

LAION

**Democratizing AI through community
organizing**

»KI-Forschung darf nicht hinter verschlossenen Türen stattfinden«

KI sollte öffentliches Gut sein: So sieht es das KI-Forschungsnetzwerk LAION, das Datensätze zum Trainieren großer KI-Modelle bereitstellt und auch selbst offene KI-Modelle entwickelt. LAION setzt sich vehement für einen offenen Zugang zu KI ein. Was treibt die Gründer an?

von [Silke Hahn](#)



© FALLING WALLS FOUNDATION (AUSCHNITT)

Wie offen sollte KI sein? Die einen wollen alles offenlegen, die anderen ihre Geschäftsgeheimnisse nicht preisgeben. Open-Source-Verfechter wie das gemeinnützige Forschungsnetzwerk LAION haben Sicherheitsbedenken, und die aktuelle Führungskrise bei OpenAI scheint ihnen Recht zu geben.

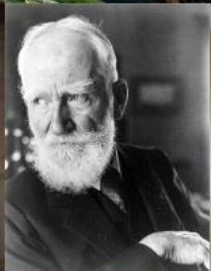
WHY ADVANCE AI AT ALL?

Intelligence = Ability to solve Problems

→ Advancing AI is about increasing our ability to control problems & to solve them when we wish to

WHAT WE ACCOMPLISHED SO FAR

- LAION - 400M (e.g. used for Imagen & Make a Video)
- LAION - 5B
- LAION - Aesthetics (e.g. used for SD)
- OpenCLIP (up to CLIP G so far)
- LeoLM







Poster

LAION-5B: An open large-scale dataset for training next generation image-text models

Christoph Schuhmann · Romain Beaumont · Richard Vencu · Cade Gordon · Ross Wightman · Mehdi Cherti · Theo Coombes · Aarush Katta · Clayton Mullis · Mitchell Wortsman · Patrick Schramowski · Srivatsa Kundurthy · Katherine Crowson · Ludwig Schmidt · Robert Kaczmarczyk · Jenia Jitsev




Hall J #1012

Keywords: [reproducibility] [large-scale datasets] [CLIP] [open source] [multi-modal learning]



Outstanding Paper

[[Abstract](#)]

[ Slides] [ Poster] [ OpenReview]

Wed 30 Nov 2:30 p.m. PST – 4 p.m. PST ([Bookmark](#))

[[Paper Metadata for Authors \(e.g. Poster and Slide Uploads...\)](#)]



LeoLM: Igniting German-Language LLM Research

by: Björn Plüster, 28 Sep, 2023

We proudly introduce LeoLM (Linguistically Enhanced Open Language Model), the first comprehensive suite of German-language Foundation Language Models trained in collaboration with HessianAI on their new supercomputer 42! Built on Llama-2 and trained on a large-scale, high-quality German text corpus,...

**WE ARE A
FALLING
WALLS
BREAK—
THROUGH**

LAION Triumphs at the Falling Walls Science Breakthrough of the Year 2023 Awards

by: Christoph, Jenia, Robert, 14 Sep, 2023

We happily announce that we, LAION, won the Falling Walls Science Breakthrough of the Year 2023 Award in the category Science and Innovation Management for “democratizing AI research by providing open access to advanced AI models, tools, and datasets, fostering public engagement and awareness, ...

PROJECT ALEXANDRIA

**Making humanities knowledge accessible
for everyone**

WHY?

- Converting all knowledge from all books, scientific papers, websites, ... into a copyright-friendly data format that can be shared & used for training LLMs
- **Making all knowledge** in the database **accessible for semantic search**
- Allowing it people from the whole world to quickly get
 - **pro & contra arguments for hypotheses**
 - **instructions**

→ that are **grounded with links to the original source texts** with exact positions

→ TL;DR: **Build an open source search engine for AI trainings & apps**



IT-RECHT - DATENSCHUTZ - COMPLIANCE

HEIDRICH RECHTSANWÄLTE

- The storage of texts for analysis should be covered by the exceptions of §§ 44b and 60d UrhG (German Copyright Law) and thus be lawful, provided that the copies are deleted after the analysis.
- The analysis and creation of knowledge graphs themselves, as well as the publication of the graphs, should not infringe on the rights of the copyright holders and thus also be lawful.
- The extraction and permanent storage of sentence fragments for setting anchor points is also likely to be non-infringing.

STEP 1 - CONVERT TEXTS INTO KNOWLEDGE GRAPHS

CURRENT INPUT:

Improved Baselines with Visual Instruction Tuning

Haotian Liu¹ Chunyuan Li² Yuheng Li¹ Yong Jae Lee¹

¹University of Wisconsin–Madison ²Microsoft Research, Redmond

<https://llava-vl.github.io>

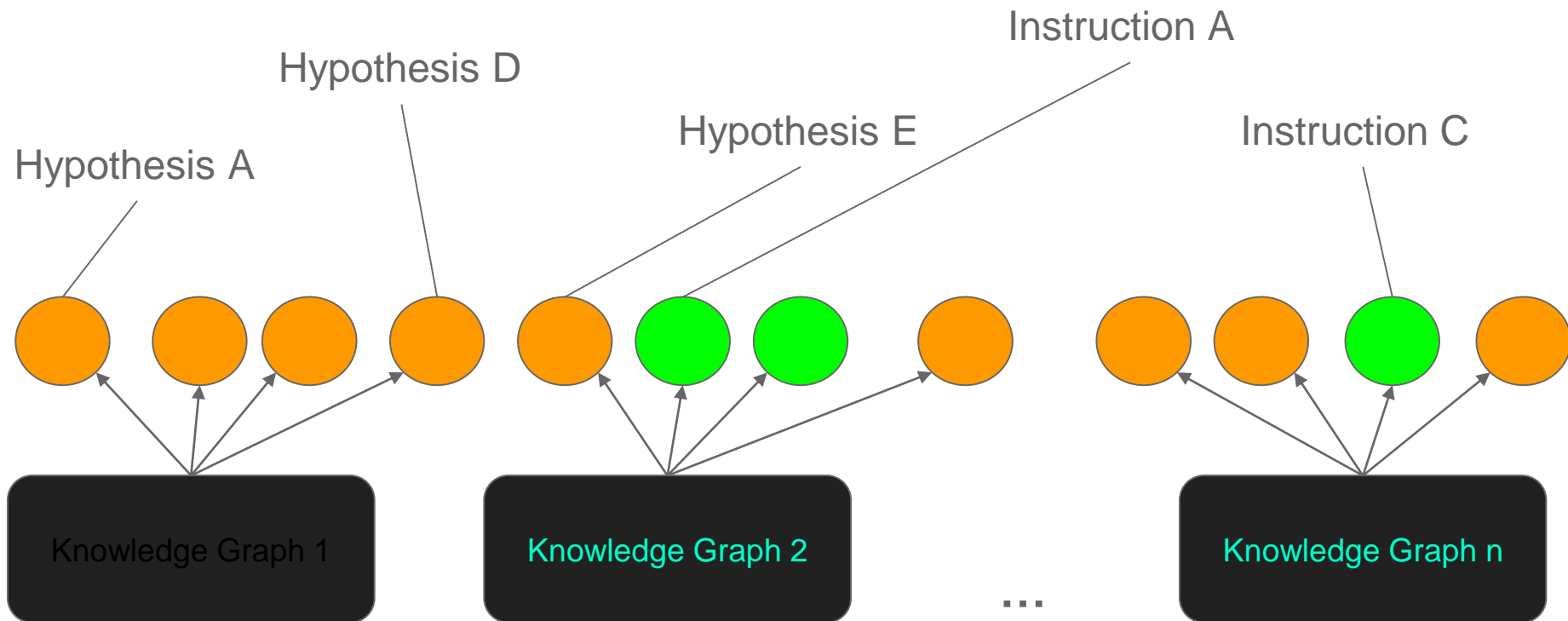
Abstract

Large multimodal models (LMM) have recently shown encouraging progress with visual instruction tuning. In this note, we show that the fully-connected vision-language cross-modal connector in LLaVA is surprisingly powerful and data-efficient. With simple modifications to LLaVA, namely, using CLIP-ViT-L-336px with an MLP projection and adding academic-task-oriented VQA data with simple response formatting prompts, we establish stronger baselines that achieve state-of-the-art across 11 benchmarks. Our final

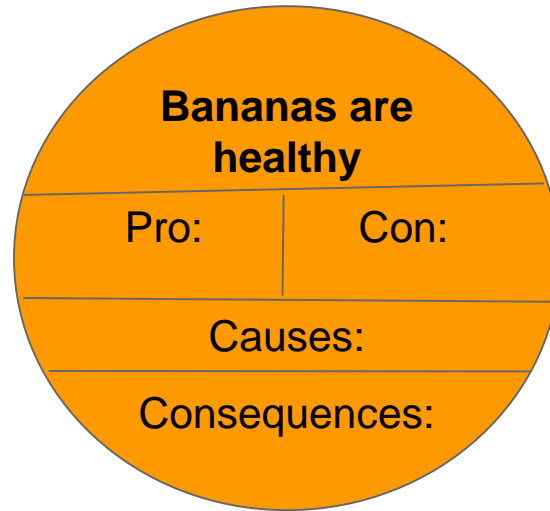
STEP 1 - CONVERT TEXTS INTO KNOWLEDGE GRAPHS

```
'LLaVA': {
  'relations': {
    'improves_baselines_with': 'Visual Instruction Tuning' # TEMPORAL_ID 0 ANCHOR "Improved Baseli
  },
  'attributes': {
    'authors': ['Haotian Liu', 'Chunyuan Li', 'Yuheng Li', 'Yong Jae Lee'], # TEMPORAL_ID 1 ANCHOR
    'affiliations': ['University of Wisconsin-Madison', 'Microsoft Research, Redmond'], # TEMPORAL
    'arxiv_link': 'https://llava-vl.github.io', # TEMPORAL_ID 3 ANCHOR "https://llava-vl.github.io"
  }
},
'Visual Instruction Tuning': {
  'relations': {
    'improves': 'LLaVA' # TEMPORAL_ID 0 ANCHOR "Improved Baselines with Visual Instruction Tuning"
  },
  'attributes': {
    'description': 'A method for improving the performance of large multimodal models on visual re
  }
},
'CLIP-ViT-L-336px': {
  'relations': {
    'used_in': 'LLaVA' # TEMPORAL_ID 5 ANCHOR "With simple modifications to LLaVA"
  },
  ...
}
```

STEP 2 - HYPOTHESES & INSTRUCTIONS GET EXTRACTED FROM KNOWLEDGE GRAPHS



HOW A HYPOTHESIS OBJECT LOOKS LIKE



HOW AN INSTRUCTION OBJECT LOOKS LIKE



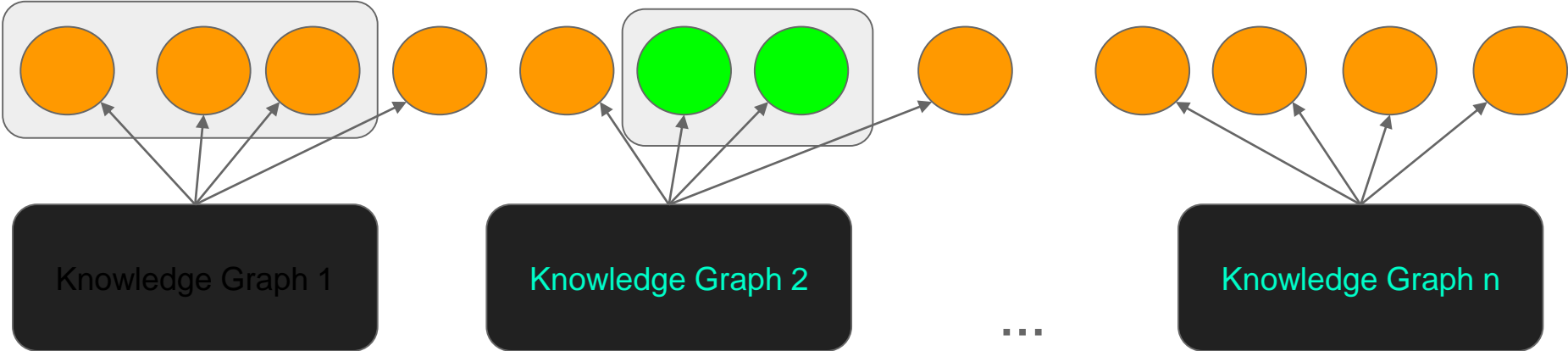
**How to make
banana ice**

Instructions:

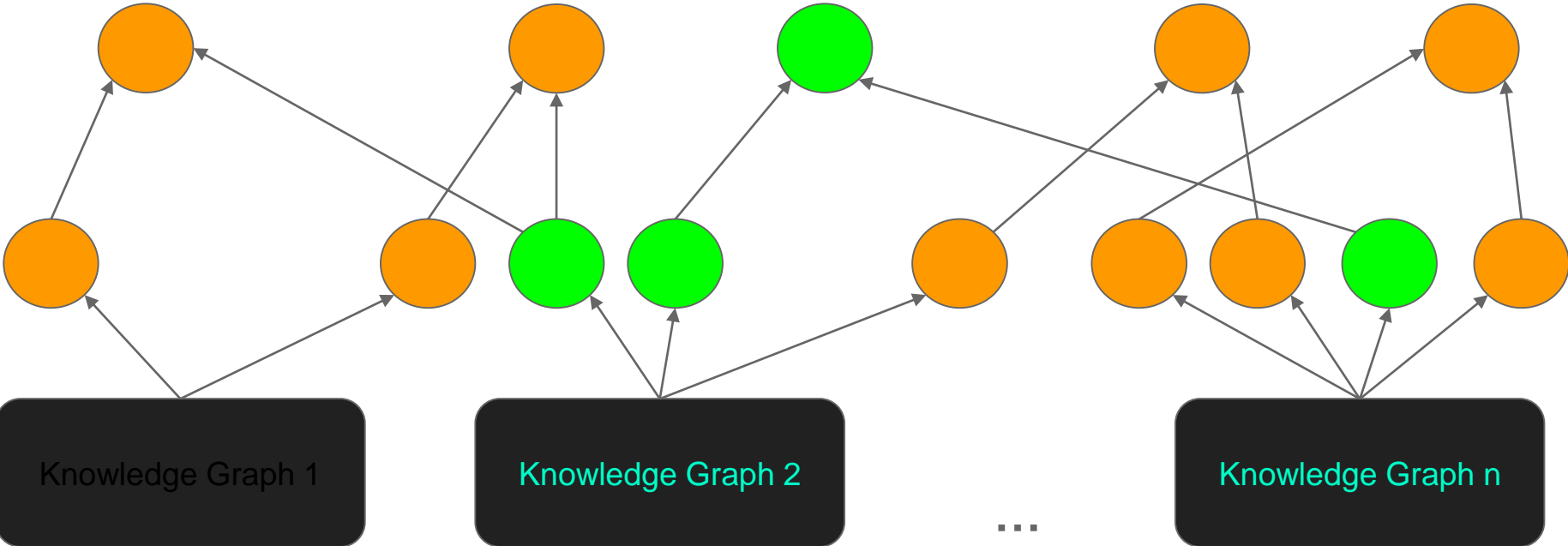
Purpose:

Necessary
conditions:

STEP 3 - MERGE HYPOTHESES & INSTRUCTIONS THAT MEAN THE SAME



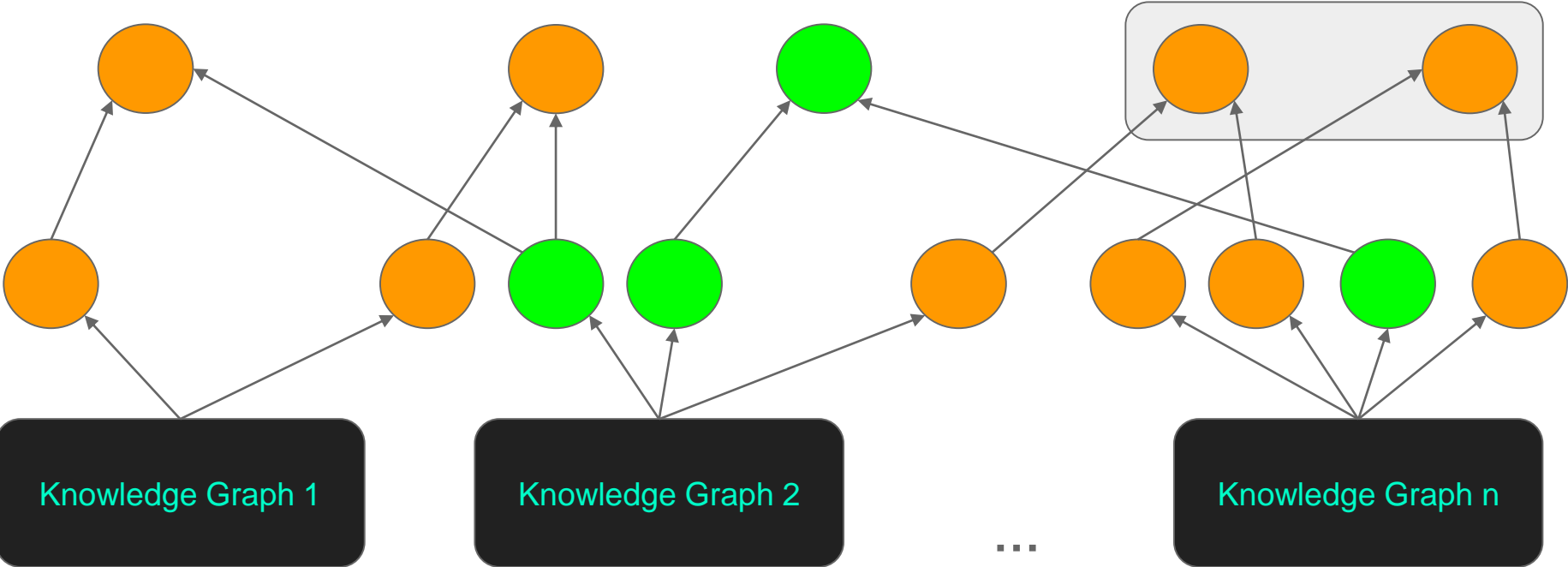
STEP 4 - HYPOTHESES & INSTRUCTIONS GET CONSTRUCTED ON A 2ND LAYER OF ABSTRACTION



STEP 4 - HYPOTHESES & INSTRUCTIONS GET CONSTRUCTED ON A 2ND LAYER OF ABSTRACTION

- For a given node on the 1st abstraction layer, we **retrieve from a vectorDB** with all nodes in the 1st abstraction layer:
 - Neighbors to the antithesis of the current node (e.g. “Banana’s are unhealthy”)
 - Neighbors to the causes & the potential consequences
 - Neighbors to the pros & cons
- Put them in context of the LLM together with the given node and let it formulate new diverse hypotheses and instructions that combine the information in these nodes in an intelligent manner.

STEP 5 - MERGE HYPOTHESES & INSTRUCTIONS THAT MEAN THE SAME



CONTINUE UNTIL A STOP-CRITERIA IS MET

- E.g. n layers of abstraction
- When the resulting hypotheses are very similar to the ones it references to as anchors
- When a LLM thinks the resulting hypotheses seem “useless”

RECONSTRUCTION FROM THE KNOWLEDGE GRAPH

The development of LLaVA, an innovative approach in the realm of Large Multimodal Models (LMMs), marks a significant advancement in the field. This project stands out for its application of Visual Instruction Tuning, a method specifically designed to enhance the performance of LMMs on visual reasoning tasks. The introduction of this tuning technique has been pivotal in improving baselines, setting LLaVA apart in its category.

Behind the success of LLaVA is a team of accomplished researchers: Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. Their collaborative effort, bridging the expertise from the University of Wisconsin–Madison and Microsoft Research, Redmond, has been fundamental in driving the project forward. The diverse academic and research backgrounds of these individuals have synergistically contributed to the model's development.

META-DATA

CATEGORY/GENRE: Academic Research and Technological Development


KEYWORDS: LLaVA, Large Multimodal Models, Visual Instruction Tuning, Technology, Innovation, Research Collaboration, Academic Data, AI Efficiency, Scalability, Visual Reasoning Tasks, Interdisciplinary Research

SUMMARY: The text describes the development of LLaVA, a significant advancement in Large Multimodal Models (LMMs), highlighting its innovative use of Visual Instruction Tuning to improve performance on visual reasoning tasks. It details the collaborative efforts of researchers from the University of Wisconsin–Madison and Microsoft Research, Redmond, and outlines key technical aspects like the integration of CLIP-ViT-L-336px and MLP projection. The text underscores LLaVA's efficiency, its impact on advancing LMM research, and its potential to set new benchmarks in AI and multimodal learning.

AGE-RATINGS

The provided text contains no violence, physical violence, or psychological violence. It is entirely focused on the development of a technological model and the collaborative efforts of researchers. There is no sexual content in the text. The content is technical and academic, discussing the development of a Large Multimodal Model (LMM) named LLaVA, and is devoid of any themes or language that would be inappropriate for children and non-adult teenagers. Considering the nature and content of the text, it is "Suitable for kids & people of all ages." However, the technical complexity might be more appreciated and understood by older children or teenagers with an interest in technology and science.

The Rise And Fall... And Rise Of Sam Altman Has Grave Implications For AI Research And Humanity

Hessie Jones Contributor 

Strategist, Investor, Advocating for Human-Centered AI, Privacy

Follow



Nov 30, 2023, 12:04pm EST

CONNECT

Our LAION Discord Server

<https://discord.gg/nGuc6rGdqP>

Mail

contact@laion.ai

Website

<https://laion.ai>