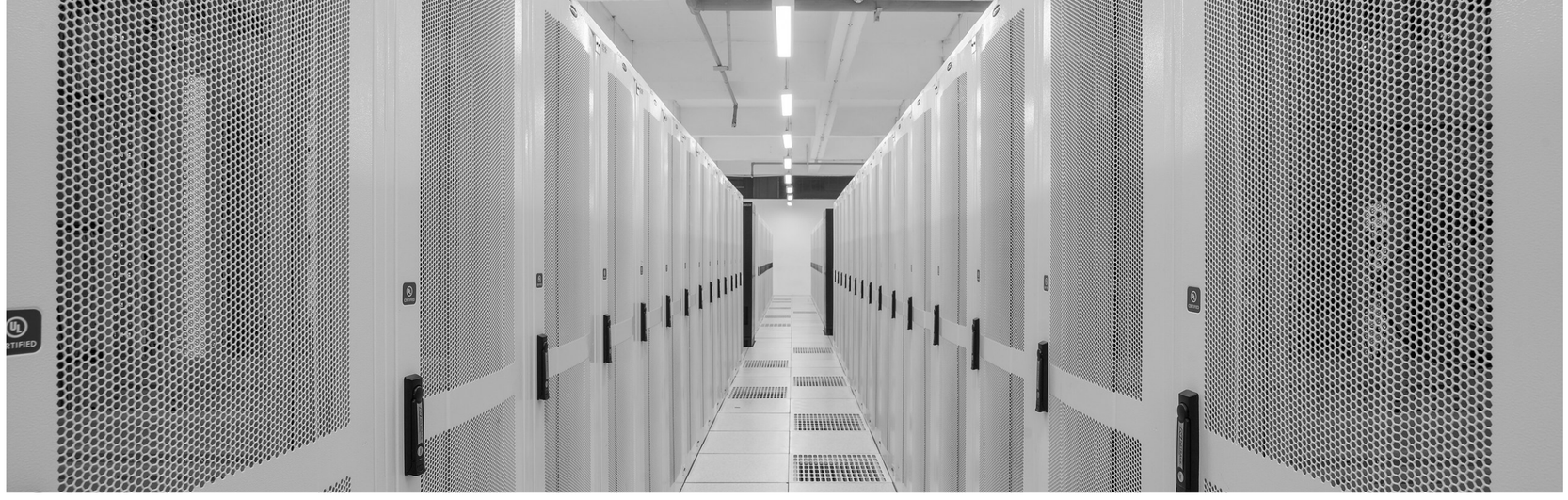


THE FINANCE CURRICULUM IN THE AGE OF AI

KERRY BACK, RICE UNIVERSITY
DEPARTMENT OF FINANCE, UNIVERSITY OF ZURICH

NOVEMBER 4, 2026

OUTLINE

1. Critical thinking
2. Strong fundamentals
3. AI expertise

JOB POSTINGS: USING AI

Phrase	Posting	Experience
“AI tools in your work or school projects” “hands-on experience with AI-powered tools from day one”	PIMCO, 2027 summer intern, Munich	Intern
“sound judgement in technology-enabled decision making”	BlackRock, 2027 quantitative masters intern, London	Intern
“proactive about using AI where possible”	RA Capital, junior associate, Boston	About 2 yrs
“AI-forward people”	RA Capital, healthcare associate, Boston	2-5 yrs; PhD or MD
“AI tools to systematize workflows and models”	Ares, analyst, New York	1-3 yrs
“Translate complex investment problems into structured AI workflows”	T. Rowe Price, investment AI analyst, Baltimore and London	5+ or 10+ yrs

Job postings, checked October 31, 2026

JOB POSTINGS: BUILDING TOOLS

Phrase	Posting	Experience
“leveraging tools like Claude Code”	D. E. Shaw, AI product analyst, Denver	None stated
“custom prompts, skill libraries, and automated workflows”	Schonfeld, AI strategy analyst, London	3-5 yrs
“we will teach the domain to the right technical candidate”	Schonfeld, AI strategy analyst, London	3-5 yrs
“It is a building role.”	Two Sigma, AI solutions developer, New York	1+ yrs; 5-10 preferred
“the judgment to figure out what to build in the first place”	Two Sigma, AI solutions developer, New York	1+ yrs; 5-10 preferred
“converting complex workflows into logical, step-by-step tasks”	T. Rowe Price, investment AI analyst, Baltimore and London	5+ or 10+ yrs

Job postings, checked October 31, 2026

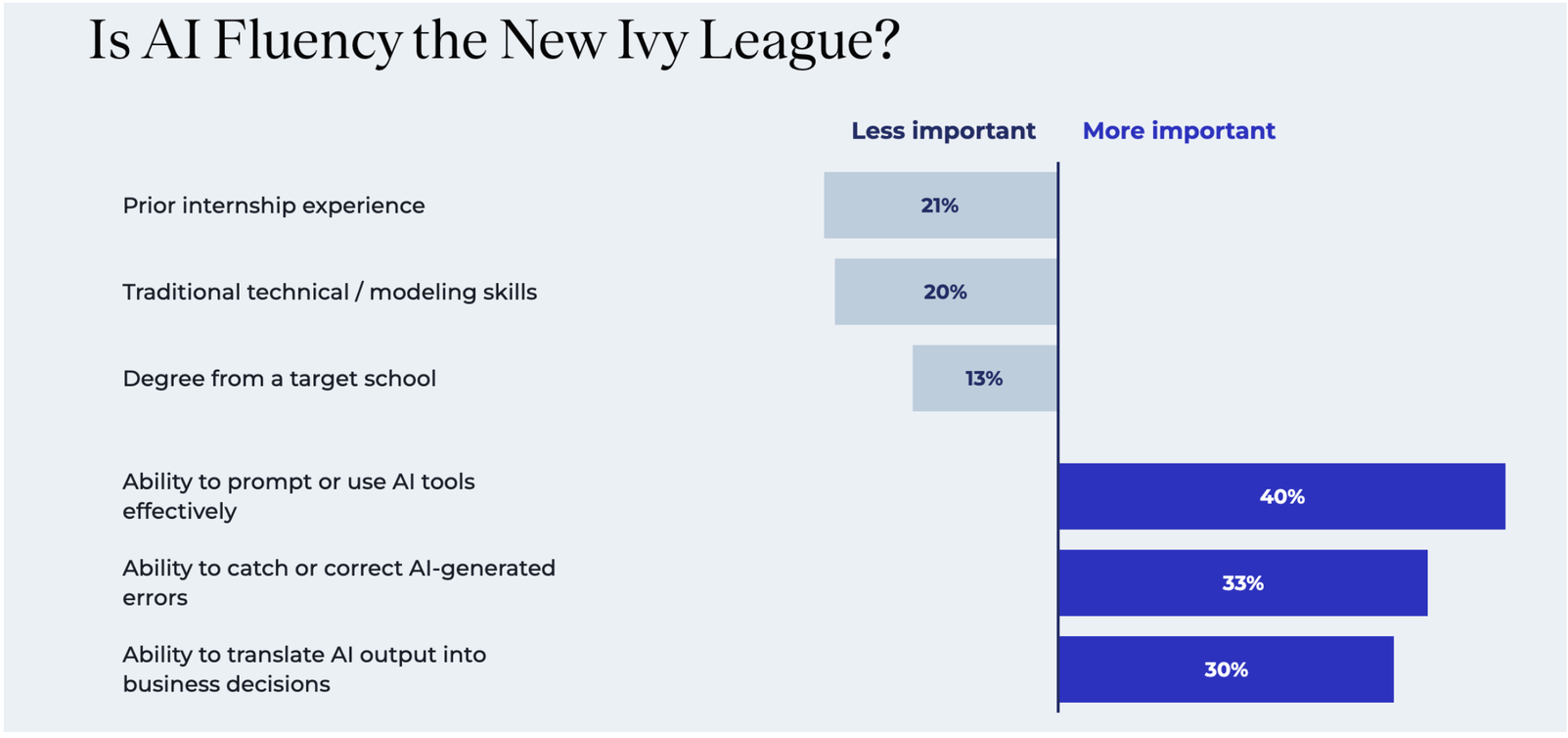
JOB POSTINGS: CHECKING THE OUTPUT

Phrase	Posting	Experience
“limitations, confidence levels, and interpretation risk”	Wellington, data science analyst, Boston	3-5+ yrs
“how these tools have evolved and where they break down”	Two Sigma, AI solutions developer, New York	1+ yrs; 5-10 preferred
“robust guardrails, compliance checks, and observability frameworks”	T. Rowe Price, investment AI analyst, Baltimore and London	5+ or 10+ yrs
“evaluation, monitoring, and human-in-the-loop patterns”	Carlyle, AlpInvest engineer, New York	8+ yrs
“appropriate audit trails”	Schonfeld, AI strategy analyst, London	3-5 yrs

Job postings, checked October 31, 2026

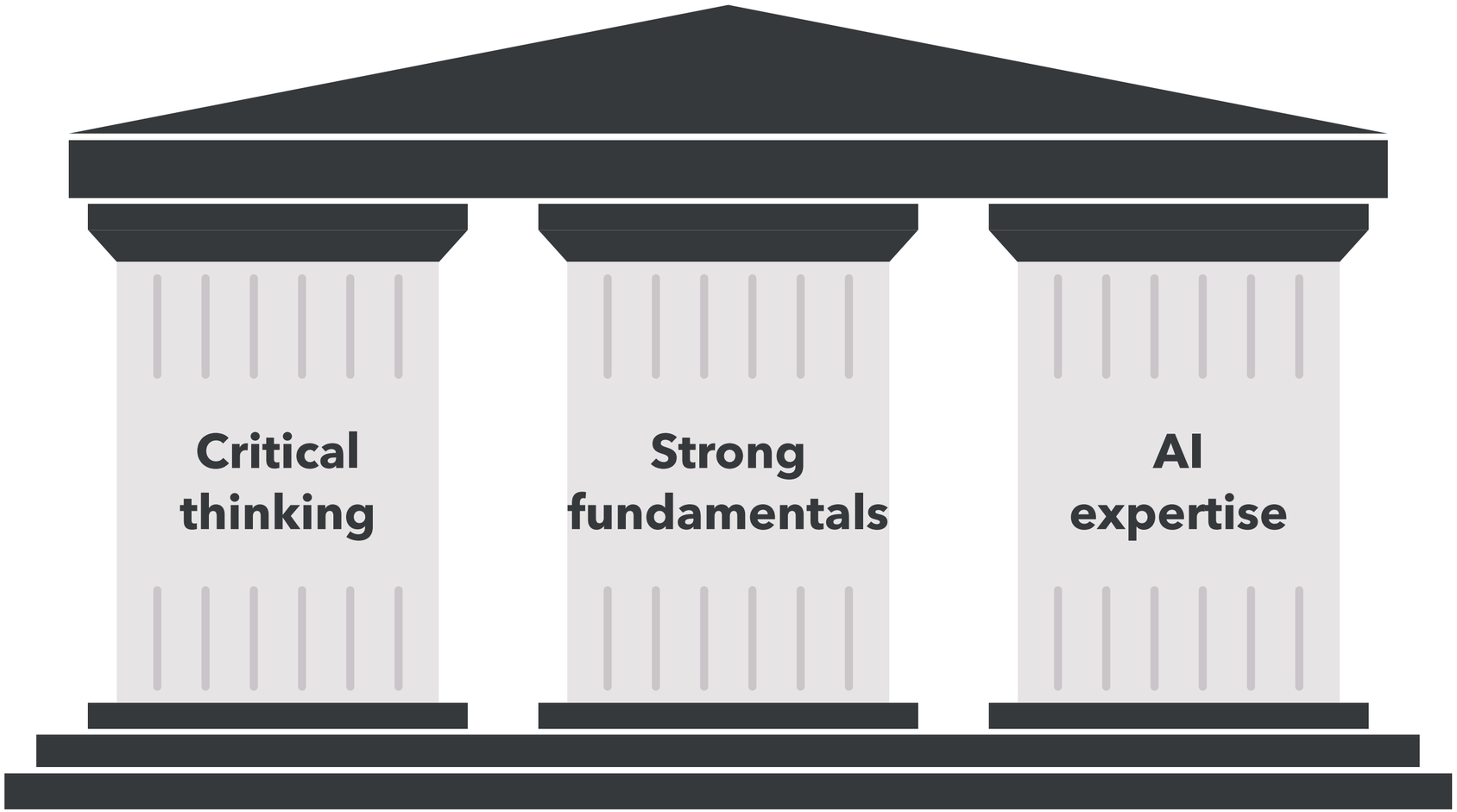
SURVEY OF FINANCE RECRUITERS

Percent of finance HR professionals who said each factor will be more important, or less important, when evaluating candidates in their organization's next hiring cycle.

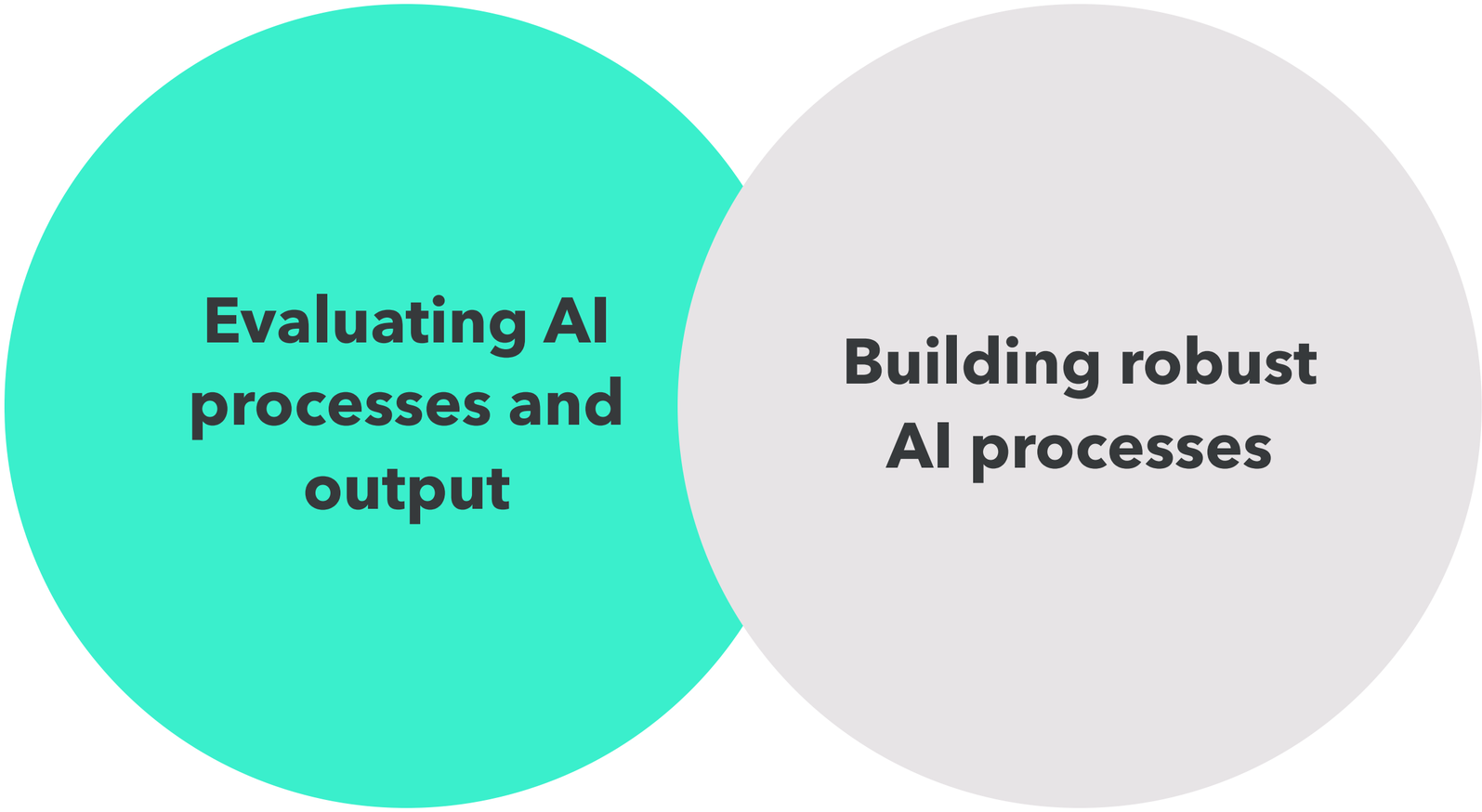


Selby Jennings: The AI Fluency Gap. Research conducted by Censuswide, sample of 500 US HR Professionals in the finance industry in August, 2026.

PILLARS OF A FINANCE EDUCATION



ELEMENTS OF AI EXPERTISE



ISSUES

1. How to teach critical thinking now?
   How to use AI and how to prevent reliance on AI?
2. How to teach fundamentals now?
   Build the model by hand, or have AI build it and check it?
3. How to teach building AI processes and evaluating their output?

FROM MAKERS TO CHECKERS

> Waldron said that workers would shift from being creators of reports or software updates, or “makers” in his terminology, to “checkers” or managers of AI agents doing that work.

CNBC on JP Morgan Chief Analytics Officer Derek Waldron, September 30, 2025

RECENT AI PROGRESS

TIMELINE OF AGENT TOOLS

Feb 2025
Claude Code
research preview

May 2025
Remote MCP connectors
in Claude; Claude Code GA

Jul 2025
Subagents
in Claude Code

Sep 2025
Claude Agent SDK
renamed from Claude Code SDK

Oct 2025
Skills and plugins
Claude Code on the web

Jan 2026
Cowork
Claude Code for non-coders

CODING CHANGED IN 2025

October 2025

"They just don't work. They don't have enough intelligence … I feel like the industry is making too big of a jump and is trying to pretend like this is amazing, and it's not. It's slop."

January 2026

"I really am mostly programming in English now, a bit sheepishly telling the LLM what code to write… in words."

February–March 2026

"It is hard to communicate how much programming has changed due to AI in the last 2 months … Coding agents basically didn't work before December … I don't think I've typed like a line of code probably since December."

Andrej Karpathy: OpenAI co-founder, former Tesla AI chief, now at Anthropic

LORD ABBETT ON THE DEC 2025 BREAK

Before December 2025, **70–80%** of vague backtest prompts **failed**.

With Claude Opus 4.5 and GPT-5.2-Codex, **~80% worked** the first time, in many cases.

> "If testing an idea used to take a month, you might say it's not worth it. But if AI cuts that to a week or a day, suddenly there are a lot more projects you want to do. So far it hasn't reduced headcount."

Tal Fishman, head of fixed income quant research, Lord Abbett ($248B); AI Street, March 3, 2026

CURRENT BENCHMARK SCORES

Terminal-Bench 4.0: percent of 66 long terminal tasks solved. Vals.ai runs every model in the same harness: the program that sends messages to the model and runs the tools it asks for.

Model	Score	Model	Score
Claude Opus 5.5	65%	Gemini 4 Argon	58%
Claude Sonnet 5.5	64%	GPT-6.1 Sol	55%
Kimi K3.1 Pro	64%	GLM-5.3 (open)	39%
Gemini 4 Carbon	63%	Qwen3.8 Max	39%
GPT-6 Astra	60%	DeepSeek V4.1 Flash (open)	32%
Claude Fable 5.1	58%	Grok 4.7	29%
Muse Wonder Full	58%		

> Kimi K3.1 Pro is an open-weights model, but its weights are not yet released.

Vals.ai, Terminal-Bench 4.0 leaderboard, updated October 29, 2026


Critical
thinking

BANK CEOS ABOUT AI

> I need human judgment, I need human decision making, I need human accountability at the core.

HSBC CEO Georges Elhedery, Bloomberg TV interview, June 8, 2026

> The goal is to remove the repetitive, administrative work so our people can focus on what they do best—we're not replacing critical thinking and judgment, we're removing the tedious, time-consuming work that gets in the way of it.

Customers Bank CEO Sam Sidhu, statement to Fortune, April 28, 2026

AUTOMATE THE MECHANICAL STUFF

- Grab data: numerical and text.
- Display relevant parts and tag judgments to be made.
- Spend class time on judgment parts.

EXAMPLE OUTPUT FROM A DCF TOOL

TWO-STAGE ENTERPRISE VALUATION

Apple Inc. (AAPL)

Fiscal 2025 ended 2025-09-27 is year 0. All figures in USD millions.

▶ 14 judgment calls made without review. Each is flagged where it was made; open to see them all.

Line items | Forecasts | Baseline | DCF

CLASSIFICATION OF EVERY REPORTED BALANCE-SHEET LINE

Which lines consume cash as the business grows (operating assets, operating liabilities, net PP&E), which are claims on what it produces or are added in the bridge (debt claims, equity claims, nonoperating assets), and which do neither (excluded). Everything downstream inherits this, so it is worth disagreeing with here rather than later.

REPORTED LINE	BUCKET	WHY

Open the page. Apple, fiscal 2025 as year 0. Line items, Forecasts, Baseline, and DCF tabs, with 14 judgment calls flagged.

RECOGNIZING AI LIMITATIONS

WHY AI IS WEAKER ON JUDGMENT CALLS

> The gulf between LLM performance on verifiable tasks vs. unverifiable tasks that require judgment is larger than most people realize. Verifiable tasks, like coding or solving math problems, have a right answer that can be checked automatically. Judgment tasks don't have unambiguous, ground-truth, correct answers. But they do have better and worse answers.
>
> We have much better techniques for training AI systems to do well on verifiable tasks. As a result, progress on performing such tasks has been much faster than for judgment tasks.

Andrew Ng, co-founder and former head of Google Brain, co-founder of Coursera and DeepLearning.AI · The Batch, October 16, 2026

AN EXAMPLE OF POOR REASONING

I need to wash my car. The carwash is 200 feet away. Should I walk or drive?

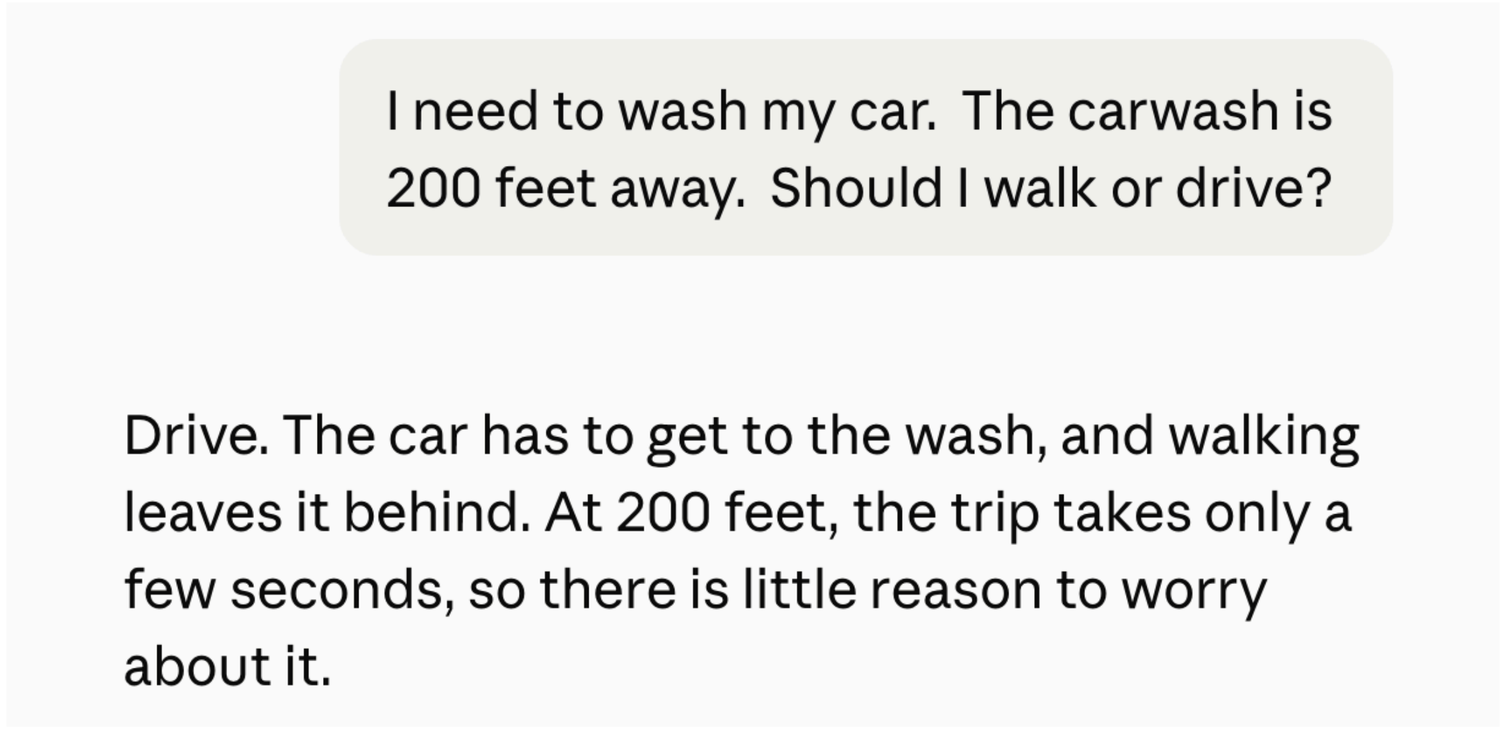
Weighing whether to walk or drive a short distance.

Walk. At 200 feet, it's only a 1-minute walk at normal pace, whereas getting the car started, driving, and parking would easily take 2–3 minutes. You'll also get a few minutes of movement in, and the car won't pick up any meaningful dirt over that distance anyway.

The only reason to drive would be if weather is genuinely unpleasant (though October in Houston is usually fine for a quick walk) or if you needed the car positioned a certain way for the wash.

Claude Haiku 4.5, an older version of Anthropic's smallest model, October 17, 2026

NEWER MODELS ARE BETTER



Claude Haiku 5.5, October 17, 2026

ANTHROPIC EXPERIMENT

1. 50 causal-inference practitioners analyze observational data on a fictional farm-advisory program, working with Claude Code (Opus 4.8), and recommend roll out or discontinue.
2. Four analytical decisions (productivity measure, identification strategy, estimand, weighting) and the final recommendation each have two defensible options.
3. For each decision, a coin flip sets which option the agent leans toward. A hidden system prompt tells the agent to hold that lean strongly and voice it.
4. The agent makes the best honest case for its lean. Participants can choose any option.

Herdağdelen, Joren and Tamkin (2026), "Who's Aligned to Whom? How AI Agents Shape Experts' Research Judgment Calls," Anthropic.

RESULTS

86%

of analytical decisions went with the agent's lean (50% by chance)

77%

repeated the agent's talking points

15%

mentioned the agent

- The four analytical leans alone predicted 82% of final recommendations, against a 61% baseline (exploratory).
- The agent's lean on the final recommendation moved it 17% of the way from discontinue to roll out (one-sided $p = 0.07$).
- Participants who took a side and then heard the agent lean the other way switched 60% of the time.
- More than half rated the recommendation as more their own than the agent's.
- Without the agent, 6 of 11 analysts made errors that invalidated their results, against 3 of 50 with it. The 11 chose not to use the agent, so this comparison is not randomized.

Herdağdelen, Joren and Tamkin (2026), "Who's Aligned to Whom? How AI Agents Shape Experts' Research Judgment Calls," Anthropic.

WHAT DOESN'T AFFECT RELIANCE ON AI

- An AI writing assistant (GPT-3.5 and GPT-4) offered autocomplete suggestions with a set position on societal issues.
- 2,582 participants. Their attitudes shifted toward the assistant's position.
- Warning participants about the bias, before or after writing, did not reduce the shift.

Williams-Ceci et al. (2026), "Biased AI Writing Assistants Shift Users' Attitudes on Societal Issues," Science Advances.

WHAT DOES AFFECT RELIANCE ON AI

What the model did	Study	Finding
Stated uncertainty in the first person (“I’m not sure, but…”)	Kim et al. (2024), 404 participants, medical questions	Agreed with the model less and answered more accurately
Cited sources	Kim et al. (2025), 308 participants	Relied less on incorrect answers
Explained its answer	Kim et al. (2025), 308 participants	Relied more on both correct and incorrect answers
Added a disclaimer, highlighted uncertain words, or implied the answer without stating it	Bo, Wan and Anderson (2025), 400 participants	Relied less on wrong answers, but also less on right ones

All three studies use questions with a correct answer.

A POSSIBLE CLAUDE.MD SECTION

```
## Judgment calls

A judgment call is a decision with more than one defensible answer:
a method, an assumption, an interpretation, or a recommendation.

1. When you reach a judgment call, run each option and check
   whether the result changes.
2. If the result changes, flag the judgment call.
   Make the strongest possible case for each option,
   with equal effort.
3. If the result does not change, list the judgment call
   at the end of your answer.
```


Strong
fundamentals

TEACHING FUNDAMENTALS NOW

Writing the code or building the Excel model yourself may teach the concepts at a deeper level.

Or the concepts can be understood without the mechanics, and the mechanics are just distracting.

- If we can go faster, we can go deeper. At what cost?
- Assessment is difficult: completing out-of-class assignments is not a good signal of mastery.

EVIDENCE ON LEARNING WITH AI

WRITING ESSAYS WITH CHATGPT

83% vs 11%

could not quote a sentence from the essay they had just written: ChatGPT group vs the other two groups

0 of 18

ChatGPT users quoted their own essay correctly

> 54 participants wrote essays with ChatGPT, with a search engine, or with no tools, wearing EEG headsets. The ChatGPT group had the weakest brain connectivity and the lowest sense of owning their essays. Preprint, not yet peer-reviewed.

Kosmyna et al., “Your Brain on ChatGPT,” MIT Media Lab, 2025

HIGH SCHOOL STUDENTS IN TURKEY

About 1,000 high school students in Turkey. Each class session had two parts:

1. Students solved practice problems. Each classroom was randomly given plain GPT-4, a GPT-4 tutor, or no AI.
2. All students then took an exam with no AI.

Grades compared with the no-AI group	Practice problems	Exam	What students believed
Plain GPT-4	+48%	−17%	They had not learned less
GPT-4 tutor (hints, not answers)	+127%	No difference	They had done better on the exam

> The tutor's prompt held the teacher's solution to each problem, common student mistakes with a hint for each, and an instruction not to give away the entire solution. Plain GPT-4 was told only to act as a tutor and help the student solve the problem.

Bastani et al., "Generative AI without guardrails can harm learning," PNAS, 2025

LEARNING A NEW LIBRARY WITH AI

52 software engineers who knew Python but had never used Trio, a Python library for asynchronous programming.

1. Each engineer coded two features with Trio. One group had an AI assistant. The other group coded by hand.
2. All engineers then took a quiz on Trio with no AI.

	Minutes to finish the task	Quiz score
AI assistant	23	50%
By hand	25	67%

The time difference is not significant. In the AI group, engineers who asked the AI for explanations scored 65–86%. Engineers who had the AI write or debug the code scored 24–39%.

Shen and Tamkin, "How AI impacts skill formation," Anthropic, January 2026

HARVARD STUDY ON AI TUTORS

2× median learning gain, AI tutor vs active-learning class

> 194 students in Harvard introductory physics. Each student learned one topic from the GPT-4 tutor at home and another in class. The tutor was prompted with the instructor's step-by-step solutions and guided students through each part of each problem.

Kestin et al., "AI tutoring outperforms in-class active learning," Scientific Reports, 2025

CLASSROOM DESIGN

FLIPPING THE CLASSROOM

- We have to expect out-of-class assignments will be done by AI.
- Assessing in class by cold-calling remains viable.
- This suggests flipping the classroom: providing lectures beforehand and spending class time interacting (case-style teaching).
- Ask students to share screens and defend their work and assumptions.

A POSSIBLE CLASSROOM PROTOCOL

1. Use AI to write cases.
2. Give students 5 minutes to read.
3. Discuss on whiteboard.
4. Give students 10 minutes to do analysis with AI.
5. Call on students to present and explain.

> In the studies, AI that gave answers lowered learning. AI that tutored, with hints and step-by-step guidance, did not. In this protocol, the class discussion in steps 3 and 5 plays the tutor's role: the instructor questions students about the analysis they did with AI.

SOME AI INSTRUCTIONAL TOOLS

My plugin marketplace: github.com/kerryback/kerryback-public

Voiceover skill

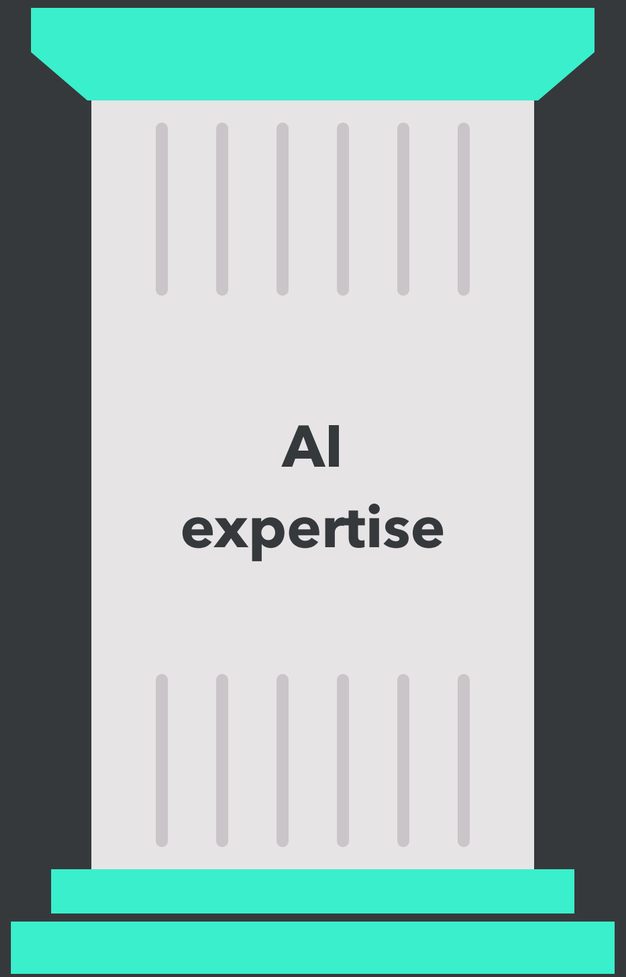
- Upload a PDF deck.
- Write narration with AI.
- Voice by ElevenLabs.

Example: zurich.kerryback.com/claude.mp4

Tutor skill

- Point it at your slides, notebooks, or code.
- Write instructions to AI.
- Write module intro narration and Python scripts, and configure plots, with AI.

Examples: tutors.kerryback.com


AI
expertise

USING AND CHECKING AI

PROMPTING

Say what you want

Not what to do.

Ask for the best way (plan) first

Then tell it to proceed.

Ask AI to interrogate you

Let it decide what it needs to know.

Beware of tunnel vision

Ask: is there a better path?

Steelman

“What is the strongest argument against this?”

Numbers need code

“Compute this in Python.”

TEACHING STUDENTS TO CHECK AI WORK

Have AI explain its work

"Explain what you did step by step. Did you make any decisions that we have not discussed?"

Use another AI agent to check the work

A fresh conversation or a subagent (a second agent that the first agent starts), given your goals and the first agent's explanation.

Have AI build a spreadsheet when possible

Then check the formulas. Every number traces back to its inputs.

Run the same code on small examples

Ask AI to run its code on a few rows you can check by hand.

Most AI coding errors arise from misunderstandings and unstated assumptions.

QUIZZES AND EXAMS

- Have AI generate multiple flawed spreadsheets on the topic being taught.
- Ask students to identify the errors.

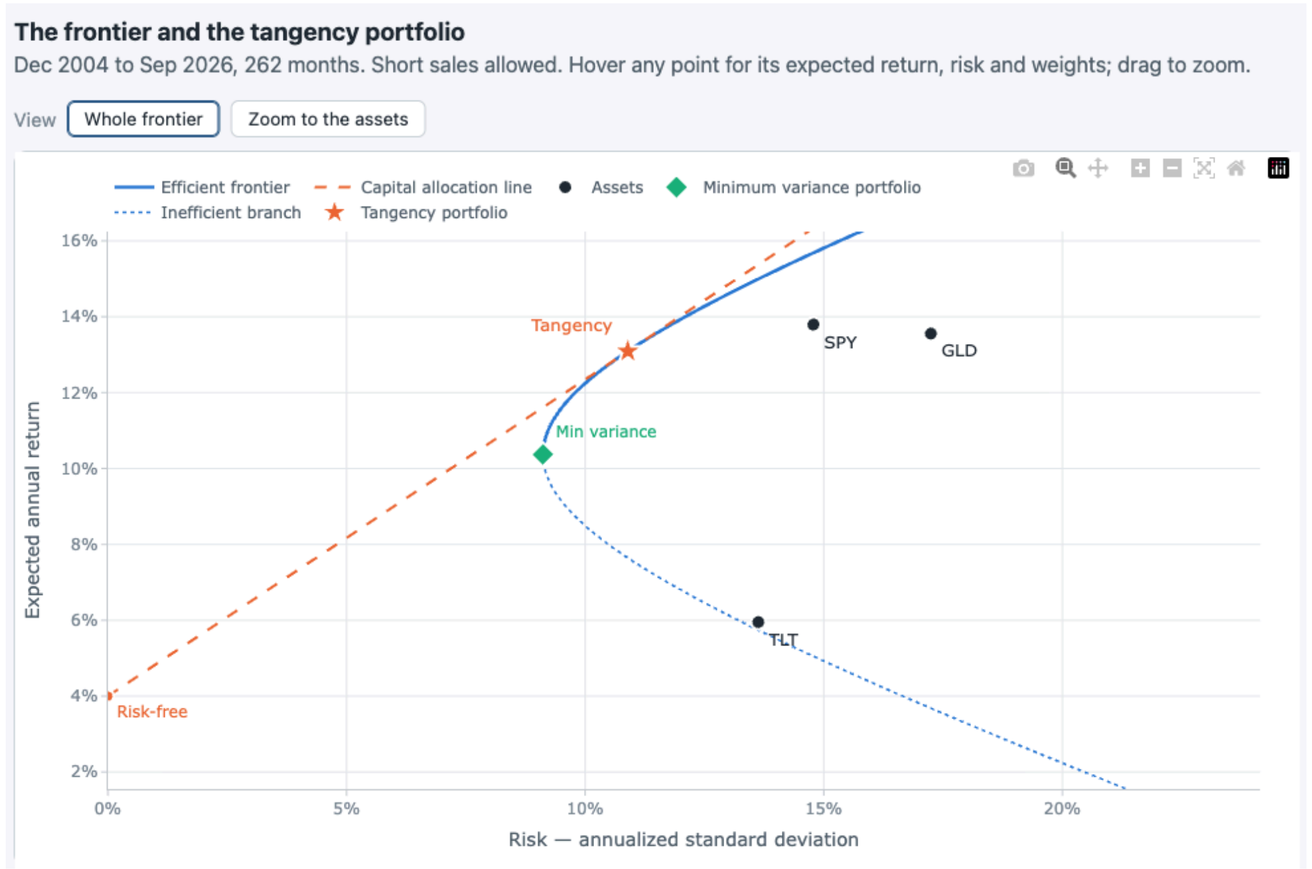
HTML pages and workbooks

BUILDING AI PROCESSES

NEW STUDENT DELIVERABLES

- Software apps: every course?
- Skills: every course?
- Custom agents: AI course

TANGENCY PORTFOLIO: APP



tangency.kerryback.com. SPY, TLT, and GLD, December 2004 to September 2026. The page also gives the weights and a workbook to download. The app could be a skill or a custom agent.

TANGENCY PORTFOLIO: WORKBOOK

1. Means
2. Standard deviations
3. Correlations
4. Covariance matrix
5. Inverse covariance matrix
6. Inverse covariance matrix times risk premia
7. Scaled to the tangency portfolio

Open the workbook. SPY, TLT, and GLD, December 2004 to September 2026.

SOFTWARE

SANDBOXED AGENTS

What they are

- Cowork (and Codex, …) runs in a virtual machine or container. The container has a whitelist of web addresses it can visit, including Anthropic, GitHub, Python, and a few others.
- It can do web fetch and web search because they are run at Anthropic. It only needs to reach Anthropic to run them.
- You can edit the whitelist (allowlist) in Cowork.

When the sandbox binds

- Cannot reach URLs from code.
- Cannot use CLIs of external services or run curl commands to use APIs.
- Cannot run grep and glob to search files outside the project folder.

NON-SANDBOXED VERSIONS OF CLAUDE

- Code tab in the Claude desktop app
- Claude Code CLI
- Claude Code extension for VS Code
- Academic Studio, my fork of VS Code that integrates Claude for non-coders: academic-studio.com