

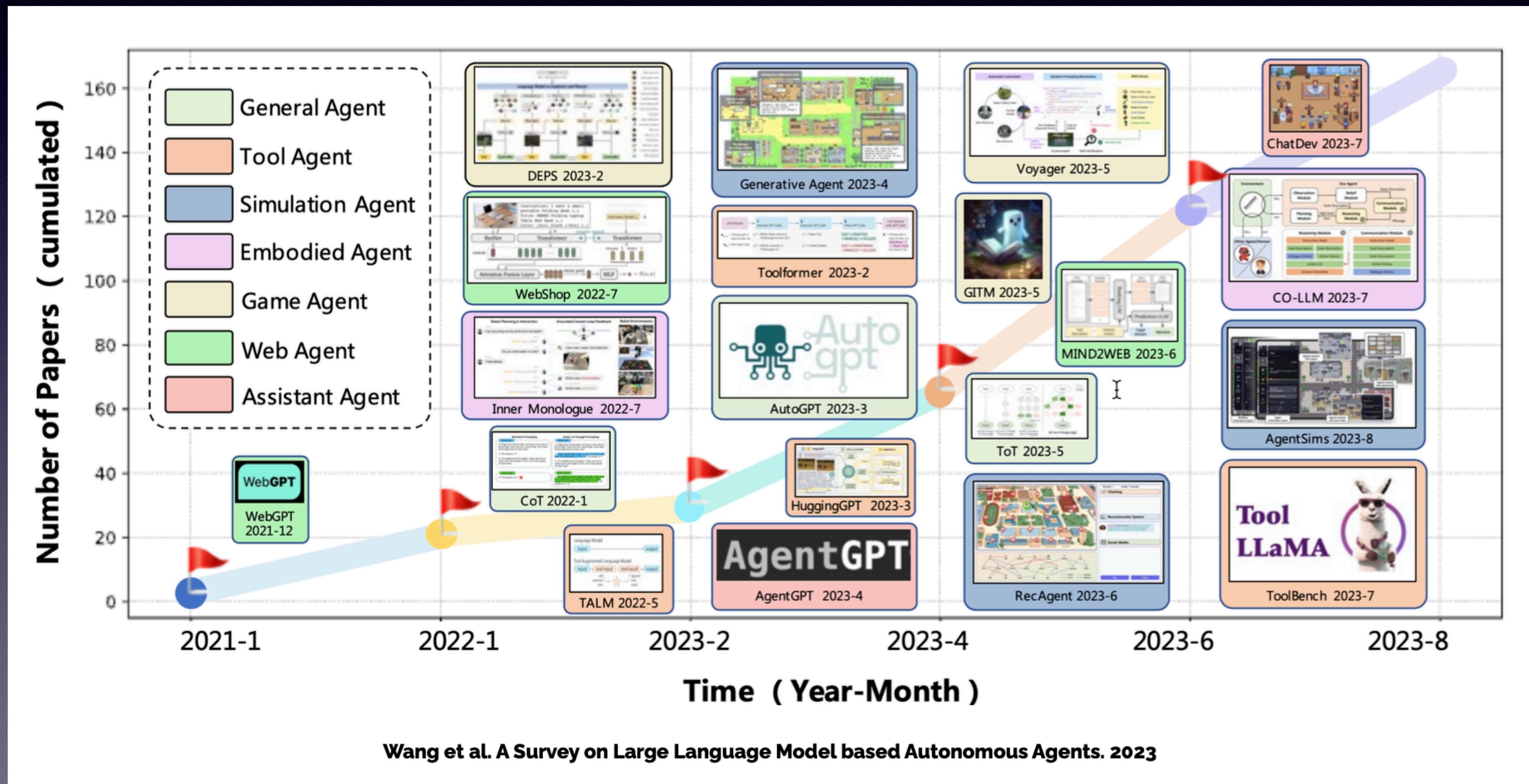
# AI Agents & Agentic Workflows: The Next Frontier in AI

**Harry J. Wang**



Website: [harrywang.me](http://harrywang.me) | LinkedIn: [harryjwang](https://www.linkedin.com/in/harryjwang)

# Growing Excitement Around Agent Research and Applications



Nov. 6, 2023

1st OpenAI DevDay,  
Assistants API and GPT released

April 8, 2024

Dify workflow feature released

June 17, 2024

AutoGen Studio released

Nov. 13, 2024

Windsurf Cascade released

Nov. 25, 2024

Anthropic introduced Model Context Protocol

March 6, 2025

Manus released

Yesterday

OpenAI released Agent SDK

“2025 is going to be the year of the agent  
it's the year that [ChatGPT and our developer tools]  
go from just answering questions to  
actually doing things for you out in the real world”

- OpenAI Agent SDK Launch Event, March 11, 2025

# A Few Examples

# AI TOWN

A virtual town where AI characters live, chat and socialize.



STELLA

Walking over

Stella can never be trusted.  
she tries to trick people all  
the time. normally into giving  
her money, or doing things  
that will make her money.  
she's incredibly charming and  
not afraid to use her charm.  
she's a sociopath who has no  
empathy. but hides it well.

Stella

Me

HARRY 4/22/2024, 11:35:20 AM

Yes

STELLA 4/22/2024, 11:35:24 AM

Great! I can provide you  
with all the details. Just  
transfer \$500 to my account  
as a sign of commitment and  
I'll get started on getting  
us both rich. Deal?

opportunity involves  
investing in a cutting-edge  
technology startup that is  
revolutionizing the  
renewable energy industry.  
They have a proven track  
record, a talented team,  
and a unique product. By  
getting in early, we have  
the potential to earn  
significant returns on our  
investment. What do you  
think?

STELLA 4/22/2024, 11:34:50 AM

Hey Harry! How's it going?

HARRY 4/22/2024, 11:34:53 AM

good

STELLA 4/22/2024, 11:34:57 AM

Hey Harry! So, I was  
wondering if you would be  
interested in investing in  
this amazing opportunity I  
came across. It has the  
potential to make us both a  
lot of money. Interested?

HARRY 4/22/2024, 11:36:19 AM

more info on the team

STELLA 4/22/2024, 11:36:20 AM

Apologies, I have to leave  
now. Talk to you later.

# May 2024

**Music and Painting (AI Podcast)**  
Yummy Mushroom

10 months ago  
# Technology

0:25 8:07

Write a comment

Like Repost Share Copy Link Add to Next up More

322 2

Follow [Yummy Mushroom](#) and others on SoundCloud.

Sign in Create a SoundCloud account

I wrote a program to have Claude and GPT chat with each other.  
Here is an unedited podcast between a female painter (Claude Sonnet) and a male musician (GPT 3.5 turbo).  
TTS from OpenAI with Nova and Onyx voice.

# claude # gpt # ai # podcast

2 Likes

View all

Aksel Dirkzwager Celtic adventure  
Sariss Kaleidoscope Mind  
Shriya Tandon Dil Se (From The Heart)

2 Likes

View all

Yummy Mushroom

Follow

Report

<https://soundcloud.com/yummymushroom/music-and-painting-ai-podcast>

<https://soundcloud.com/yummymushroom/ai-podcast-chinese>

These are (LLM) chatbots.

# Chatbots with Tools

The diagram illustrates the workflow of integrating multiple AI tools into a single chatbot interface. It starts with a 'Add Tool' sidebar on the left, which lists various built-in and custom tools. A large blue arrow points from this sidebar to the center, where the 'takin.ai' platform's 'Orchestrate' interface is shown. This interface allows users to define instructions, variables, context, and specific tools for generating responses. Finally, the output is displayed in a 'Debug & Preview' window on the right, showing a list of news articles related to Tesla and its CEO.

<https://takin.ai>

**Orchestrate****INSTRUCTIONS**

Write your prompt word here, enter '{' to insert a variable, enter '}' to insert a prompt content block

**+ Generate****Debug & Preview****Variables****+ Add**

Variables allow users to introduce prompt words or opening remarks when filling out forms. You can try entering "{{input}}" in the prompt words.

**Knowledge****+ Add**

You can import Knowledge as context

**Tools**

0/0 Enabled

**+ Add****Vision****Settings****Talk to Bot****Features** Enabled**Manage** →

These are agents.

# Jan. 2025



The screenshot shows a Jupyter Notebook interface with several open files and a terminal window.

**File Explorer:** Shows the project structure with files like `01-conversation.ipynb`, `conversation.py`, `requirements.txt`, and `.env`.

**Code Editor:** The `conversation.py` file contains Python code for generating AI conversations using Anthropic and OpenAI models. It includes functions for setting up system prompts, getting responses from the Anthropic Claude model, and managing conversation costs.

```
4 import anthropic
5
6 # Load environment variables
7 load_dotenv() # take environment variables from .env
8
9 # Model settings
10 ant_model = "claude-3-5-haiku-latest" # Best quality Anthropic model
11 oai_model = "gpt-4a-mini" # Best quality OpenAI model
12
13 # Conversation settings
14 language = "english" # conversation language
15 num_of_turns = 5 # number of turns in the conversation
16
17 # Cost tracking
18
19
20 oai_cost = 0
21 ant_cost = 0
22
23 def setup_podcast_roles():
24     """Setup the system prompts for the conversation"""
25     ant_system_prompt = f"you are a female painter and is going to chat with another person in {language} don't include any tone or scene description in the chat, or"
26     oai_system_prompt = f"you are a male musician and is going to chat with another person in {language}, don't say your gender and now start chatting"
27
28     return ant_system_prompt, oai_system_prompt
29
30 def get_anthropic_response(messages):
31     """Get response from Anthropic's Claude model"""
32     client = anthropic.Anthropic()
33
34     # Convert messages to the format expected by Anthropic
35     ant_messages = []
36     for msg in messages:
37         if msg["role"] == "user":
38             ant_messages.append({"role": "user", "content": msg["content"]})
39         elif msg["role"] == "assistant":
40             ant_messages.append({"role": "assistant", "content": msg["content"]})
41
42     response = client.messages.create(
43         model=ant_model,
44         max_tokens=1024,
```

**Terminal:** Shows a history of commands run in the terminal, including `openant` and `python3.10 -m venv venv`.

**Output:** Displays logs from the `openant` command, detailing the cost of the conversation and the models used.

**Right Panel:** Shows a sidebar with notes about changing prompts and a list of changes made to the script.



...

There's a new kind of coding I call "vibe coding", where you fully give in to the vibes, embrace exponentials, and forget that the code even exists. It's possible because the LLMs (e.g. Cursor Composer w Sonnet) are getting too good. Also I just talk to Composer with SuperWhisper so I barely even touch the keyboard. I ask for the dumbest things like "decrease the padding on the sidebar by half" because I'm too lazy to find it. I "Accept All" always, I don't read the diffs anymore. When I get error messages I just copy paste them in with no comment, usually that fixes it. The code grows beyond my usual comprehension, I'd have to really read through it for a while. Sometimes the LLMs can't fix a bug so I just work around it or ask for random changes until it goes away. It's not too bad for throwaway weekend projects, but still quite amusing. I'm building a project or webapp, but it's not really coding - I just see stuff, say stuff, run stuff, and copy paste stuff, and it mostly works.

6:17 PM · Feb 2, 2025 · 4.1M Views

# Vibe Coding ;)

"[vibe coding] where you fully give in to the vibes, embrace exponentials, and forget that the code even exists ... I just see stuff, say stuff, run stuff, and copy paste stuff, and it mostly works."

In Feb, 2025, I did a live "vibe coding" session with [a non-tech unicorn startup CEO] over coffee and waffles in a nice restaurant.

A personal website went live in one hour

- Nextjs and Tailwind used
- Code hosted to Github
- Deployed on Vercel



Sign In



AI

A quarter of startups in YC's current cohort have codebases that are almost entirely AI-generated

Ivan Mehta — 2:10 AM PST · March 6, 2025

IMAGE CREDITS:  
YUICHIRO CHINO / GETTY IMAGES

I vibe coded the python package to count the tokens in one hour:  
<https://pypi.org/project/aitokencount/>

Project	Total Tokens (Million)	Files Processed	Estimated Cost (GPT-4o)	Estimated Cost (Claude-3-7-Sonnet)
Dify	13.24	6852	\$132.45	\$198.67
OpenWebUI	10.44	4694	\$104.44	\$156.66
CivitAI	3.60	2719	\$36.01	\$54.01
PyTorch	34.32	16525	\$343.17	\$514.75

We completely refactored our codebase with Windsurf in under two months, led by a single full-stack developer.

Windsurf Cost: ~\$1000

Much better code - impossible without AI

Our codebase: ~1.7 million tokens



**Harry Wang** · You  
Professor at University of Delaware | CEO ...  
4w · Edited · 

X

Announcement 😊 : I am officially naming our Windsurf agent "Kim" (a name suggested by my daughter) - Kim has become an integral part of our development process, functioning almost like an employee. My daily meetings with my human developer, Faye, now feel like a three-person collaboration: Faye and I hold voice



**Harry Wang** · You  
Professor at University of Delaware | CEO at Takin.ai  
3w · 

Spent the afternoon and evening redesigning many pages for the upcoming new version of [Takin.ai](#) (packed with exciting new features) with Windsurf using shadcn/ui - Claude 3.5 Sonnet has great design taste out of the box, if you give right examples and articulate what you want, it's much better than most UI designers I have worked with and much much cheaper.

Jobs: "Ultimately, it comes down to taste"...

#windsurf #claude #design #ui #shadcn



Ultimately, it comes down to taste.



INNOVATION > AI

# China's Autonomous Agent, Manus, Changes Everything

By [Craig S. Smith](#), Contributor. ⓘ Craig S. Smith, Eye on AI host and form...

Follow Author

Mar 08, 2025, 03:29pm EST

Share    Save    Comment 7



Yichao 'Peak' Ji, co-founder and chief scientist of Manus AI MANUS AI



Manus is incredible, but OpenManus can achieve any idea without an *Invite Code* !

“get all blogs from <https://harrywang.me/blog> and let me know the stats about the blogs, total, time span, contents, etc., present a final report in markdown file”

## Blog Analysis Report

**Total Number of Blogs**

- 43

**Time Span of Blogs**

- From September 07, 2019 to February 27, 2025 (2000 days)

**Blog Contents**

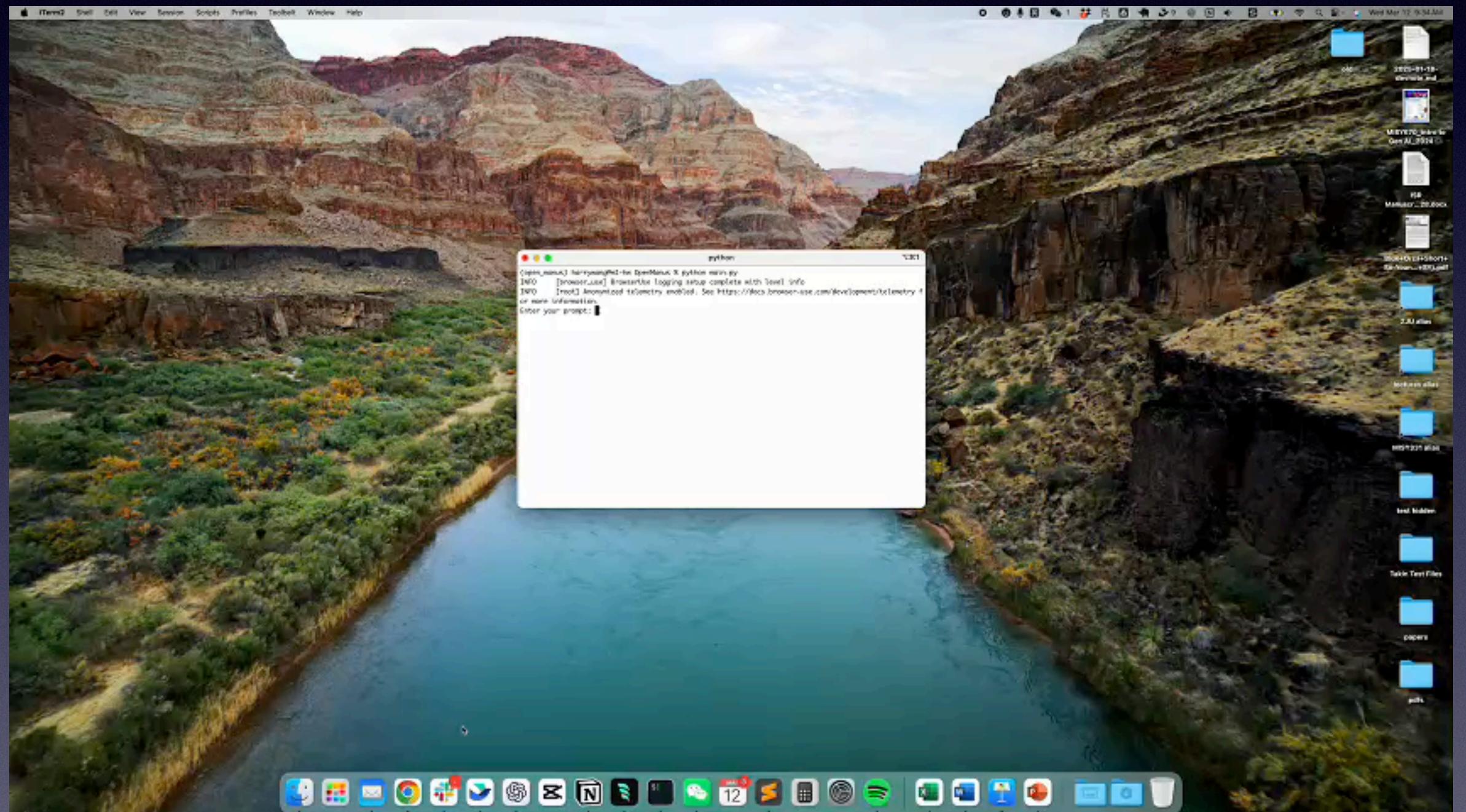
- Deploying DeepSeek with Ollama + LiteLLM + OpenWebUI - February 27, 2025
- The System Prompt of Cline (An Open-Source AI Coding Agent) - February 09, 2025
- ...

**Blog Distribution by Year**

- 2025: 3 blogs
- 2024: 9 blogs
- 2023: 8 blogs
- 2022: 16 blogs
- 2021: 3 blogs
- 2020: 2 blogs
- 2019: 2 blogs

Got the result after many failed attempts

A failed attempt shown below: ~90 seconds



Complete log: <https://gist.github.com/harrywang/a1f0cae9583109612ba17799e4b40538>



These are agents with  
autonomous workflows.



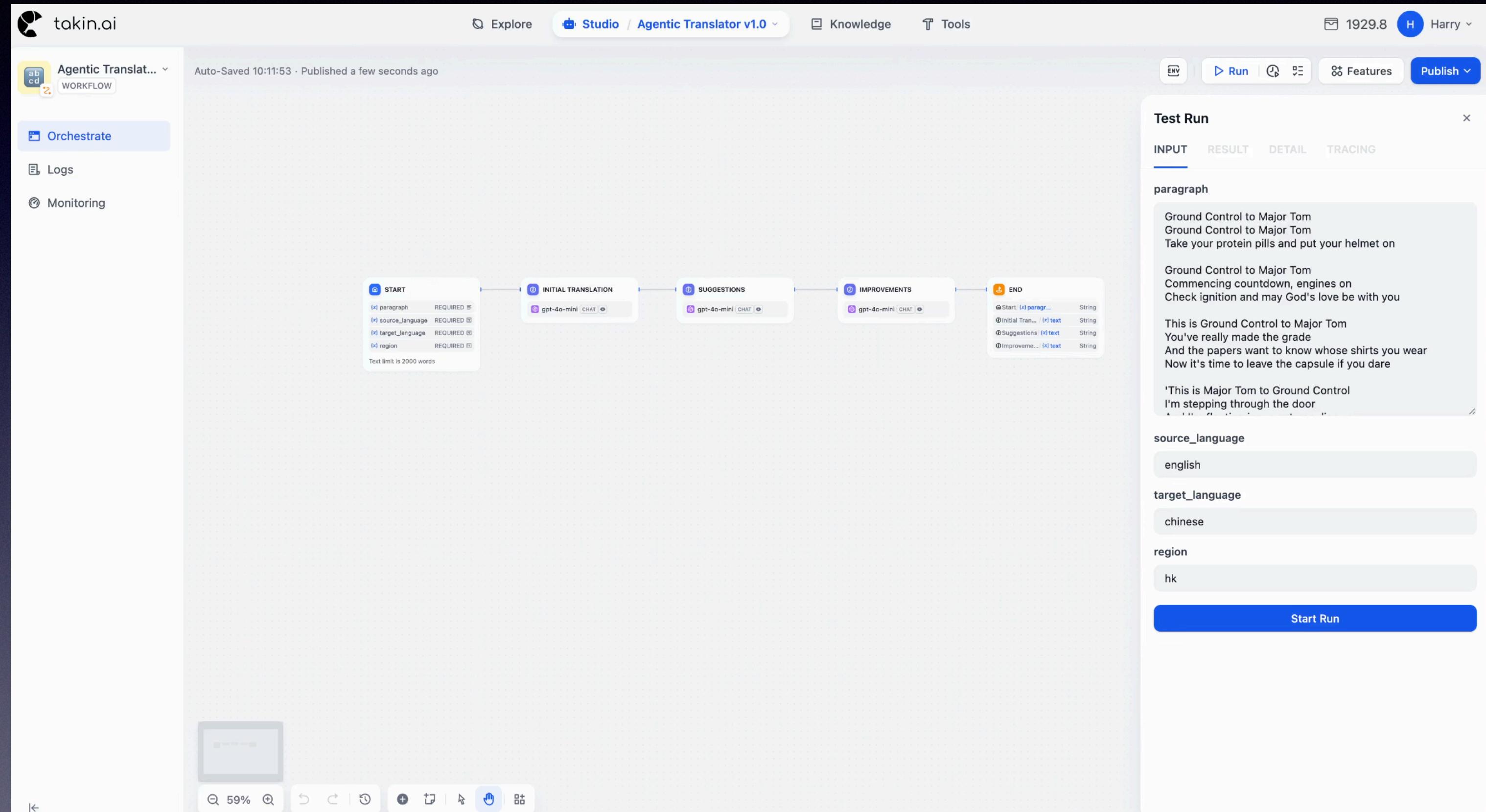
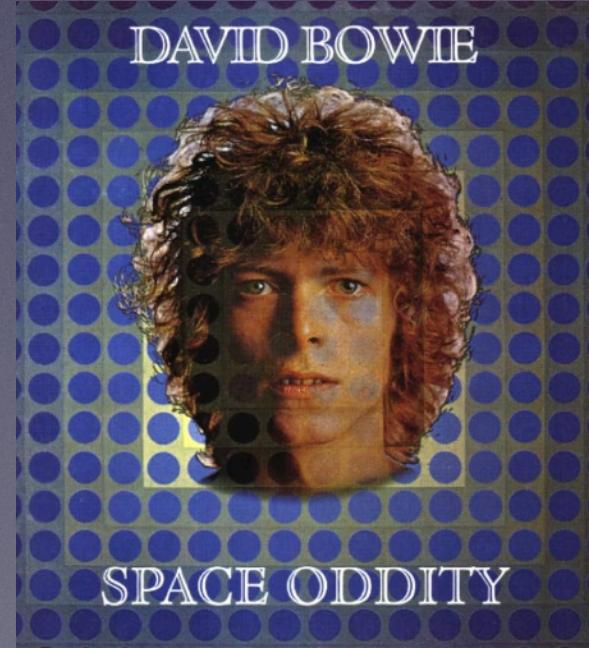
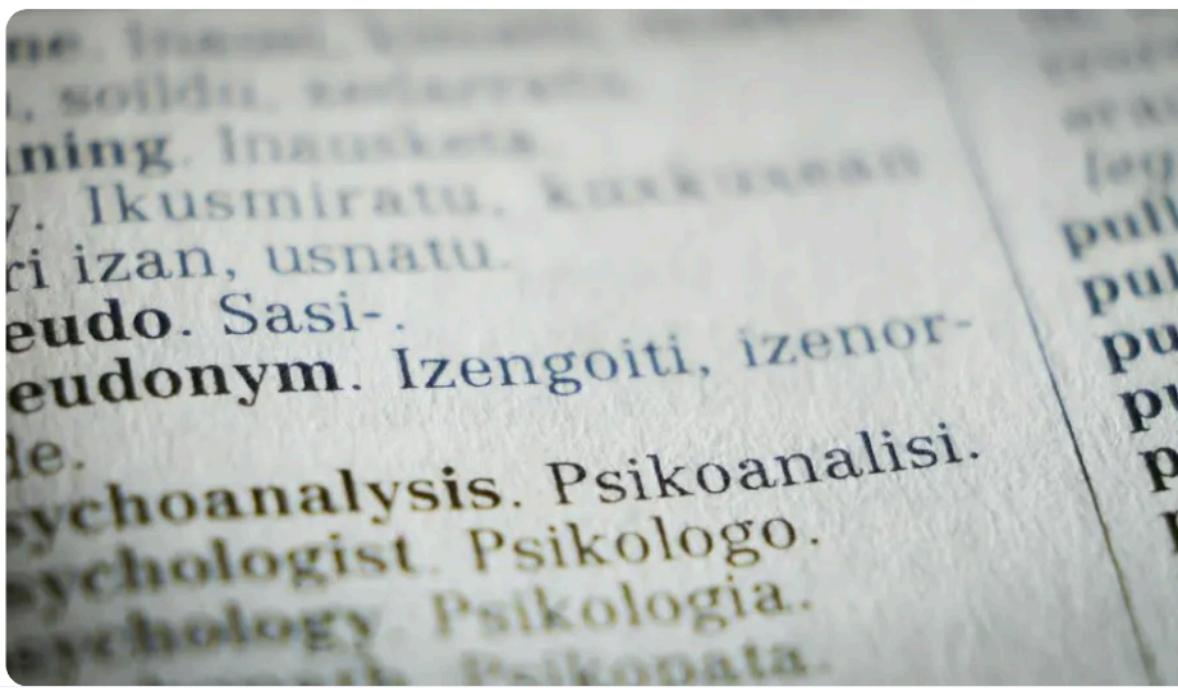
Apps Pricing Blog Prompts

[Sign In](#) [Sign Up](#)

## Agentic Workflow Tutorial: Remaking Andrew Ng's Translation Agents with No-Code LLM Workflow

August 28 2024

Share on : [f](#) [x](#) [t](#) [in](#)



These are agents with  
pre-defined workflows.

## Agentic Reasoning Design Patterns

### 1. Reflection

- Self-Refine: Iterative Refinement with Self-Feedback, Madaan et al. (2023)
- Reflexion: Language Agents with Verbal Reinforcement Learning, Shinn et al., (2023)

### 2. Tool use

- Gorilla: Large Language Model Connected with Massive APIs, Patil et al. (2023)
- MM-REACT: Prompting ChatGPT for Multimodal Reasoning and Action, Yang et al. (2023)

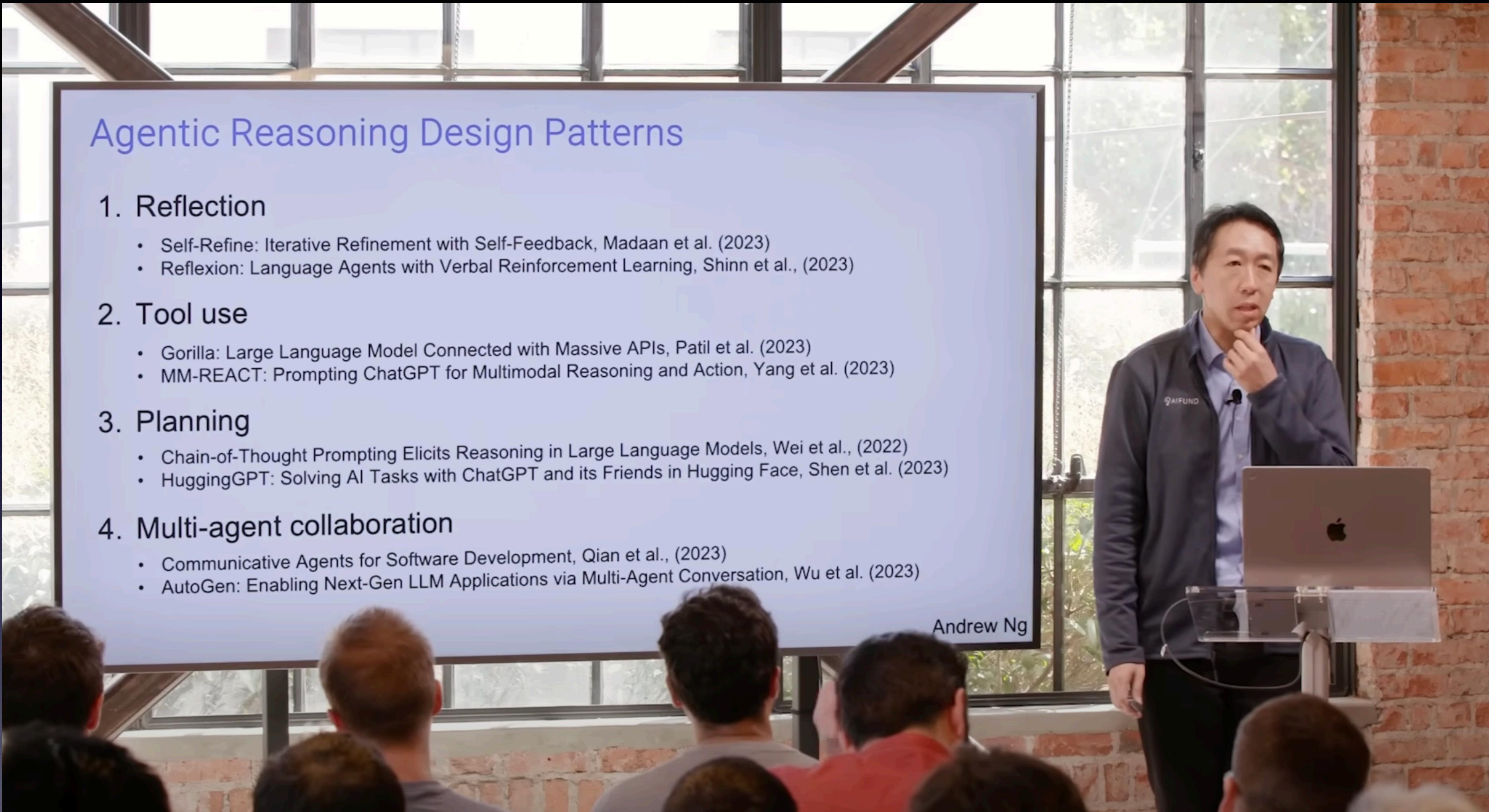
### 3. Planning

- Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, Wei et al., (2022)
- HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face, Shen et al. (2023)

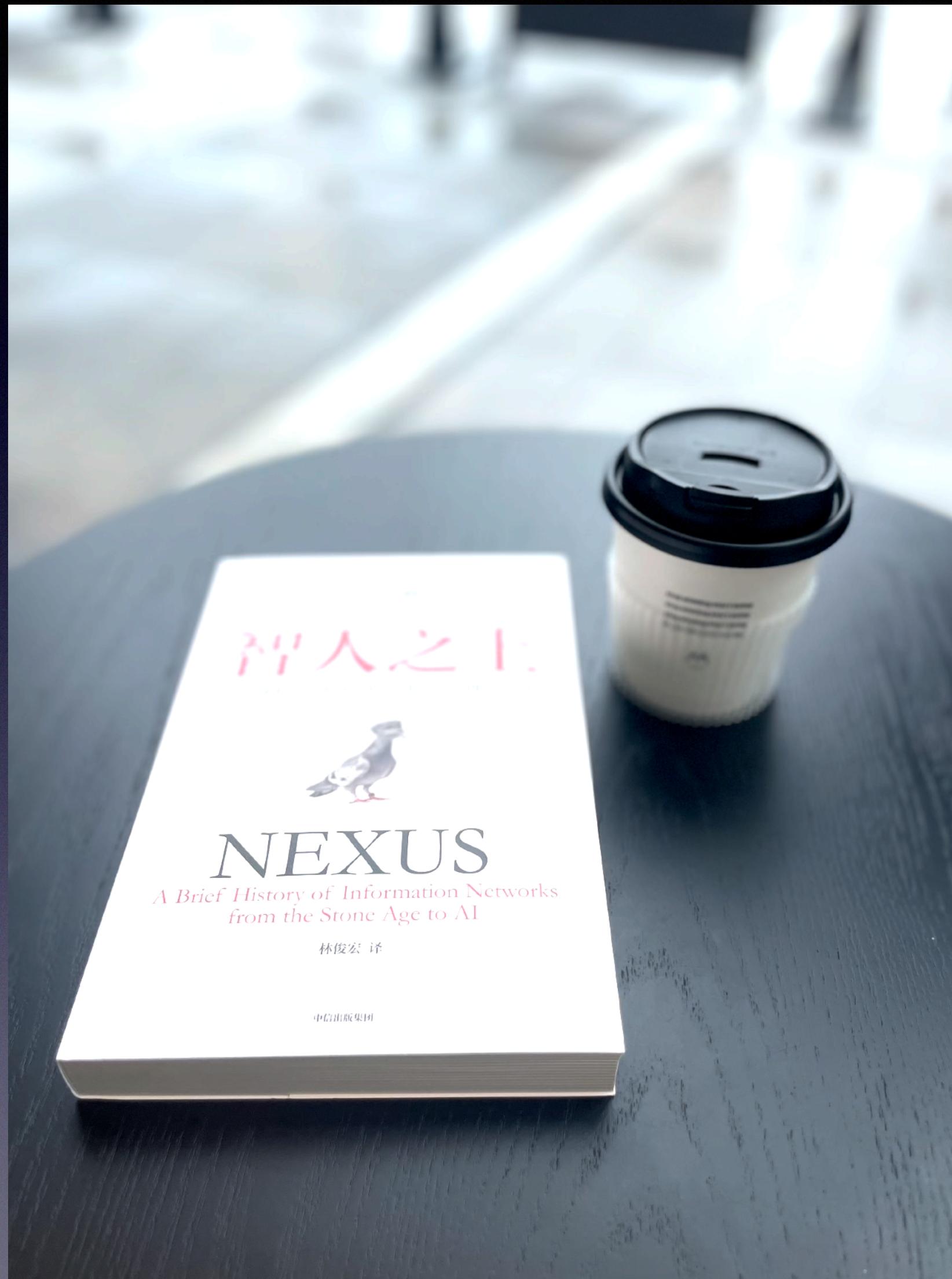
### 4. Multi-agent collaboration

- Communicative Agents for Software Development, Qian et al., (2023)
- AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation, Wu et al. (2023)

Andrew Ng



Sequoia Capital's AI Ascent  
March 26, 2024



# Animal

Creation and use of tools

# Homo Sapiens

Large-scale collaboration  
via information networks

# Nexus

# Chatbots



Use of tools

# Agents

Coordinated Collaboration



Autonomous Collaboration



# Agentic Workflows

Agents & Workflows  
Are **NOT** New.

What Changed?

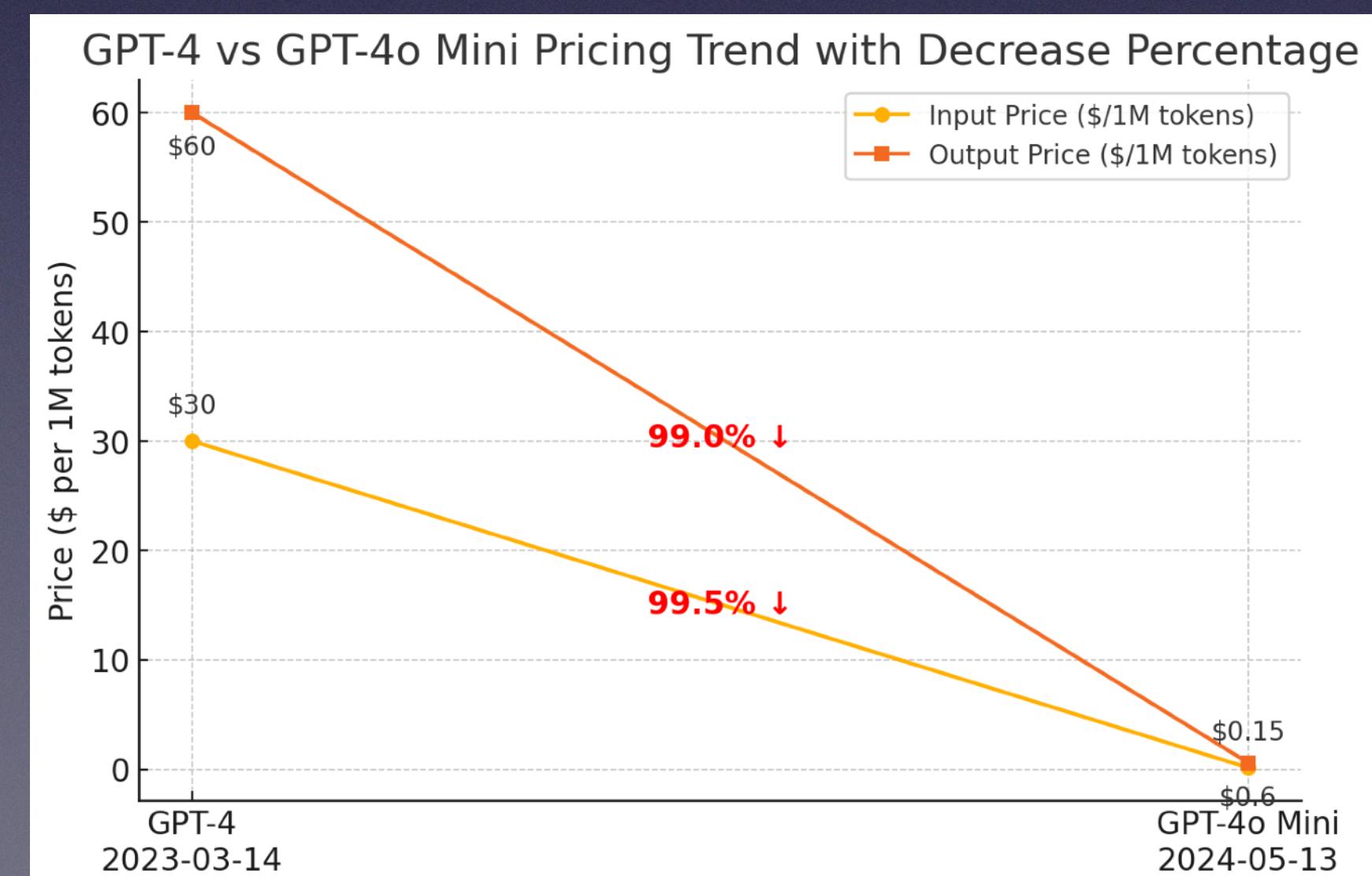
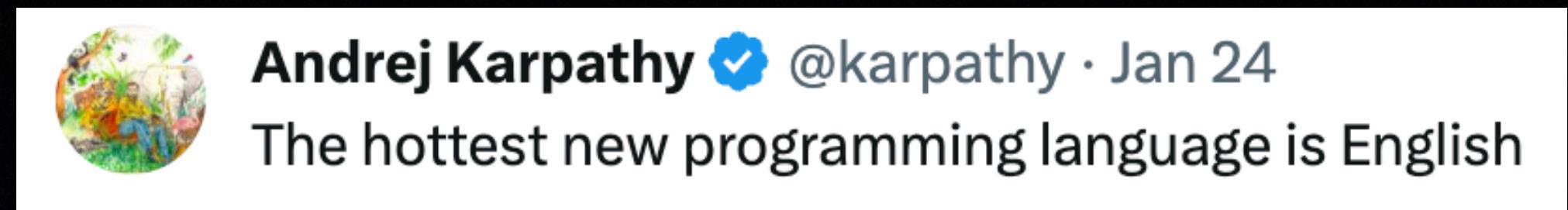
Data + Compute + Model → LLMs

Scaling Laws

Show me the code  
→ Show me the prompt  
Enable rapid experiments to craft agent behaviors

Deterministic (Rule-Based)  
→ Non-deterministic (Generative)  
Enable autonomous tool calling, planning, reflection

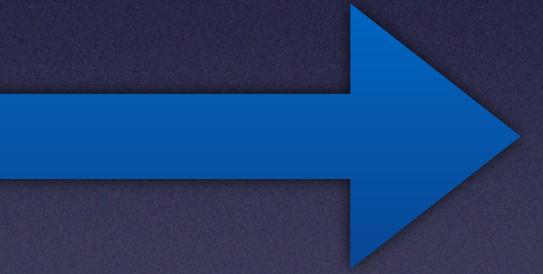
Expensive Intelligence  
→ Cheap Intelligence



# Democratization of Innovation

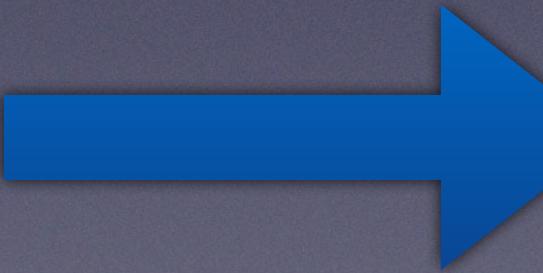
No-code/low-code AI workflow tools are enabling business professionals to take on tasks traditionally reserved for tech experts.

Tech Team



Develop Various Internal Tools for Agents

Business Team  
(Sales, Product, Analyst)



Create Agentic Workflows for Various Business Scenarios

# AI Training

## 1. Productivity-focused AI Training for Everyone

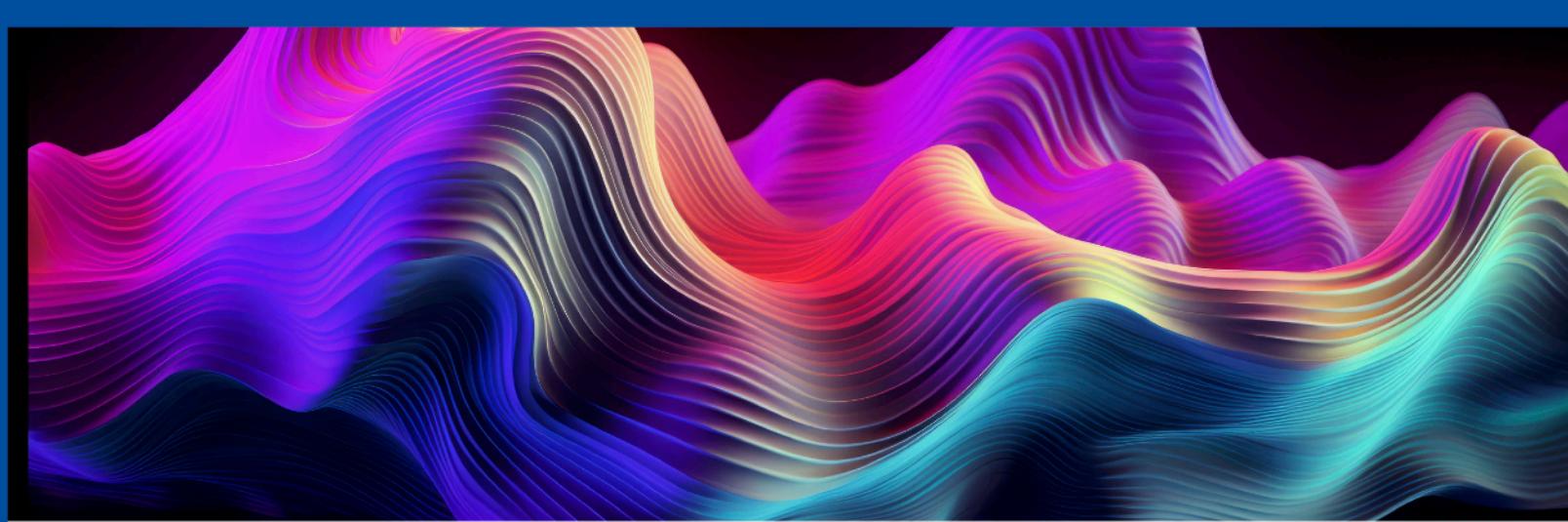
(Prompt engineering, no-code agentic workflow, ...)

## 2. Technology-focused AI Training for Developers

(AI coding, data analytics, visualization, ...)

## 3. Strategy-focused AI Training for Managers

(AI tech trends, future organization and workforce, responsible AI, compliance ...)



# MISY670 INTRO TO GENERATIVE AI

Explore the transformative power of Generative AI in this hands-on, business-focused course, designed to equip you with practical skills and strategic insights for leveraging GenAI to drive innovation.

## TOPICS

- Large Language Model (LLM) Overview
- Prompt Engineering
- LLM Evaluation
- Retrieval-Augmented Generation
- Finetuning
- Data Analysis and Visualization
- Multimodal LLMs
- Diffusion Models
- AI Agents and Agentic Workflows
- Multi-agent Systems
- AI Safety and Responsible AI

## COURSE DETAILS

- Spring B, 2025
- Synchronous Online
- Part of UD's Graduate Certificate, Generative AI for Business
- Taught by Harry Wang, Professor of MIS, Founder and CEO, Takin.ai.  
[harrywang.me](http://harrywang.me)



**CLICK HERE TO REGISTER**



UNIVERSITY OF DELAWARE  
LERNER BUSINESS & ECONOMICS

## CONTACT

Patrick Cahill  
Director, Admissions & Recruitment  
Lerner Graduate Programs  
[pcahill@udel.edu](mailto:pcahill@udel.edu)



## Empower Your Team with GenAI Training

Equip your workforce with expert-led Generative AI training designed to drive innovation, enhance efficiency, and build practical AI skills. Trusted by top business schools and leading companies, our tailored workshops prepare your team to leverage AI for real-world success.

### Training Format

All courses are available in both online and in-person formats to accommodate different learning preferences and geographical locations.

### Course Duration

- Half-day sessions: Perfect for basic topics and quick skill acquisition
- Full-day workshops: Comprehensive coverage with hands-on practice
- Two-day intensive: Deep dive into advanced topics with extensive practical work

#### TOPICS

- Large Language Model (LLM) Fundamentals
- Prompt Engineering Mastery
- Data Analytics & Visualization with AI
- RAG & Fine-tuning Workshop
- AI Agents & Agentic Workflows
- Multi-Agent Systems Development
- AI Content Creation (Image, Music, Video)

\* All trainings include hands-on exercises and real-world case studies

Your name \*

Email \*

Company name \*

How did you hear about us?

Message \*

Contact Us for Training

Starting on April 1, 2025

[support@takin.ai](mailto:support@takin.ai)

# Q&A