

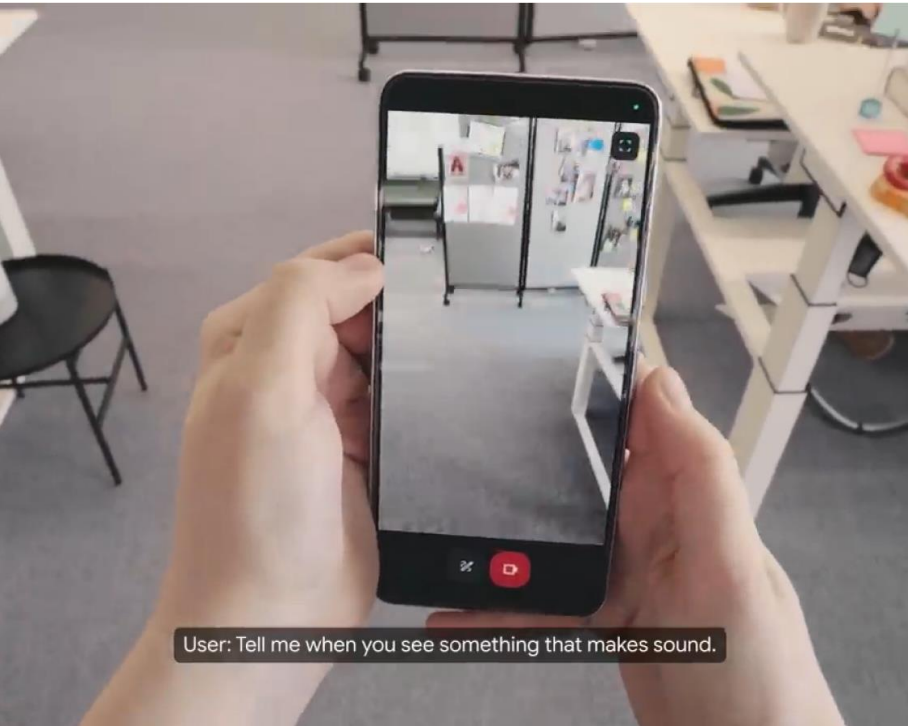
Artificial intelligence – A tale of two meanings

On useful tools, existential threat, and the imperative of hope

Thilo Stadelmann, ZHAW Centre for AI

Zurich University
of Applied Sciences

zhaw School of
Engineering



PROF. DR. THILO STADELMANN

About Me

- Director & Professor of AI & ML @ ZHAW Centre for AI
- Founder & member of the Board @ AlpineAI AG
- Studied computer science in Giessen & Marburg (DE)
- PhD voice recognition, industry years in automotive
- Worked on ecosystem buildup (associations, conferences, networks, ...)

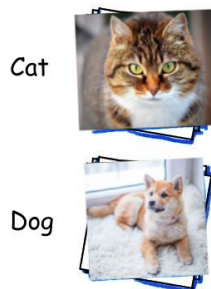


What is artificial intelligence?

The mimicking of *intelligent behavior* with a computer



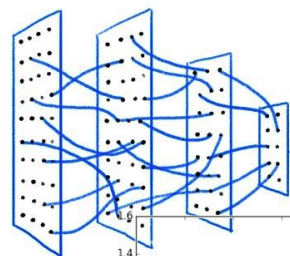
Since 2012: Machine Learning
(especially *deep learning*)



Cat

Dog

Image: Rishu Shrivastava



Output
Cat

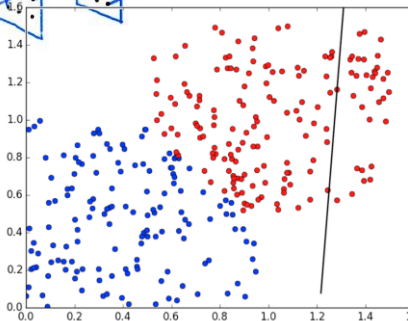


Image: Sai Chandra Nerella

Since 2022: Large
language models

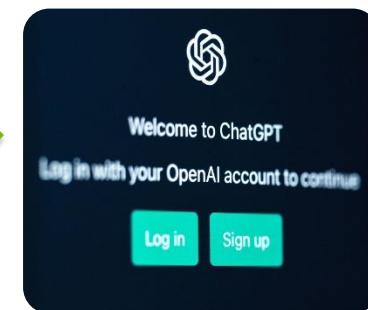
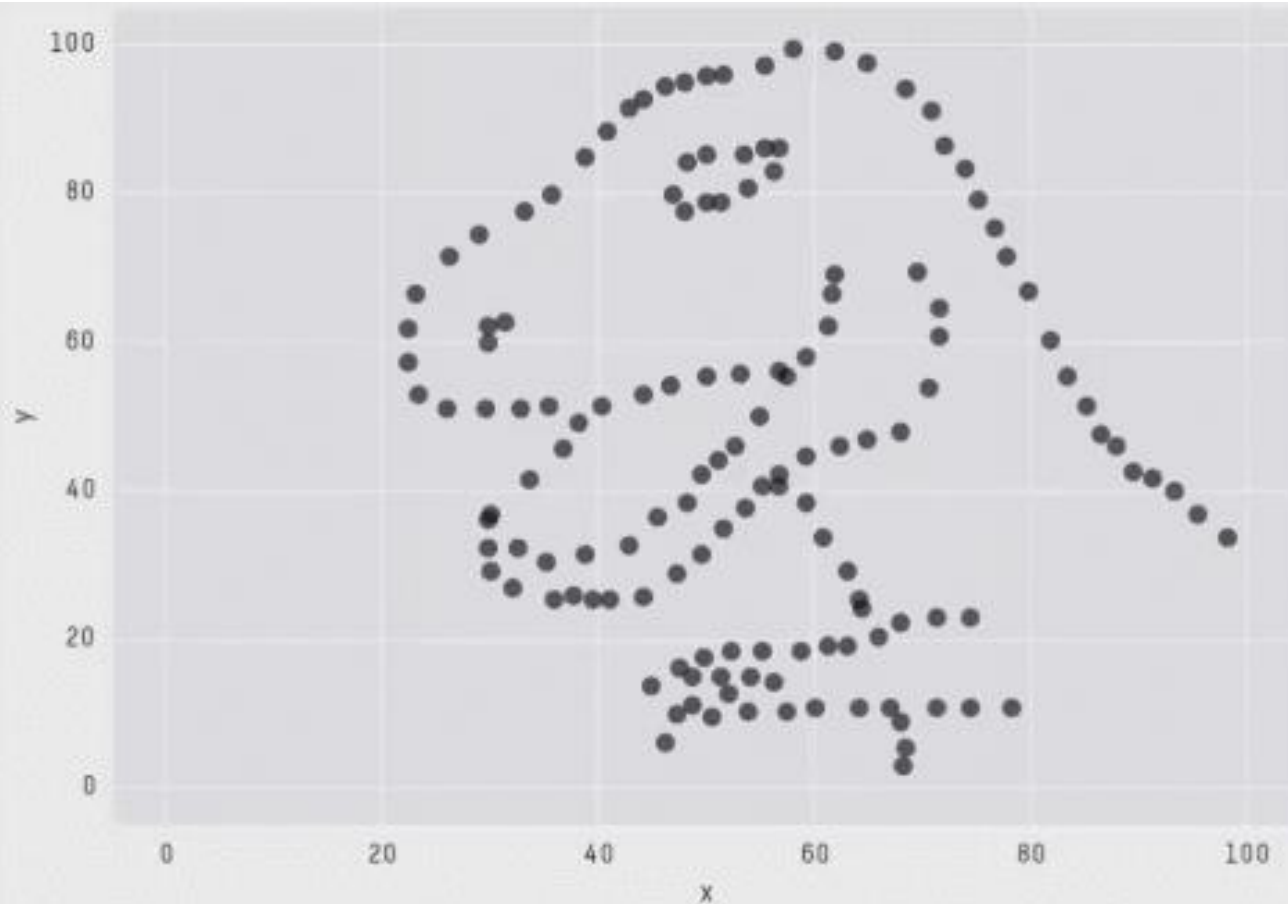


Image: Wikipedia

The result: Unlike humans

Statistical results - better on average, may be wrong in individual cases



X Mean: 54.2659

Y Mean: 47.8313

X SD : 16.7649

Y SD : 26.9342

Corr. : -0.0642

The result: Unlike humans

Statistical results - better on average, may be wrong in individual cases



You

A man and a goat are standing at a river. The man has a boat. How can both be brought to the other side?



ChatGPT

To solve the riddle, let's break down the scenario and figure out the steps the man needs to take to get both himself and the goat across the river using the boat.

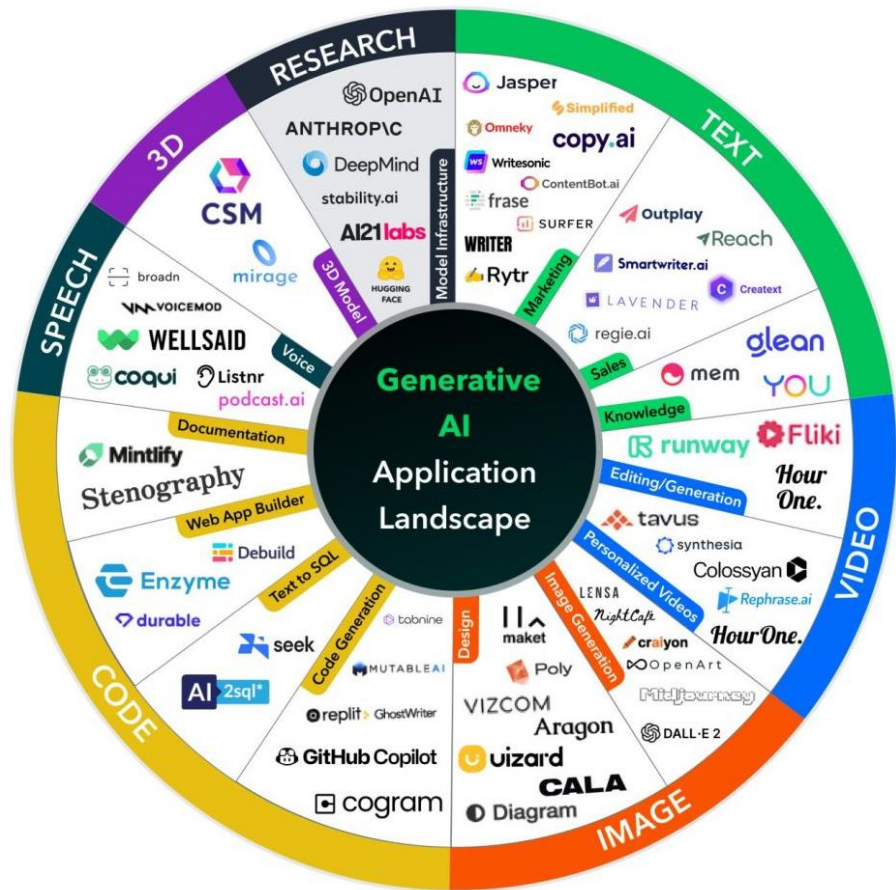
1. **Initial Situation:** The man and the goat are on the starting side of the river. The man has a boat.
2. **Constraints:** The boat can carry only the man and the goat at the same time. We need to ensure both reach the other side safely.

Steps to Solve the Riddle

1. **First Trip:** The man takes the goat across the river and leaves the goat on the other side.
2. **Return Trip:** The man returns to the starting side alone.
3. **Second Trip:** The man crosses the river alone once more.

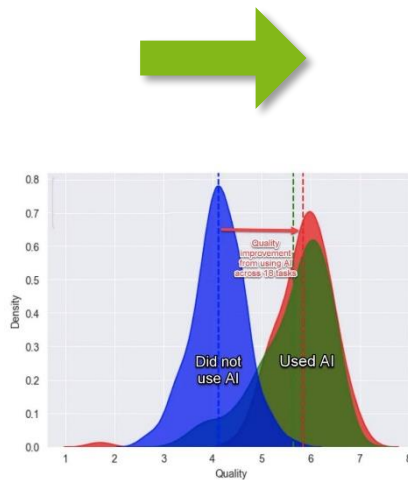
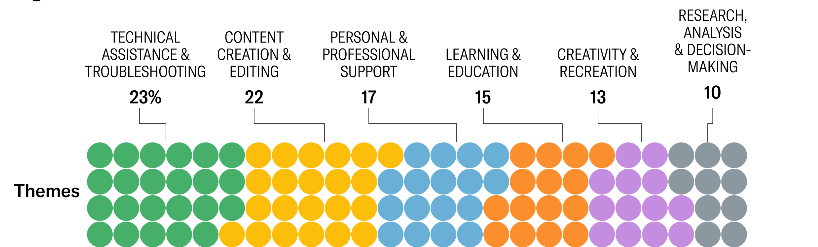


Usefulness → spread



How People Are Using GenAI

Have people found ways for generative AI to help lighten their workloads, increase their productivity, or think through problems in new ways? To understand how individuals are using the technology, researchers mined web forums like Quora and Reddit, filtering through tens of thousands of posts to identify 100 different use-case categories, which they then organized into six themes.



Distribution of output quality across all the tasks. The blue group did not use AI, the green and red groups used AI, the red group got some additional training on how to use AI.

- | Categories | |
|------------|------------------------------------|
| 1 | Generating ideas |
| 2 | Therapy/companionship |
| 3 | Specific search |
| 4 | Editing text |
| 5 | Exploring topics of interest |
| 6 | Fun & nonsense |
| 7 | Troubleshooting |
| 8 | Enhanced learning |
| 9 | Personalized learning |
| 10 | General advice |
| 11 | Drafting emails |
| 12 | Simple explainers |
| 13 | Writing/editing CV/resumé |
| 14 | Excel formulas |
| 15 | Adjusting tone of email |
| 16 | Evaluating copy |
| 17 | Enhanced decision-making |
| 18 | Language translation |
| 19 | Improving copy (pros) |
| 20 | Drafting a document |
| 21 | Reconciling personal disputes |
| 22 | Summarizing content |
| 23 | Making a complaint |
| 24 | Recommending movies, books, etc. |
| 25 | Cooking with what you have |
| 26 | Generating appraisals |
| 27 | Creativity |
| 28 | Medical advice |
| 29 | Generating a legal document |
| 30 | Fixing bugs in code |
| 31 | Drafting a formal letter |
| 32 | Writing & editing a cover letter |
| 33 | Personalized kid's story |
| 34 | Explaining technical documents |
| 35 | Preparing for interviews |
| 36 | Critique & counterargument |
| 37 | Knowledge checks |
| 38 | Coding for amateurs |
| 39 | Meeting summaries |
| 40 | Cleaning up notes |
| 41 | Explaining legalese |
| 42 | Spotting logical fallacies |
| 43 | Creating a holiday itinerary |
| 44 | Editing a legal document |
| 45 | Business advice |
| 46 | Replying to emails |
| 47 | Generating code (pros) |
| 48 | Getting past writer's block |
| 49 | Generating a lesson plan |
| 50 | Rubber ducking (debugging code) |
| 51 | Negotiating a deal |
| 52 | Fact-checking |
| 53 | Career advice |
| 54 | Practicing difficult conversations |
| 55 | Seeing blind spots |
| 56 | Data entry |
| 57 | Legal research |
| 58 | Writing job postings |
| 59 | Strengthening an argument |
| 60 | Jumping to the useful info |
| 61 | Generating video |
| 62 | Safe space to ask |
| 63 | Interpreting song lyrics |
| 64 | Dungeons & Dragons |
| 65 | Generating relevant images |
| 66 | Data manipulation |
| 67 | Homework |
| 68 | Writing social media copy |
| 69 | Translating code (pros) |
| 70 | Realistic web copy |
| 71 | With MS Office apps |
| 72 | Understanding movie plots |
| 73 | Coding for a basic video game |
| 74 | Tracking medical symptoms |
| 75 | Healthier living |
| 76 | Preparing for meetings |
| 77 | Explaining idioms |
| 78 | UX/user story writing |
| 79 | Generating code libraries |
| 80 | Writing poems |
| 81 | Work buddy |
| 82 | Editing video transcript |
| 83 | Motivating yourself |
| 84 | Packing for travel |
| 85 | Sampling data |
| 86 | Technical use of software |
| 87 | For people with ADHD |
| 88 | Ad/marketing copy |
| 89 | Special needs education |
| 90 | Spotting anomalies |
| 91 | Building a business plan |
| 92 | Refining prompts |
| 93 | For entrepreneurs/startups |
| 94 | Building a website/app |
| 95 | Writing blog posts |
| 96 | Writing a funding proposal |
| 97 | Writing a press release |
| 98 | Editing digital images |
| 99 | Planning workouts |
| 100 | Project management |

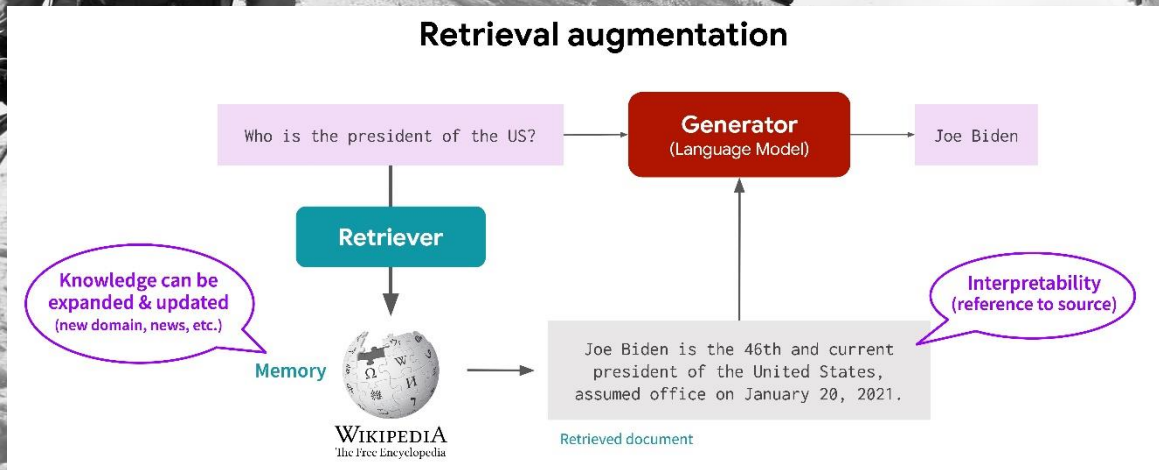
Use Case for organisations: Chat-based access to proprietary content



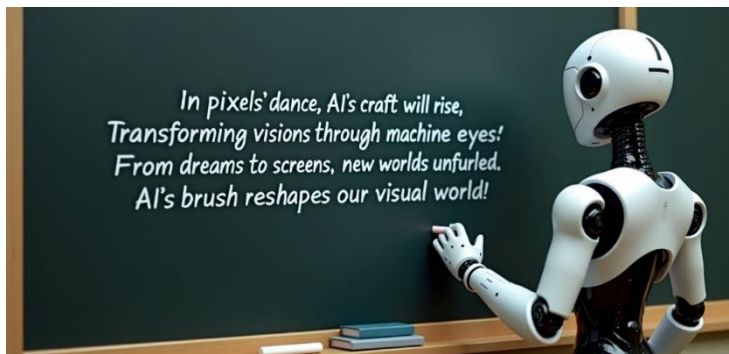
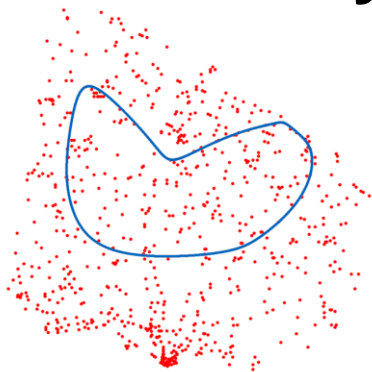
AI agents with **safe access to proprietary data** for reliable answers

Design principle of **retrieval augmented generation**:

- Data lake for own data **embeddings** in vector database
- LLM pre/post-processing **adds more semantic understanding to the search**



Other types of AI



AlphaFold 3 predicts the structure and interactions of all of life's molecules

May 08, 2024
6 min read

Introducing AlphaFold 3, a new AI model developed by Google DeepMind and Isomorphic Labs. By accurately predicting the structure of proteins, DNA, RNA, ligands and more, and how they interact, we hope it will transform our understanding of the biological world and drug discovery.

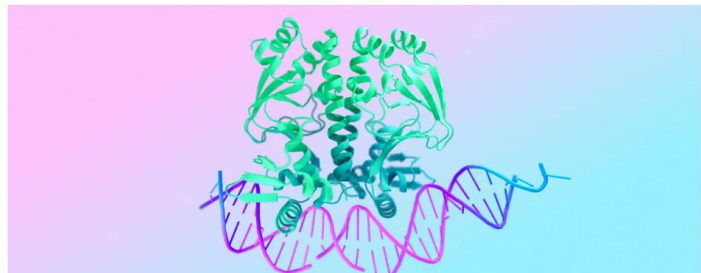


Google DeepMind
AlphaFold team



Isomorphic Labs

Share



THE GAME HAS CHANGED

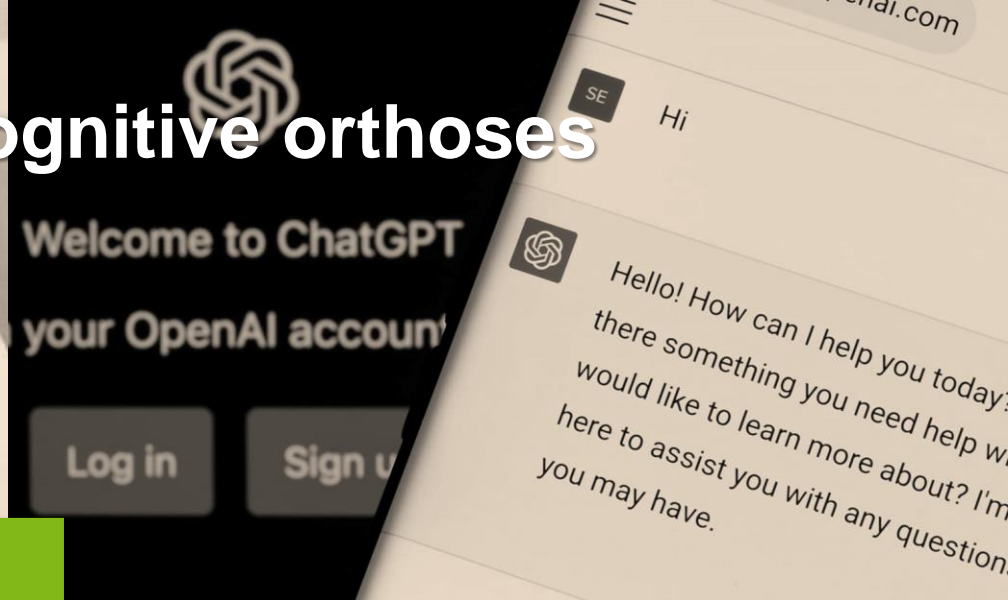
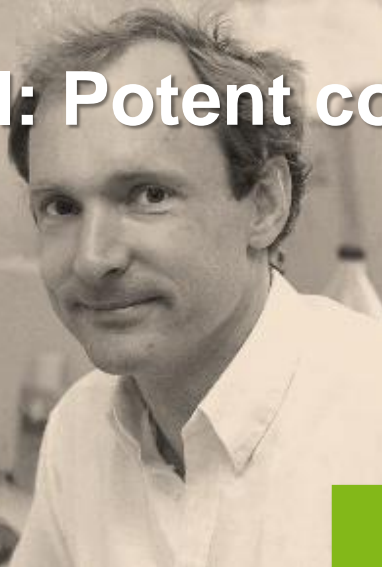
Top 200 shot locations in the NBA, 2001-02 versus 2019-20

2001-02

2019-20



The future of AI: Potent cognitive orthoses

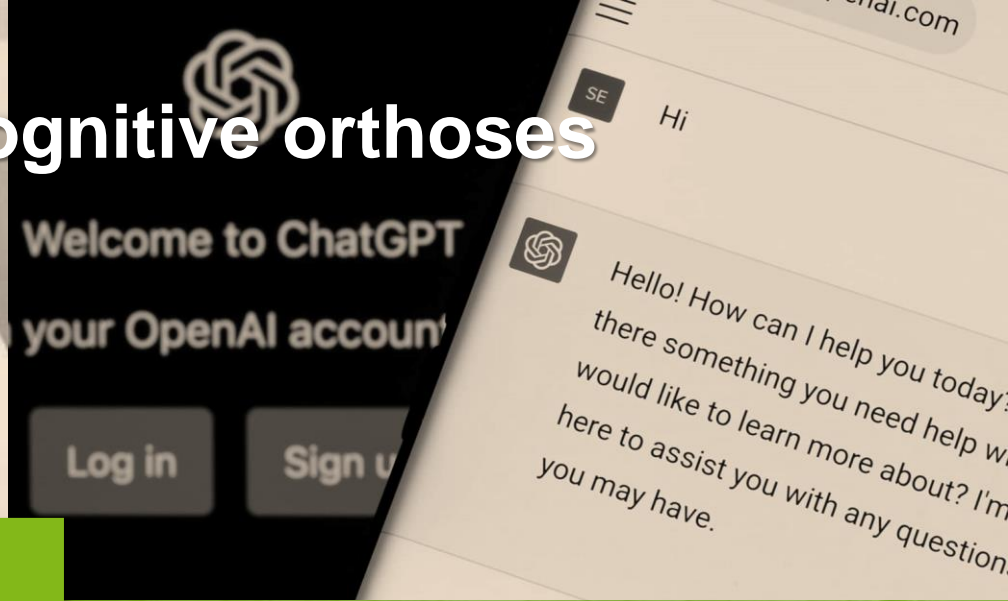
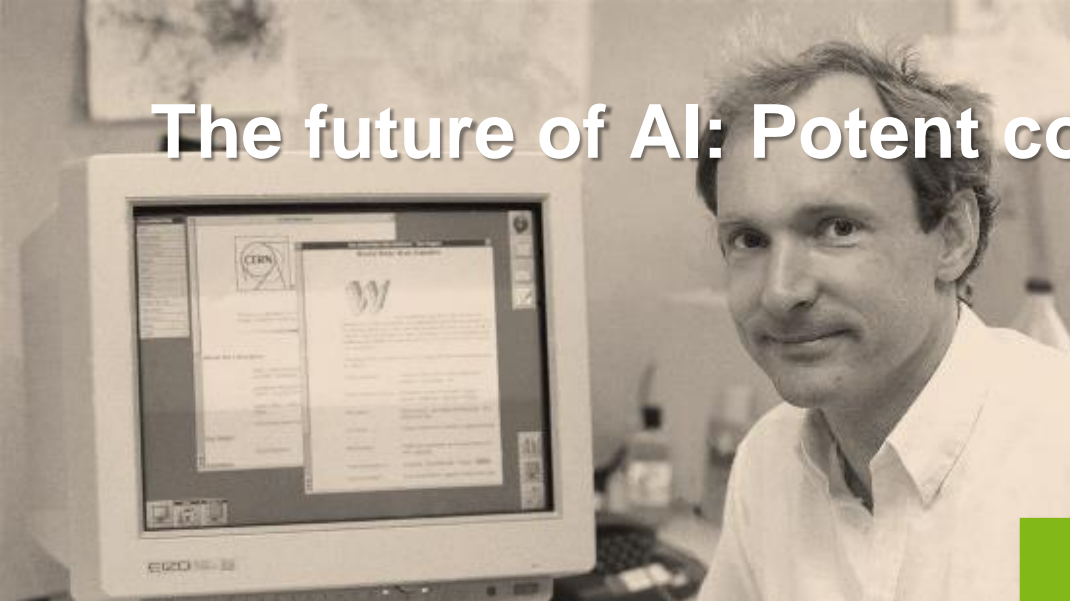


AI agents as personal assistants
(no superintelligence)

Web became the unified portal to
media & relationships



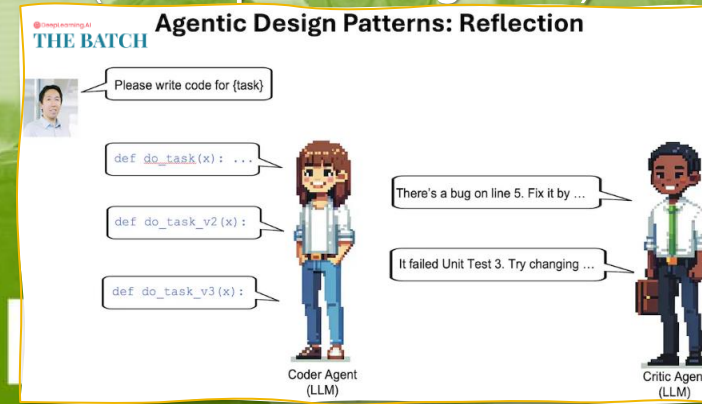
The future of AI: Potent cognitive orthoses



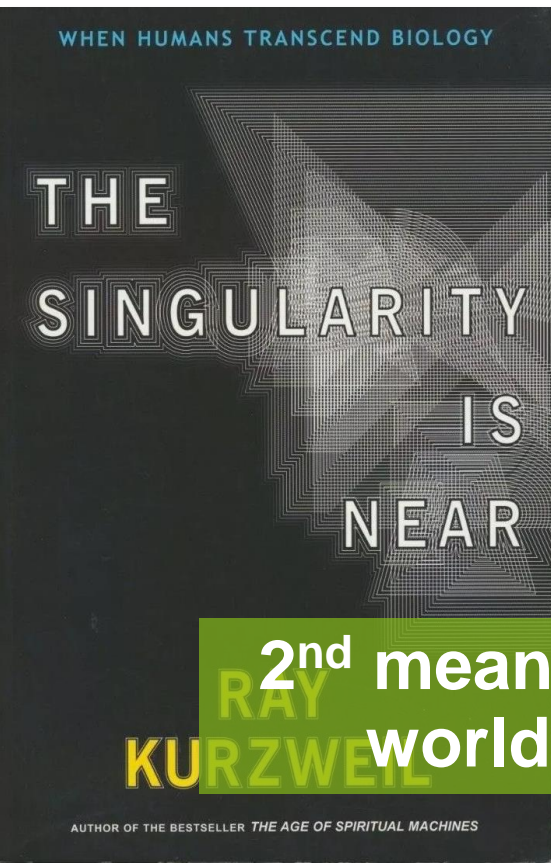
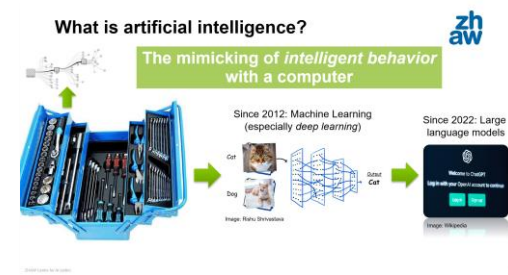
Web became the unified portal to media & relationships



AI agents as personal assistants
(no superintelligence)



But... is AI more than we have discussed so far?



THE IRISH SUN News Sport Fabulous TV Showbiz Money Travel

Philosophies to adopt:

- transhumansim
- extropianism
- singularitarianism
- cosmism
- rationalism
- effective altruism
- longtermism

Tech > News Tech

AI 'GOLD RUSH' reveals secret to becoming 'Pill' of any future

Intelligence will be 10,000x smarter than us in 10yrs

OpenAI frontman Sam Altman said the AI chatbot will be smarter than any human in 10 years.

Published: 16:32, 3 Oct 2024 | Updated: 16:34, 3 Oct 2024

tdy Sign in

DIG JUN 15, 2023

The Acronym Behind Our Wildest AI Dreams and Nightmares

To understand the deepening divide between AI boosters and doomers, it's necessary to unpack their common origins in a bundle of ideologies known as TESCREAL.

TESCREAL

Twenty-First Century: New Names, Old Ideas' Dig series

ÉMILE P. TORRES / TRUTHDIG CONTRIBUTOR

2nd meaning: empty term empowering futuristic worldviews (e.g., «TESCREAL» package)

ADVANCED artificial intelligence programs are now able to tell secrets to making more money, a top banker has revealed.

How exaggerated fantasy of feasibility leads to existential fears

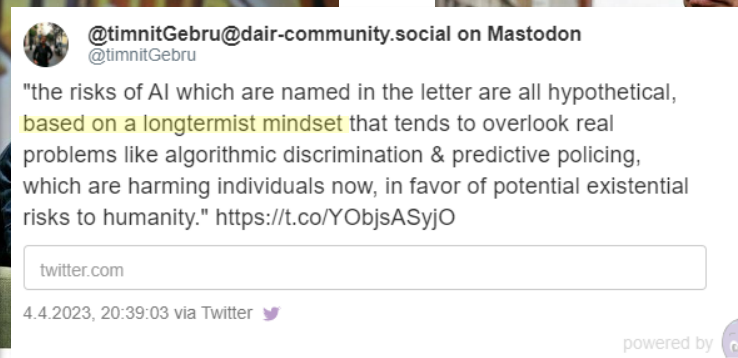
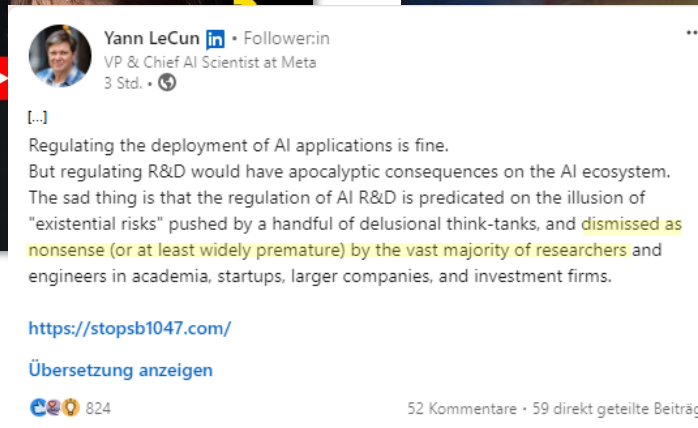
- transhumansim
 - extropianism
 - singularitarianism
 - cosmism
 - rationalism
 - effective altruism
 - longtermism
- „close match for my philosophy“

zhaw



AI risks leading humanity to 'extinction,' experts warn

It's the most recent in a series of alarms raised by experts in artificial intelligence – but also one that stoked growing pushback against a focus on what seem to be its overhyped hypothetical harms.



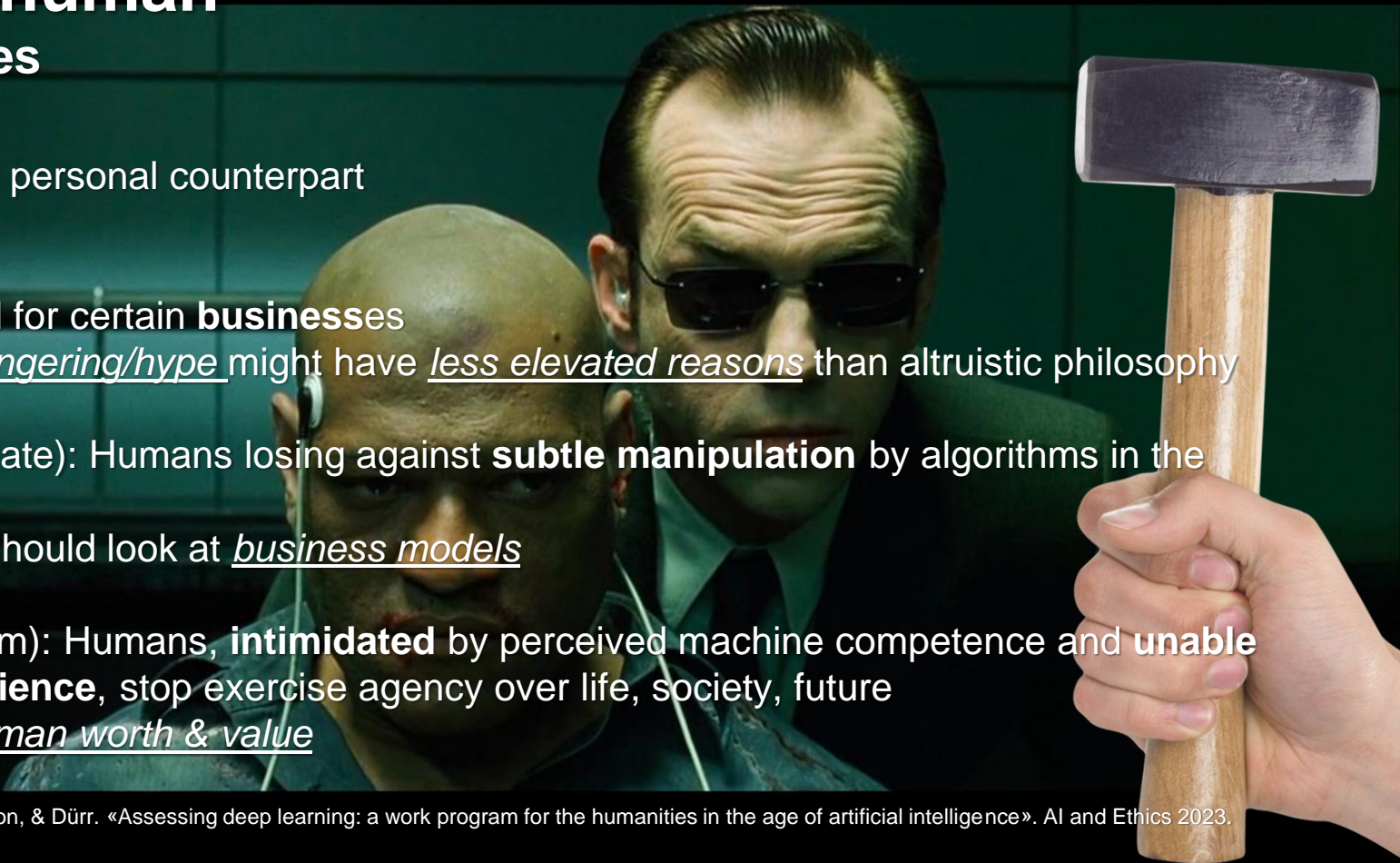
Fear is not based on scientific arguments, but on worldviews

AI and the human

Four hypotheses



- AI is a **tool**, not a personal counterpart
→ it just *isn't*.
- AI **hype** is helpful for certain **businesses**
→ some *fear-mongering/hype* might have *less elevated reasons* than altruistic philosophy
- Real risk (immediate): Humans losing against **subtle manipulation** by algorithms in the name of profit
→ AI *regulation* should look at *business models*
- Real risk (mid-term): Humans, **intimidated** by perceived machine competence and **unable to resist convenience**, stop exercise agency over life, society, future
→ strengthen *human worth & value*



Segessenmann, Stadelmann, Davison, & Dürr. «Assessing deep learning: a work program for the humanities in the age of artificial intelligence». AI and Ethics 2023.

How to strengthen human worth & value



Home > AI and Ethics > Article

Assessing deep learning: a work program for the humanities in the age of artificial intelligence

Review | Open access | Published: 21 December 2023 (2023) | Cite this article

[Download PDF](#) • You have full access to this open access article



AI and Ethics

Aims and scope →

Submit manuscript →

Jan Segessenmann, Thilo Stadelmann, Andrew Davison & Oliver Dürr

2459 Accesses | Explore all metrics

Abstract

Following the success of deep learning (DL) in research, we are now witnessing the fast and widespread adoption of artificial intelligence (AI) in daily life, influencing the way we act, think, and organize our lives. However, much still remains a mystery when it comes to how these systems achieve such high performance and what they reach the outputs they do. This presents us with an unusual combination: of technical mastery on the one hand and a striking degree of mystery on the other. This conjunction is not only fascinating, but it also poses considerable risks, which urgently require our attention. Awareness of the need to analyze ethical implications, such as bias, discrimination, and privacy, is growing. However, other dimensions of inquiry, such as the nature of intelligence, the role of the human mind, and the pervasive ways in which our dealings with AI shape our way of living and thinking, transforming our culture and human self-understanding. If we want to deploy AI positively in the long term, a broader and more holistic assessment of the technology is vital, involving not only scientific and technical aspects but also the humanities. To this end, we present outlines of a work program for the humanities, intended to contribute to assessing and guiding the potential, opportunities, and risks of further developing and deploying DL systems. This paper contains a thematic introduction (Sect. 1), an introduction to the workings of DL for non-technical readers (Sect. 2), and a main part, containing the outlines of a work program for the humanities (Sect. 3). Readers familiar with DL might want to ignore 2 and instead directly read 3 after 1.

Crazy idea: use AI's manipulator-qualities to help a human with these characteristics

Insight: being human is characterized by relationship, embracing of limits & responsibility

Foundation: created in the image of God, having an indwelling spirit



Conclusions


A = 🔍 Popular Latest Newsletters *The Atlantic* Sign In Subscribe

TECHNOLOGY

It's Time to Stop Taking Sam Altman at His Word

Understand AI for what it is, not what it might become.

By David Karpf



seongjeon Cho / Bloomberg / Getty

OCTOBER 4, 2024 SHARE ↗ SAVE ↺

OpenAI announced this week that it has raised \$6.6 billion in new funding and that the company is now valued at \$157 billion overall. This is quite a feat for an organization that reportedly burns through \$7 billion a year—far more cash than it brings in—but it makes sense when you realize that OpenAI's primary product isn't technology. It's stories.

Conclusions

- **AI changes** the way **information** is **obtained** and **organised** → becomes the user interface
 - **AI automates** complex & **recurring** tasks → very good with language, image-based to catch up
 - The technology will be **greatly refined** over the **next 2 years** → but **no fundamental leaps** in technology
- **Positive narratives** are **vital** to **attain** the **future** we want → strengthen human relations & character
- **Be outspoken about your worldview, strive to build hopeful narratives of the future**



About me:

Prof. Dr. Thilo Stadelmann
Director, ZHAW Centre for Artificial Intelligence (CAI)
Head, Machine Perception & Cognition research group
Founder & member of the board, AlpineAI AG

LinkedIn: <https://www.linkedin.com/in/thilo-stadelmann/>

Email: stdm@zhaw.ch

Phone: +41 58 934 72 08

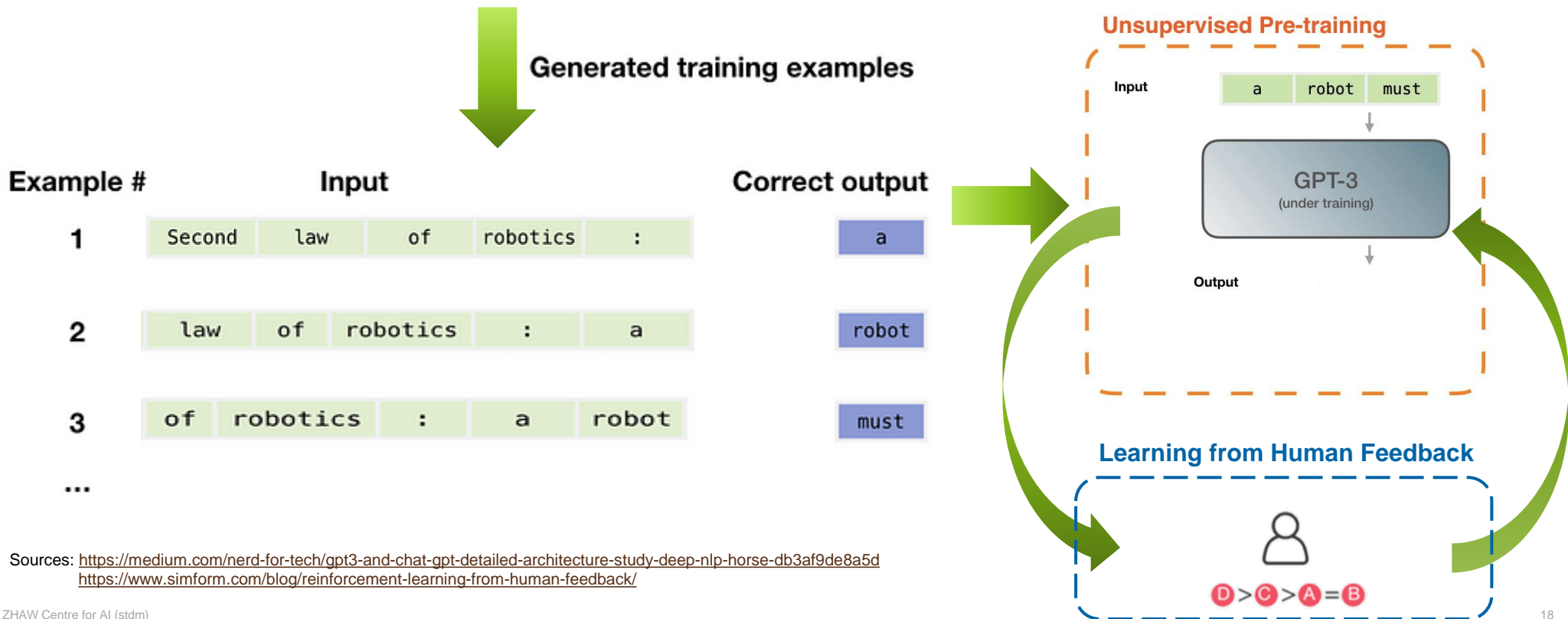


APPENDIX

How does LLM-based AI work?

Prediction of the next word based on previous context

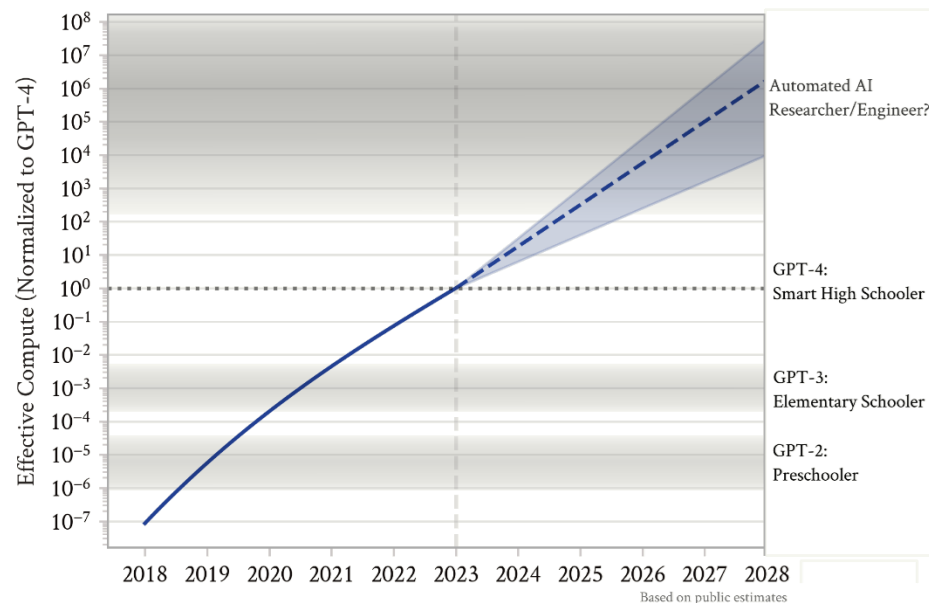
Text: Second Law of Robotics: A robot must obey the order given it by human beings.



The road to AI agents (1/2)

The power of scaling machine learning

Base Scaleup of Effective Compute



SITUATIONAL AWARENESS | Leopold Aschenbrenner

GPT-2 (2019) to GPT-4 (2024)

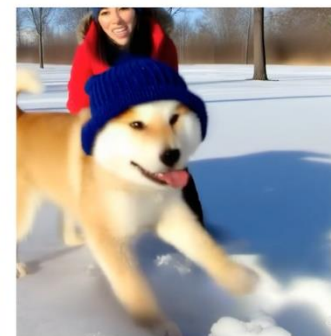
Compute	Algorithmic Efficiency	Unhobbling
3.5-4 OOMs	1-2 OOMs	2? OOMs
4.5-6 OOMs of base scaleup		Base to Chatbot

Based on public estimates.

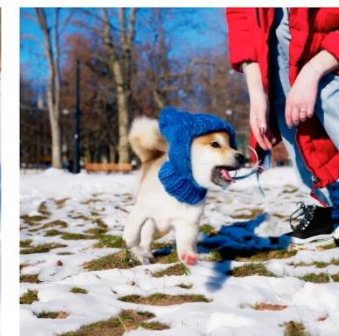
SITUATIONAL AWARENESS | Leopold Aschenbrenner



Base compute

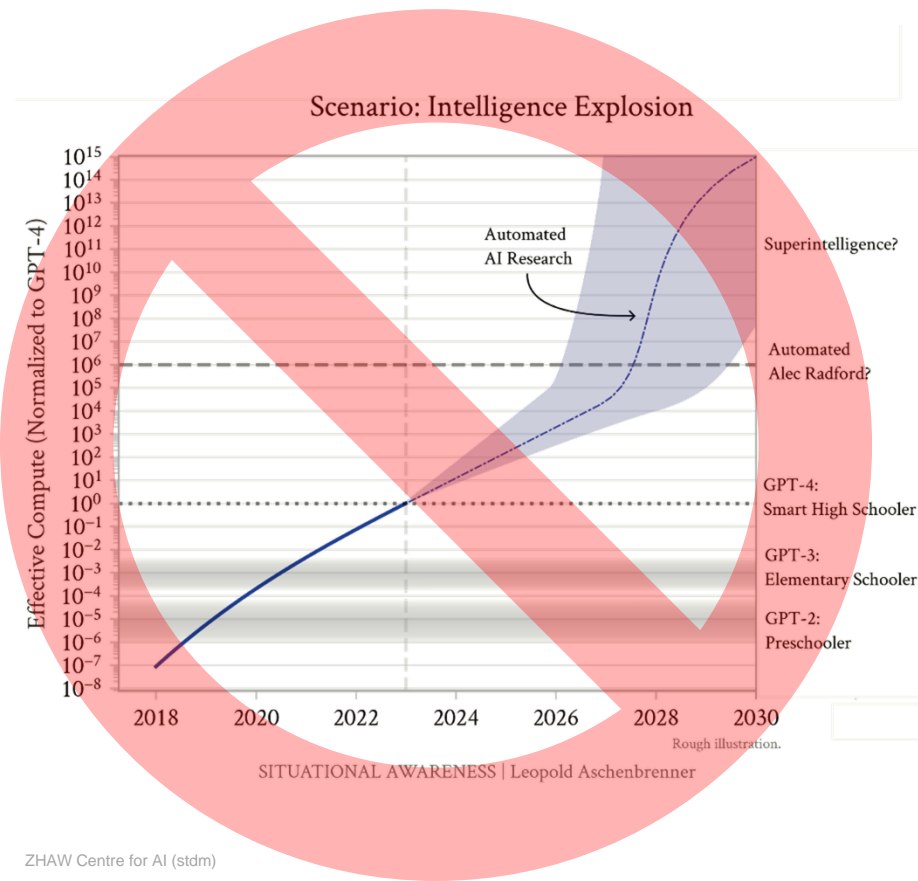


4x compute



32x compute

The road to AI agents (...but not to superintelligence



Before we reach Human-Level AI (HLAI), we will have to reach Cat-Level & Dog-Level AI.

We are nowhere near that.

We are still missing something big.

LLM's linguistic abilities notwithstanding.

A house cat has way more common sense and understanding of the world than any LLM.

[Post übersetzen](#)

7:25 nachm. · 5. Feb. 2023 · 860.007 Mal angezeigt

ts to Ditch Generative AI

ve view of AGI - advanced machine intelligence

4 Min Read



AI BUSINESS

THE BATCH

Agentic Design Patterns: Reflection



Please write code for {task}

```
def do_task(x): ...
```

```
def do_task_v2(x):
```

```
def do_task_v3(x):
```



Coder Agent (LLM)

There's a bug on line 5. Fix it by ...

It failed Unit Test 3. Try changing ...



Critic Agent (LLM)

h wants to abandon generative AI to achieve AGI. "ive."

y of AGI - Advanced Machine Intelligence or AMI. he argues.

kill us all, "don't listen to them."

of the godfathers of AI. wants to abandon generative e - in which AI assistants will have the same level of

Use Case: Creating illustrations

<https://stablediffusionweb.com/#demo>



Midjourney, spring 2022 vs. autumn 2023 vs. Sora, winter 2024

Used prompt:

- A friendly-looking cat
- A friendly-looking cat in an office environment, smiling
- A friendly, relaxed cat on a desk, peaceful, in full view
- A cat in an office, friendly and relaxed, total view, including a computer screen

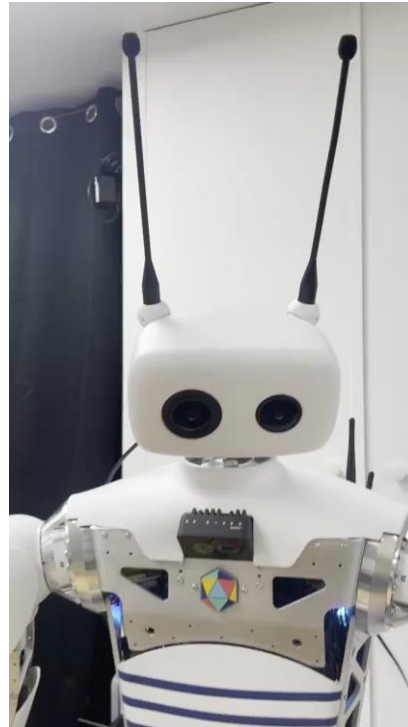
We're sharing our research progress early to get feedback from people outside of OpenAI and to give people a sense of what AI capabilities are on the horizon.

We will be taking several important safety steps before this research becomes available in any of our products.

Sora is a new AI model that can create realistic and imaginative scenes from text prompts.




Use case: Robotics

Autonomy breakthrough in 2025?



Video sources: Stanford University "Mobile ALOHA" <https://mobile-aloah.github.io/>; ANYbotics; Figure.ai; Pollen Robotics & Huggingface

Use Case: AI in Finance

 **Eddie Donmez**  • 2.
Founder at Creative Capital | LinkedIn Top Voice - Finance I +23...
[Newsletter abonnieren](#)
1 Monat • Bearbeitet • 

BREAKING: The World's Biggest Bank, [J.P. Morgan](#), Unveils 'IndexGPT' — AI Powered Thematic Investment Baskets 🤖

Last year, amidst the wild hype around ChatGPT, J.P. Morgan filed a trademark for 'IndexGPT'.

It mentioned "cloud computing software using artificial intelligence" for "analyzing and selecting securities tailored to customer needs".

But now according to Bloomberg, they are finally unveiling the product.

So, what does it do?



IndexGPT basically uses AI to identify investments that are associated with a given trend or 'theme' and creates investment baskets.

According to the BBG article here's roughly how it works:

Generates keywords associated with the theme --> Fed into a Natural Language Processing model --> Scans news articles etc --> Identifies companies associated with that theme --> Selects financial securities & creates thematic baskets

According to CEO [Jamie Dimon](#), JPM now employees in excess of 2,000 people working on AI.

One thing is clear, banks are betting BIG on AI and rightly so.


 3.105

Bloomberg

JPMorgan Unveils IndexGPT in Next Wall Street Bid to Tap AI Boom



Greg Ritchie and Justina Lee

Fri, May 3, 2024 • 4 min read



In This Article:

JPM +2.11%

MS +1.72%

(Bloomberg) -- A year after a wave of speculation broke out over its application to trademark the word "IndexGPT" in connection to an unspecified artificial intelligence-powered tool, JPMorgan Chase & Co. is finally unveiling the product that will bear the name.



AI made in Winterthur

AI in the hospital



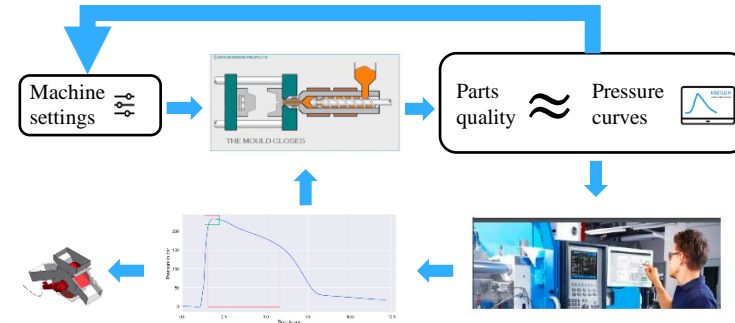
Emberger et al., «Video object detection for privacy-preserving patient monitoring in intensive care». In: SDS'23.

AI for science



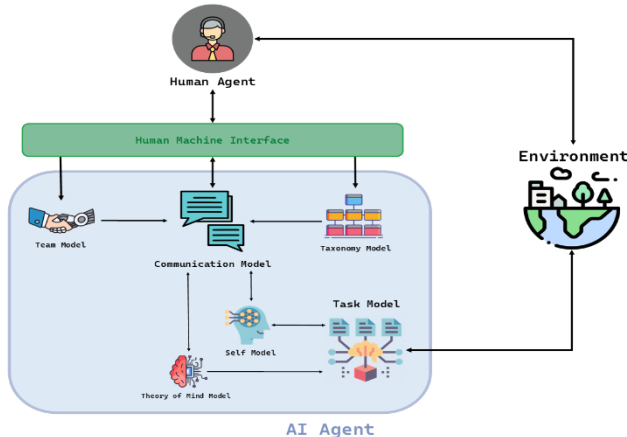
Denzel et al., «Map-to-map translation for SKA mock observations and cosmological simulations». In: Hammers & Nails 2023.

AI for process monitoring



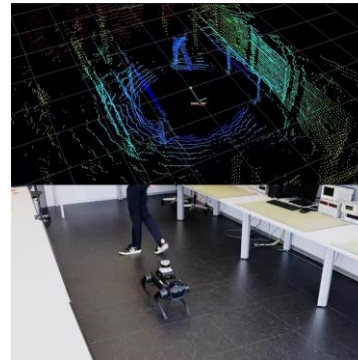
Yan et al., «Automated process monitoring in injection molding via representation learning and setpoint regression». In: SDS'24.

Human-AI teaming & AI ethics



Segessenmann et al., «Assessing Deep Learning: A Work Program for the Humanities in the Age of AI». In: AI and Ethics, 2023.

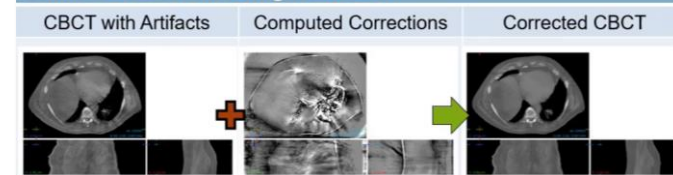
AI for autonomous systems



Pfammatter & Schweizer, «RoboDog III: Building a Vision and/or Sound-Based AI Demonstrator on a Robotic Platform». In: B.Sc. Thesis 2023

AI in medical engineering

Clinical Evaluation using Real CBCT



Amirian et al., «Mitigation of Motion-Induced Artifacts in Cone Beam Computed Tomography using Deep Convolutional Neural Networks». In: Med. Phys., 2023.