin Johnson^{1,2} Ramakrishna Vedantam⁴

Abstract

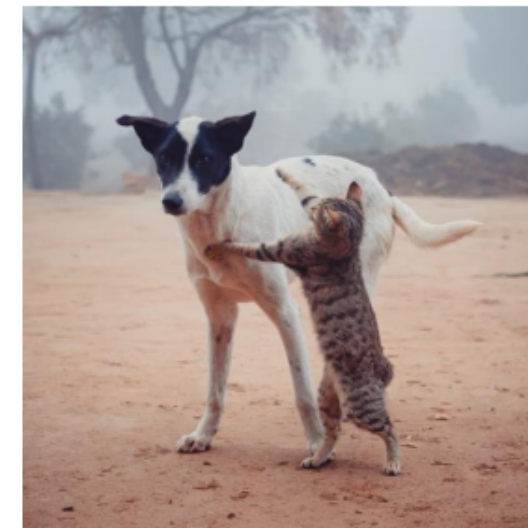
Visual and linguistic concepts naturally organize themselves in a hierarchy, where a textual concept “dog” entails all images that contain dogs.



Entailment loss (enforce partial ordering)



*pic of my labrador
in the snow*



*a cat and a dog
playing in the street*



*my cat is photogenic
look at those eyes!*

exhausted doggo

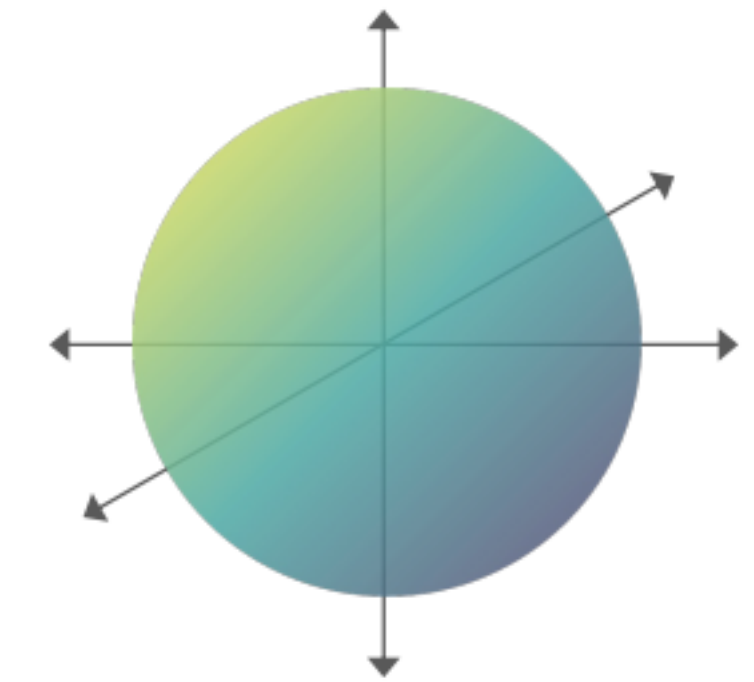
curious kitty

so cute <3

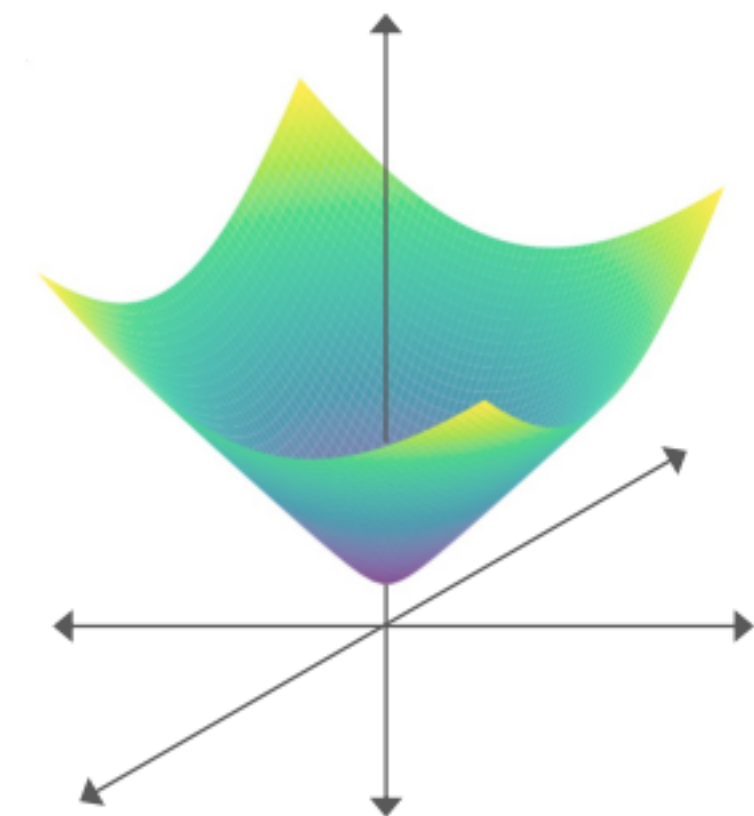


MERU	CLIP
avocado toast	avocado toast
healthy breakfast	delicious
delicious	↓
homemade	↓
fresh	↓
[ROOT]	[ROOT]

CLIP: embed images and text in a Euclidean space



MERU: embed images and text in a hyperbolic space



Chameleon: Plug-and-Play Compositional Reasoning with Large Language Models

19th April 2023

Pan Lu¹, Baolin Peng², Hao Cheng², Michel Galley²

Kai-Wei Chang¹, Ying Nian Wu¹, Song-Chun Zhu¹, Jianfeng Gao²

¹University of California, Los Angeles ²Microsoft Research, Redmond

lupantech@gmail.com {bapeng, chehao, mgalley, jfgao}@microsoft.com

<https://chameleon-llm.github.io>

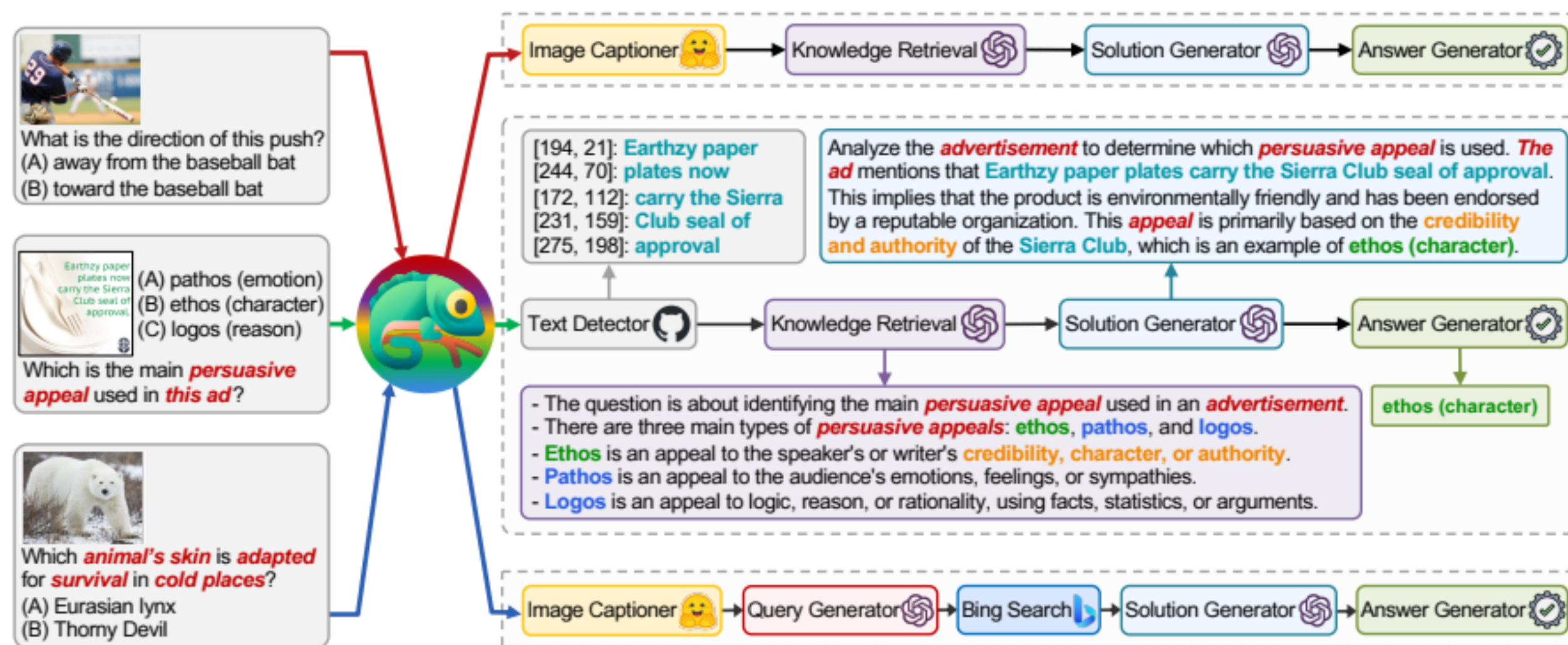


Figure 1: Examples from our Chameleon with GPT-4 on ScienceQA [27], a multi-modal question answering benchmark in scientific domains. Chameleon is adaptive to different queries by synthesizing programs to compose various tools and executing them sequentially to get final answers.

Abstract

Large language models (LLMs) have achieved remarkable progress in various natural language processing tasks with emergent abilities. However, they face inherent limitations and are not suitable for certain information utilization tasks.

"Chameleon builds on top of an LLM as a natural language planner"

"Chameleon is capable of synthesising programs to compose various tools to tackle a broad range of queries"

Strong results on ScienceQA & TabMWP

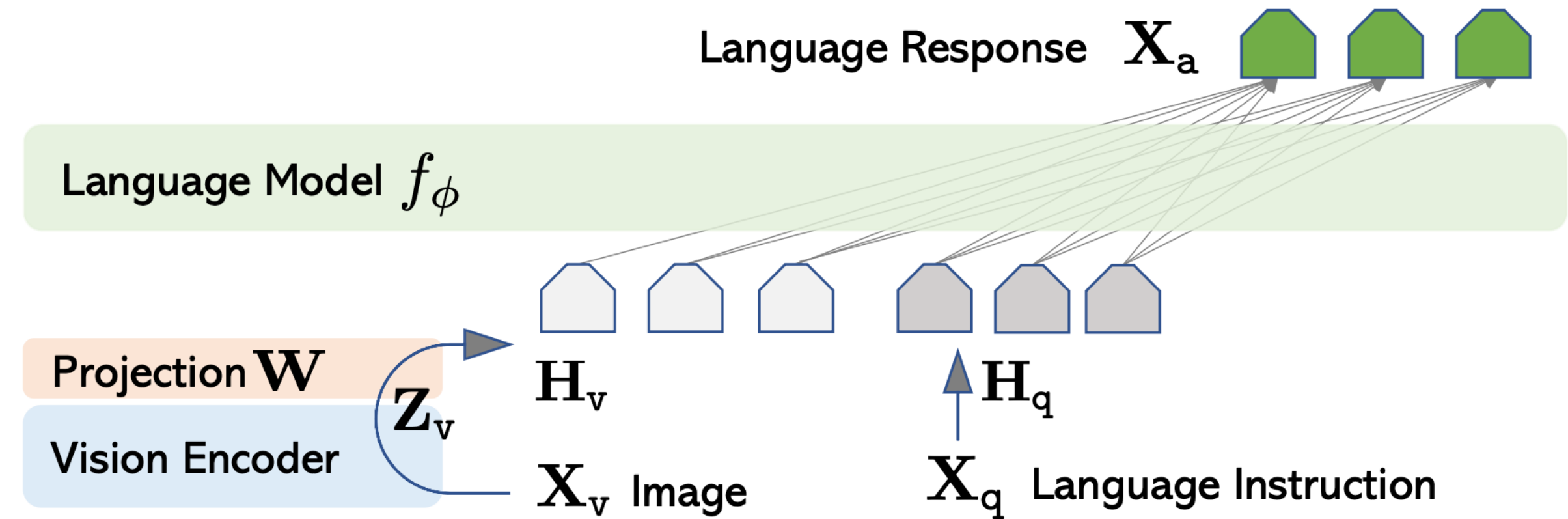
Visual Instruction Tuning

"use language-only GPT-4 to generate multimodal language-image instruction-following data"

Haotian Liu^{1*}, Chunyuan Li^{2*}, Qingyang Wu³, Yong Jae Lee¹
¹University of Wisconsin–Madison ²Microsoft Research ³Columbia University
<https://llava-vl.github.io>

Abstract

Instruction tuning large language models (LLMs) using machine-generated instruction-following data has improved zero-shot capabilities on new tasks, but the idea is less explored in the multimodal field. In this paper, we present the first attempt to use language-only GPT-4 to generate multimodal language-image instruction-following data. By instruction tuning on such generated data, we introduce LLaVA: Large Language and Vision Assistant, an end-to-end trained large multimodal model that connects a vision encoder and LLM for general-purpose visual and language understanding. Our early experiments show that LLaVA demonstrates impressive multimodal chat abilities, sometimes exhibiting the behaviors of multimodal GPT-4 on unseen images/instructions, and yields a 85.1% relative score compared with GPT-4 on a synthetic multimodal instruction-following dataset. When fine-tuned on Science QA, the synergy of LLaVA and GPT-4 achieves a new state-of-the-art accuracy of 92.53%. We make GPT-4 generated visual instruction tuning data, our model and code base publicly available.



Visual input example, Extreme Ironing:



Source: <https://www.barnorama.com/wp-content/uploads/2016/12/03-Confusing-Pictures.jpg>

User
LLaVA

What is unusual about this image?
 The unusual aspect of this image is a man ironing clothes on the back of a minivan or van. This is not a typical place to perform this activity, as one would usually iron clothes in a more stationary and safe location, such as a home, using a regular ironing board. The scene depicted in the image is peculiar as it involves a makeshift ironing setup on a vehicle, which can be both unsafe and unconventional. Additionally, it is not clear how the man is able to maintain balance and stability while ironing clothes in such an unstable environment.



LLaVA: Large Language and Vision Assistant



AIP Terminal

Investigations

Proposals

Share

File Edit

Saved

Battlefield Overview

MOCK DATA

AI processing on SAR imagery has detected potential unusual enemy activity.

AE (YOU)

Show me more details

AIP Assistant

Alert - Anomalous military activity detected

Hand-off Inspector

Here are additional details about this alert.

SAR Imagery

29 April 2023 15:35Z



Summary

This satellite image was captured 8 minutes ago. An algorithm that detects military equipment found 5 pieces of equipment in the image.

Details

Timestamp	29 April 2023 14:39Z
Satellite	StarTech N-58200
Imagery Type	SAR
Image Quality	1m
Algorithm	Military Vehicle Detector V 25.3.0

Start typing something to explore with AIP...



25th April 2023

Comparing ChatGPT vs Physicians

28th April 2023

Research

JAMA Internal Medicine | [Original Investigation](#)

Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum

John W. Ayers, PhD, MA; Adam Poliak, PhD; Mark Dredze, PhD; Eric C. Leas, PhD, MPH; Zechariah Zhu, BS; Jessica B. Kelley, MSN; Dennis J. Faix, MD; Aaron M. Goodman, MD; Christopher A. Longhurst, MD, MS; Michael Hogarth, MD; Davey M. Smith, MD, MAS

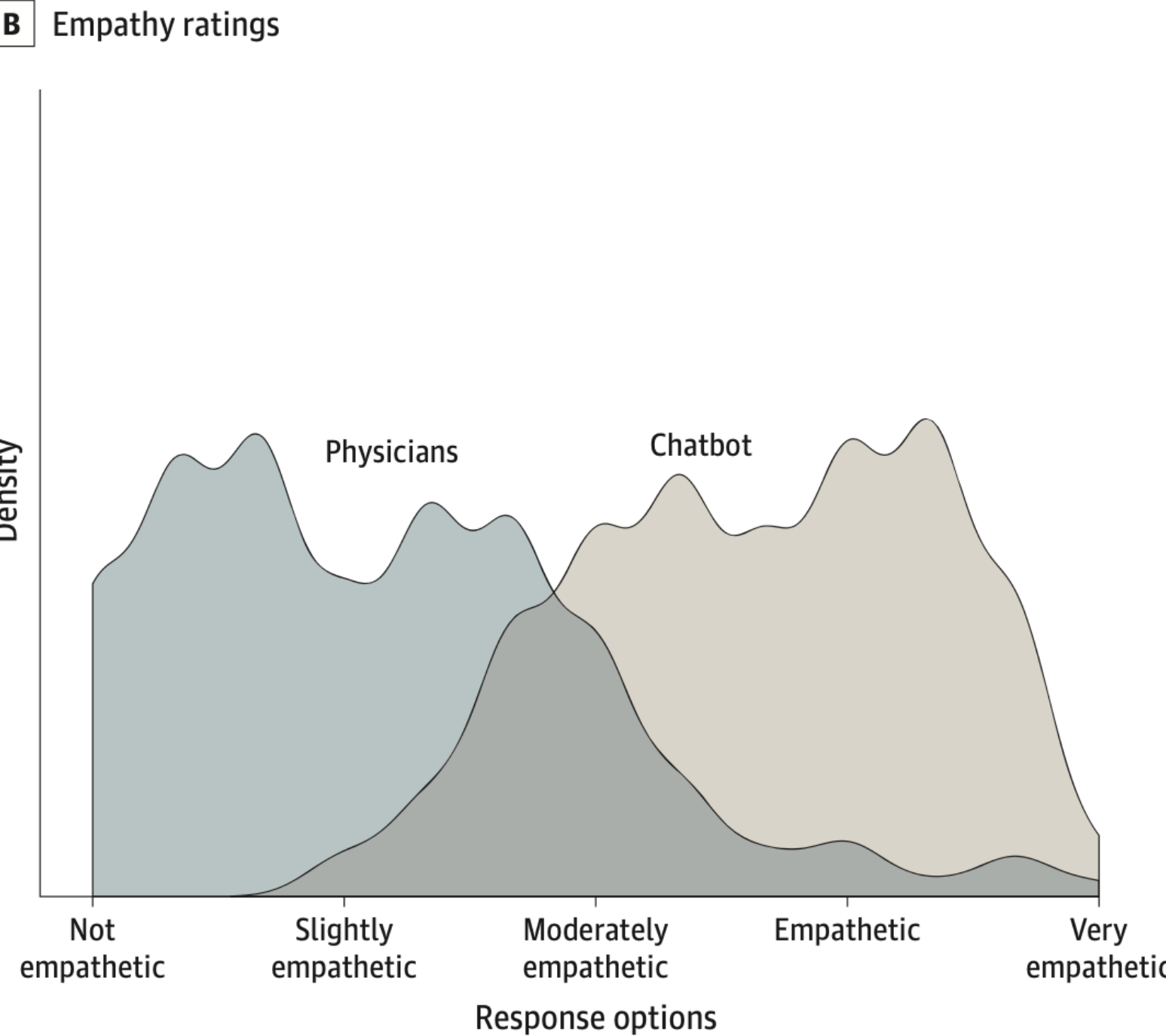
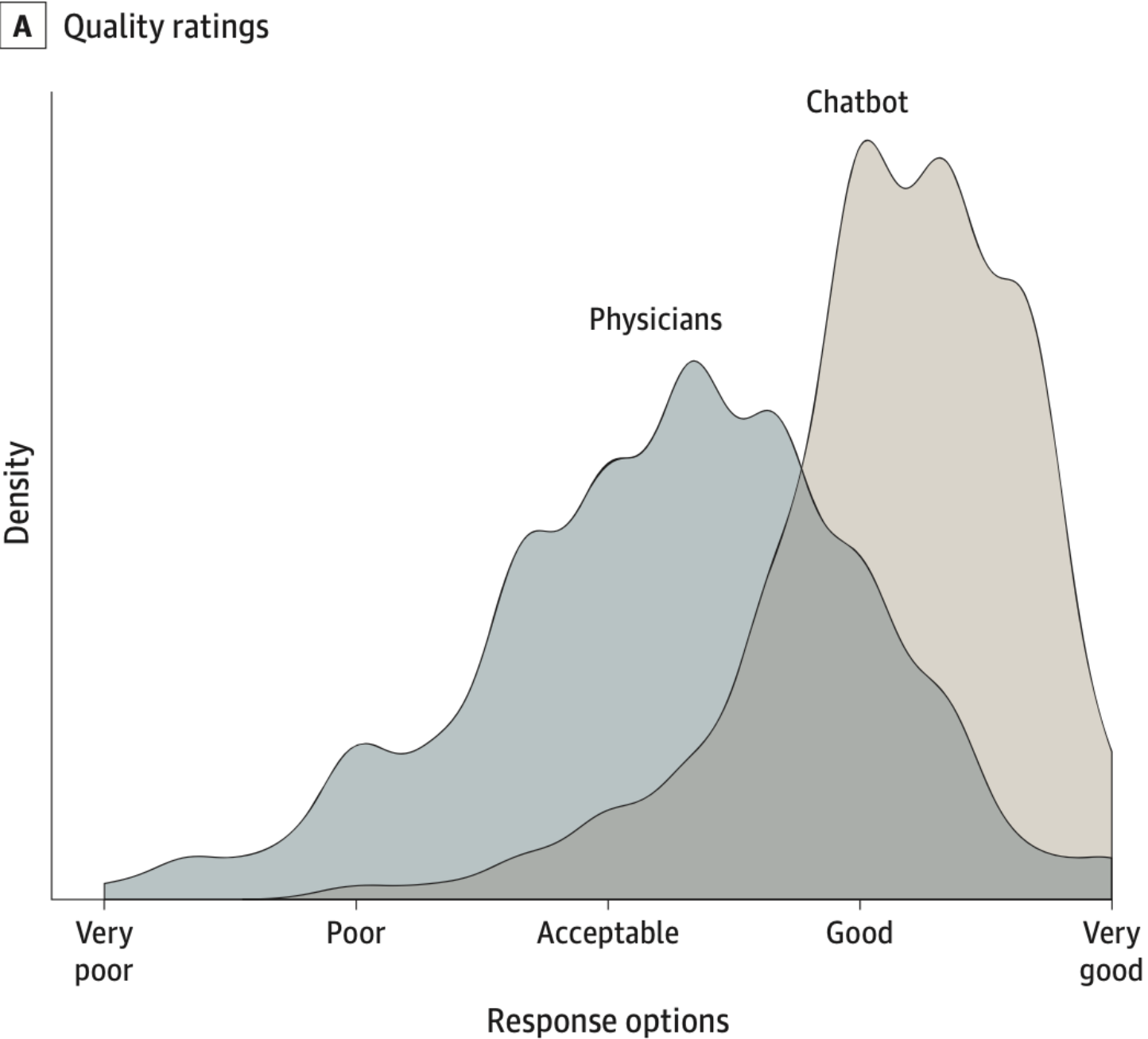
- [+ Invited Commentary](#)
- [+ Related article](#)
- [+ Supplemental content](#)

IMPORTANCE The rapid expansion of virtual health care has caused a surge in patient messages concomitant with more work and burnout among health care professionals. Artificial intelligence (AI) assistants could potentially aid in creating answers to patient questions by drafting responses that could be reviewed by clinicians.

OBJECTIVE To evaluate the ability of an AI chatbot assistant (ChatGPT), released in November 2022, to provide quality and empathetic responses to patient questions.

DESIGN, SETTING, AND PARTICIPANTS In this cross-sectional study, a public and nonidentifiable database of questions from a public social media forum (Reddit's r/AskDocs) was used to randomly draw 195 exchanges from October 2022 where a verified physician responded to a public question. Chatbot responses were generated by entering the original question into a fresh session (without prior history) on November 23, 2022. The original questions and chatbot responses were evaluated by two independent raters.

Limitation: online forums "may not reflect typical patient-physician questions"



News Round-Up

GigaChat BETA

Умеет отвечать на вопросы, вести диалог, написать код, нарисовать картинку по запросу. И всё на русском языке

HuggingChat v0

Making the community's best AI chat models available to everyone.

Current Model
OpenAssistant/oasst-sft-6-llama-30b

[Model page](#) [Dataset page](#) [Open Assistant Website](#)

EXCLUSIVE | MICROSOFT | STARTUPS | GOOGLE

Microsoft Readies AI Chip as Machine Learning Costs Surge

By Anissa Gardizy and Wayne Ma | April 18, 2023 6:00 AM PDT
Photo: Art by Mike Sullivan

PARESH DAVE | BUSINESS | APR 20, 2023 5:19 PM

Stack Overflow Will Charge AI Giants for Training Data

The programmer Q&A site joins Reddit in demanding compensation when its data is used to train algorithms and ChatGPT-style bots

 **Samuel Albanie** ✓
@SamuelAlbanie

The latest competitor to GPT-4:
A biological large language model
Available via a text completion API

 samuel-api.com
Samuel-API
A biological competitor to GPT-4.

2:36 PM · Apr 26, 2023 · 78.6K Views

46 Retweets 9 Quotes 453 Likes 115 Bookmarks



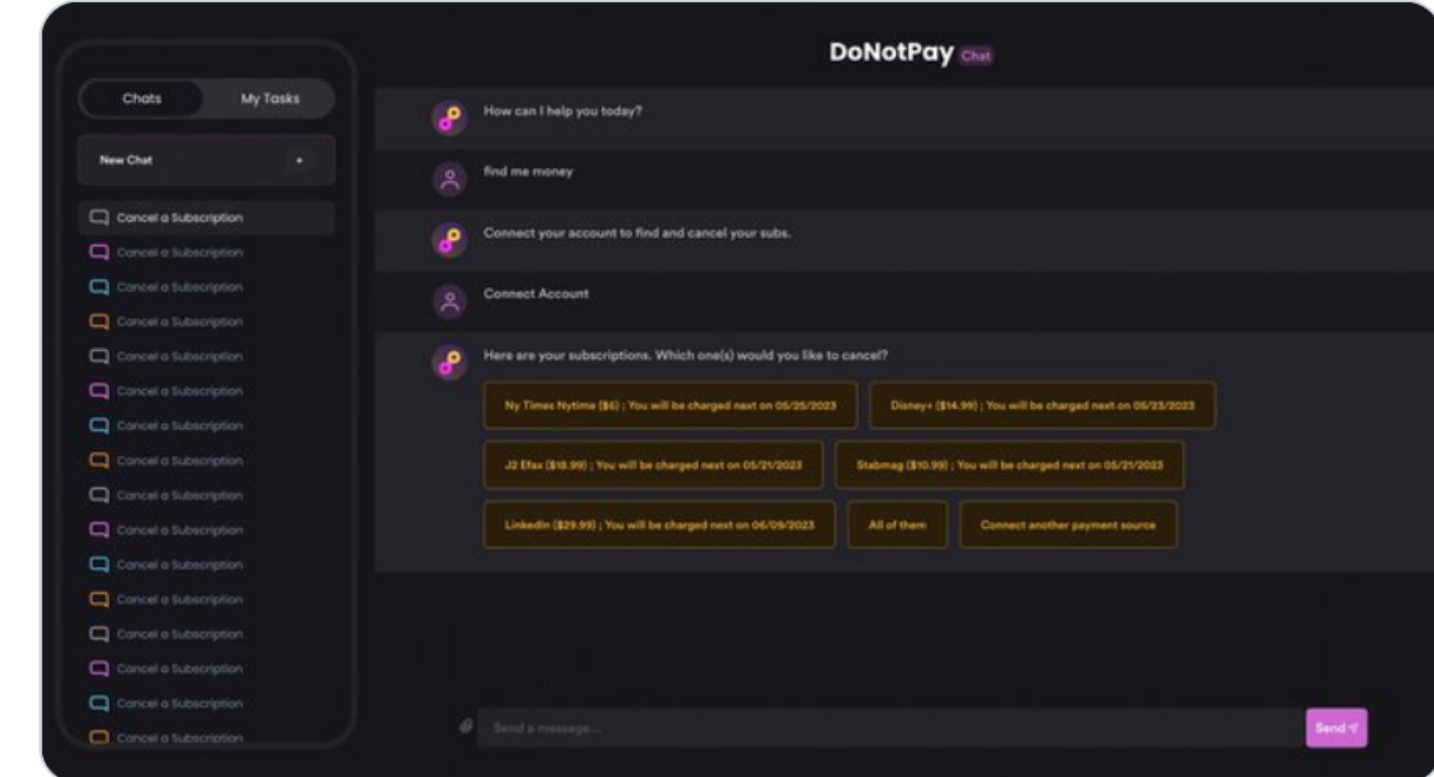
Available from samuel-api.com

 **Joshua Browder** ✓
@jbrowder1

I decided to outsource my entire personal financial life to GPT-4 (via the @donotpay chat we are building).

I gave AutoGPT access to my bank, financial statements, credit report, and email.

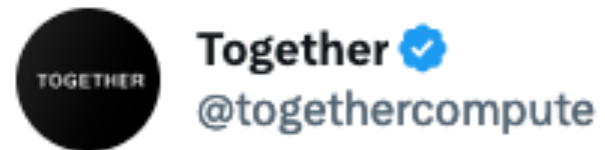
Here's how it's going so far (+\$217.85) and the strange ways it's saving money. (1/n):



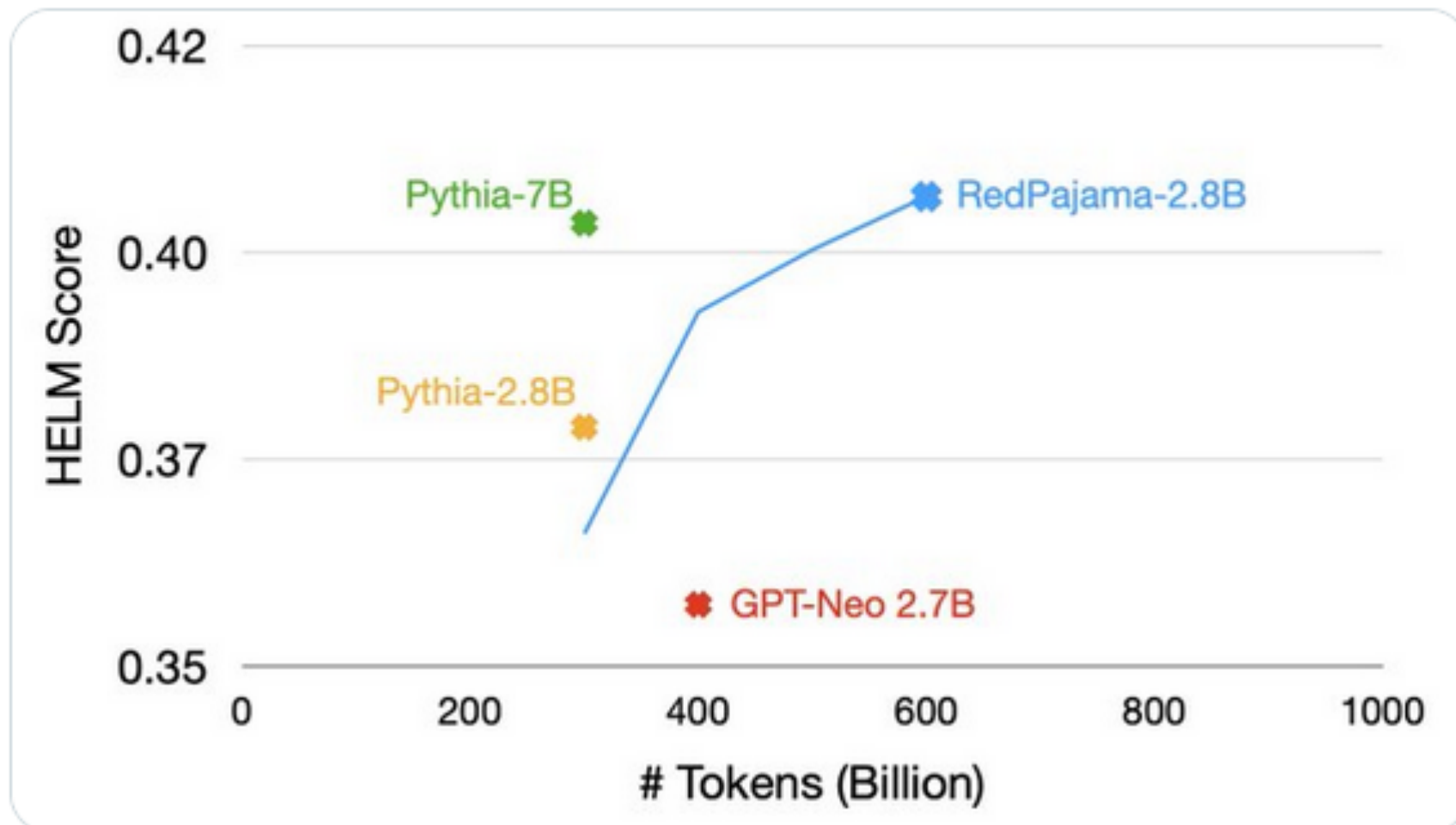
7:00 PM · Apr 29, 2023 · 5.3M Views

2,398 Retweets 787 Quotes 15.1K Likes 16.2K Bookmarks

News Round-Up



In addition to RedPajama 7B, we've also been training a 2.8B model. After 600B tokens it is exciting to see the model has higher HELM scores than the excellent Pythia-2.8B & GPT-Neo 2.7B. In fact, trained with twice the tokens, RedPajama-2.8B has comparable quality to Pythia-7B!



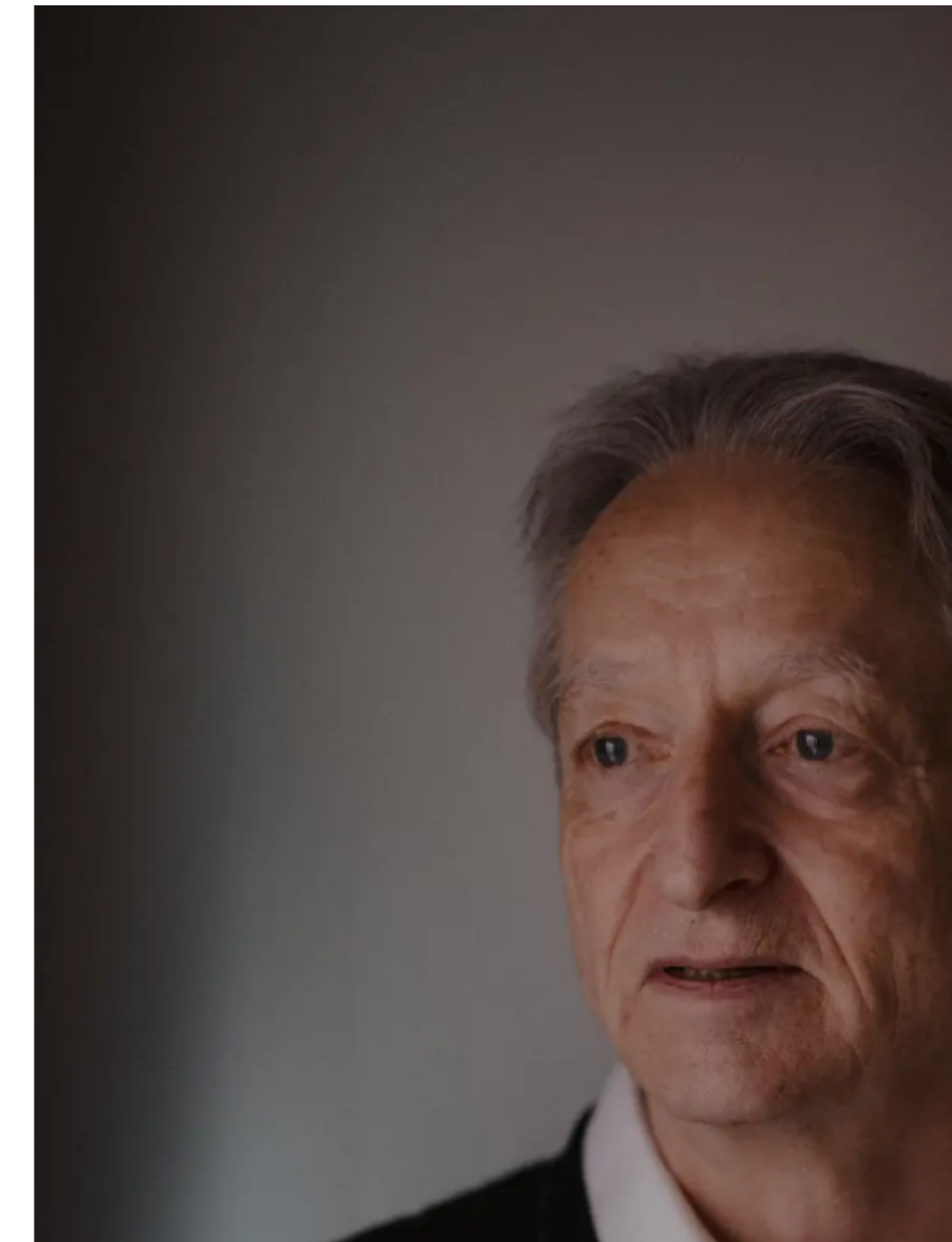
6:02 PM · Apr 30, 2023 · 211.7K Views

73 Retweets 20 Quotes 478 Likes 131 Bookmarks



'The Godfather of A.I.' Leaves Google and Warns of Danger Ahead

For half a century, Geoffrey Hinton nurtured the technology at the heart of chatbots like ChatGPT. Now he worries it will cause serious harm.



"I don't think they should scale this up more until they have understood whether they can control it"

Tools Round-Up

NVIDIA / NeMo-Guardrails Public

Watch 16 Fork 50 Star 981

Code Issues 3 Pull requests 1 Discussions Actions Projects Security Insights

main 1 branch 1 tag

Go to file Add file Code

drazvan Fix general instructions override bug (#7). 12e011d 3 days ago 9 commits

.github/workflows	Add GitHub Actions configuration for running tests.	4 days ago
chat-ui	Release 0.1.0.	last week
docs	Release 0.1.0.	last week
examples	Fix action parameters inspection bug (#2) and examples.	5 days ago
nemoguardrails	Fix general instructions override bug (#7).	3 days ago
tests	Fix general instructions override bug (#7).	3 days ago
vscode_extension	Release 0.1.0.	last week
.gitignore	Release 0.1.0.	last week
.gitlab-ci.yml	Add GitLab CI configuration for running tests.	4 days ago
.pre-commit-config.yaml	Release 0.1.0.	last week
CHANGELOG.md	Release 0.1.0.	last week
CONTRIBUTING.md	Release 0.1.0.	last week
LICENCES-3rd-party	Release 0.1.0.	last week
LICENSE-Apache-2.0.txt	Release 0.1.0.	last week
LICENSE.md	Release 0.1.0.	last week
MANIFEST.in	Release 0.1.0.	last week
README.md	Add RELEASE notice in README.md.	4 days ago
SECURITY.md	Release 0.1.0.	last week

About

NeMo Guardrails is an open-source toolkit for easily adding programmable guardrails to LLM-based conversational systems.

- Readme
- View license
- Security policy
- 981 stars
- 16 watching
- 50 forks

Report repository

Releases

1 tags

Packages

No packages published

Languages

- Python 98.7%
- HTML 1.3%

Nemo-Guardrails

Keep chatbots on topic & improve security

main 3 branches 5 tags

Go to file Add file Code

m-bain	Update setup.py	5b85c54 3 days ago	🕒 134 commits
📁 .github	Create FUNDING.yml		4 months ago
📁 figures	.wav conversion, handle audio with no detected speech		last month
📁 whisperx	allow custom model_dir for torchaudio models		2 weeks ago
📄 .gitignore	multilingual init		5 months ago
📄 EXAMPLES.md	clean up logic, use pandas where possibl		4 months ago
📄 LICENSE	update license to bsd-4		5 months ago
📄 MANIFEST.in	init commit		5 months ago
📄 README.md	update python example		last month
📄 requirements.txt	fix whisper version to 20230314 for no breaking		last month
📄 setup.py	Update setup.py		3 days ago

About

WhisperX: Automatic Speech Recognition with Word-level Timestamps (& Diarization)

- speech
- speech-recognition
- speech-to-text
- whisper
- asr

- 📖 Readme
- 📄 BSD-4-Clause license
- ☆ 2.6k stars
- 👁 58 watching
- 🍴 217 forks
- Report repository

Releases 4

v2.0.1 Latest 3 days ago

+ 3 releases

Sponsor this project

https://www.buymeacoffee.com/maxh...

Packages

WhisperX

Automatic Speech Recognition with Word-level Timestamps (& Diarisation)

☰ README.md

WhisperX

🌟 stars 2.6k
🐛 issues 71 open
📄 license BSD-4-Clause
🐦 Tweet

[What is it](#) •
 [Setup](#) •
 [Usage](#) •
 [Multilingual](#) •
 [Contribute](#) •
 [More examples](#) •
 [Paper](#)

Input audio

Cut & Merge

Batch

<|transcribe|>

main 1 branch 0 tags Go to file Add file Code

MaureenZOU Update README.md	9aa52d4 3 days ago	155 commits
assets	Add files via upload	5 days ago
demo_code	Update app.py	5 days ago
.gitignore	release demo code	5 days ago
LICENSE	update LICENSE	5 days ago
README.md	Update README.md	3 days ago
SEEM_arXiv.pdf	Add files via upload	3 days ago

About

Official implementation of the paper "Segment Everything Everywhere All at Once"

- Readme
- Apache-2.0 license
- 2.2k stars
- 49 watching
- 85 forks

SEEM

Interactive segmentation for many kinds of prompts

README.md

👁️ SEEM: Segment Everything Everywhere All at Once

We introduce **SEEM** that can Segment Everything Everywhere with Multi-modal prompts all at once. SEEM allows users to easily segment an image using prompts of different types including visual prompts (points, marks, boxes, scribbles and image segments) and language prompts (text and audio), etc. It can also work with any combinations of prompts or generalize to custom prompts!

[📄 \[Read our arXiv Paper\]](#) [🍏 \[Try Hugging Face Demo\]](#)

👉 **[New] One-Line Getting Started with Linux:**

```
git clone git@github.com:UX-Decoder/Segment-Everything-Everywhere-All-At-Once.git && cd Segment-Eve
```

🔥 **Related projects:**

Releases

No releases published

Packages

No packages published

Contributors 6

Languages

- Python 88.1%
- Cuda 10.6%
- C++ 1.2%
- Shell 0.1%