Building Software with Agentic Components

CSCI 7000-009
Mon/Wed @5:05pm

Danny Dig

Have you received the welcome email I sent yesterday?

What do you most want to get out of today’s class?

Today’s goals

Why learn how to build agents?

What is Agentic AI: practice and research examples

Discovery: learning about each other, expectations

How can I be successful in CSCI 7000-009?

Why this course now?

LLMs are no longer just:

- Chatbots
- code generators

They are increasingly embedded as:

- Long running services
- Tool-using systems
- Autonomous components

Are we actually building software, or just demos?

Industry reality check

THE AI SUMMIT NEWYORK

Industry reality check

THE AI SUMMIT NEW YORK

When AIs Team Up: The Rise of the Multi-Agent Stack
Next Generation Stage | 25 mins

The Enterprise Agent Factory: Building Scalable, Governed AI Systems
Data Excellence Stage | 25 mins

Creating Delightful Customer Experiences with Voice AI Agents
The AI Playground | 15 mins

From Assistants to Autonomy: The Rise of AI Workers
AI at Scale Stage | 25 mins

Under the Hood of The World's Most Advanced AI Data Agent
Data Excellence Stage | 25 mins

Panel: C-Suite Perspectives on Human-Centered Agentic AI
Headliners Stage | 40 mins

Panel: Rise of the AgentOps Stack: Building for Autonomy at Scale
Next Generation Stage | 40 mins

How to Build Trustworthy, High-Quality AI Agents with MLflow
Workshop Room | 60 mins

Managing the Machines That Manage Themselves: Executive Perspectives on Agentic AI
VisionAIres VIP Lounge | 40 mins

Building Smarter AI Agents with Azure Databases
Data Excellence Stage | 25 mins

Lightweight, Heavy Impact: Leveraging Software Design Principles for Multi-Agent AI Systems
Data Excellence Stage | 25 mins

Accelerating AI adoption in the enterprise: ensuring trust through governed agentic AI
Headliners Stage | 25 mins

Precision Engagement: How Agentic AI and Predictive Models Cut Through the Pharma Noise
The AI Playground | 15 mins

The Autonomous Advantage: Unleashing Your Agentic AI Future in Life Sciences & Healthcare
Health & Pharma Stage | 25 mins

The Agentic Era: What's Really Driving Enterprise AI Adoption
Headliners Stage | 25 mins

From Pilots to Production - Scaling AI Agents in Regulated Environment
Finance Stage | 25 mins

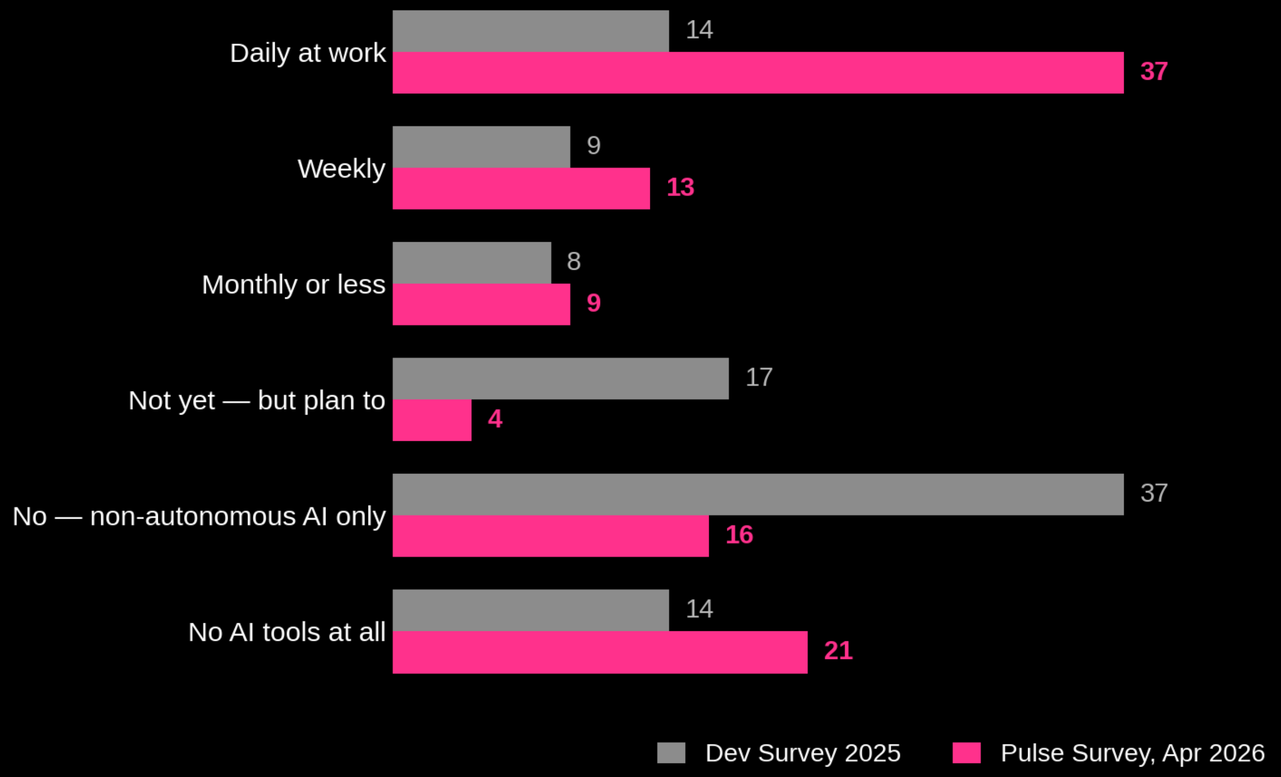
Agentic Commerce and the Future of the AI Economy
Data Excellence Stage | 20 mins

From Workflows to Agents: Transforming Financial Operations
Finance Stage | 25 mins

Industry reality check: 2026

Are you using AI agents in your work?

% of respondents · agent use at work rose from 31% to 59% in one year



But the leash stayed on

63% rarely or never let agents run on full autopilot

68% prefer predictable single-agent setups over multi-agent

60% block agents from unapproved system changes

Source: Stack Overflow pulse survey, “Agents on a leash,” May 27, 2026 · n = 1,100 developers and technology professionals, fielded late April 2026 · 2025 comparison from Stack Overflow Developer Survey 2025 (n = 49,000+)

Where have you personally seen agent-like systems in the wild?

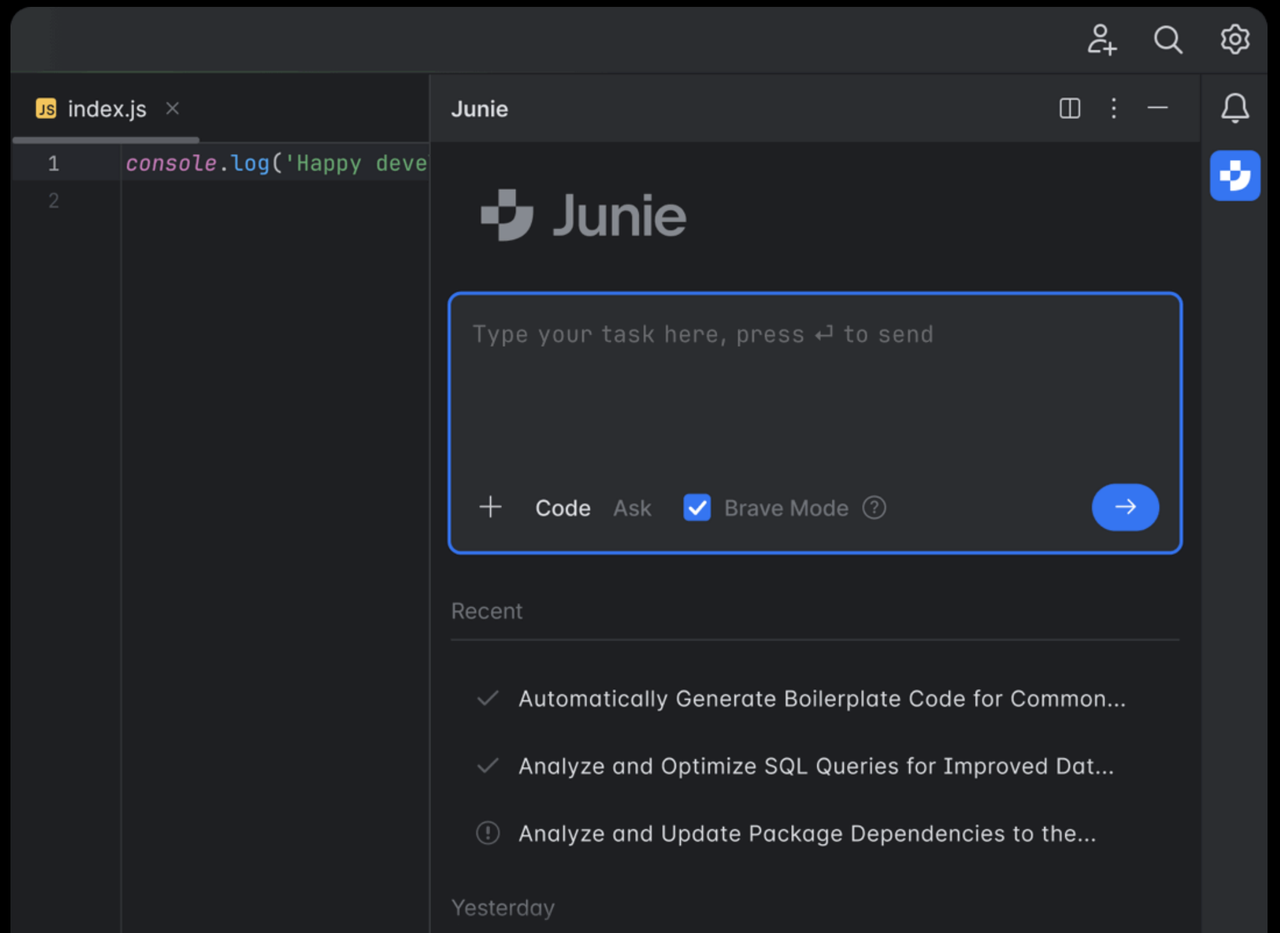
Have you used any AI-powered tools in your coding projects?

Which GenAI tools and services did you use for programming?

Junie

Smart coding agent in the IDE

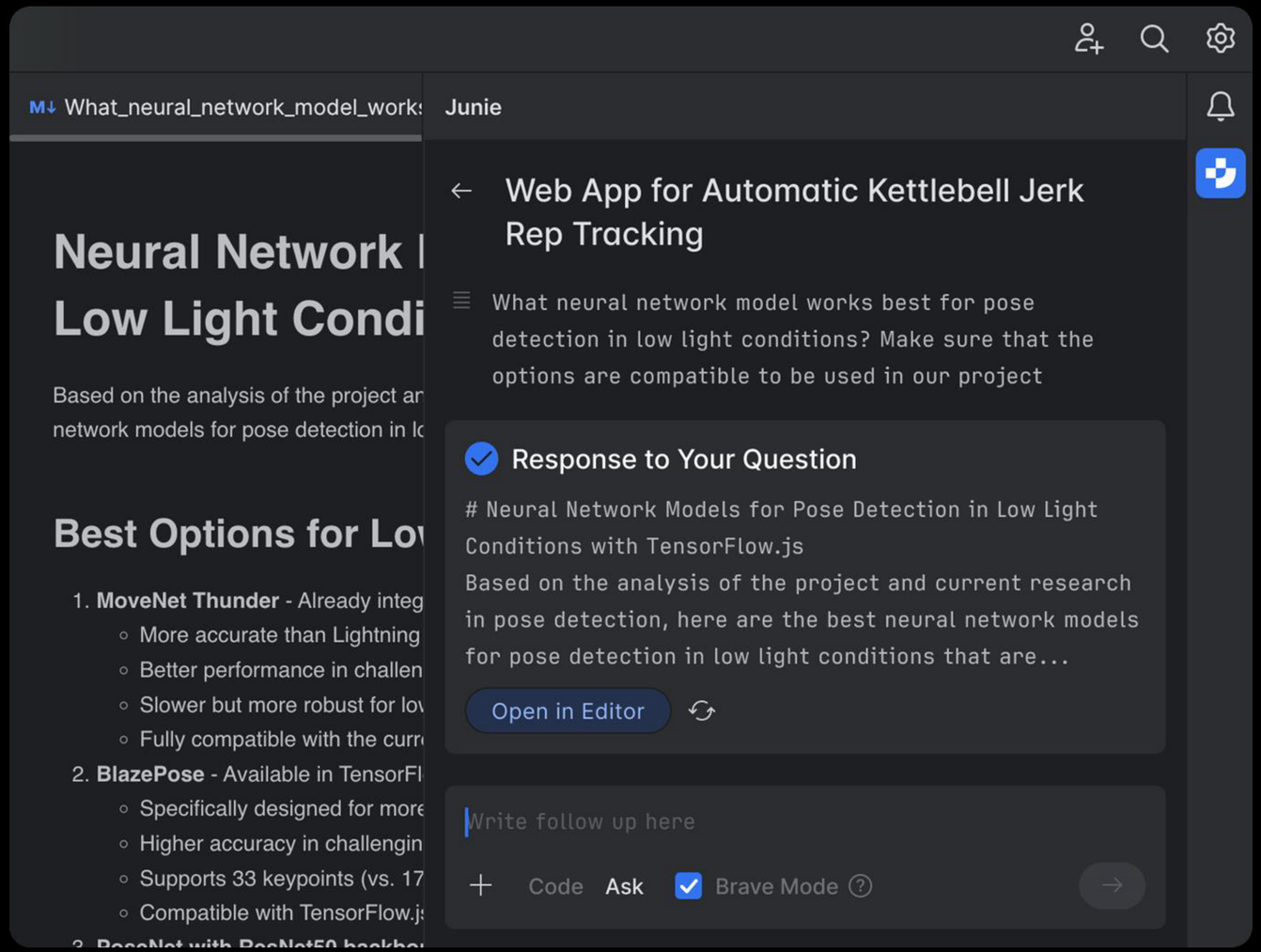
Lighten programmer’s load



Junie

Code and ASK mode

Lighten programmer’s load



Research Examples: Agentic AI for SE

AI Agents and Planning @JetBrains

Objective: advancing technologies around AI agents with the focus on locality

Infrastructure:

- Platform IDEFormer

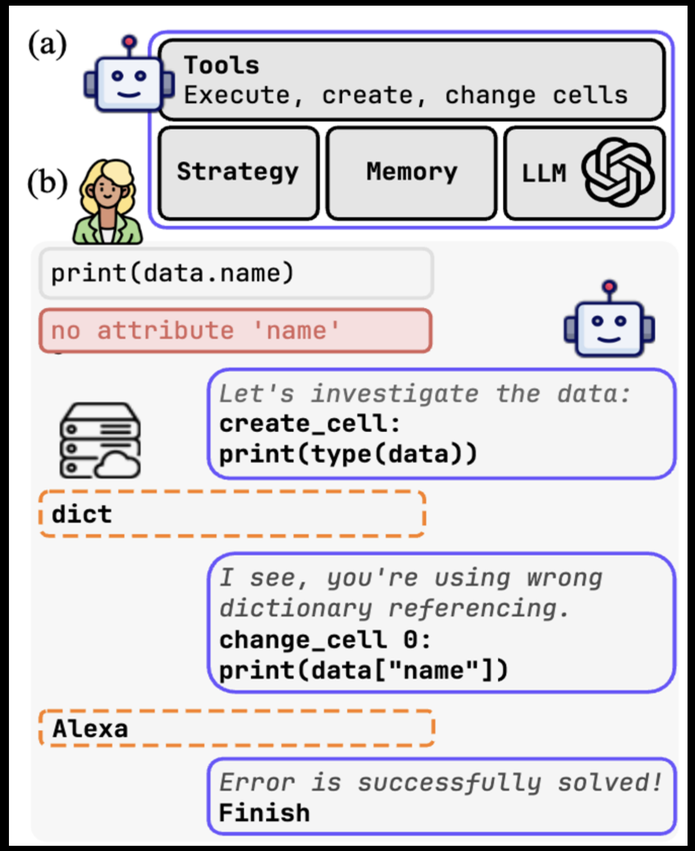
Agents:

- Debugging
- Environment Setup

Research:

- RL to tune models for Environment Setup
- Distillation for on-device models for SWE-agents

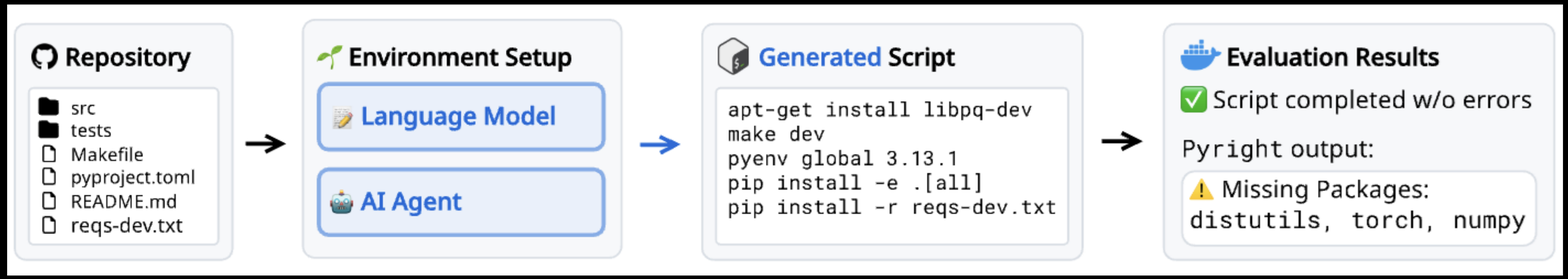
Notebook Debugging (Grotov *et. al*, 2024)



"Debug Smarter, Not Harder: AI Agents for Error Resolution in Computational Notebooks" EMNLP 2024

Setting up Environment (Eliseeva *et. al*, 2024)

EnvBench: A Benchmark for Autonomous Environment Setup



Coordinated Refactoring Agent (Bellur *et al.*, 2026)

Developers perform more than 50% of refactorings in clusters:

- e.g., 243 renames in a single commit in Apache Flink
- time consuming: ~30 minutes to perform cluster rename

Multi-agent solution with **human in the loop**

Evaluating quality:

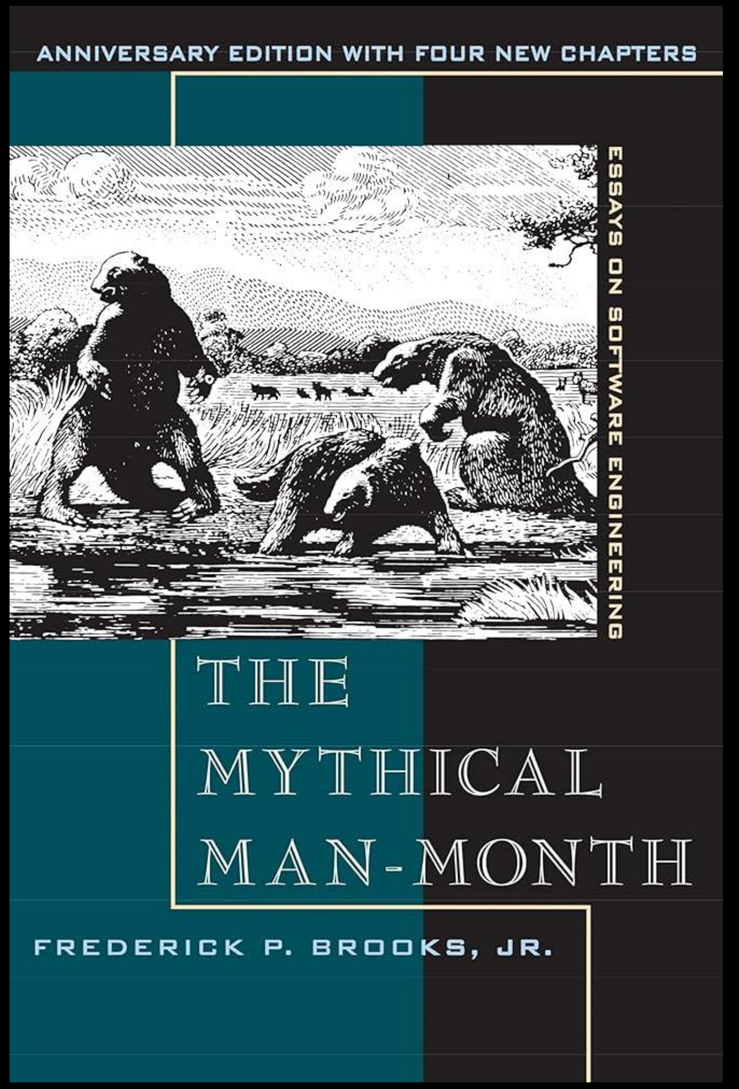
- Replicated 1500+ OSS refactorings, improvement over best baseline is 3x
- **Patches** accepted in open-source projects
- **human in the loop** is crucial to reduce cascading errors by 4x

Why learn to Build AI Agents?

Essential vs accidental complexity

Because agents introduce new software problems

- Control flow you don’t fully own
- State that evolves unpredictably
- Non-deterministic behavior
- Expensive failures at scale



Course Thesis

Agents are not prompts.
Agents are architectural commitments.

LLMs are components, not systems

Autonomy increases both power *and* risk

Robust agents require discipline

What is Agentic AI?

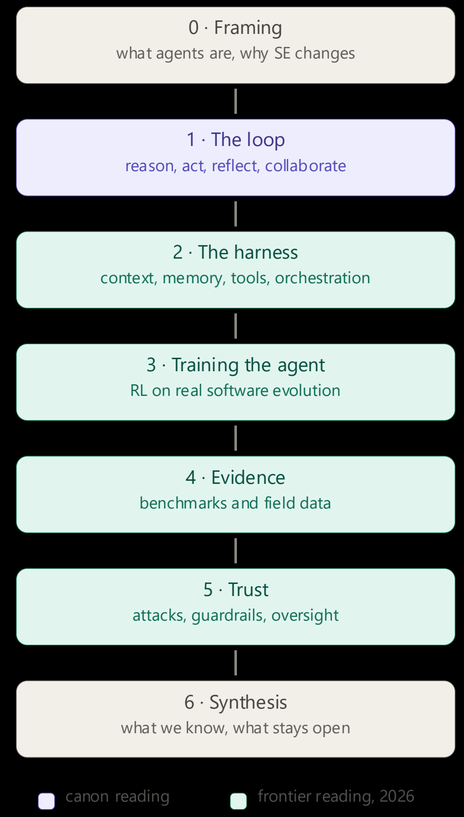
A system is agentic if it:

- Has goals
- Acts in an environment
- Uses tools
- Maintains state
- Operates over time
- Adapts based on feedback

The arc of this course

Seven units. Each one answers a single question.

- **0 · Framing** — what agents are, what SE becomes (weeks 1–2)
- **1 · The loop** — reason, act, reflect, collaborate (weeks 3–4)
- **2 · The harness** — context, memory, tools, orchestration (weeks 5–8)
- **3 · Training the agent** — RL on real software evolution (weeks 8–9)
- **4 · Evidence** — benchmarks and field data (weeks 9–11)
- **5 · Trust** — attacks, guardrails, oversight (weeks 12–13)
- **6 · Synthesis** — what we know, what stays open (week 15)



Two answers to one question

If every frontier model can already reason, what makes one agent better than another?

- Unit 2 answers: **the harness.** Engineering around the model.
- Unit 3 answers: **the weights.** Train the model to be an agent.
- Both are live in 2026. Neither has won.
- Hold the tension. It is the through-line of the semester.

How we read: canon and frontier

Every paper on the schedule carries a label.

- **[C] Canon** — published before 2026. Still assigned in 2030. Read it for concepts.
- **[F] Frontier** — published in 2026. May be wrong by December. Read it for claims under test.
- About 32 papers this term: one third canon, two thirds frontier.
- Your critique should look different depending on the label.

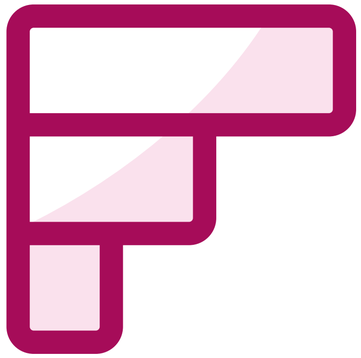
The syllabus will change — on purpose

- Two thirds of this reading list did not exist when I taught this course in the Spring.
- If a better paper appears mid-semester, we swap it in.
- The presenter argues for the replacement. **You curate, not just consume.**
- Structure is fixed. Content is dynamic. That is what a live field feels like.

Getting to know you

What are your expectations from CSCI 7000-009: Building AI Agents?

A. **Do a cool Agentic Project that helps programmers/society**
B. **Learn about exciting Agentic Applications for SE**
C. **Exposure to research on Agentic AI for SE**
D. **Learn about the Challenges in building Agentic software**
E. **How to provide guardrails for Agentic AI?**
F. How to use agents to write software that isn't overly complex + bloated
G. When NOT to use agents
H. How to enable determinism with AI Agents
I. How to efficiently integrate feedback (like RL) to train the agent
J. Better memory/context management for long running sessions or over multiple sessions
K. How we evaluate AI agents as reliable or good enough for use?
L. when should we use verification Agent
M. How we make agents reliable over long context periods without supervision
N. How to optimize cost and token usage by building better agentic architectures.



Rank these topics based on what you want to learn most in this course.

Course Administration

Check webpage:

https://danny.cs.colorado.edu/courses/csci7000-005_Fa26

Work items due before next class Wed (08/26):

- **Familiarize with class webpage**
- **sign up on Piazza (all communications through Piazza, no email after this week)**
- **Read and write critiques for two research papers (see template on webpage). Submit on Canvas before 12pm.**

Check prerequisites: computing background (either practical experience or undergrad-level knowledge of SE, Systems, ML), please check with me after the class

Boulder

CS 7000-009 is Different!!!

This class is NOT:

- an introduction to Software Engineering (take instead CSCI 5828)
- a tutorial on GenAI APIs/tools (see free resources on the internet)
- a prompt-engineering class
- a LangChain/MCP tutorial
- a chatbot-building workshop
- an asynchronous course (you need to participate live each class)

This class IS:

- a grad level seminar on agentic software systems
- a place to build serious agentic prototypes
- a space for critique and experimentation

Boulder

CS 7000-009 is Different!!!

Research-based course:

- **at times it would feel it is not ”organized”**
- **there are lots of choices, you need to select**
- **structure is fixed, but content is dynamic**

Complete a research or industrial-novel project of your choice (teams of 3-4 students) [50% of grade]

- **follow the iterative steps of open-ended/risky research (proposal, fit in literature, evaluate empirically)**

- **at the end of semester you produced a research paper that you can submit to conference, showcase portfolio**

- **WHY: equips you to lead novel R&D**

Boulder

Projects on Agentic AI

Showcase in your job application portfolio

For new grad student, project gives ideas for dissertation

For experienced grad student, project advances your research

Technological shifts/opportunities for Agentic SE:

- any process & tool we use in software development lifecycle (requirements, design, coding, testing, evolution)

Academic-research: generate new insights and knowledge

Industrial-innovation: application of established practices to produce a novel agentic software

Research projects (not an app), teams of 3-4 people

Boulder

CS 7000-009 is Different!!!

Participate in live class discussion and activities. [10% of grade]

Read 1-2 research papers for every class meeting (11 pages each, double column => total of 500+ research pages)

- later on, you choose papers that match your project

Paper Critiques: for each class meeting, for each research paper, submit before class (by 12pm) [20% of grade]

- WHY: equips you with critical thinking

Research presentation: you prepare and deliver for the selected research papers [20% of grade]

- WHY: equips you to communicate your ideas

Boulder

What Happens Next

Next class:

- **What is an agent, really?**
- **Architectures vs prompts**
- **First research paper discussion**

Before next class:

- **Read the webpage, signup on Piazza**
- **Read two papers**
- **Submit Paper Critique + Questions in Canvas (by Wed at 12pm)**

Office hours: right after the class

Boulder

Testimonials

“It was difficult to balance implementing and testing solutions and reading papers. In the end, I think both are really useful. I appreciated the emphasis on papers in this course.”

— Spring 2026 student

“Final project felt a little too demanding at points, but overall a great class. Super happy with the outcomes.”

— Spring 2026 student

“I enjoyed when I had to present them myself because that led to an even deeper understanding. Discussions after the paper presentation were very fruitful. The project made me confident in my ability to build agentic systems!”

— Spring 2026 student

“The first time when I’ve felt that I am learning the current, or more like emerging, technologies. Very cool feeling.”

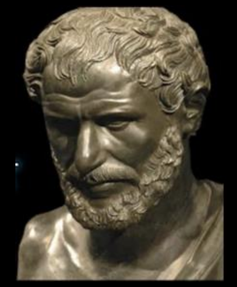
— Spring 2026 student

“For someone who hates sitting in class, this was something I did look forward to week after week.”

— Spring 2026 student

Boulder

Executive Summary



Change is the only guaranteed constant

Agentic AI + Human

“...Game changer! Enter the AI assistant, the great equalizer.” -Lenovo CTO

“AI **will not** displace programmers. But programmers who use AI **will** displace programmers who do not use AI.” – Prof Danny Dig

Together We Go Further