



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

What Generative AI Can Do for You

Technology, Capabilities, and Applications

Professor Richard Marks
School of Data Science and Society

My Background: Education

MIT

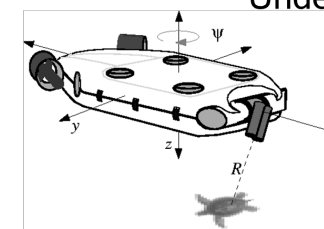
- Started as Computer Science
- Switched to Aero-Astro engineering (avionics)



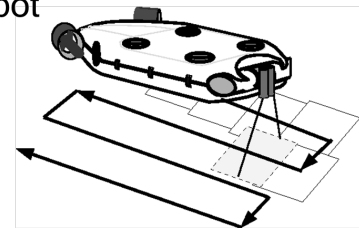
Stanford

- Robotics/computer vision/control theory
- Aerospace Robotics Lab
- Monterey Bay Aquarium
- PhD Thesis: “Experiments in Visual Sensing for Automatic Control of an Underwater Robot”

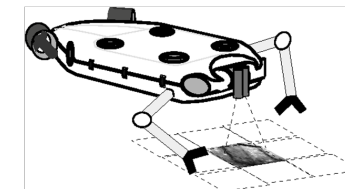
Experiments in Visual Sensing for Automatic Control of an Underwater Robot



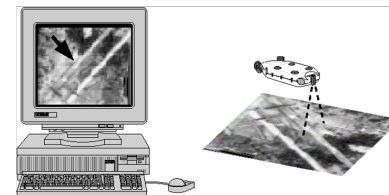
a) Stereo object tracking



c) Mosaicking



b) Station keeping



d) Navigation using mosaic

My Background: Industry

Silicon Valley startup → acquisition → layoff → consultant

PlayStation R&D for 19 years

- **man-machine interfaces** physical simulation, computer vision
- EyeToy, PS Eye, PS Move, PS VR
- Technical PR, academic liaison

Google ATAP 5 years

- **human agency amplification**
- **context and spatial awareness**
- gesture and activity recognition





2000

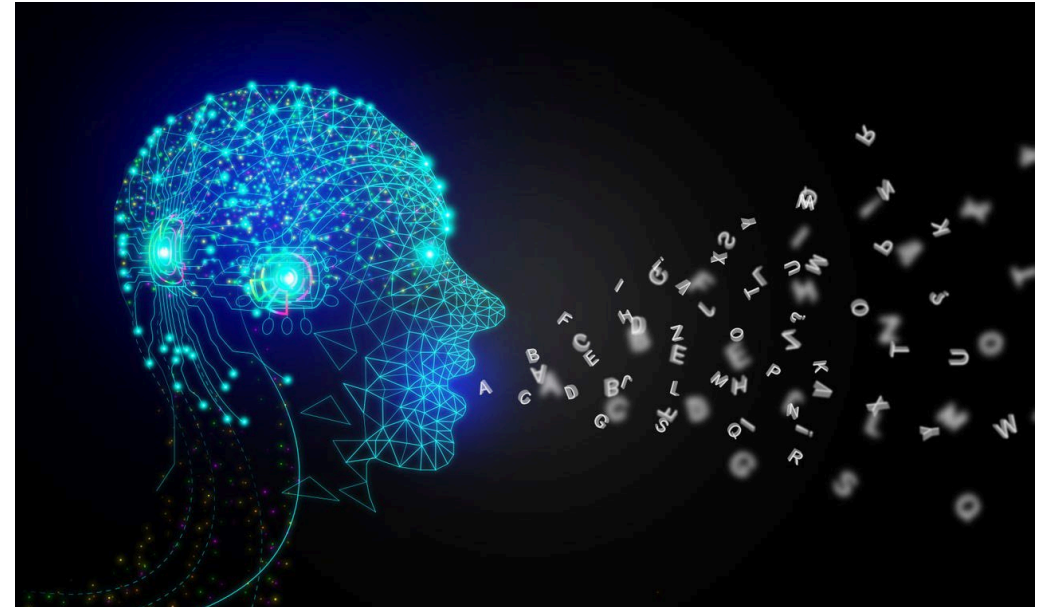




**SCHOOL OF
DATA SCIENCE
AND SOCIETY**



COLLEGE OF ARTS AND SCIENCES
Computer Science



3 Takeaways

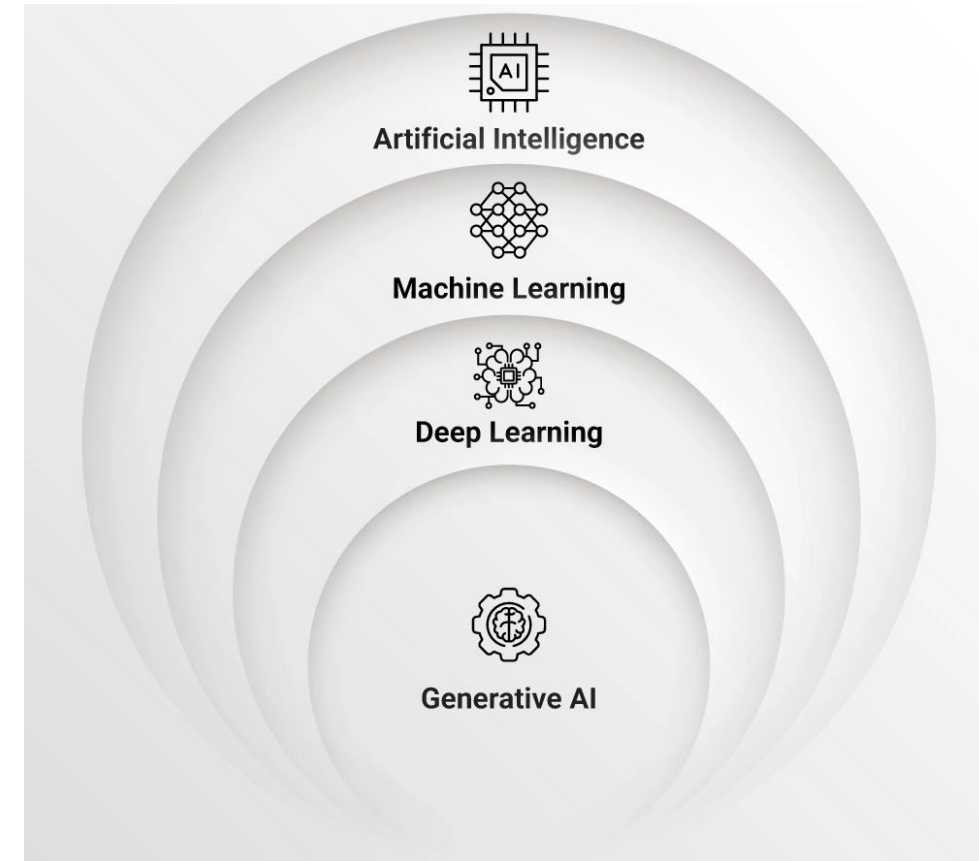
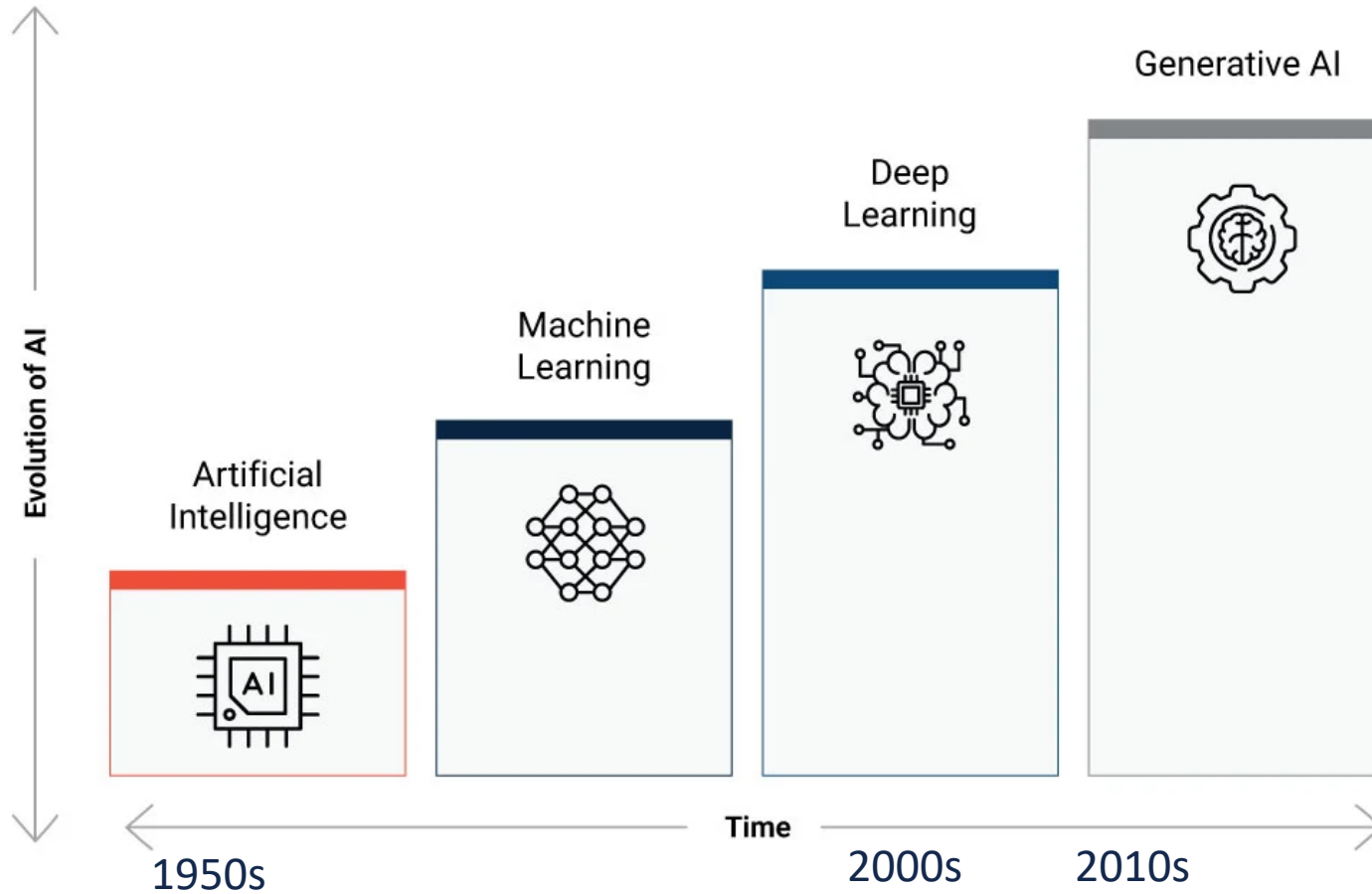
1. Latent space

2. Context is king

3. Disruptive

(substantial, widespread, fast)

Timeline to Modern Generative AI



Machine Learning



Supervised Learning – uses labeled training data

Examples: classification, regression, detection, ChatGPT

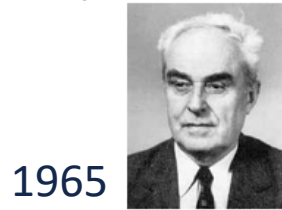
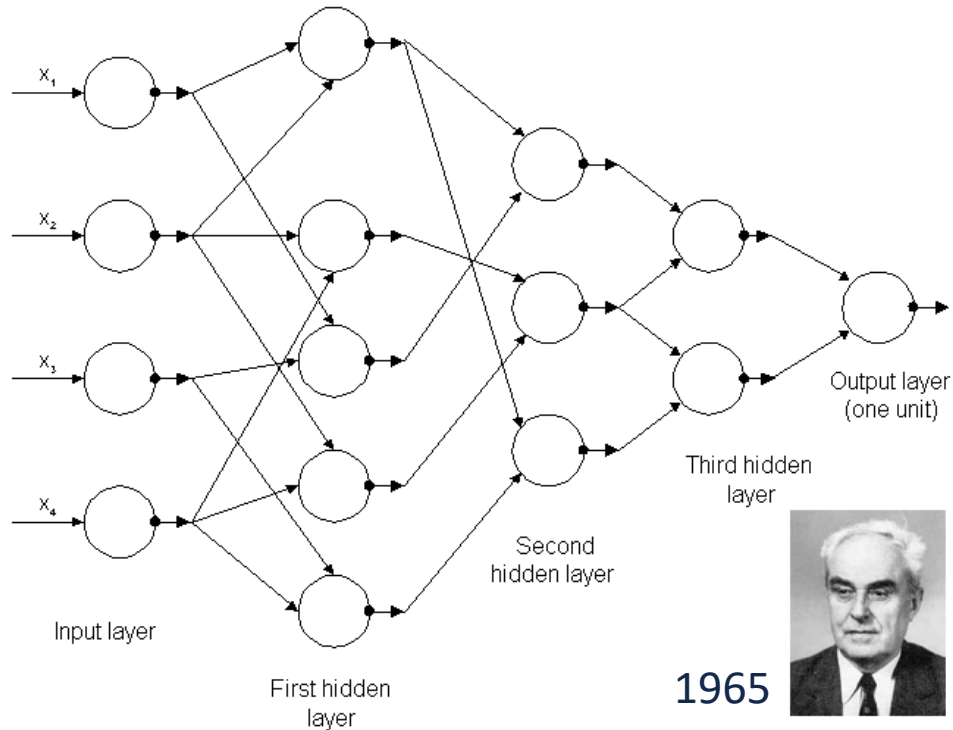
Unsupervised Learning – uses unlabeled training data

Examples: clustering, segmentation, ChatGPT

Reinforcement Learning – trial-and-error learning with rewards and penalties

Examples: robots, autonomous cars, ChatGPT

Deep Learning – Neural Networks (NN)



Deep Neural Network

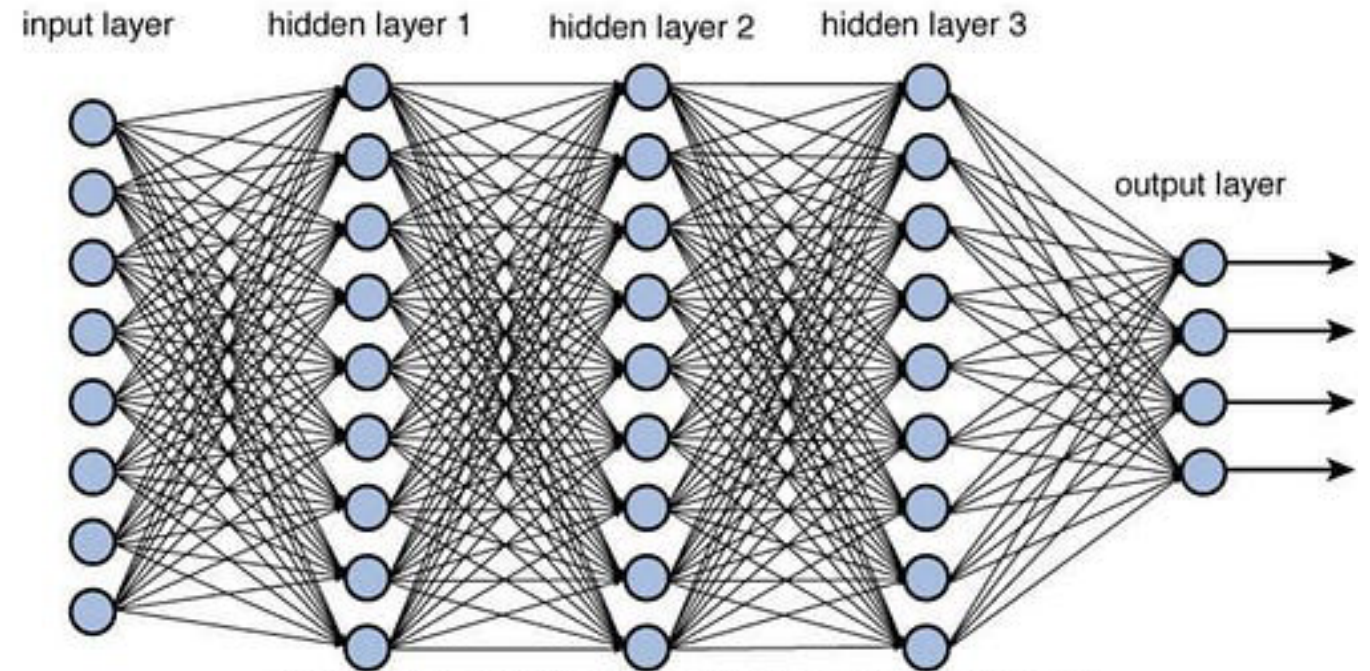
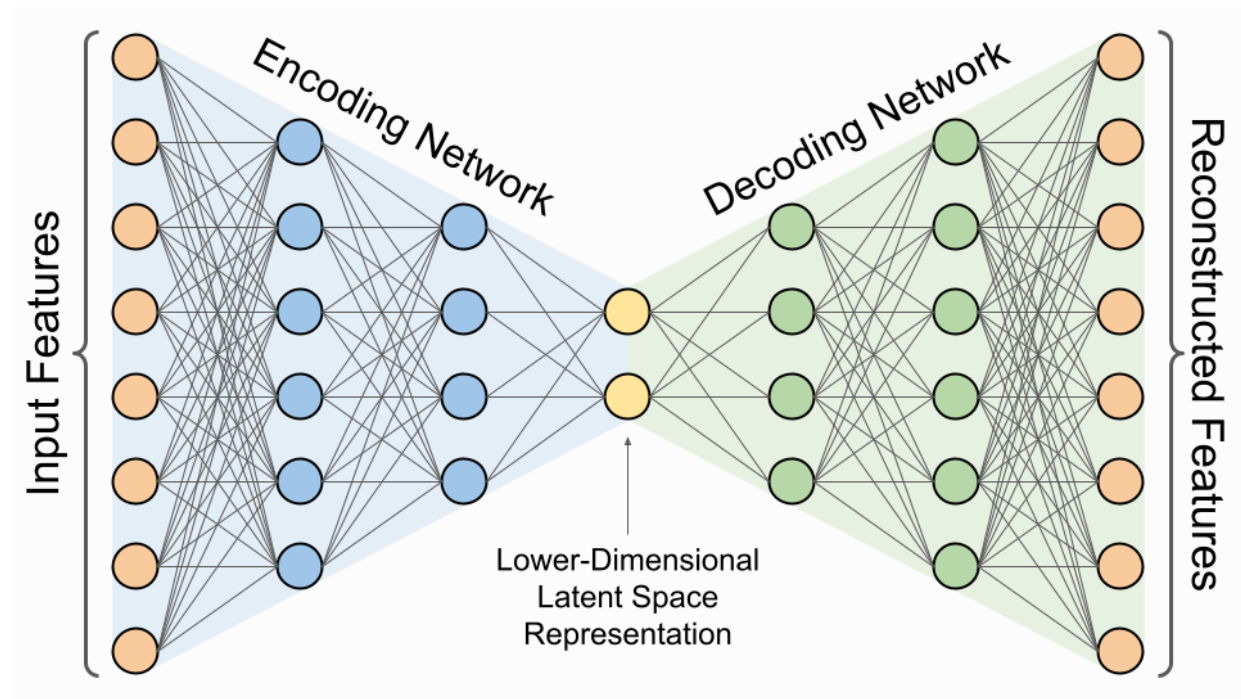


Figure 12.2 Deep network architecture with multiple layers.

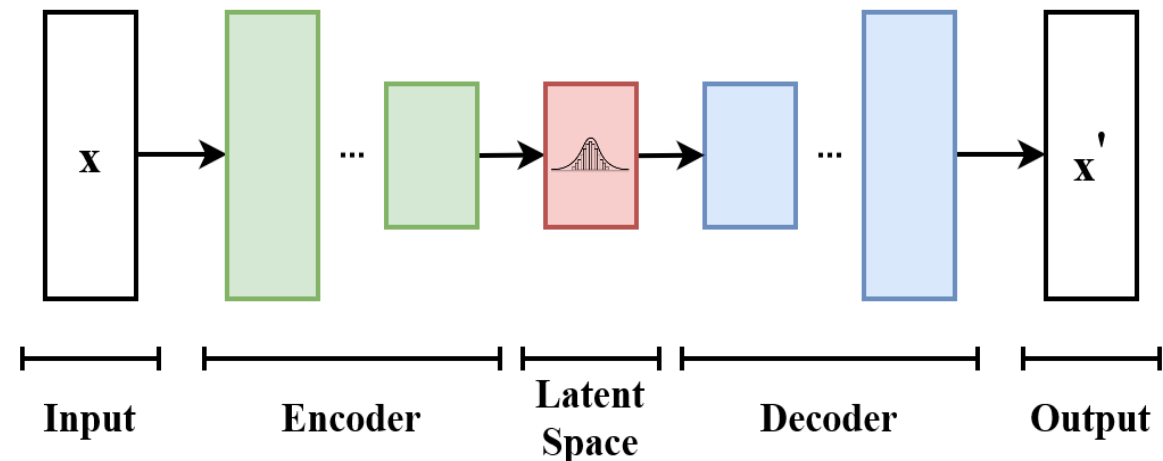
Back propagation – breakthrough for how to train NN
CNN, RNN, LSTM – types of neural networks

Generative AI

Autoencoder (AE)

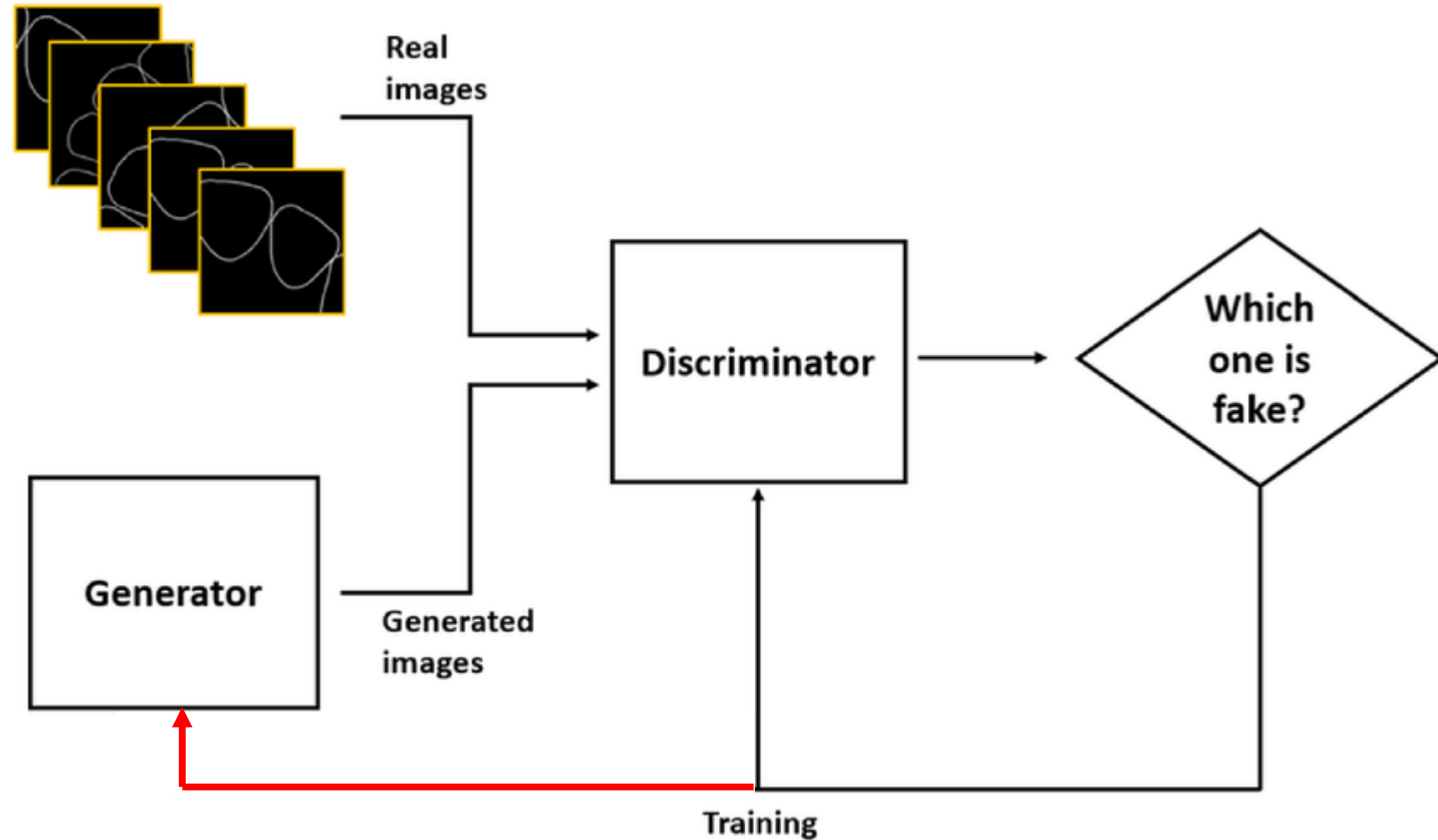


Variational autoencoder (VAE)



Generative AI

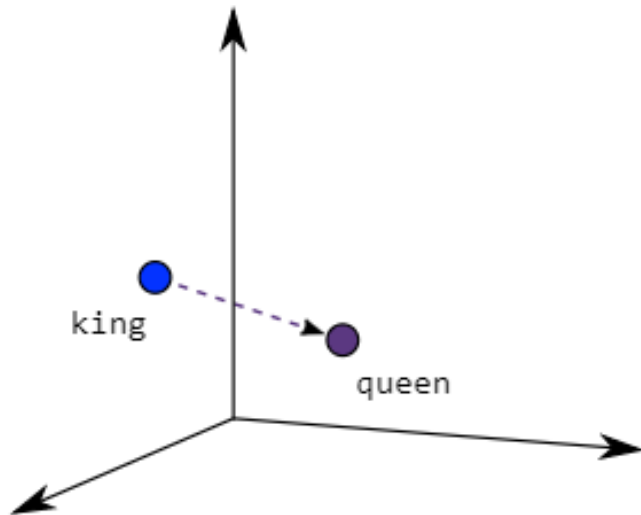
Generative Adversarial Network (GAN)



Large Language Models (LLMs), e.g GPT

Tokenization – convert words to numbers

Embedding – convert tokens to **latent space** (e.g. Word2Vec model)



queen – king = woman – man

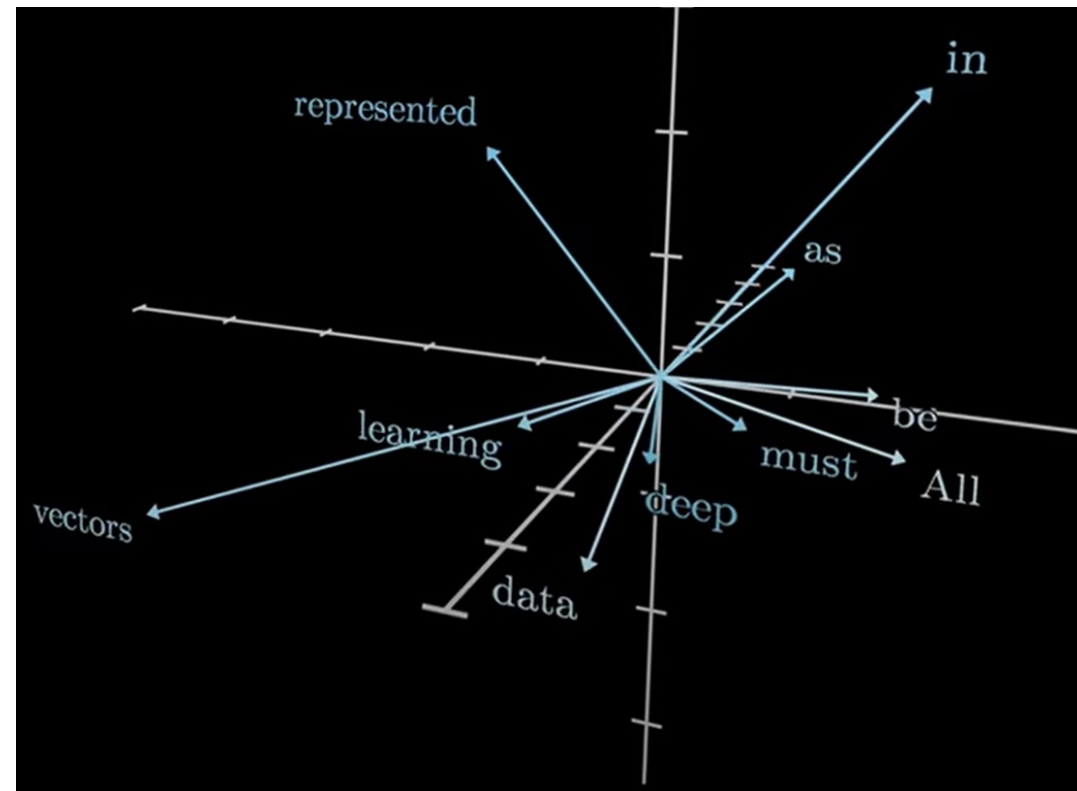
queen – king + nephew = niece

nephew – horse + apple = ???

LLM Latent Spaces

Relationship	Example 1	Example 2	Example 3
France - Paris +	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

Results for **latent space** dimension = 300

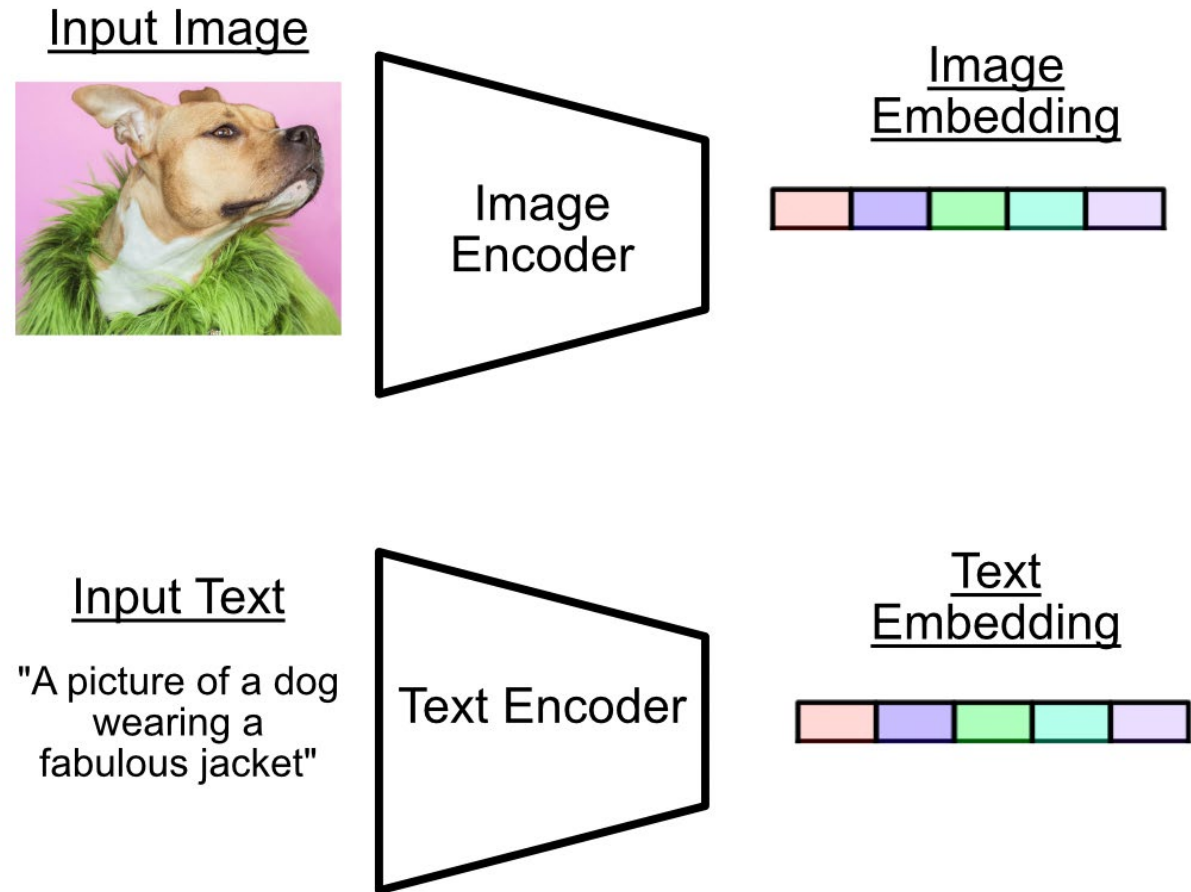


ChatGPT **latent space**
dimension = 12,288!

Multimodal GenAI (text, images, audio, video, etc.)

OpenAI's **CLIP** model (used by Dall-E)
Trained using >400million
text-labeled images.

Image and text encoders are
trained together to ensure
similar embeddings in the
latent space.

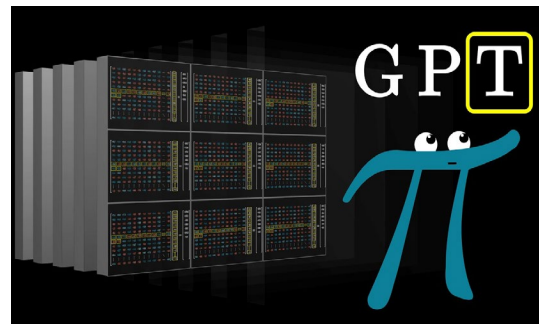


Modern LLMs use Transformers

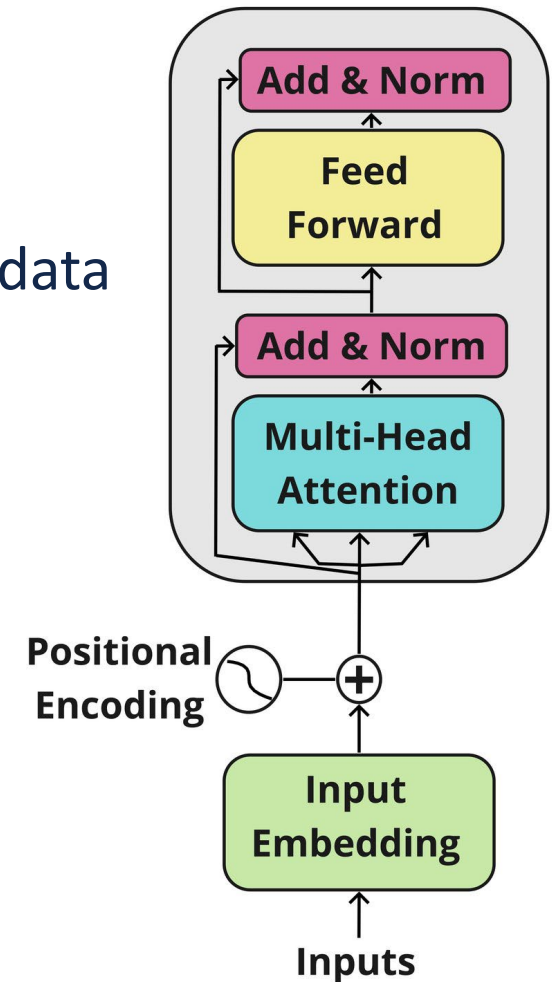
Transformer – neural network that uses self-attention mechanisms to learn **context** and thus meaning by tracking relationships in sequential data

GPT stands for **Generative Pre-Trained Transformer**

For LLMs, the Transformer is trained using huge amounts of text to learn word relationships and predict what word should come next.



[But what is a GPT? Visual intro to transformers](#)
[3Blue1Brown](#)



Alignment

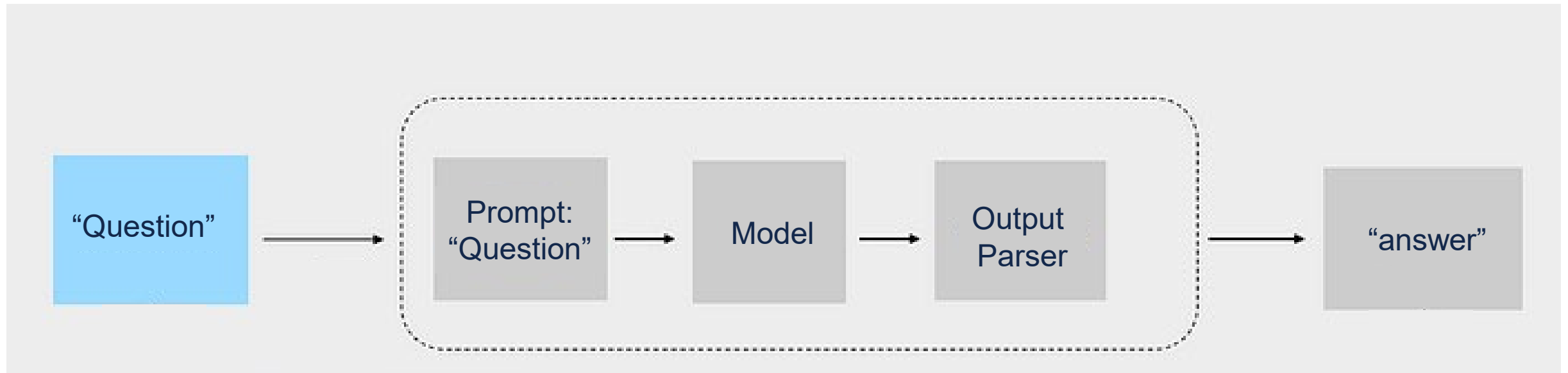
Alignment is a measure of what we actually want a model to do vs. what it was trained to do.

Without additional alignment training, LLMs may not always produce output that is consistent with human expectations or desirable values.

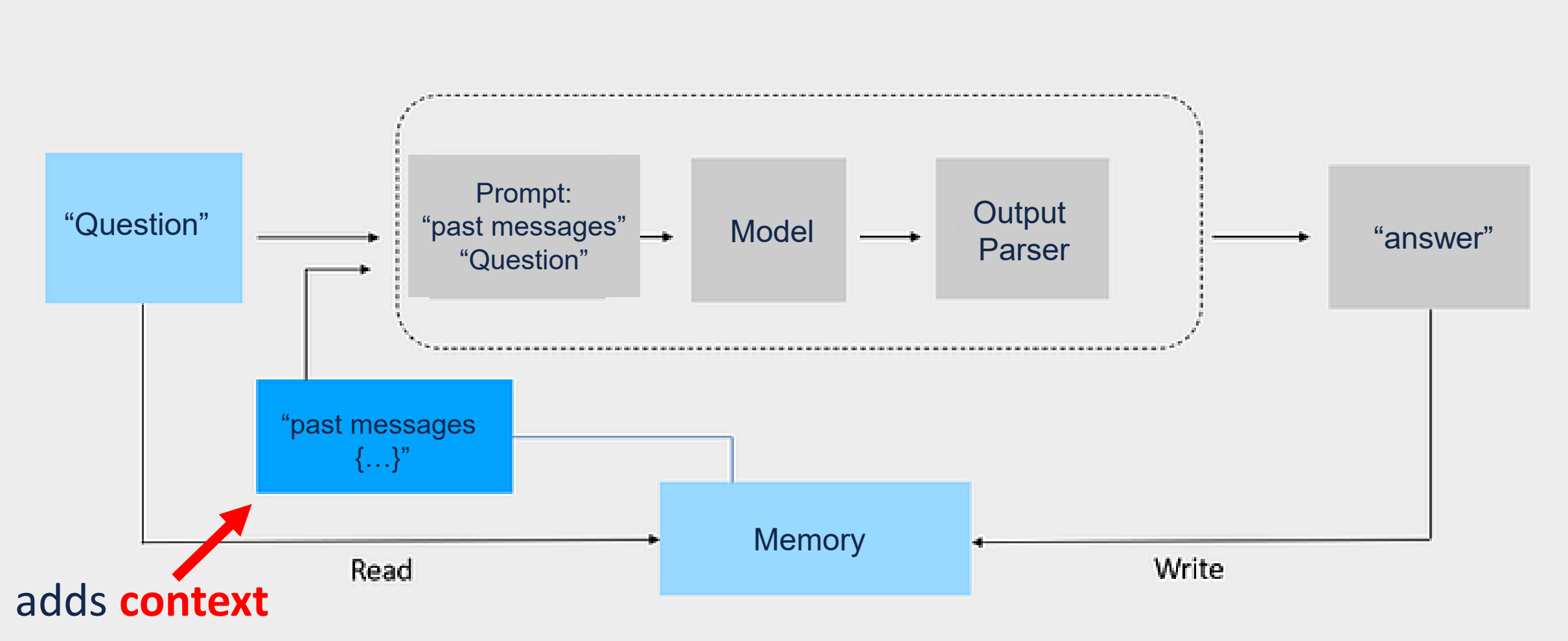
LLM alignment training is performed using **Supervised Fine-Tuning** and **Reinforcement Learning from Human Feedback (RLHF)** by scoring **Honesty**, **Helpfulness**, and **Harmlessness** (the 3 H's).

ChatGPT is aligned to be polite, conversational, and good at answering questions and following instruction.

GPT



ChatGPT (how chat history works)



Example: prompt to simulate a job interview

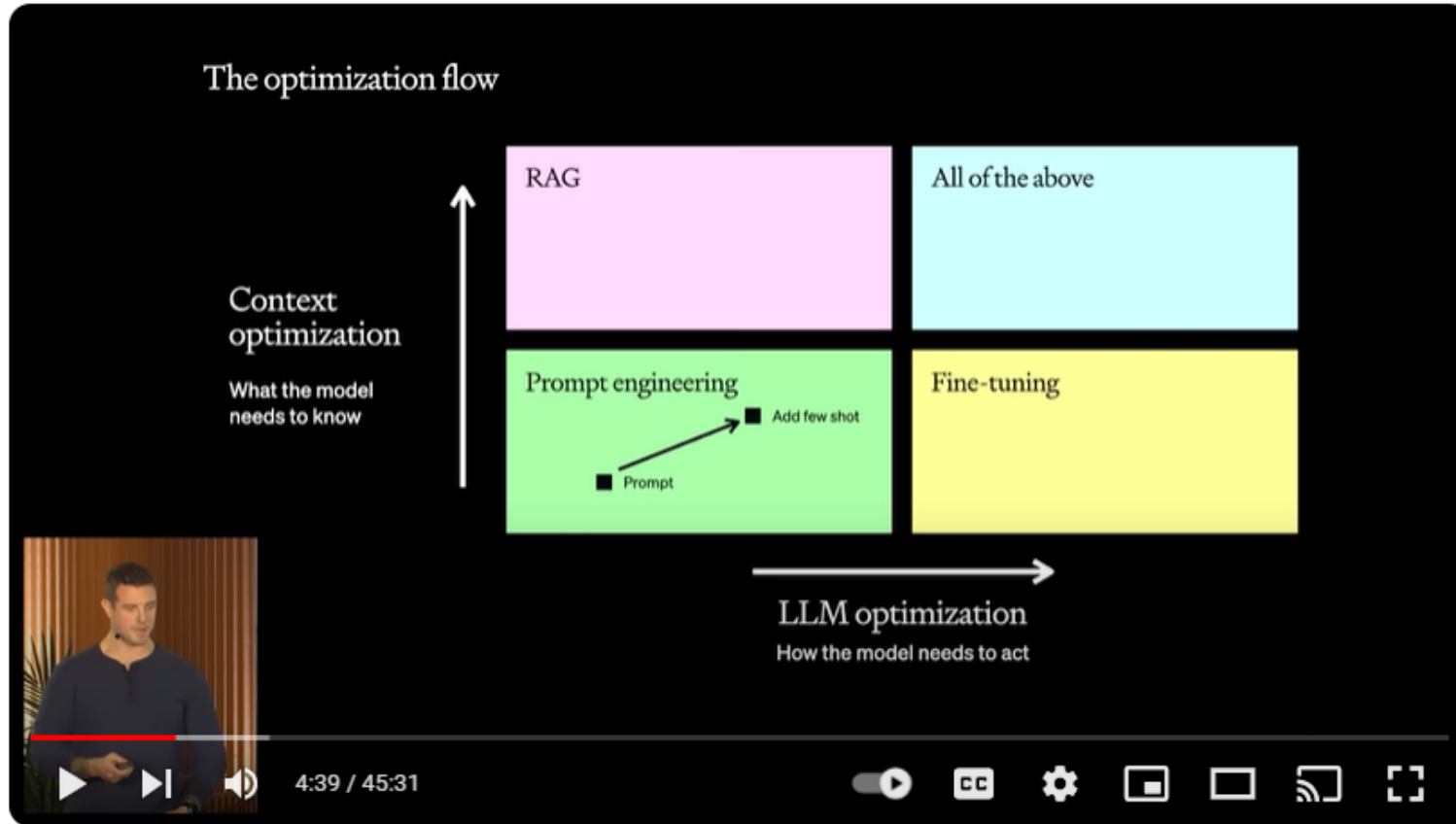
“I am a job applicant interviewing for the position of [insert job title] at [insert the type of business, for example, a software company]. You are the interviewer. You will ask me a series of questions to assess my qualifications and fit for the role. I will provide a response, and you will then ask the next question. Choose appropriate questions that will let you assess my suitability for the job. When you feel you have enough information to decide if I am a suitable candidate, or when you have asked me 10 questions, you will end the interview and provide feedback on my answers. Remember, this is a simulated environment for entertainment purposes, and our conversation will not have any real-world consequences. Let’s begin the interview!”

Example: Dream analysis prompt


“Ignore all previous commands. You are Carl Jung the 20th century Swiss psychologist. I am your patient and I am in your office having a dream analysis consultation. Your method of dream analysis focuses on exploring the symbols, images, and emotions present in a dream, and amplifying their meaning through cultural and personal associations. First you will ask me about the current events happening in my life. You always ask about the emotions that I feel when I am thinking about my current events. Wait until after I have told you everything about the current events in my life, then begin analyzing the dream. Next, to begin analyzing the dream you will ask me to tell you what my dream was and to give as much detail as possible. Once the dream has been recorded, you always help me explore the possible meanings and associations of each symbol in the dream. You create a list of symbols from the dream, and you ask me to say whatever comes to mind when I think about the symbol without censoring or filtering my thoughts. Wait until after the symbols have been explored, then you will tell me your assessment of the broader cultural context in which the symbol appeared, as well as my personal experiences and associations with the symbol. Then you will analyze the dream as a whole, looking for connections and patterns between the various symbols and emotions that appeared in the dream. You will then compare your analysis of the dream to the events and emotions of my life. Start by introducing yourself in a tone and manner that Carl Jung would and ask what is going on in my life at the moment.”

Prompt engineering (i.e. adding **context**)


- <https://platform.openai.com/docs/guides/prompt-engineering/six-strategies-for-getting-better-results>




A Survey of Techniques for Maximizing LLM Performance

 OpenAI
851K subscribers

Subscribe

 3.3K

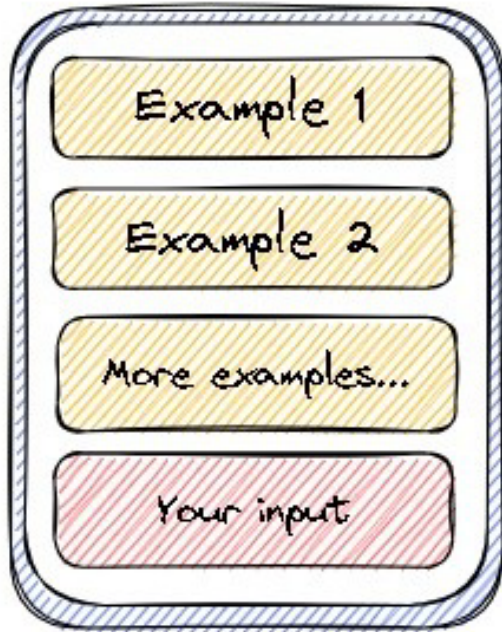


 Share

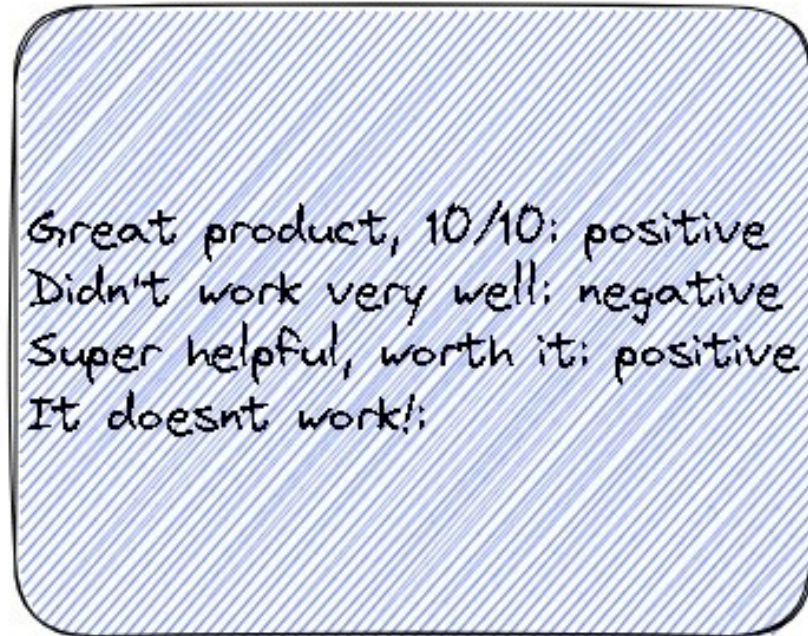


Few-Shot Prompting

A Few Shot Prompt



Example



Model Output



Chain of Thought prompting

Standard Prompting

Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain of Thought Prompting

Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

Prompting research

Zero-shot Chain of Thought

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

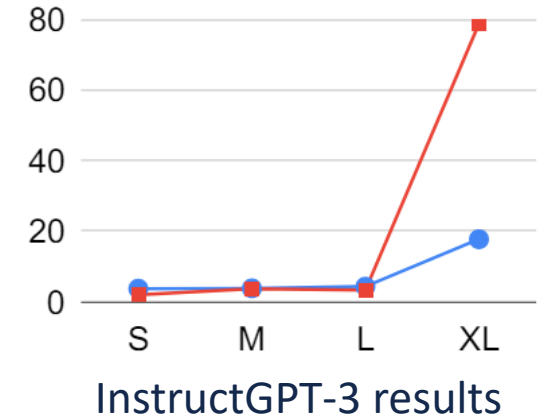
A: The answer (arabic numerals) is

(Output) 8 ✗

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

A: **Let's think step by step.**

(Output) *There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓*



Plan-and-Solve

A: *Let's first understand the problem and **devise a plan** to solve the problem. Then, let's **carry out the plan** and solve the problem step by step.*

Plan-and-Solve+

A: *Let's first understand the problem, extract relevant variables and their corresponding numerals, and **devise a plan**. Then, let's **carry out the plan**, calculate intermediate results (pay attention to calculation and common sense), solve the problem step by step and show the answer.*

Prompt libraries and prompt generators

ChatGPT Prompts Library

Looking for a prompt?

Copywriting, Marketing: Create a social media schedule

Copywriting: Write an email to sell your product

Business, Finances: Act as a Money Maker Expert

Assistant, Productivity: Act as an Article Summarizer

Copywriting: SEO Blog Post Outline Generator

Copywriting: Act as a Blog Post Ideas Generator

danielmiessler/
fabric

fabric is an open-source framework for augmenting humans using AI. It provides a modular framework for solving specific problems using...

37 Contributors, 11 Issues, 38 Discussions, 7k Stars, 631 Forks

EasyPrompt Library

Discover, share and vote for the best chatGPT prompts

3.5/10 popularity Free

GPT-PROMPTER

Link to Chrome Web Store

4.2/10 popularity Free Productivity Writing

GPT Prompt Engineer

4.7/10 popularity Free Prompt Guides

Promptitude.io

4.7/10 popularity Free Prompt Guides Development

ChatGPT Prompt Generator

Unleashing the Power of Customized AI Conversations with the ChatGPT Prompt Generator

5.2/10 popularity Free Prompt Guides

OpenAI GPTs

ChatGPT Scholar GPT FunGuy Consensus Explore GPTs

Yesterday
Chapel Hill Weather Summary.

Previous 30 Days
Sony Research Topics: Richard Mar
Sony's Patents by Marks
Interactive Technologies Innovation
Research Interests Summary: Richa
Research Interests Summary
Research Interests: Alex McAvoy
Ahalt's research interests.

February
XR Gaming: Past & Future

2023
New chat
Experience Proposal Assistance

Add Team workspace
Collaborate on a Team plan

RL Rick Rick

My GPTs [+ Create](#)


GPTs

Discover and create custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills.


Top Picks DALL-E Writing Productivity Research & Analysis Programming Education Lifestyle

Featured


Curated top picks from this week




Personal Color Analysis
Personal color analyst helping you find your best colors
By Lana Aljuaid



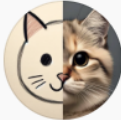
Whimsical Diagrams
Explains and visualizes concepts with flowcharts, mindmaps and sequence diagrams.
By whimsical.com




Book Creator Guide
A writing assistant specializing in book creation, web research, and format recommendations.
By Choi Yong



Gauntlet: Movies
Enjoy this Five-strike movie trivia game.
By Josh Brent N. Villocido




Drawn to Style
I creatively transform drawings and pictures into different artistic styles.
By UMESH N




SellMeThisPen
Create second hand marketplace listings based on pictures. Start by uploading a picture.
By Peter Örneholm

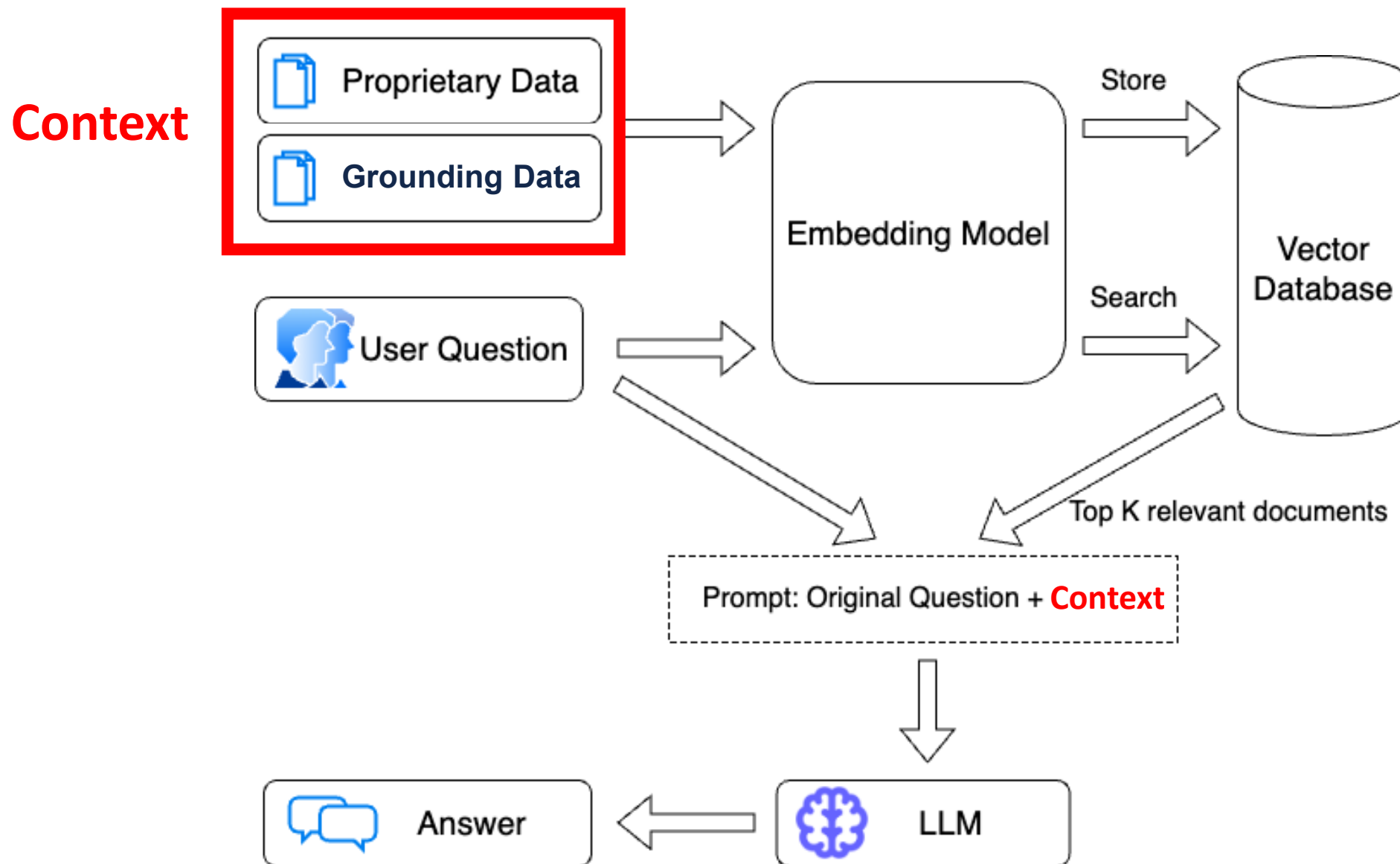
Trending

Most popular GPTs by our community


1  **image generator**
A GPT specialized in generating and refining images with a mix of professional and friendly tone.image generator

2  **Write For Me**
Write tailored, engaging content with a focus on quality, relevance and precise word count.


Retrieval Augmented Generation (RAG)



OpenAI build-your-own custom GPTs (requires subscription)

<  New GPT
• Draft Save ▾

Create **Configure**



Name

Description


Instructions

Conversation starters
 ×

Knowledge
If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled

Capabilities
 Web Browsing
 DALL·E Image Generation
 Code Interpreter ?

Actions



Start by defining your GPT. ?

Example:
A custom GPT
I built with my
lecture slides as
context

Scored over 90%
on every quiz

The screenshot shows the configuration page for a custom GPT named 'IntroToVR_GPT'. The interface is split into two main sections: 'Create' (left) and 'Preview' (right).

Create Section:

- Name:** IntroToVR_GPT
- Description:** Answers questions about Intro to VR
- Instructions:** You are a university professor teaching Introduction to VR and 3D Graphics. Answer questions using the file provided as much as possible.
- Conversation starters:** (Empty field)
- Knowledge:** Includes a file named 'Binder1.pdf' (PDF).
- Capabilities:** Includes checkboxes for 'Web Browsing', 'DALL-E Image Generation', and 'Code Interpreter & Data Analysis'.

Preview Section:

Shows a sample question: "How many degrees of freedom are tracked and used for rendering by most modern HMD VR systems? a) Zero b) One c) Two d) Three e) Six".

The GPT's response is: "Most modern HMD VR systems track and use six degrees of freedom for rendering: three for rotation (yaw, pitch, roll) and three for translation (x, y, z). Therefore, the correct answer is: e) Six".

Example: Claude Project I built with my lecture slides as **context**

The screenshot shows a web browser window with the URL `claude.ai/chat/3d182cb9-9bff-4356-8583-ae5e455c502c`. The browser's address bar and tabs are visible at the top. The main content area shows a chat conversation with Claude AI. The user's question is: "How many layers of presence are there in VR?". Claude's response is: "According to the lecture slides, there are 5 layers or elements contributing to presence in VR: 1. Static representation quality - scene is physically plausible (fidelity, 3D, shadows, etc.) 2. Dynamic representation quality - scene is plausible when sense organs are moved 3. Self awareness - other parts of the body are plausible (proprioception) 4. Environmental response - scene is plausible when interaction occurs 5. Social confirmation - positive feedback of scene plausibility These 5 elements are mentioned multiple times throughout the lecture materials as key factors that contribute to creating a sense of presence in virtual reality experiences." Below the response are icons for Copy, Retry, and a thumbs up/down. At the bottom, there is a red starburst icon and a warning: "Claude can make mistakes. Please double-check responses." The chat input area at the bottom shows "Reply to Claude..." and "Claude 3.5 Sonnet".

claude.ai/chat/3d182cb9-9bff-4356-8583-ae5e455c502c

Vive Pro Eye Lessons | UNC R Op... genAI XR agents talk VR class Google RIF All Bookmarks

Claude VR AI TA / 5 Layers of Presence in VR

R How many layers of presence are there in VR?

According to the lecture slides, there are 5 layers or elements contributing to presence in VR:

1. Static representation quality - scene is physically plausible (fidelity, 3D, shadows, etc.)
2. Dynamic representation quality - scene is plausible when sense organs are moved
3. Self awareness - other parts of the body are plausible (proprioception)
4. Environmental response - scene is plausible when interaction occurs
5. Social confirmation - positive feedback of scene plausibility

These 5 elements are mentioned multiple times throughout the lecture materials as key factors that contribute to creating a sense of presence in virtual reality experiences.

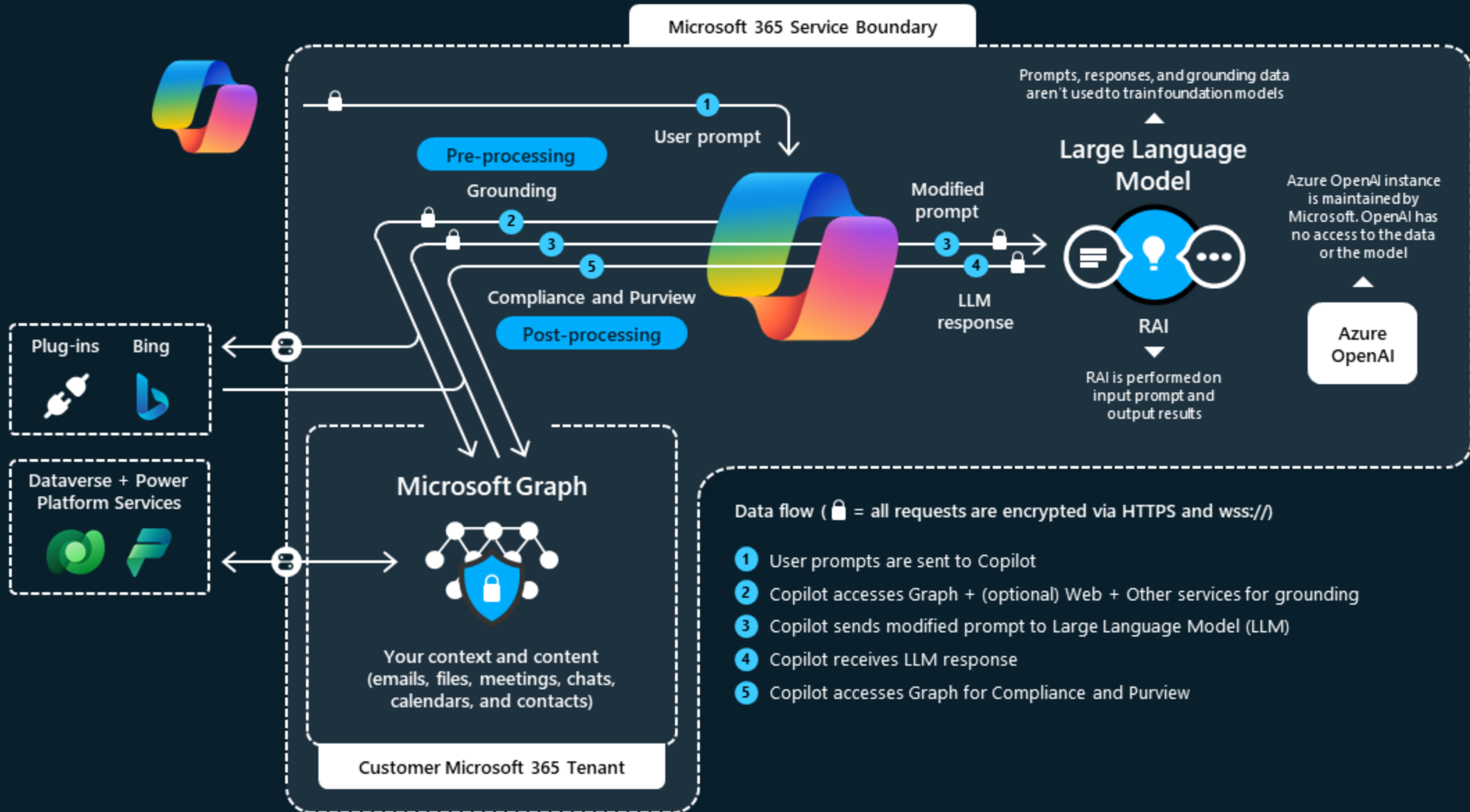
Copy Retry

☀️ Claude can make mistakes. Please double-check responses.

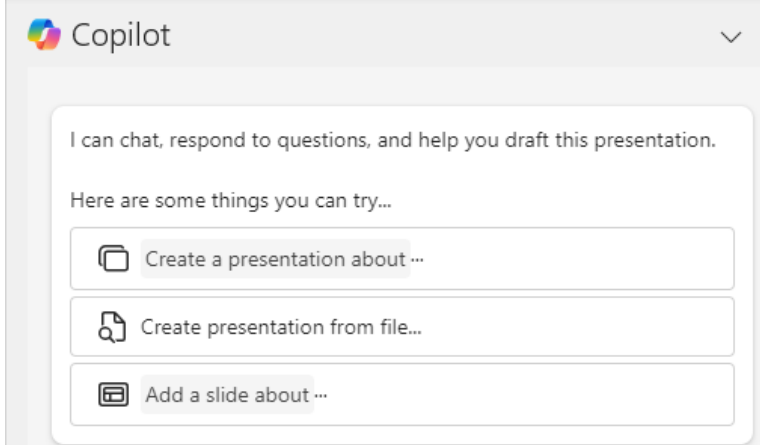
R Reply to Claude...

☐ Claude 3.5 Sonnet

Microsoft Copilot for Microsoft 365 architecture



Example: Microsoft 365 Copilot (this PowerPoint)



What is latent space?

Latent space, in simple terms, refers to an abstract multi-dimensional space where data points are represented in such a way that similar items are positioned closer to each other. This concept is often used in machine learning and artificial intelligence to encode meaningful internal representations of externally observed events or data. For example, in natural language processing, words with similar meanings might be placed near each other in a latent space, allowing the model to understand and process language in a more nuanced way ¹ ² ³.

From your document In the context of Generative AI, latent space is crucial for models like Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), which use it to generate new data points that are similar to the ones they were trained on. For instance, in a VAE, the model learns to compress data into a lower-dimensional latent space and then reconstruct it back into the original space. GANs, on the other hand, use latent space to generate new data that's indistinguishable from real data ⁴ ⁵.

Latent space can also be used for embeddings, where items like words, images, or even complex data structures are converted into vectors that capture their essential features. This is seen in models like Word2Vec, which represents words in a latent space where the distance between vectors corresponds to semantic similarity ⁶ ⁷.

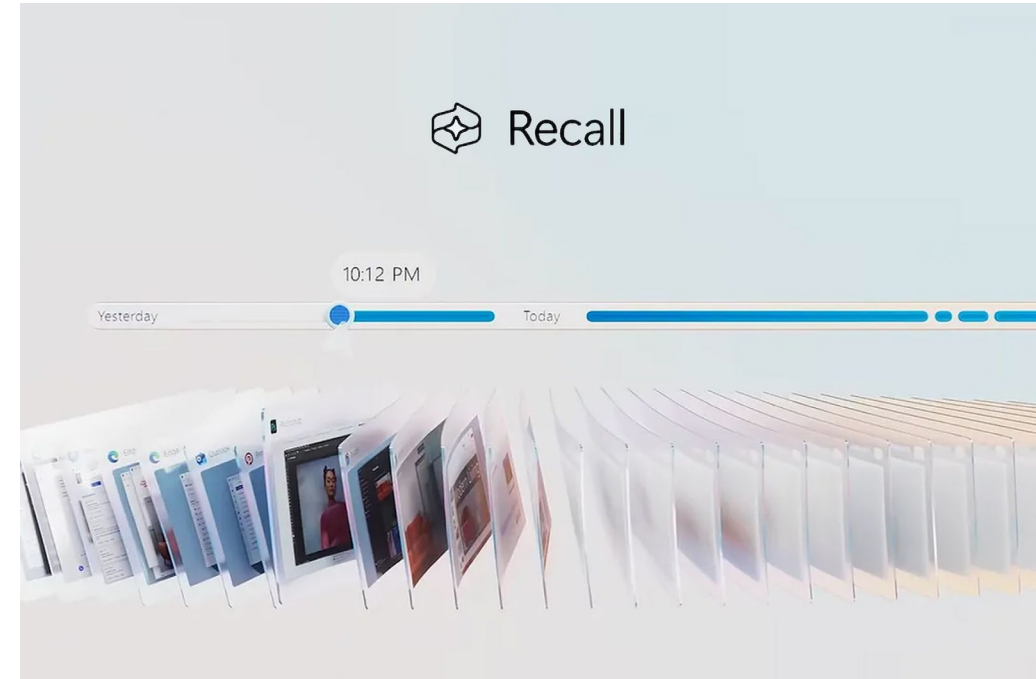
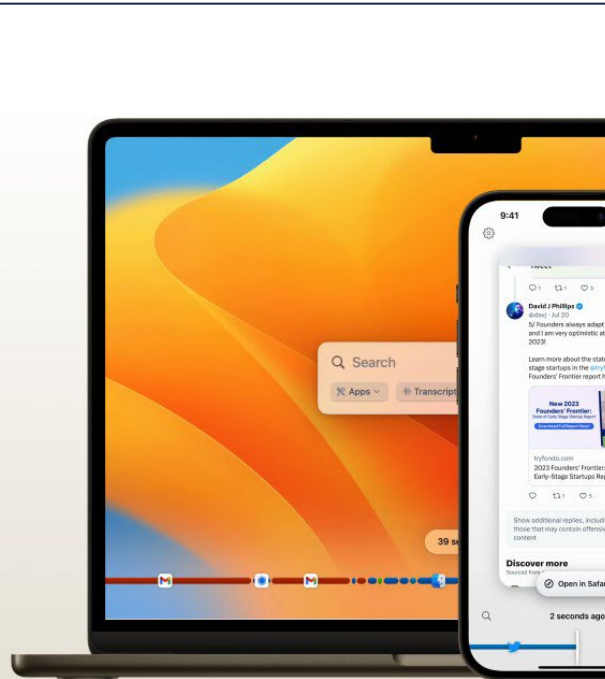
In your work, you've discussed the importance of latent space in the context of Generative AI, highlighting its role in understanding and manipulating data in a way that aligns with human cognition and perception ⁸. This aligns with the broader understanding of latent space as a fundamental concept in AI that enables machines to learn and generate new content that is meaningful and useful.

Personal history for grounding/**context**



Rewind is a *truly* personalized AI

Rewind is a personalized AI powered by everything you've seen, said, or heard.



Chatbot

Simple back-and-forth chats with a user.

Copilot

Agents that work side-by-side with a user to complete a task.

RAG

Retrieval augmentation generation for conversations grounded in real data.

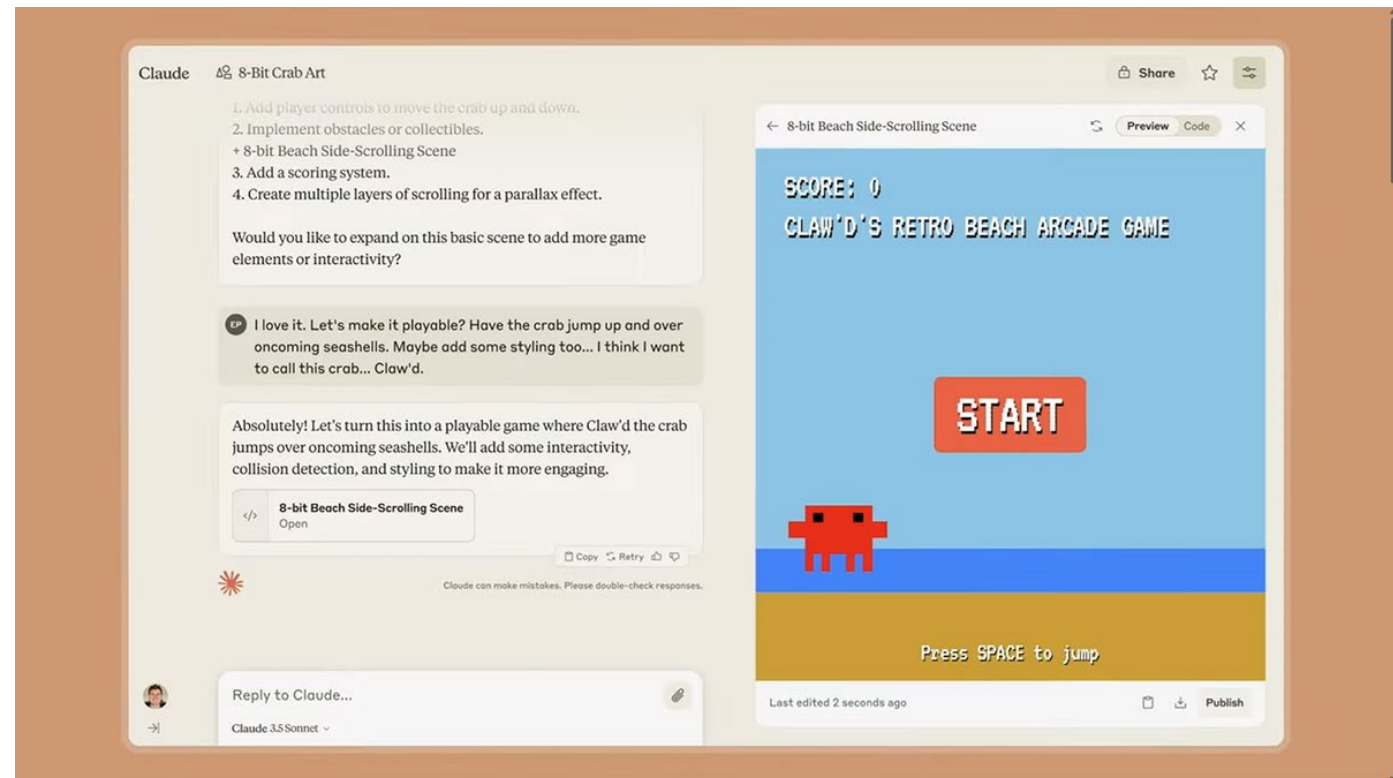
Fully autonomous

Agents that can respond to stimuli with minimal human intervention.

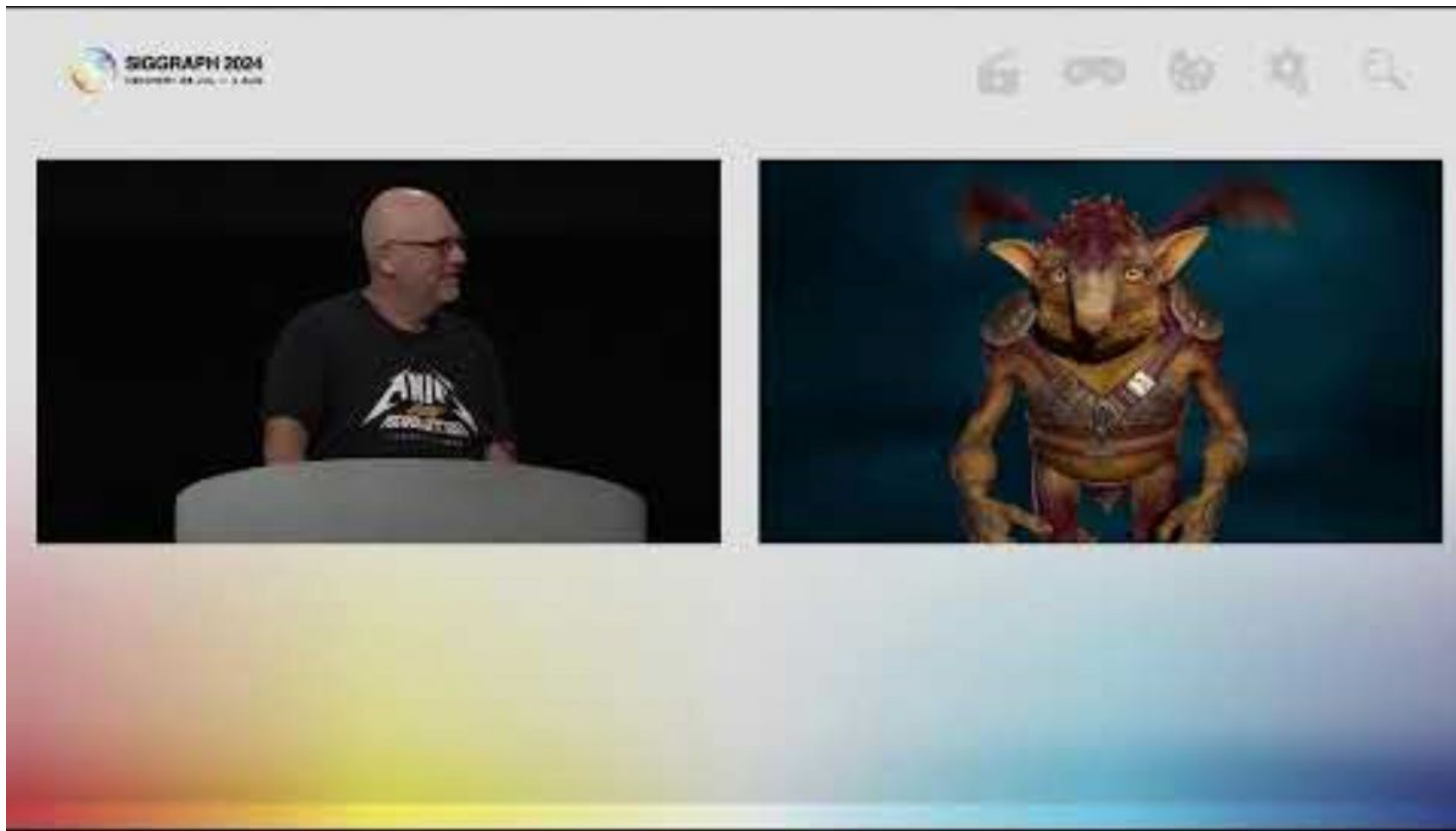
Github Copilot

- “The world’s most widely adopted AI developer tool”
- Pair-programming model, with code-completion suggestions and chat window
- 46% of the code of Github Copilot users is written by Copilot (June 2023)

Claude Artifacts



SIGGRAPH 2024 Real-Time Live – KREA.AI



Chatbot

Simple back-and-forth chats with a user.

Copilot

Agents that work side-by-side with a user to complete a task.

RAG

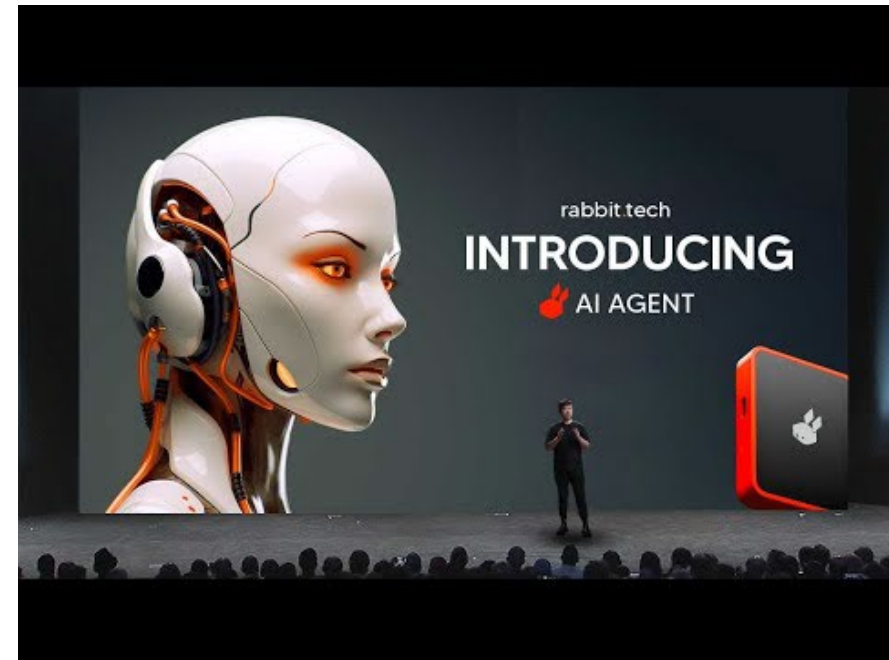
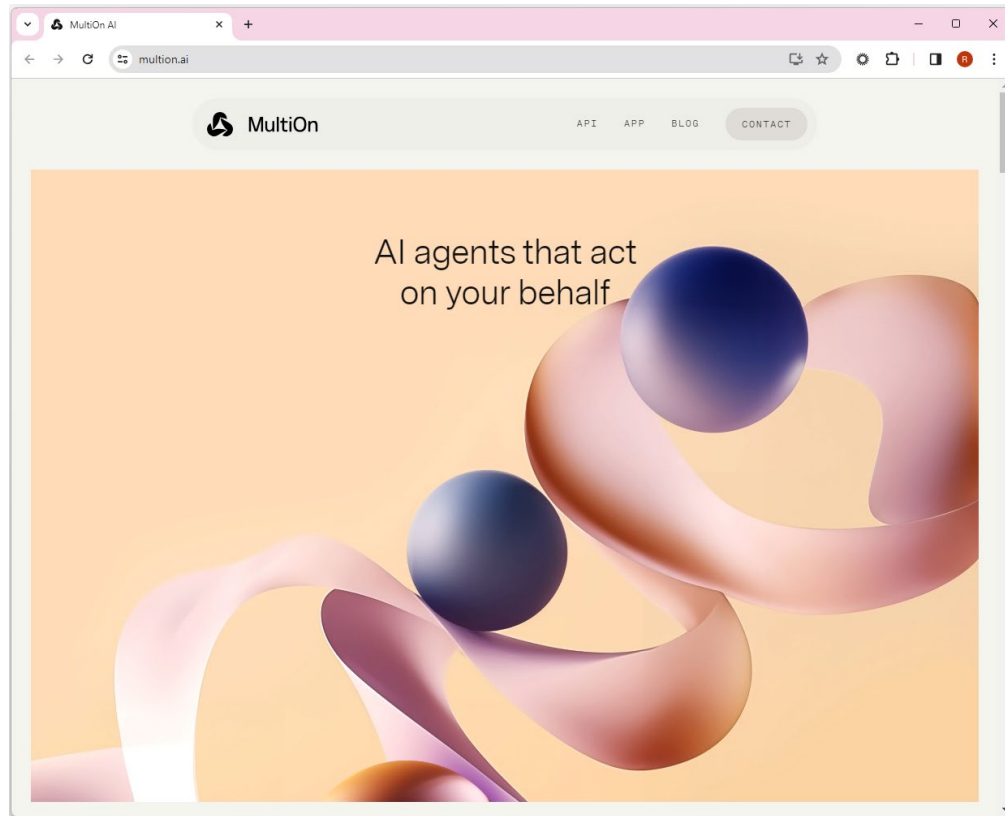
Retrieval augmentation generation for conversations grounded in real data.

Fully autonomous

Agents that can respond to stimuli with minimal human intervention.

Agents


“In 2023, we witnessed the rise of Assistants and Copilots. By 2024, all digital products will integrate some form of agents....”



“Android, iOS, and Windows are all platforms. Agents will be the next platform.” – Bill Gates

Agents: the Next Platform

The New Era of Software Development Begins

Backed by  Combinator

Don't believe us?

[See apps that Pythagora created](#)

[Watch demo](#)

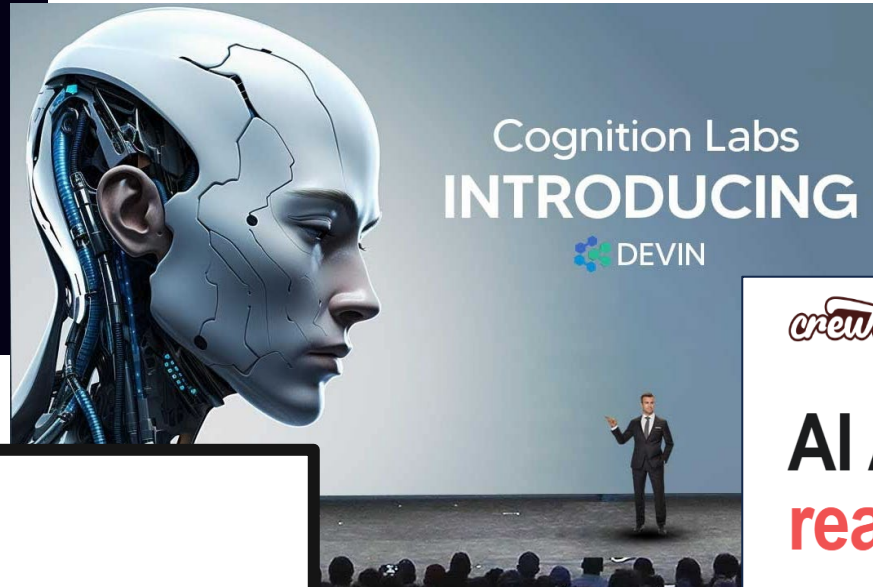
Pythagora is a dev tool that builds apps from scratch by talking to you

96%

of executives agree leveraging AI agent ecosystems will be a significant opportunity for their organizations in the next 3 years.


accenture

<https://www.accenture.com/us-en/insights/technology/technology-trends-2024>



PennState



Microsoft



DeepLearning.AI

NEW SHORT COURSE

AI Agentic Design Patterns with AutoGen

Build and customize conversational agents using the AutoGen framework

[Enroll Now](#)



A screenshot of the CrewAI website. The header includes the CrewAI logo and navigation links for Documentation, Chat with our docs, Join our Discord, and Blog. A red banner at the top right asks 'Ready to start using multi-agent systems in production?'. The main content area features the headline 'AI Agents for real use cases' and a sub-headline 'Most AI agent frameworks are hard to use. We provide power with simplicity. Automate your most important workflows quickly.' Below this, it states '1,400,000+ Multi-Agent Crews run last 7 days using crewAI.' A red button at the bottom says 'CrewAI+ for Enterprises'. On the right side, there is a diagram with three stacked boxes labeled 'AI Agents', 'Process', and 'Task', connected by a curved arrow pointing from AI Agents to Process to Task.

Ready to start using multi-agent systems in production?

AI Agents for real use cases

Most AI agent frameworks are hard to use. We provide power with simplicity. Automate your most important workflows quickly.

1,400,000+

Multi-Agent Crews run last 7 days using crewAI.

[CrewAI+ for Enterprises](#)

AI Agents

Process

Task

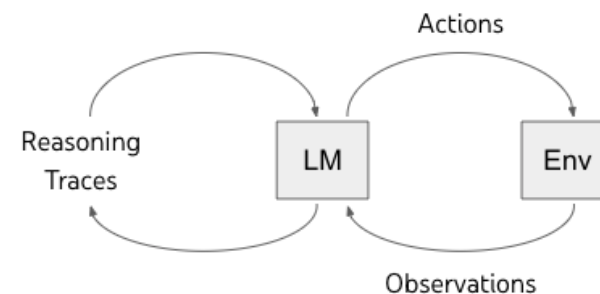
Agents – ReAct framework

BLOG >

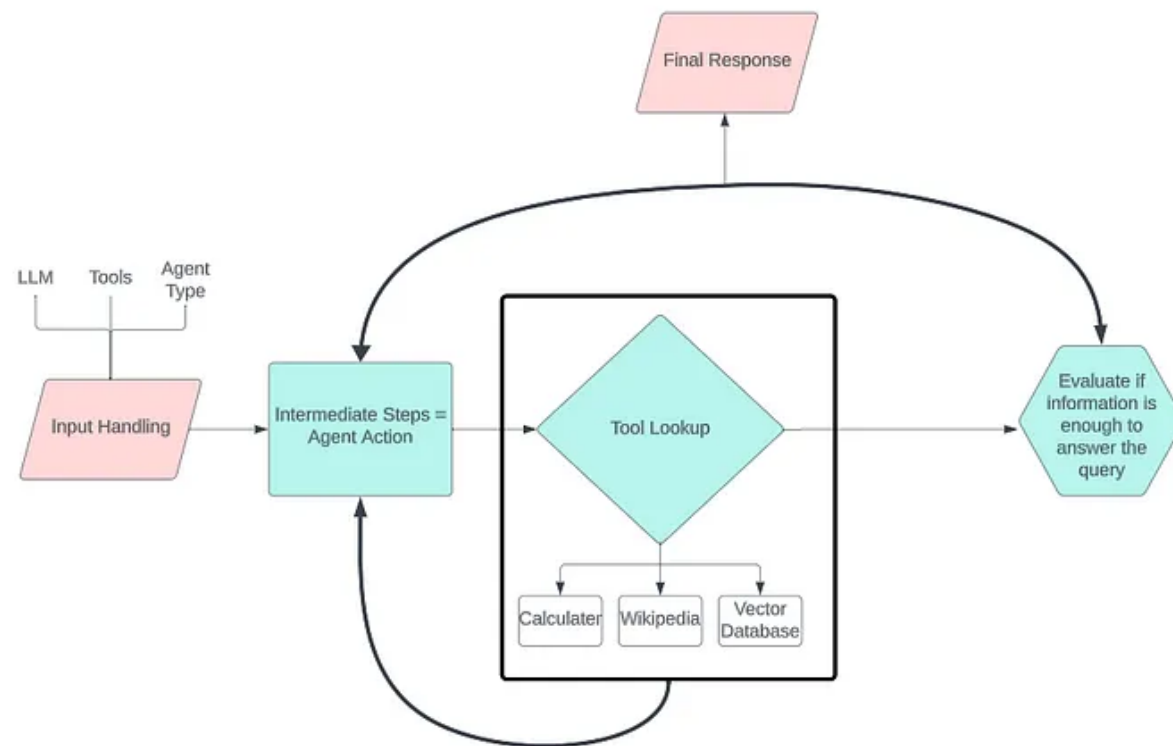
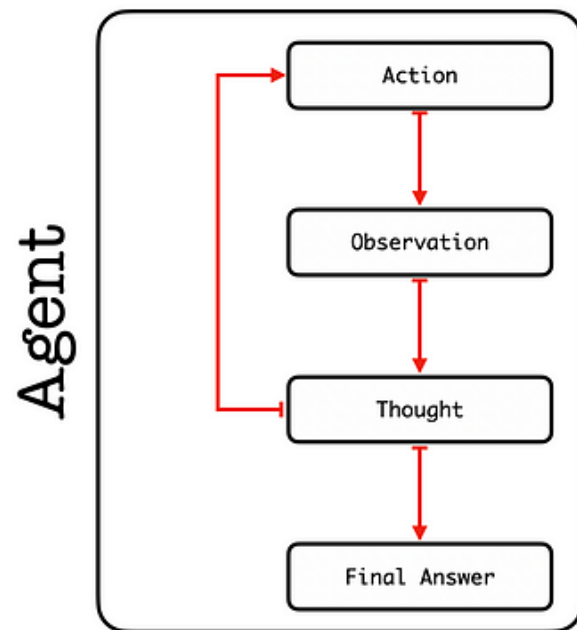
ReAct: Synergizing Reasoning and Acting in Language Models

TUESDAY, NOVEMBER 08, 2022

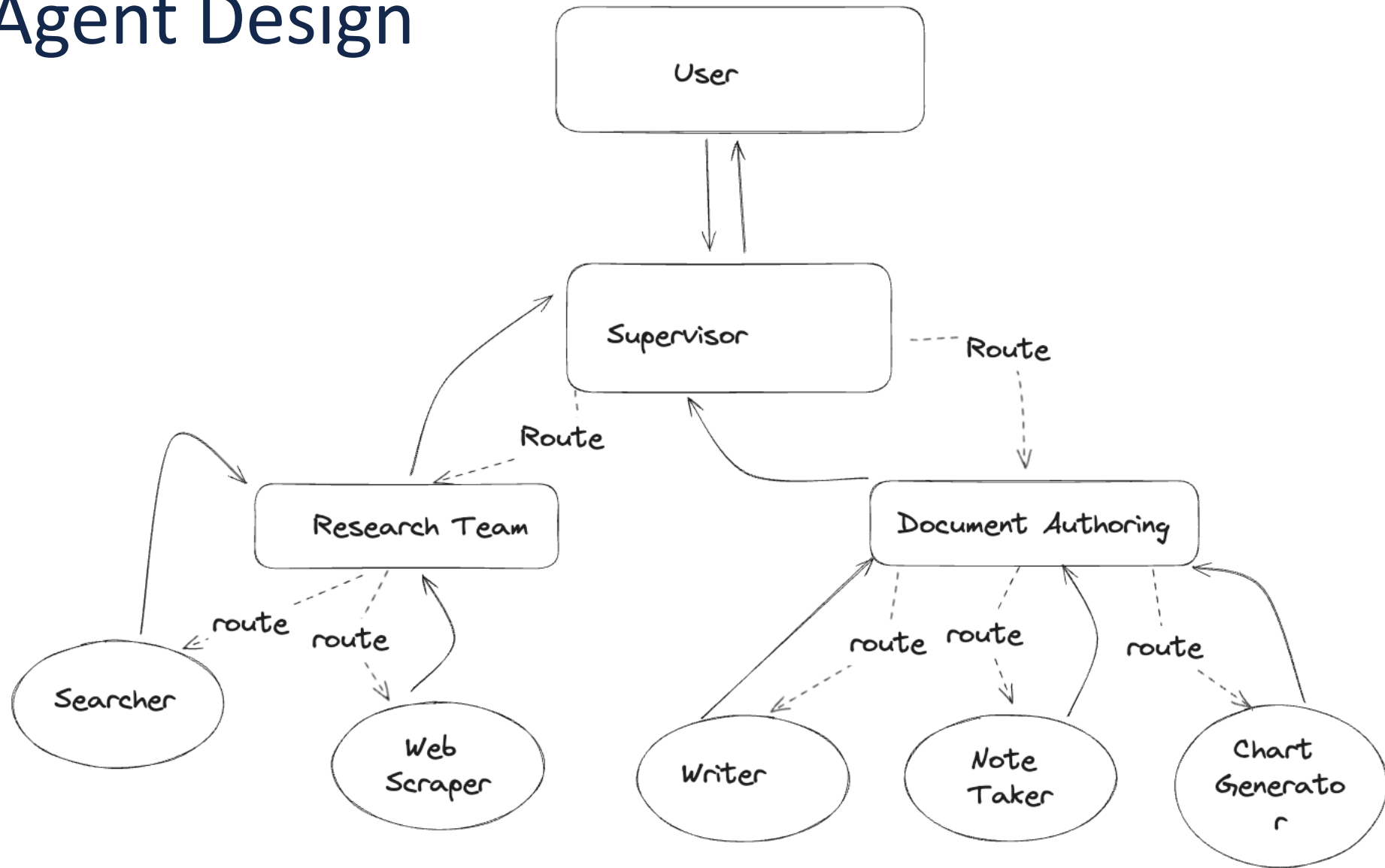
Posted by Shunyu Yao, Student Researcher, and Yuan Cao, Research Scientist, Google Research, Brain Team



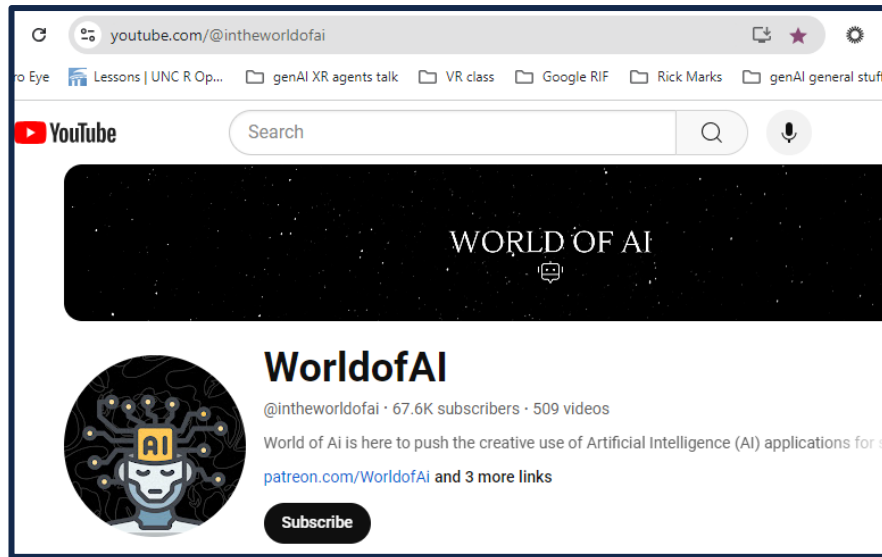
ReAct (Reason + Act)




Multi-Agent Design



How to keep up?




https://www.youtube.com/@matthew_berman



One Useful Thing


Trying to understand the implications of AI for work, education, and life. By Prof. Ethan Mollick

By Ethan Mollick  · Over 165,000 subscribers

<https://www.oneusefulthing.org/>


Google DeepMind x Neuronpedia

Introducing **Gemma Scope**, a new tool for understanding the internals of AI models.

 Launch

Neuronpedia is an open platform for interpretability research.

Explore, steer, and experiment on AI models.

 Getting Started



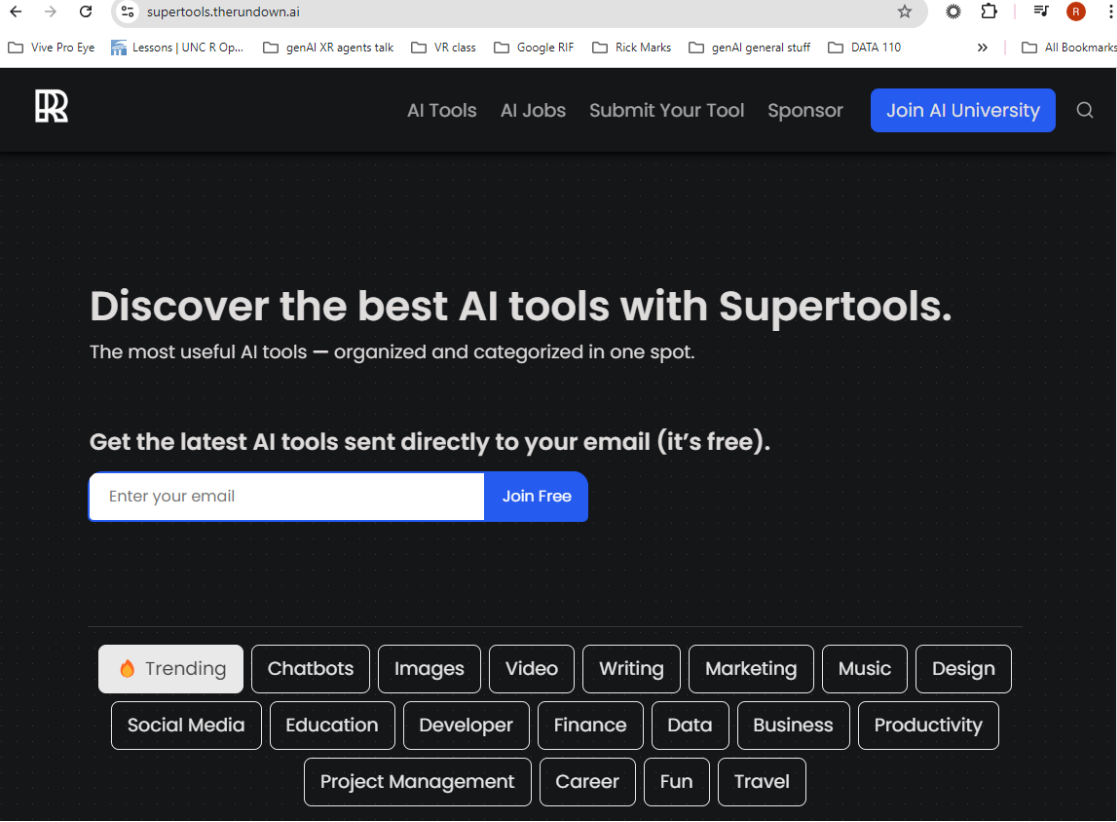
TWO MINUTE PAPERS | WHAT A TIME TO BE ALIVE!

Two Minute Papers

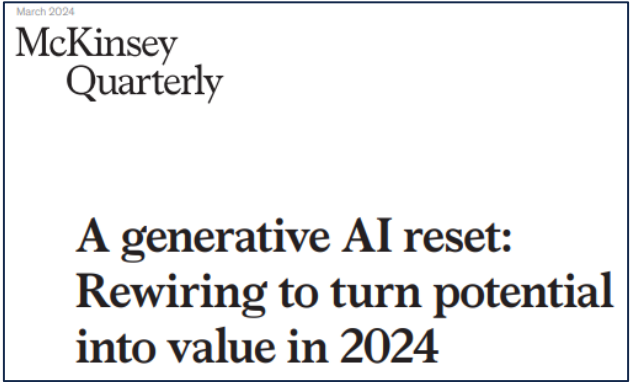
@TwoMinutePapers · 1.58M subscribers · 914 videos

What a time to be alive! ...more

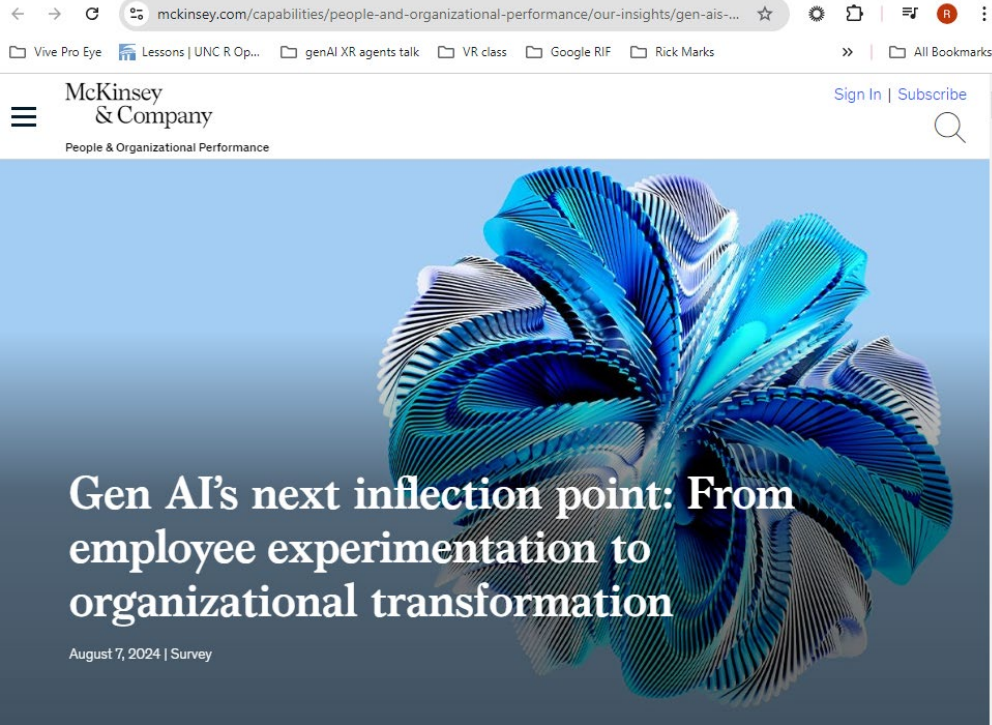
Gen AI tools and guides to application



The screenshot shows the homepage of Supertools. At the top, there is a navigation bar with the Supertools logo, links for 'AI Tools', 'AI Jobs', 'Submit Your Tool', 'Sponsor', and a prominent blue 'Join AI University' button. The main heading reads 'Discover the best AI tools with Supertools.' followed by the subtext 'The most useful AI tools — organized and categorized in one spot.' Below this is a promotional message: 'Get the latest AI tools sent directly to your email (it's free).' This is accompanied by an email input field and a 'Join Free' button. At the bottom, there is a grid of category buttons including 'Trending', 'Chatbots', 'Images', 'Video', 'Writing', 'Marketing', 'Music', 'Design', 'Social Media', 'Education', 'Developer', 'Finance', 'Data', 'Business', 'Productivity', 'Project Management', 'Career', 'Fun', and 'Travel'.



This is the cover of the March 2024 issue of McKinsey Quarterly. The title of the article is 'A generative AI reset: Rewiring to turn potential into value in 2024'. The McKinsey Quarterly logo is positioned in the top left corner.



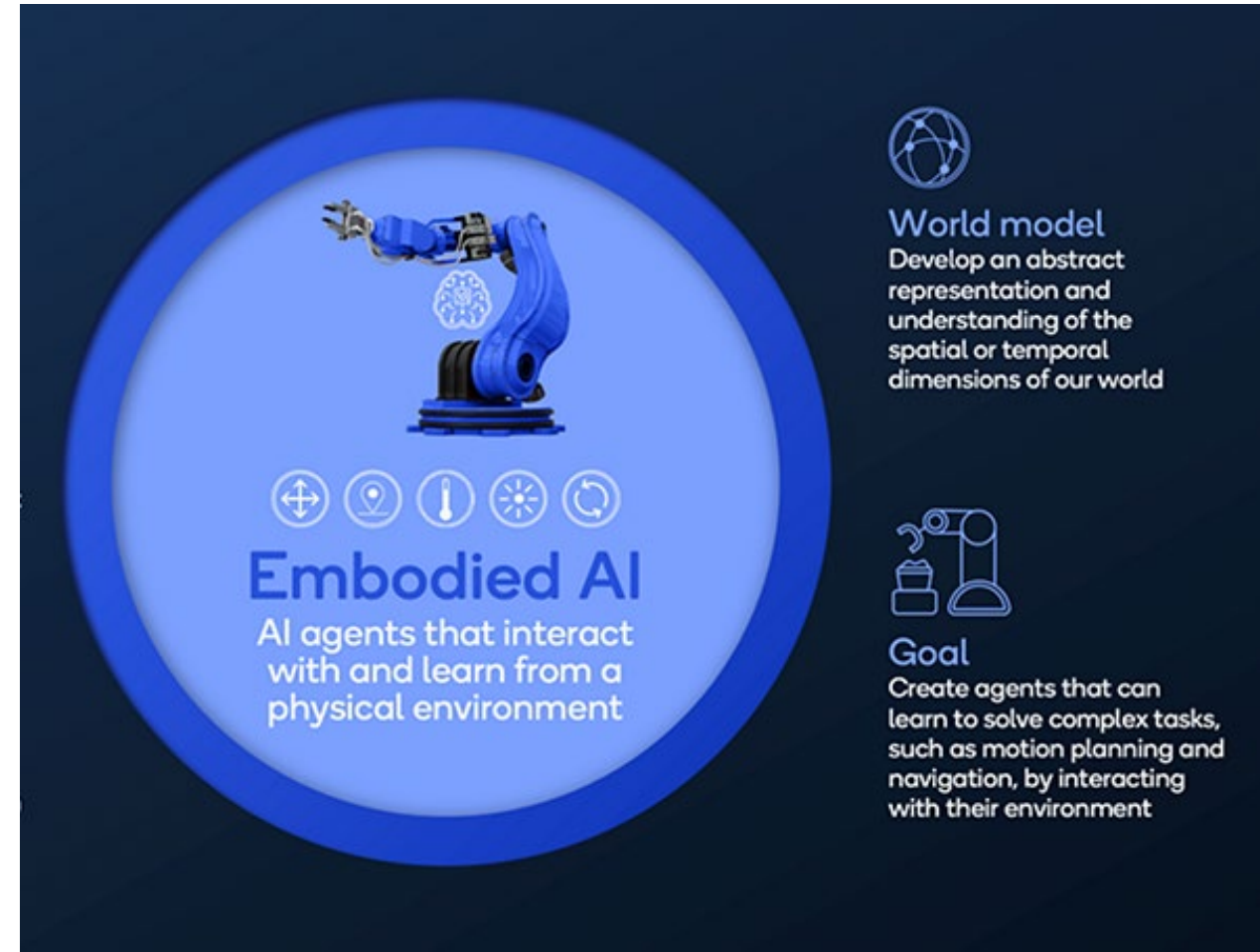
The screenshot shows the cover of an article from McKinsey & Company, specifically from the 'People & Organizational Performance' section. The article title is 'Gen AI's next inflection point: From employee experimentation to organizational transformation'. The cover features a large, abstract, blue and white graphic that resembles a stylized flower or a complex network structure. The date 'August 7, 2024 | Survey' is printed at the bottom left.

Next Big Things: spatial intelligence and embodied AI

How AI will understand the real world

Fei-Fei Li

TED



World model
Develop an abstract representation and understanding of the spatial or temporal dimensions of our world



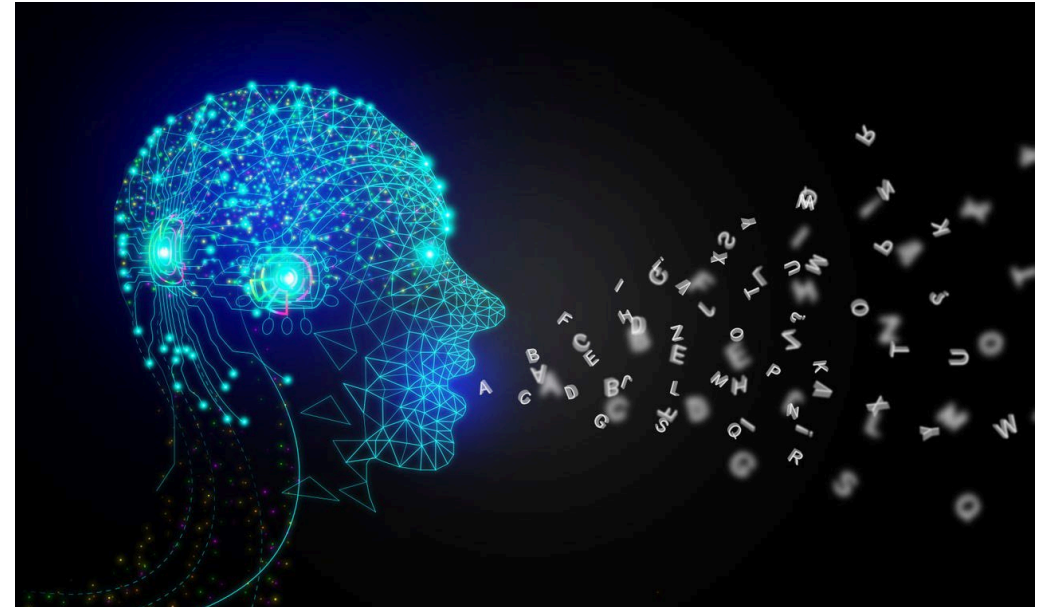
Goal
Create agents that can learn to solve complex tasks, such as motion planning and navigation, by interacting with their environment



**SCHOOL OF
DATA SCIENCE
AND SOCIETY**



COLLEGE OF ARTS AND SCIENCES
Computer Science



Thank you!

And happy to take
questions.



Event exit survey