

---

**CS 294/194-280:  
Advanced Large Language Model Agents**

# Teaching Staff

---

- **Instructor: Prof. Dawn Song**
- **(guest) Co-instructors: Dr. Xinyun Chen, Dr. Kaiyu Yang**
- **Head TA: Alex Pan**
- **Readers: Tara Pande, Ashwin Dara, Jason Yan**

# Fall 2024: broad coverage of topics about LLM agents

## Large Language Model Agents MOOC



**Dawn Song**



**Xinyun Chen**



**Denny Zhou**



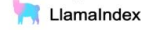
**Shunyu Yao**



**Chi Wang**



**Jerry Liu**



**Burak Gokturk**



**Omar Khattab**



**Graham Neubig**



**Nicolas Chapados**



**Yuandong Tian**



**Jim Fan**



**Percy Liang**



**Ben Mann**



Berkeley RDI

# Fall 2024: LLM Agents MOOC Hackathon Overview

- **Participation:** Close to 3000 developers and AI enthusiasts from around the globe
- **Five Tracks:**
  - Applications: Build cutting-edge LLM agents
  - Benchmarks: Create innovative AI agent evaluation benchmarks
  - Fundamentals: Strengthen core agent capabilities
  - Safety: Address critical safety challenges in AI
  - Decentralized & Multi-Agents: Push the boundaries of multi-agent systems
- **Sponsorship:** Supported by industry leaders such as OpenAI, Google, AMD, Intel, and Amazon with more than \$200,000 in prizes and resources.
- *Hackathon winners to be announced very soon*

## Partners & Sponsors



OpenAI



Google AI



Lambda



intel.tiber.  
AI Cloud



Open  
Philanthropy



Schmidt Sciences



SIERRA

ORBY

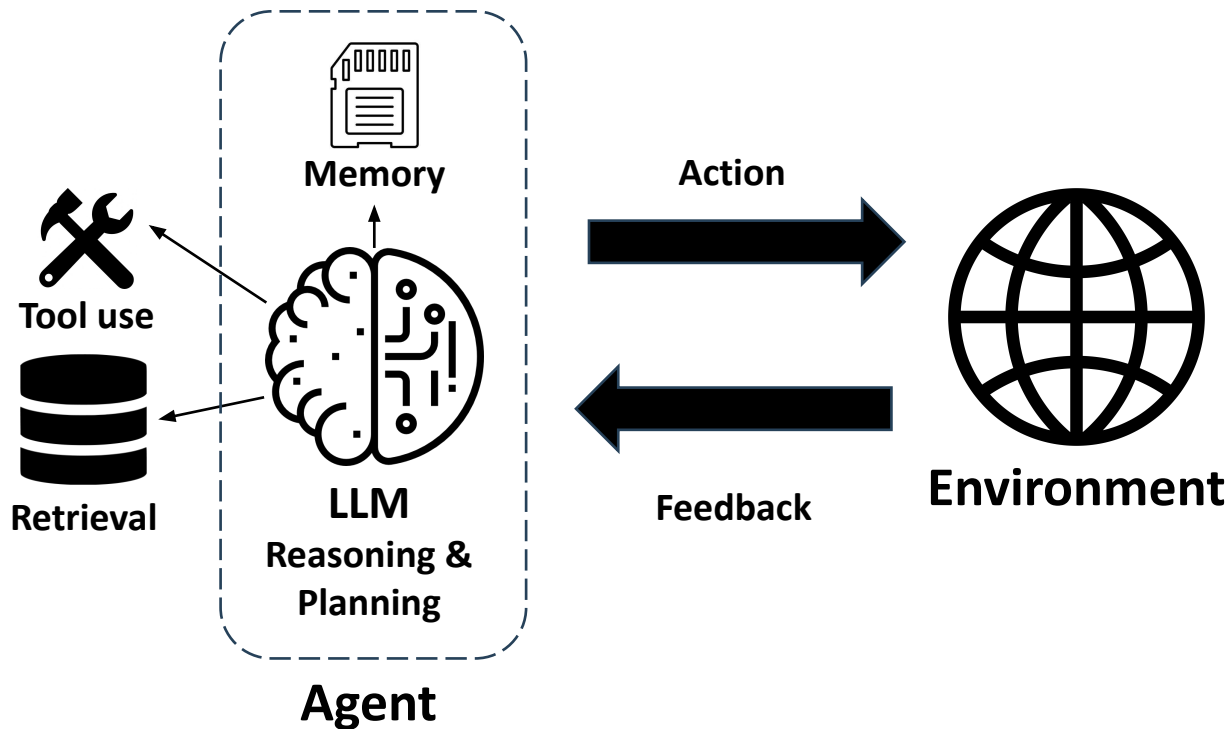
Enterprise AI Automation. No Rules.

servicenow

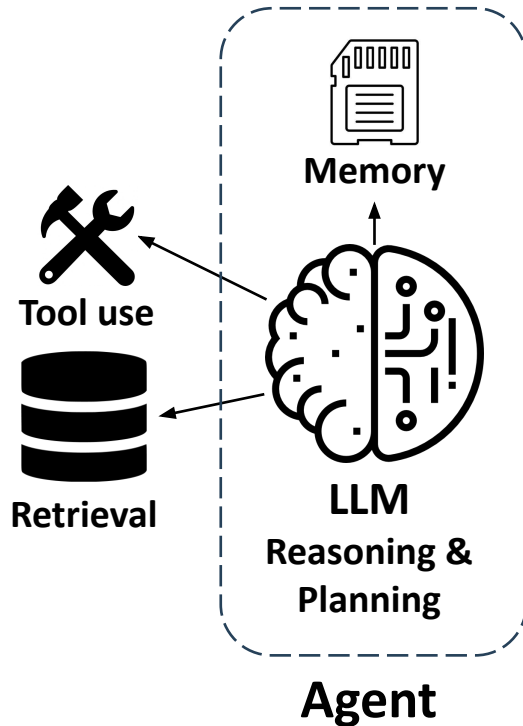
amazon

science

# LLM agents: enabling LLMs to interact with the environment

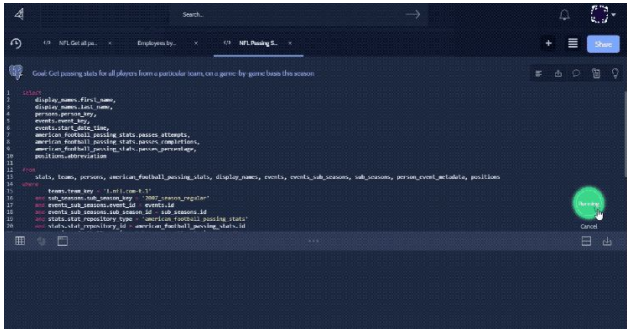


# Why empowering LLMs with the agent framework



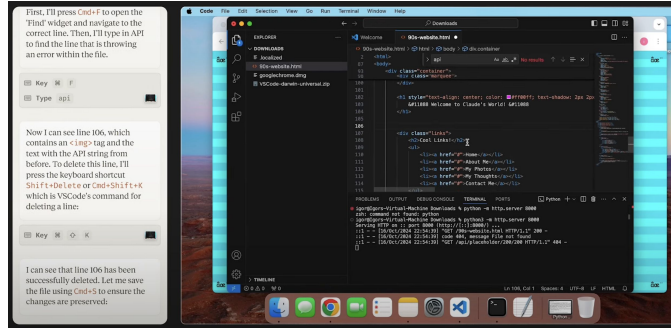
- Solving real-world tasks typically involves a trial-and-error process
- Leveraging external tools and retrieving from external knowledge expand LLM's capabilities
- Agent workflow facilitates complex tasks
  - Task decomposition
  - Allocation of subtasks to specialized modules
  - Division of labor for project collaboration
  - Multi-agent generation inspires better responses

# LLM agents transformed various applications



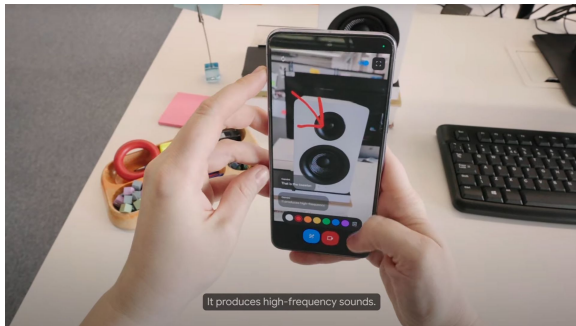
## Code generation

Cursor, GitHub Copilot, Devin, Google Jules...



## Computer use

Anthropic Claude, Google Jarvis, OpenAI Operator



## Personal assistant

Google Astra, OpenAI GPT-4o,...



## Robotics

Figure AI, Tesla Optimus, NVIDIA GROOT...

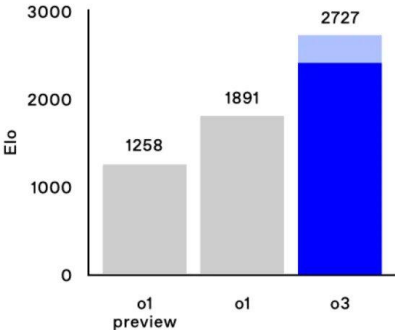
- Education
- Law
- Finance
- Healthcare
- Cybersecurity

...

# Recent rapid progress of reasoning models

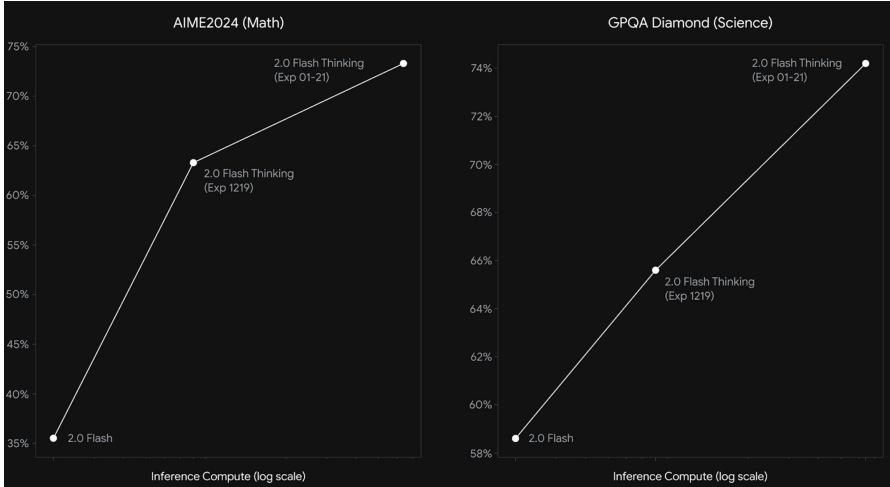
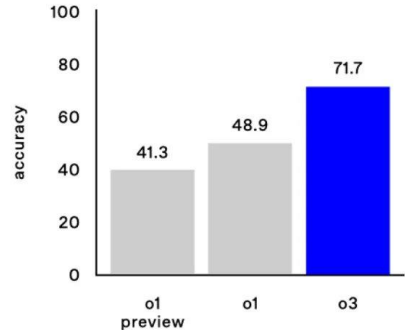
- Sep 2024: OpenAI o1
- Dec 2024: Gemini 2.0 Flash Thinking, OpenAI o3
- Jan 2025: DeepSeek-R1, Kimi k1.5

Competition Code (Codeforces)



Progress from o1 -> o3

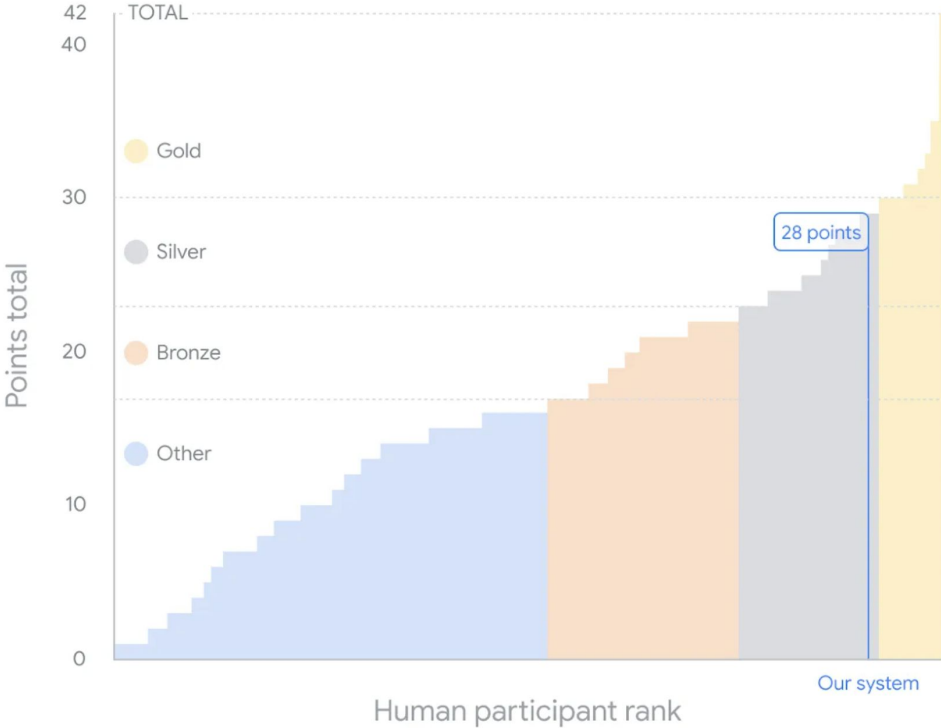
Software Engineering (SWE-bench Verified)



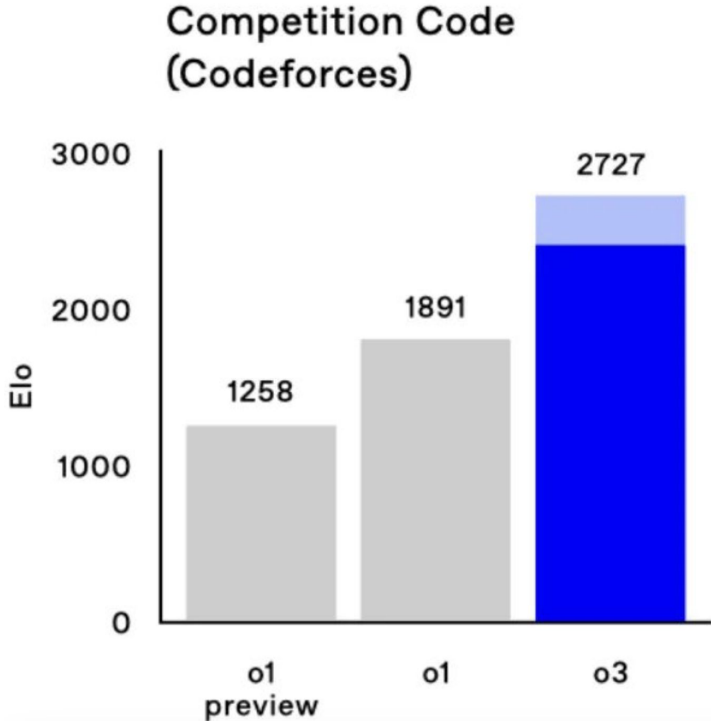
Progress from Gemini Thinking 1219 -> 0121



# Impressive performance on competitive math and coding



Google DeepMind AlphaProof and AlphaGeometry 2 achieve silver-medal performance in IMO 2024



OpenAI o3 is ranked top 200 in Codeforces competitive programming contests

# Topics covered in this course

---

- Fundamental reasoning techniques
  - Inference-time techniques
  - Training techniques
  - Search and planning
- LLMs for software engineering
  - Code generation
  - Code verification
  - Web applications
- LLMs for mathematics
  - Fundamental training techniques
  - Autoformalization and theorem proving
- Agentic workflows, real-world applications
- Safety and ethics

# Course Work

---

- Weekly reading assignments & take-home quizzes
  - Due midnight PT Sunday before the next Monday's lecture
- 1 hands-on lab (released later in the semester)
- Semester-long course project
  - Applications Track
  - Research Track

# Grading

---

lecture attendance & weekly readings/quizzes

+

- 1 unit: article about the topic of a lecture (at least 2 pages)
- 2 units: lab + project (implementation not required)
- 3 units: lab + project with implementation
- 4 units: lab + project with significant implementation and end-to-end demo

# Grading

|                         | 1 unit | 2 units | 3/4 units |
|-------------------------|--------|---------|-----------|
| Participation           | 40%    | 16%     | 8%        |
| Reading Summaries & Q/A | 10%    | 4%      | 2%        |
| Quizzes                 | 10%    | 4%      | 2%        |
| Article                 | 40%    |         |           |
| Lab                     |        | 16%     | 8%        |
| Project                 |        |         |           |
| <i>Proposal</i>         |        | 10%     | 10%       |
| <i>Milestone 1</i>      |        | 10%     | 10%       |
| <i>Milestone 2</i>      |        | 10%     | 10%       |
| <i>Presentation</i>     |        | 15%     | 15%       |
| <i>Report</i>           |        | 15%     | 15%       |
| <i>Implementation</i>   |        |         | 20%       |

# Class Project

---

- **Applications Track:**

- 3-4 students per group
- Focus on applied use cases of LLMs
- Does not necessarily need to contribute novel research

- **Research Track:**

- 2-3 students per group
- Conduct novel research under the supervision of postdocs and graduate students
- Goal of publishing in a workshop or conference
- Students must apply to participate via a forthcoming Google Form

# Timeline

|                            | Released | Due  |
|----------------------------|----------|------|
| Project group formation    | 1/27     | 2/17 |
| Project proposal           | 2/3      | 2/17 |
| Project milestone #1       | 2/17     | 3/24 |
| Project milestone #2       | 3/24     | 4/28 |
| Lab                        | 3/31     | 4/28 |
| Project final presentation | 4/28     | 5/9  |
| Project final poster       | 4/28     | 5/9  |
| Project final report       | 4/28     | 5/16 |

# Course Website

---

<https://rdi.berkeley.edu/adv-ilm-agents/>