



AI at Your Side

The Student's Guide to Smarter Learning

Forthcoming from Oxford University Press

Sebastian Galiani
Raul A. Sosa

Kyiv School of Economics
Lecture 1 · What AI Is, and What It Is Not



Book website

AI Is Already Everywhere

Nearly everyone already uses it. The open question is how

95%

of UK university
students use AI
HEPI, 2026

85%

of US college students
used it for coursework
Inside Higher Ed, 2025

75%

of knowledge workers
use AI at work
Microsoft & LinkedIn, 2024



AI can make you *sound* smarter without making you **think** better, and from the outside the two look the same.



This course is six sessions on the harder use, where the work also improves your own thinking.

The Dilemma: *Apocalittici vs. Integrati*

Umberto Eco, 1964: mass culture split intellectuals in two, as AI does now

✗ The doomsayers

Mass unemployment, algorithmic control and the collapse of human agency. Every new technology is a threat.

The fear takes two shapes: **Huxley**, a soft tyranny that pacifies through pleasure, and **Orwell**, a hard one built on fear.

✓ The enthusiasts

Cured disease, climate change stopped and knowledge democratized. Innovation is unambiguous progress.

Every previous technology created more than it destroyed, so this one will too.

Between those poles, the one sure thing is that we do not yet know.

That uncertainty is a reason to run careful experiments and to be honest about what they show.

The Third Path: AI on Our Side

The four principles the book is built on



Augmentation, not replacement

Build on what you can do. Never skip the learning that matters most.



Appropriate automation

Decide, task by task, when to rely on it and when to engage directly.



Ethical AI literacy

Know its limits and its biases. Declare your use. Ask what you are missing.



Intellectual ownership

Say why you asked, why you trust the answer, and which parts are yours.



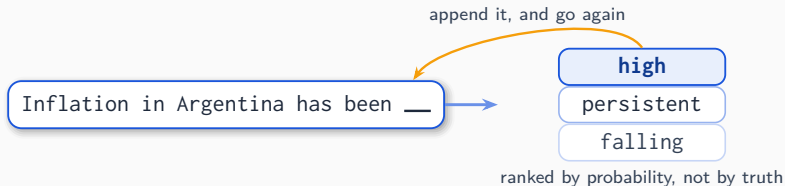
All four come down to one decision you make every time you open a chat: **which part of the work you are not willing to give up.**

What It Is

And where it stops working

What a Language Model Actually Is

One operation, repeated



That is the whole operation: pick the next word given everything before it, append it, and do it again. Reasoning, code and citations are all that same loop, run for longer.

✓ Why it is powerful

Trained on more text than you will ever read, it links fields you would not have connected.

⚖️ Where the ranking comes from

From patterns in the text it was trained on. Nothing in the ranking knows which word is true.

Why It Sounds Sure When It Is Wrong

Two reasons its confidence is not evidence



1. It was trained to assist

A second round of training, graded by people and by automatic tests, turns the text predictor into an assistant. The side effect: that grading tends to reward agreeing with you, and a confident guess over "I don't know."



2. It writes without checking

Where the text it learned from is dense, the most probable word is usually the true one. Where the text runs thin, it writes something just as fluent, and wrong. That is a hallucination.

Where the errors cluster

citations

numbers

names

dates

anything recent

anything obscure



It sounds just as sure either way, so asking it whether it is sure is not a check. If you want it to push back, ask for it. Web search helps, because it puts real sources in front of the model, but it can still misread them.

The Help Is Real

Three studies, three very different kinds of work, with and without an AI assistant

BCG consultants

+30%

better performance, 25% faster

Dell'Acqua et al., 2026

Customer support

+15%

more cases resolved per hour

Brynjolfsson et al., 2025

Software developers

+26%

more tasks completed

Cui et al., 2026

→ But the first study did something the other two did not. It included, on purpose, a task just outside what the AI could do well.

The Jagged Frontier

Where the help reverses, and nothing tells you it has

🚧 **The jagged frontier** is the edge of what AI does well. Tasks that look equally hard can fall on either side, and past the edge people do worse with AI than without it.

BCG, the task just outside the frontier

19 points

lower share of correct answers with AI than without

Dell'Acqua et al., 2026

Developers who knew their code well

19% longer

to finish their tasks with AI, in a randomized trial

Becker et al., 2025

⚠️ They could not tell which side they were on. At BCG, the wrong answers still read as convincing. The developers:

Expected: **24% faster**

→ Took: **19% longer**

→ Believed afterward: **20% faster**

What Gets Cheap, and What Gets Scarce

When one input gets cheaper, its complements become more valuable

AI lowers the cost of execution.
It raises the value of judgment.

↓ The cost collapses

- Writing the code
- Searching and summarizing
- Algebra and derivations
- Simulating
- Drafting

↑ The value rises

- Choosing the question
- Choosing the assumptions
- Choosing the model
- Interpreting the result
- Knowing whether it is right

 **Cognitive debt** (Kosmyna et al., 2025): the cost of work done in your name but not in your head. It comes due when the tool is gone, or when it is wrong and you cannot tell.

How We Will Work

The method, the six sessions and the rules

How Every Session Works

Five steps, and the two at the ends are never delegated



You think before the model does and verify after it is done: those two ends are **yours**. The three steps in between involve the model, and they take most of the time.

→ You will not see this diagram again. Every session runs it, and by the third one you should be able to say which step you are in without being told.

What We Build, Session by Session

Six sessions, and what each one is built around

01 Asking

The same question asked two ways, and the difference checked against the data.

02 Thinking

An idea turned into a question worth researching and a map of what is already known.

03 Learning

Becker's model of time, understood first and then broken on purpose, six ways.

04 Programming

A neighborhood-choice model written from the economics down, simulated and then estimated.

05 Deciding

A valuation where the single number is often the least interesting output.

06 Research

All of it at full scale: four papers that use or test these tools in real research.

The first five show the work step by step, with real output, so you can repeat it on your own. The first two problem sets need nothing installed. **Problem Set 3** needs an agent and Python, and the installation guide comes with the course materials.

How We Work in This Course

The rules from our own economics courses, for every problem set

 **The AI Process Log:** a record of intellectual ownership One row per exchange that shaped the work

What you asked

The exact prompt

What it proposed

The part that mattered,
pasted whole

What you kept or changed

And why

What you verified

Against which source

✓ Always

- ✓ Name the model and the date
- ✓ Check one claim against another source
- ✓ Flag where the model takes a side

✗ Never paste

- ✗ Unpublished work that is not yours to share
- ✗ Personal data about anyone
- ✗ Anything under a confidentiality agreement

● ● ● **A window on a slide is a real session:** what we typed and what came back, verbatim, with cuts marked [...]. A dark one shows an agent, defined on slide 17. Our reading sits outside it.

Agents

When AI executes instead of answering

From Chatbots to Agents to the API

Three ways of working with these models



Chatbot

You write,
it answers.

ChatGPT, Claude, Gemini



Agent

It plans and runs the
work. You supervise.

Claude Code, Cursor, Codex



API

You write code that calls
the model, at scale.

OpenAI, Anthropic, Google

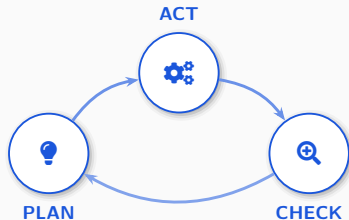
>_ Agent example: “Inspect this dataset, run a fixed-effects regression, produce a chart.” The agent does it. You verify every result.

>_ API example: a Python script sends 200 abstracts through the same strict prompt and extracts hypothesis, method and findings into a CSV.

What Is an Agent?

The unit of work goes from a reply to an assignment

An agent is an AI that pursues a **goal**: it plans the steps, uses tools (files, the web, code) and checks what they return, with minimal human intervention.



Chatbot and agent, side by side

	Chatbot	Agent
The unit	a reply	an assignment
Works from	what you paste	what it goes and finds
Ends in	text	a file, an analysis, a report



An agent can check that its code runs and that the file exists. Whether the answer is right is still your check.

The Golden Rule

Where to start, what the rule says and what is never delegated

✓ Where to start

- ✓ Tasks with documents and clear rules
- ✓ Work you repeat across many cases
- ✓ Anything you can check against a source

⚖ The golden rule

People **design the instrument**, a prompt with rules for one concrete task, and **audit the result**. The agent does the steps in between.

🔒 Never delegated

- 🔒 The goal and the standard of quality
- 🔒 The final judgment on what gets published
- 🔒 Owning the result

🎓 **You can only judge work you know how to do.** So the golden rule is an expert's rule: a student first does the work, often enough to judge it, and then hands it over.

Deep Research on a Loop (DRIL)

The golden rule at scale (Afonso, Galiani, Gálvez & Sosa, 2026)

 Building datasets from primary sources is one of the costliest tasks in empirical economics. **DRIL** turns the rules into a frozen **instrument** and runs it on each case separately.

How it works

Design. The researcher sets the objective and brings the reference material. An agent turns them into a codebook, the rules for what to measure and how, and flags the ambiguities. The researcher resolves them and freezes the codebook.

Implementation. Agents run it on every case and log sources, gaps and uncertainty.

One real run

A 2025 partial update of the Global Tax Expenditures Database: **86 jurisdictions**, about 40% of the 218 in the database, for about **\$22.50** at API prices. 66 got estimates, and the other 20 a recorded gap instead of a guess.

 **A log is not a check:** a separate verification stage has to run before anyone uses the data.

Who Does It Help?

Evidence from a randomized experiment

Does AI Narrow Productivity Gaps?

Cruces, Fernández Mejjide, Galiani, Gálvez & Lombardi (2026)

Most of the evidence studies AI *inside* firms, where hiring has already selected who is in the room. A **randomized online experiment outside firms** was run on adults in Argentina, to ask whether AI narrows productivity gaps by **education**.

The design

- **1,174 Argentine adults**, aged 25 to 45
- An incentivized, workplace-style **problem-solving task**
- Randomized: a generative-AI assistant and a short tutorial, or **neither**

The follow-up, without AI

After the main task, everyone completes a second module **without AI**. This reveals whether any of the gain carries over, or whether it was mere **delegation** that vanishes once the tool is taken away.

AI Narrows the Gap Without Closing It

Larger gains for those with less education

75%

of the education gap in task performance closed, with AI access and a short tutorial

✓ AI improves performance for **everyone**, with much larger gains for those with **less education**. In task execution, it closes most of the gap.

⚠ **The nuance:** without AI, a sizable gap re-emerges. Those with more education used the tool somewhat more effectively, and follow-up scores were high where people put real effort into the task, and low where heavy AI use took its place. Both patterns are associations: only the AI treatment was randomized.

Same Question, Two Ways

A worked demonstration: why is rent in Berkeley so expensive?

The Way Everybody Asks

One line, sent to a new chat

Claude Sonnet 5 · clean session · Sep 23, 2026 condensed

Give me some research ideas on why rent in Berkeley is so expensive.

 Claude

Research Ideas: Why Is Rent in Berkeley So Expensive?

1. Supply constraints [...] **2. Demand-side drivers [...]** **3. Rent regulation and market structure [...]**
4. Institutional and political economy questions [...] **5. Measurement and comparative questions [...]**

Data sources to consider [...] **Suggested starting points [...]**

If you tell me your discipline, your timeframe, and whether you want a class paper, a thesis, or a policy brief, I can narrow this to one or two well-scoped projects with a proposed identification strategy and a literature list.

The first of two runs: the second is on the next slide. Clean session: a new chat, with no tools, no web search and no memory of other chats.

 It gave twenty angles in five families. Nothing is wrong, and it is a good menu. But it cannot choose for you: it does not know your field, your deadline or your format, so it asks.

The Same Line, a Second Time

Same prompt, same model, a new chat: one fact came back wrong

Claude Sonnet 5 · clean session · Sep 23, 2026 condensed

 Claude
[...] Berkeley's rent control was partially preempted by Costa-Hawkins (1995). [...]

✓ **Run 1: right.** The 1995 state law exempted newer buildings, single-family homes and condos, and let landlords reset the rent between tenants.

Claude Sonnet 5 · clean session · Sep 23, 2026 condensed

 Claude
[...] Berkeley had rent control until Costa-Hawkins (1995), and Berkeley's post-1995 regime has left large parts of the stock partly decontrolled. [...]

✗ **Run 2: its first clause is wrong.** Berkeley still has rent control: its Rent Board regulates the rent of over 20,000 units.

Top: the run on slide 24. Bottom: the same line in a second new chat, which gave seventeen angles in six families, then sources and a request to know more about you.

⚠ Neither run marked a doubt, and nothing in the request asked it to. Ask once, and nothing in the answer tells you which one you got. Both runs have the date right. One is wrong about what happened then.

Five Things a Model Can Only Guess

Unless you tell it, it assumes you are the average user



Context

Who you are, what you can do, and what you already suspect.



Task

The one thing you want done.



Constraints

Time, data, and what is out of scope.



Output

What the answer must contain, item by item.



Verification

What it must flag so you can check it, and what you will check yourself.



A prompt is an **instruction set**, not a search query. The more of your own thinking you put in at the start, the more the model gives back.

→ Written once and saved for a task you repeat, that instruction set is a *harness*. The other sessions practice writing one.

The Prompt, Rebuilt

The same question, with the five fields written out



Claude Sonnet 5 · clean session · Sep 23, 2026

reply on the next slide

Context: I'm a second-year economics major at Berkeley. I've taken intermediate micro and intro econometrics, so I can run OLS and a simple difference-in-differences in R. My hunch is that the new student housing near campus hasn't brought rents down.

Task: propose research designs that could explain why rent in Berkeley is so high.

Constraints: one semester, about four hours a week. Only free data I can download myself. No surveys, no scraping, nothing beyond OLS and difference-in-differences.

Output: two designs. For each one: the question in one sentence, the dataset (name and link), the regression, and the main threat to it.

Verification: for every dataset, tell me the first year and the smallest geography it covers for Berkeley. If you are not sure of something, write UNSURE instead of guessing. I will open the data myself.

Highlights are ours: the three phrases that do the most work.

What It Returns Now

Same model, a new chat: the second of its two designs, condensed

Claude Sonnet 5 · clean session · Sep 23, 2026

condensed

 Claude

[...] **Design 2: How much of near-campus rent is student demand?**

Question: When UC Berkeley went remote in 2020, did rents in near-campus ZIP codes fall more than in comparable ZIPs farther from campus?

Data: Zillow Observed Rent Index (ZORI) [...] Treated ZIPs are probably 94704 and 94709. [...]

Main threat: Dense, renter-heavy urban ZIPs all fell during COVID for non-student reasons [...]

Verification table [...]

ZORI (Zillow) The national series starts in 2015. UNSURE for each Berkeley ZIP, since ZIPs with thin data may start later or be missing. [...]

We ran the prompt from slide 27 in two new chats. Both answers had a design on this index, and both marked its Berkeley ZIPs UNSURE.



What each field did

The **context** put the hunch to the test: design 1 is about the new buildings.

The **output** made each design arrive with its data and its main threat.

The **constraints** kept both designs to files you download and a regression you can run.

The **verification** line made it say where it was unsure.

Open a File It Was Unsure About

Three minutes, and the step no field can do for you

Zillow's rent index, Berkeley

First month with data. Bold: treated.

Red: data only from 2026.

94703 Nov 2018

94704 Nov 2019

94702 Jan 2020

94710 Mar 2021

94709 Jan 2022

94705 Mar 2022

94707 Jan 2026

94708 Aug 2026, one month

 Our check: Zillow ZORI by ZIP, Sep 2026.

The doubt was in the right place

The index exists, and its Berkeley coverage was worth doubting.

Design 2 has almost no “before”

94704 starts four months before classes went remote. The other one, 94709, starts in 2022, after the students were back.

 The five fields **made the answer checkable**, and the model flagged its doubt where you asked for it.

What you get from the model depends on what you bring to it.

Same model, same question. A hunch and a few limits changed the answer.
The check at the source was still yours.

AI at Your Side

Sebastian Galiani · Raul A. Sosa
Forthcoming from Oxford University Press

Thank you!



Book website