STEVENS
INSTITUTE OF TECHNOLOGY
1870

LLM Agents and Agentic Workflows

FA690 Machine Learning in Finance

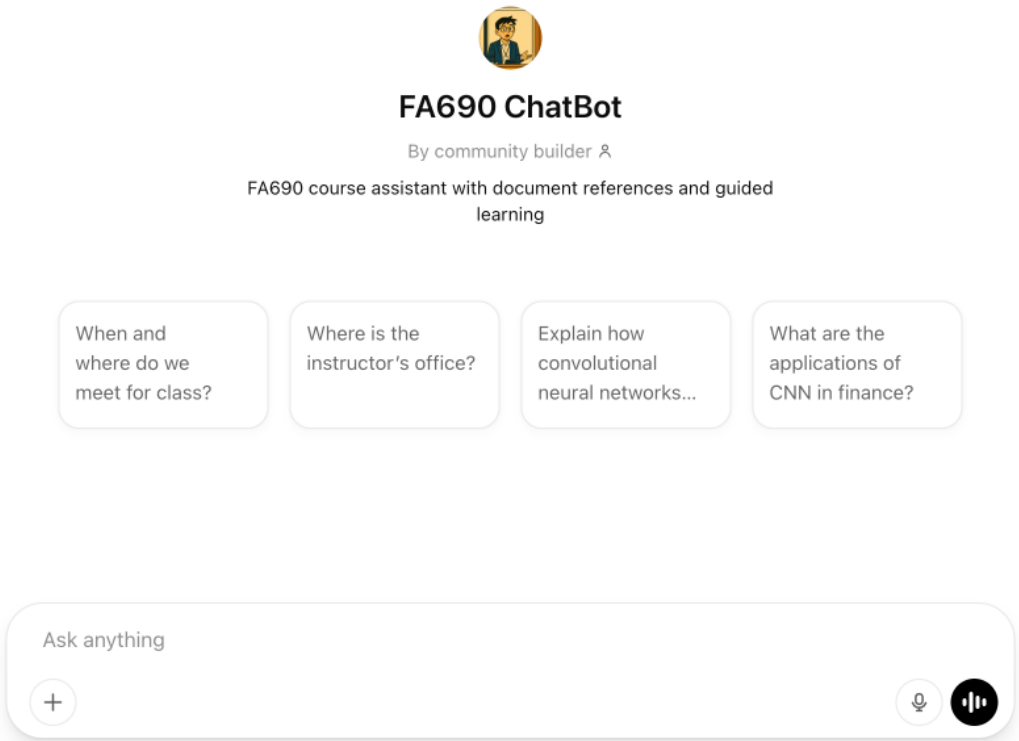
Dr. Zonghao Yang

2025 Spring

LLM Agents

FA690 ChatBot

FA690 ChatBot is a customized GPT, ready to answer questions related to FA690 Machine Learning in Finance 24/7 [Link to FA690 ChatBot]



Key Configuration of FA690 ChatBot

- **Objective:** Its goal is to help students better understand the course by
  - Answering questions about logistics directly, such as syllabus content, requirements, deadlines
  - Guiding exploration of lecture content and technical topics through questions, examples, or references to relevant materials
- **How to respond:** Does the chatbot directly provide the answer or guide the student to discover answers? How to preempt hallucination?
- **Knowledge base:** Course syllabus, lecture notes, project descriptions, etc.
- **Tool use:**
  - Web search ✅
  - Canvas ✅
  - DALLE image generation ❎
  - Code interpreter & data analysis ❎

From ChatBot to LLM Agents

- An LLM agent is like an AI assistant that can think, decide, and act
  - Set goals
  - Make decisions
  - Use tools or APIs
  - Work in steps or loops
- **FA690 ChatBot** is not a fully-fledged LLM agent yet
- What functions should we expect if **FA690 ChatBot** were an LLM agent?
  - Autonomous Task Execution
  - Goal-Oriented Behavior
  - Reflection & Iteration
  - Memory & Personalization

FA690 Agent

Autonomous Task Execution

- Proactively check deadlines
- Track assignment progress for each student
- Send nudges like: “You haven’t started Assignment 3. Want help?”

Goal-Oriented Behavior

Students can give it broad tasks like: “Help me get an A in FA690.”

- Plan a week-by-week study strategy
- Schedule reminders

Reflection & Iteration

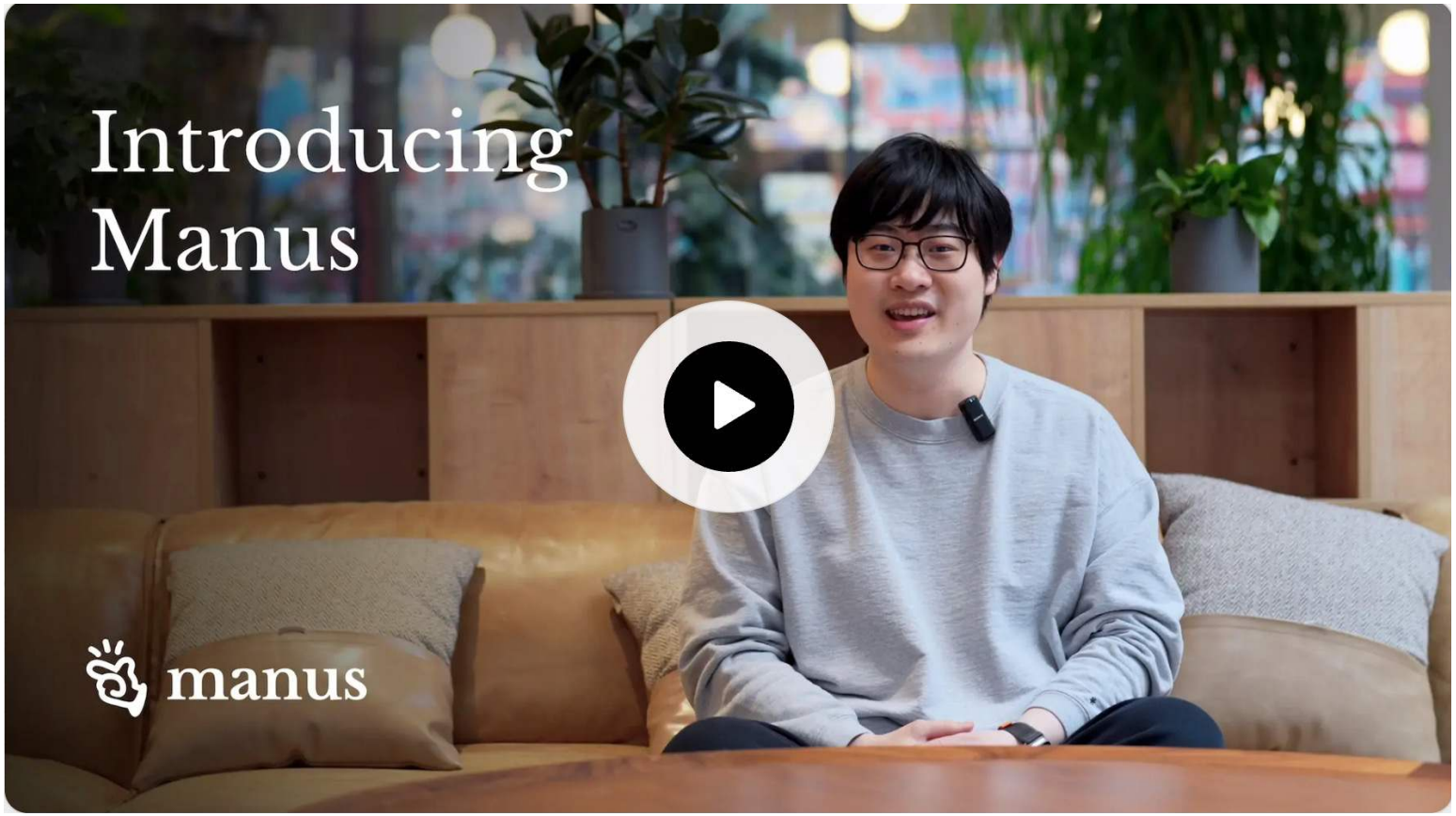
- Notice a student keeps getting questions wrong
- Suggest different study techniques
- Track learning over time and adapt

Memory & Personalization

- Knows student preferences, goals, struggles
- Remembers past conversations
- Could even say: “Last week you struggled with generative models. Want to review before the quiz?”

Agentic Workflows

manus The General AI Agent (March 6, 2025)



Source: Manus AI (https://manus.im/)

LLM-based Agents

Non-agentic workflow (zero-shot)

Please type out an essay on topic X from start to finish in one go, without using backspace

Agentic workflow

Write an essay outline on topic X

Do you need any web research?

Write a first draft.

Consider what parts need revision or more research.

Revise your draft.

…

Non-Agentic Workflow

- Completes a task in one go without iteration
- No revisions or refinements
- The use of tools and knowledge base

Agentic Workflow

- Breaks tasks into stages (e.g., planning, research, drafting, revising)
- AI iterates based on human guidance (e.g., Feedback Loop)

Truly Autonomous AI Agent

- Fully independent AI agent that determines steps, tools, and iterates without human involvement
- AI adjusts workflow dynamically to achieve the best outcome

Source: AI Agents Fundamentals In 21 Minutes [Link to YouTube]

OpenAI Agent SDK (March 11, 2025)

"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

Agentic Reasoning Design Patterns

- Reflection
- Tool use

robust technology

- Planning
- Multi-agent collaboration

emerging technology

Reflection

Please write code for {task}

def do_task(X): …

def do_task_v2(X): …

def do_task_v3(X): …

Coder Agent
(LLM)

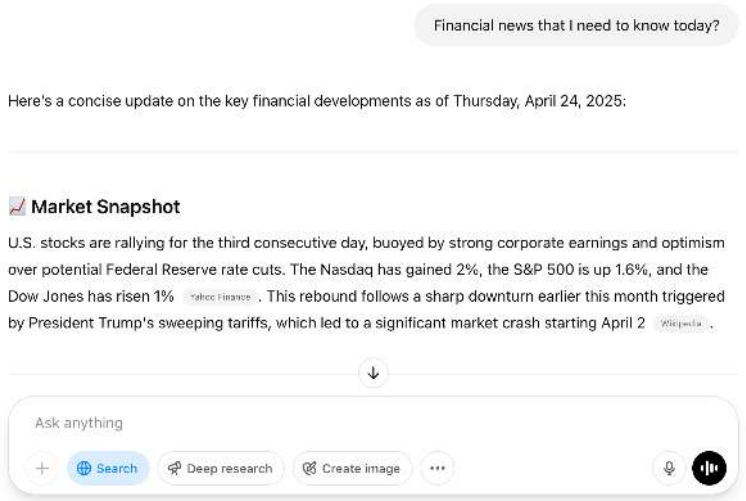
Here's code intended for {task}:
def do_task (x):
…
Check the code carefully for correctness, style and efficiency, and give constructive criticism for how to improve it.

There's a bug on line 5. Fix it by …

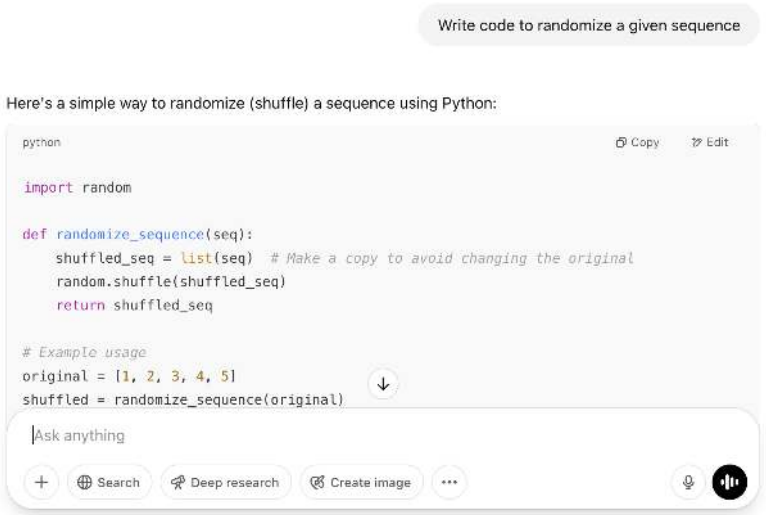
It fails Unit Test 3. Try changing …

Critic Agent
(LLM)

Tool Use



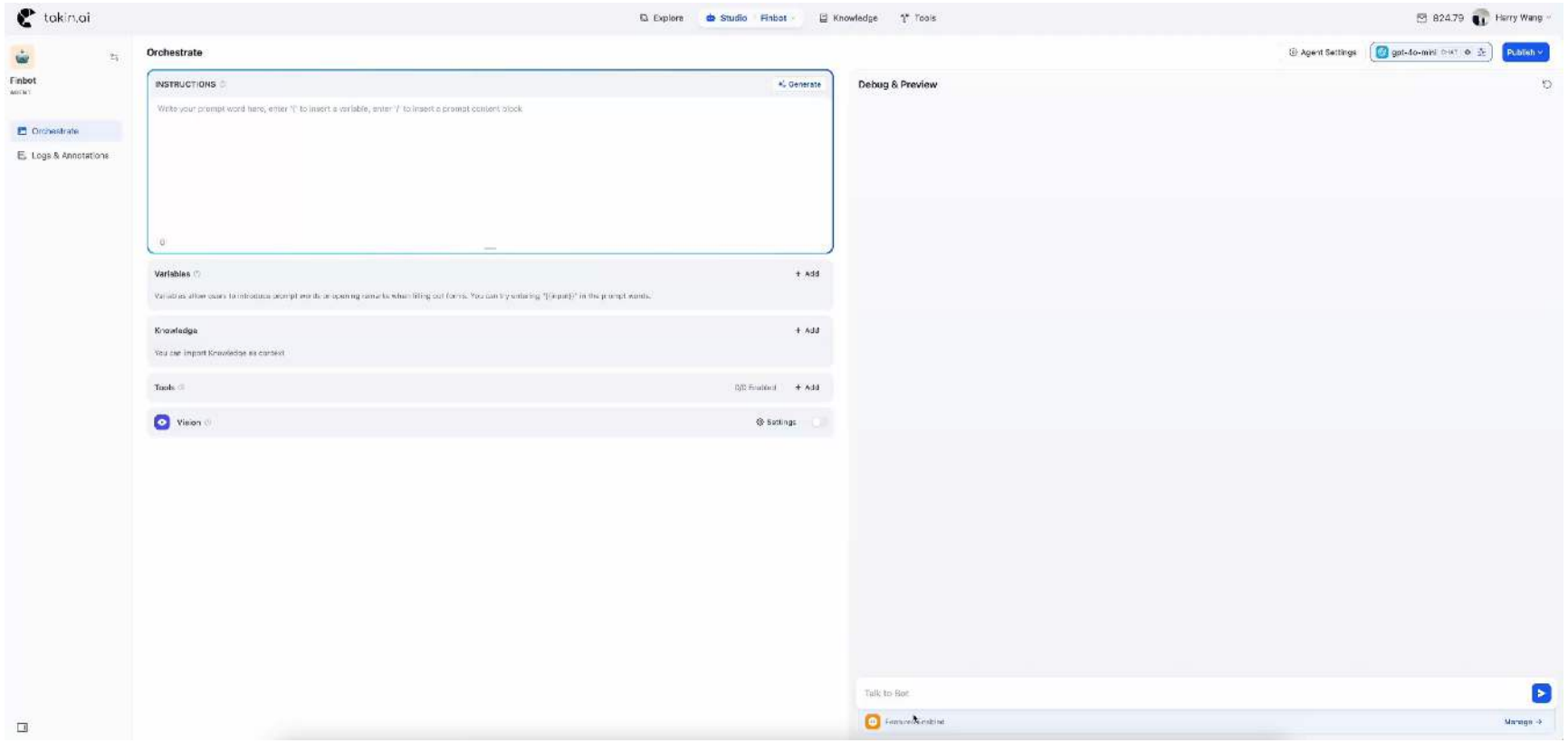
Web search tool



Code execution tool

Analysis	Research	Productivity	Images
• Code Execution • Wolfram Alpha • Bearly Code Interpreter	• Search engine • Web browsing • Wikipedia	• Email • Calendar • Cloud storage	• Image generation (e.g., DALLE) • Image captioning • Object detection

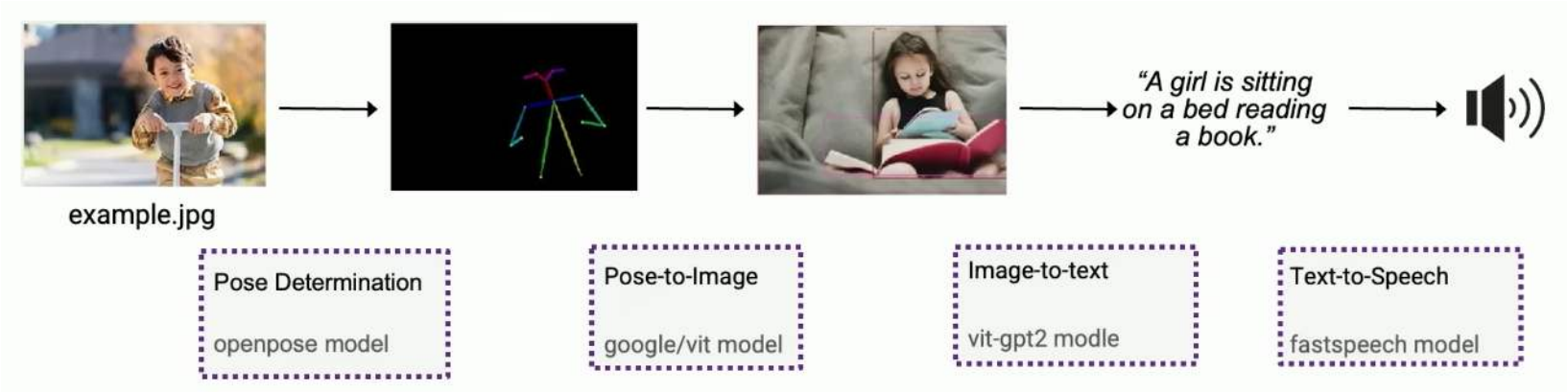
Tool Use Example



Source: Blog on “AI Agents & Agentic Workflows: The Next Frontier in AI” from Harry Wang

Planning

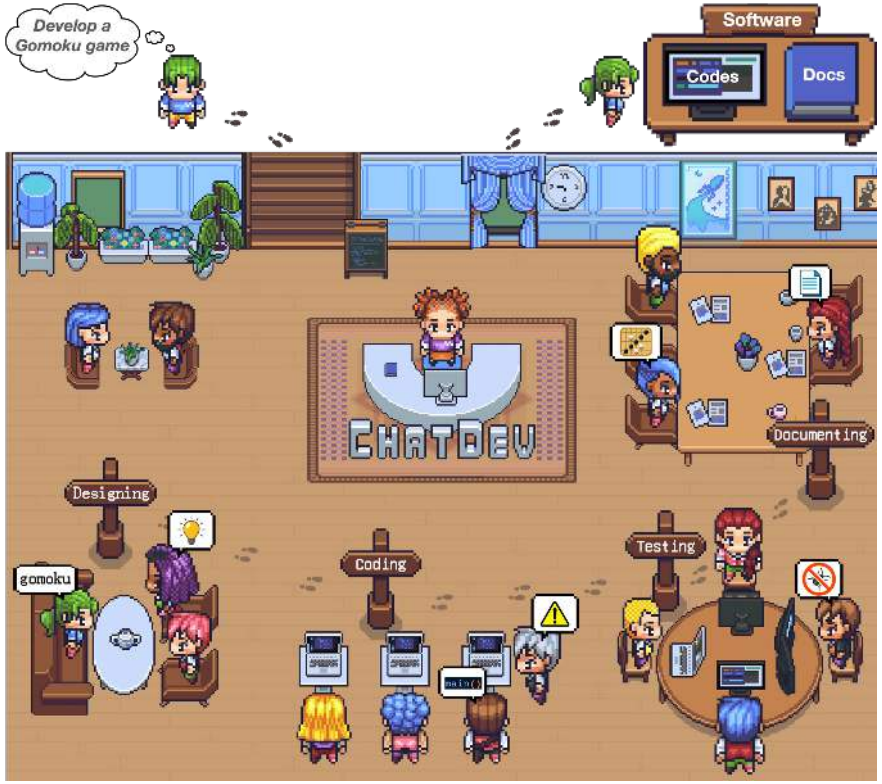
Request: Please generate an image of a girl reading a book in the same pose as the boy in *example.jpg*. Then, describe the new image with your voice.



Sources:

- "What's Next for AI Agentic Workflows," Andrew Ng, March 26, 2024
- Shen, Y., Song, K., Tan, X., Li, D., Lu, W., & Zhuang, Y. (2023). HuggingGPT: Solving AI tasks with ChatGPT and its friends in Hugging Face. Advances in Neural Information Processing Systems, 36, 38154-38180.

Multiagent Collaboration



Model	Arithmetic (%) ↑	Grade School Math (%) ↑	Chess (ΔPS) ↑
Single Agent	67.0 ± 4.7	77.0 ± 4.2	91.4 ± 10.6
Single Agent (Reflection)	72.1 ± 4.5	75.0 ± 4.3	102.1 ± 11.9
Multiagent (Majority)	75.0 ± 3.9	81.0 ± 3.9	105.2 ± 5.9
Multiagent (Debate)	**81.8 ± 2.3**	**85.0 ± 3.5**	**122.9 ± 7.6**

Multiagent Debate: Du et al. (2023)

Sources:

- Qian, C., Liu, W., Liu, H., Chen, N., Dang, Y., Li, J., ... & Sun, M. (2023). Chatdev: Communicative agents for software development.
- Du, Y., Li, S., Torralba, A., Tenenbaum, J. B., & Mordatch, I. (2023). Improving factuality and reasoning in language models through multiagent debate. In Forty-first International Conference on Machine Learning.

Predicted Trend from Andrew Ng in Mar 2024

- The set of tasks that AI can do will expand dramatically because of agentic workflows
- We have to get used to delegating tasks to AI agents and patiently wait for a response
- Fast token generation is important. Generating more tokens even from a lower quality LLM can give good results
- If you're looking forward to running GPT-5/Claude 4/Gemini 3 (zero-shot) on your application, you might already be able to get better performance with agentic reasoning on an earlier model

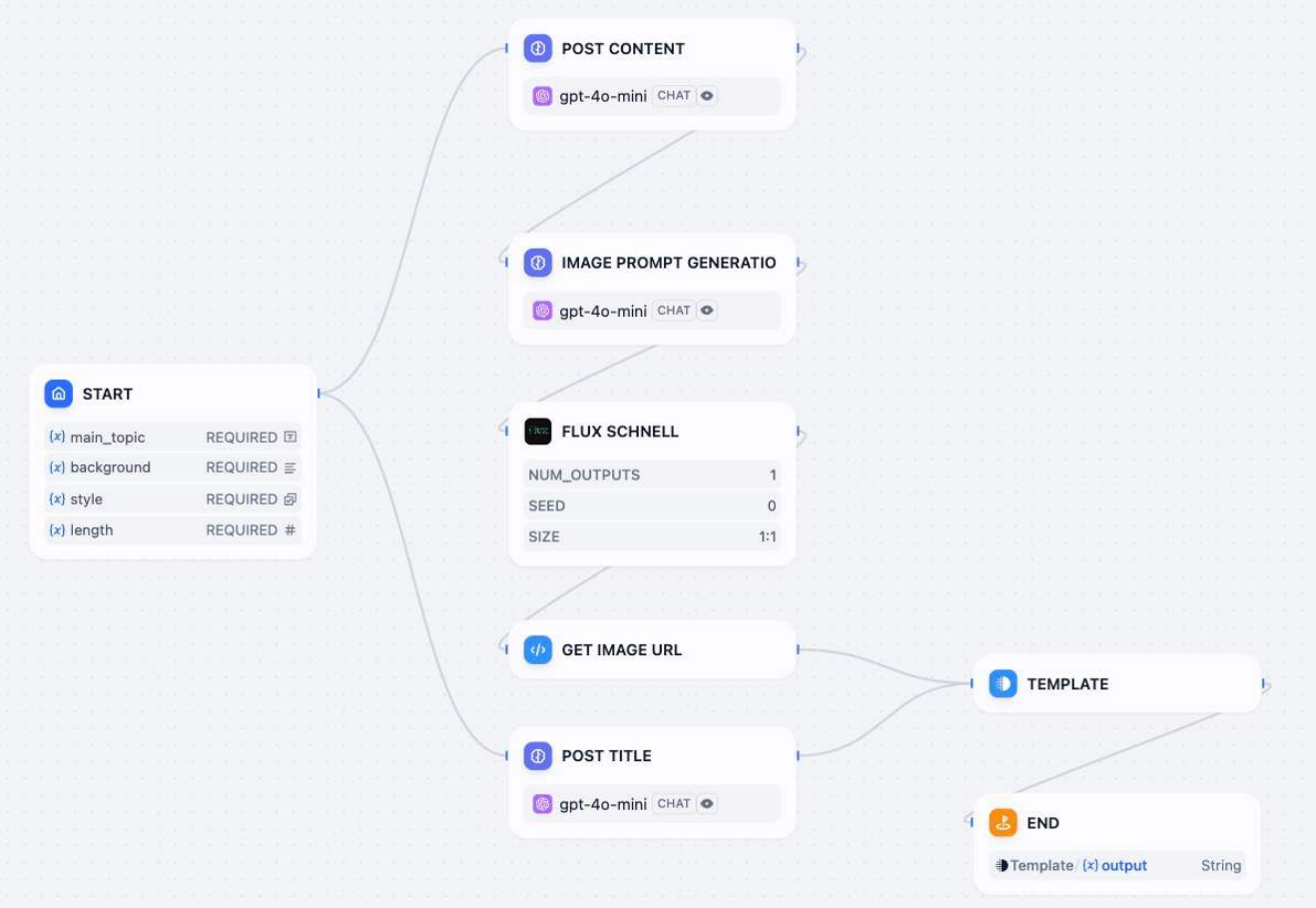
Practical Use Cases

Translation Agents based on No-Code LLM Workflow

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

Instagram Post Generator

Agentic Workflow Tutorial: Instagram Post Generator [Link]



State-of-the-Art: OpenAI Deep Research

OpenAI

An agent that uses reasoning to synthesize large amounts of online information and complete multi-step research tasks for you [OpenAI Deep Research]

- Why Deep Research is an LLM Agent
  - Requires multiple steps (e.g., gathering data, comparing sources, summarizing findings)
  - Benefits from reasoning (e.g., evaluating source credibility, spotting contradictions)
  - Needs memory and iteration (e.g., refining questions based on early results)
- Example: Stock Analysis in the Age of Generative AI [Link to Deep Research]

Predict the expected movement of S&P 500 Index (SPX) over the next month as bullish, bearish, or uncertain. Predict as:

- Bullish if you predict a potential increase in the stock price,
- Bearish if you predict a potential decrease,
- Uncertain if the prediction is neutral

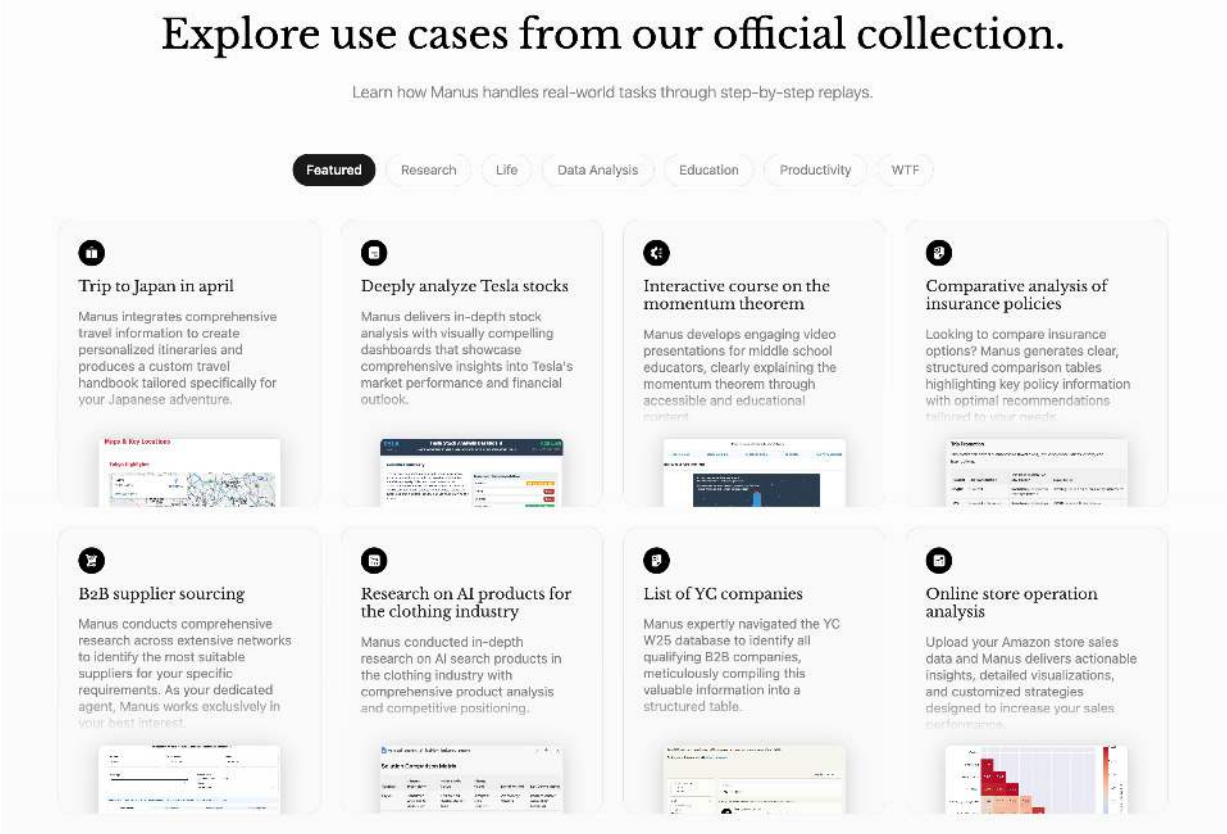
State-of-the-Art: Windsurf Cascade

A code agent, equipped with knowledge, tools, and human collaboration, operates in full sync with the developer while also being empowered to act with maximum independence [Cascade]

- Agentic abilities
  - Suggest and run commands with full contextual awareness
  - Web search
  - Multimodal reasoning: Images and text
- Example: Build a course website for FA690 Machine Learning in Finance

State-of-the-Art: Manus AI

Manus is a general AI agent that bridges minds and actions: it doesn't just think, it delivers results [Manus AI]



Final Words on Generative AI

- **GenAI-First Mindset**
  - There’s no doubt that generative AI increases productivity
  - Today’s workforce requires you to know how to work with generative AI
- **Do Your Own Due Diligence**
  - Generative AI can make mistakes (hallucination)
  - You are responsible for the final delivery
- **Stay Open-Minded and Curious**
  - Generative AI is a fast-evolving field
  - Read and try what’s new and trendy

Acknowledgement

The lecture note has benefited from various resources, including those listed below. Please contact Zonghao Yang (zyang99@stevens.edu) with any questions or concerns about the use of these materials.

- “AI Agents & Agentic Workflows: The Next Frontier in AI,” March 12, 2025, Harry J. Wang [Blog]
- “What’s Next for AI Agentic Workflows,” Andrew Ng, March 26, 2024 [YouTube]

STEVENS
INSTITUTE OF TECHNOLOGY
1870

THANK **YOU**

Stevens Institute of Technology
1 Castle Point Terrace, Hoboken, NJ 07030