

AI for Economic Theorists and Mathematicians

Pietro Ortoleva
& Fedor Sandomirskiy

Princeton University

AI for Economic Theorists and Mathematicians



Pietro Ortoleva
Princeton



Fedor Sandomirskiy
Princeton

Who we are, and who this is for

WHO WE ARE

Pietro Ortoleva and Fedor Sandomirskiy, economic theorists at Princeton

WHO THIS IS FOR

Economic theorists at all levels

WHAT WE ASSUME

Any level of AI use: from occasional chat to agents in a project folder

Five parts, built around broad questions

- 1 Core uses of AI for economic theory** the jobs these tools do in the theory workflow
- 2 Can models be creative?** why frontier math results matter for our expectations
- 3 Which model, and when to run it** intelligence, stamina, tiers, browser versus agents
- 4 Prompting and practices** prompting basics, then agents, LaTeX, restarts, proofreading
- 5 Other AI Tools & Main Takeaways** takeaways, writing, voice, meeting memory

PART 1 · CORE USES

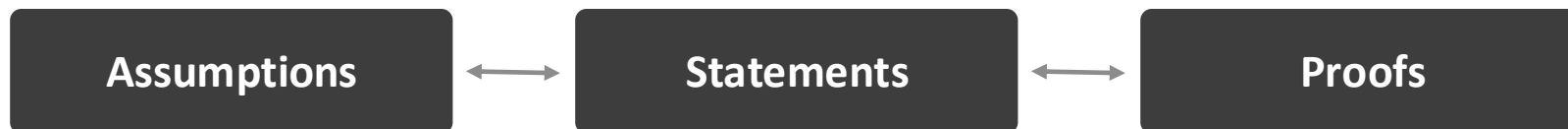
Core uses of AI for economic theory



Five parts, built around broad questions

- 1 **Core uses of AI for economic theory** the jobs these tools do in the theory workflow
- 2 **Can models be creative?** why frontier math results matter for our expectations
- 3 **Which model, and when to run it** intelligence, stamina, tiers, browser versus agents
- 4 **Prompting and practices** prompting basics, then agents, LaTeX, restarts, proofreading
- 5 **Conclusions and tools** takeaways, writing, voice, meeting memory

Economic theory as a fixed point



A good theory paper searches for a fixed point: defensible assumptions, interesting statements, proofs that certify them

- 1 Model sketching** identify the forces, formulate the primitives
- 2 Failure diagnosis** counterexample → adjust one corner, not the whole project
- 3 Cheap iteration** with AI, each loop costs minutes, not afternoons

Use cases for AI in economic theory

1 Sketching and brainstorming models

2 Reviewing literature

3 Suggesting proofs

4 Checking proofs

5 Extensions, microfoundations, simplifications

6 General proofreading

7 Simulations

From intuition to primitives

1

Primitives

states, signals, actions, preferences

2

Timing

what is known, chosen, observed

3

Solution concept

Bayesian, maxmin, rational inattention, dynamic

4

Predictions

comparative statics, tests, examples

PROMPT

"Give me three minimal models that capture this intuition. For each, say what is elegant, what is fragile, and what theorem would be worth proving."

Spin off variants

Change one primitive at a time: preferences, information, timing. Which comparative static flips?

Strip to the minimum

What is the smallest model that still delivers the prediction? Remove every idle assumption

Proofs

Frontier models can generate reliable proofs → still your responsibility to check

Non frontier models are less reliable → decompose big tasks into smaller ones

ATTACK

Hostile referee checks for weak steps

REPAIR

Assumptions that rescue a false statement

INSPIRATION

Even a wrong proof identifies the attractive route, the hidden assumption, the missing lemma. Mine it, do not trust it

Extensions, microfoundations, simplifications

Once the machinery exists, ask what else it can do, and what it does not need

MORE GENERAL

Extensions

Change information, preferences, timing, menu structure, equilibrium concept. Which comparative statics survive, which flip?

DEEPER

Microfoundations

Can the reduced form assumption come from search, attention, ambiguity, private signals, institutions?

CLEANER

Simplifications

Can you do it with two types and two periods?
With binary states and two actions?

PROMPT

“Strip the model to the smallest example that still delivers the main prediction. Then list every assumption that survives and what breaks when it is removed.”

Five parts, built around broad questions

- 1 **Core uses of AI for economic theory** the jobs these tools do in the theory workflow
- 2 **Can models be creative?** why frontier math results matter for our expectations
- 3 **Which model, and when to run it** intelligence, stamina, tiers, browser versus agents
- 4 **Prompting and practices** prompting basics, then agents, LaTeX, restarts, proofreading
- 5 **Conclusions and tools** takeaways, writing, voice, meeting memory

AI for Economic Theorists and Mathematicians

Part 2/5: Can models be creative?

Pietro Ortoleva
& Fedor Sandomirskiy

Princeton University

Five parts, built around broad questions

- 1 **Core uses of AI for economic theory** the jobs these tools do in the theory workflow
- 2 **Can models be creative?** why frontier math results matter for our expectations
- 3 **Which model, and when to run it** intelligence, stamina, tiers, browser versus agents
- 4 **Prompting and practices** prompting basics, then agents, LaTeX, restarts, proofreading
- 5 **Conclusions and tools** takeaways, writing, voice, meeting memory

Is it real creativity?

- Creative content: clever proof strategies, useful conjectures, counterexamples? Or only fluent, plausible looking text?
- Little direct evidence in economic theory → we lean on mathematics
- **Our experience: yes, with limitations**
- Trajectory: high school math → frontier results, in about three years. The public story is noisier than the trend

The ten Erdős problems episode

- OpenAI claims a model solved ten open Erdős problems. Collapses within a day
- Thomas Bloom traced what the model did: found existing references, presented them as its own



Paul Erdős



Thomas Bloom

POOR AT ATTRIBUTION

Cannot tell its own ideas from its training data

SUPERB AT AGGREGATION

No lemma from an appendix gets lost

[Thomas Bloom on X](#) · [TechCrunch, Oct 2025](#)

Real progress, carefully catalogued

- Terence Tao curates AI contributions to Erdős problems: graded, sourced, no leaderboard
- **July 2026 count: 19 fully AI generated standalone solutions, many machine checked in Lean**
- Around ten genuinely new; the rest matched results later found in the literature; one beats the known 1978 proof
- Mostly specialized pipelines, rarely novel arguments → weak evidence on creativity



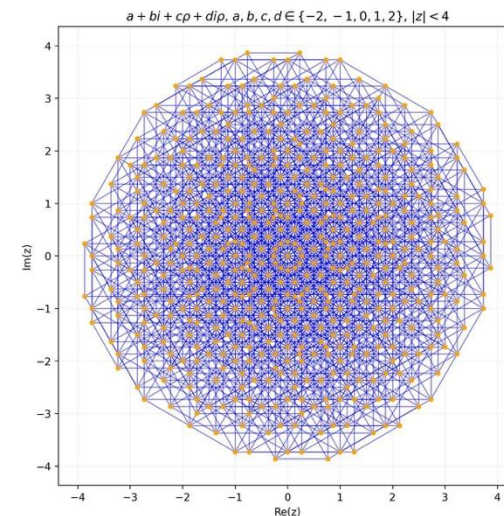
Terence Tao

For creativity, the next three data points are stronger

[teorth/erdosproblems.wiki: AIcontributions](https://teorth.erdosproblems.wiki/AIcontributions)

The Erdős unit distance conjecture falls

- Since 1946: n points in the plane, how many pairs at distance exactly one? Conjecture: the grid is optimal
- Eighty years open; tied to deep algebraic geometry and number theory
- Settled by an internal OpenAI model, from a single prompt. Prompt and transcript public
- **And the conjecture was wrong: you can beat the grid**



The construction that beats the grid. Figure: A. Lozano-Robledo

[OpenAI: prompt and proof \(PDF\)](#) · [arXiv:2605.20695](#)

Reception



“an outstanding achievement, settling a long-standing open problem ... it is hard to overestimate the full potential impact of this change.”

Noga Alon, Princeton



“I would accept it for any journal without hesitation. I actually briefly worked on this problem and tried to make a counterexample, but failed to make progress.”

Jacob Tsimerman, Toronto

No new theory: known tools, combined in a way specialists did not expect to work

Fable refutes the Jacobian conjecture

- Since 1939: polynomial map $\mathbb{C}^n \rightarrow \mathbb{C}^n$ with constant nonzero Jacobian must be one to one. Number 16 on the 1998 Smale list, beside Riemann and P vs NP
- **Levent Alpöge asked Fable 5. An explicit counterexample in $\mathbb{C}^3$ came back the same evening**

$$F(x, y, z) = ((1+xy)^3 z + y^2(1+xy)(4+3xy), \quad y + 3x(1+xy)^2 z + 3xy^2(4+3xy), \quad 2x - 3x^2y - x^3z)$$

$\det DF \equiv -2$ everywhere, yet three points share the image $(-1/4, 0, 0)$

Hard to find, trivial to check

[The announcement on X](#)

Cycle double cover: proved by Sol Ultra

- Since 1973 (Szekeres, then Seymour 1979): every bridgeless graph carries cycles covering each edge exactly twice
- **Proved by GPT-5.6 Sol Ultra in Codex: 64 subagents, told to work “at least 8 hours”, done in under one hour**
- The complete prompt is public: a masterclass in managing an agent army. Dissected in Part 4
- Thomas Bloom: “a very nice proof”, with a caveat: central 1983 ideas appear without citation

[The complete prompt \(PDF\)](#) · [The proof \(PDF\)](#)

What this means for economic theory

JUNE

One frontier result, with caveats

JULY

Three frontier results, two labs, subscription products

Why economic theorists may gain even more than mathematicians

- 1 **Lower mathematical bar** → models clear it comfortably
- 2 **Conjectures are negotiable** → adjust assumptions until the statement is true and interesting
- 3 **Fixed point search** → AI accelerates every loop of assumptions, statements, proofs

Five parts, built around broad questions

- 1 **Core uses of AI for economic theory** the jobs these tools do in the theory workflow
- 2 **Can models be creative?** why frontier math results matter for our expectations
- 3 **Which model, and when to run it** intelligence, stamina, tiers, browser versus agents
- 4 **Prompting and practices** prompting basics, then agents, LaTeX, restarts, proofreading
- 5 **Conclusions and tools** takeaways, writing, voice, meeting memory

AI for Economic Theorists and Mathematicians

Part 3/5: Which model, and when to run it

Pietro Ortoleva
& Fedor Sandomirskiy

Princeton University

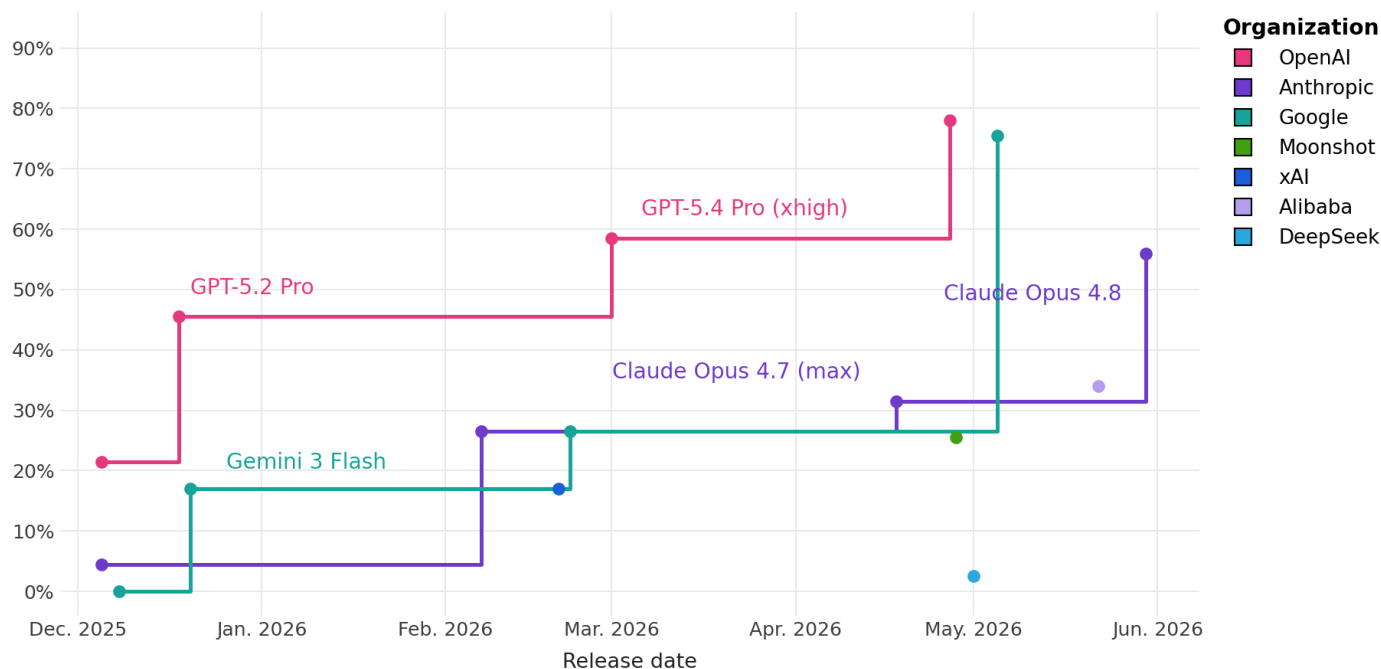
Five parts, built around broad questions

- 1 **Core uses of AI for economic theory** the jobs these tools do in the theory workflow
- 2 **Can models be creative?** why frontier math results matter for our expectations
- 3 **Which model, and when to run it** intelligence, stamina, tiers, browser versus agents
- 4 **Prompting and practices** prompting basics, then agents, LaTeX, restarts, proofreading
- 5 **Conclusions and tools** takeaways, writing, voice, meeting memory

The pre Fable era, in one picture

AI performance on the most challenging expert-level mathematics problems

FrontierMath Tier 4 accuracy



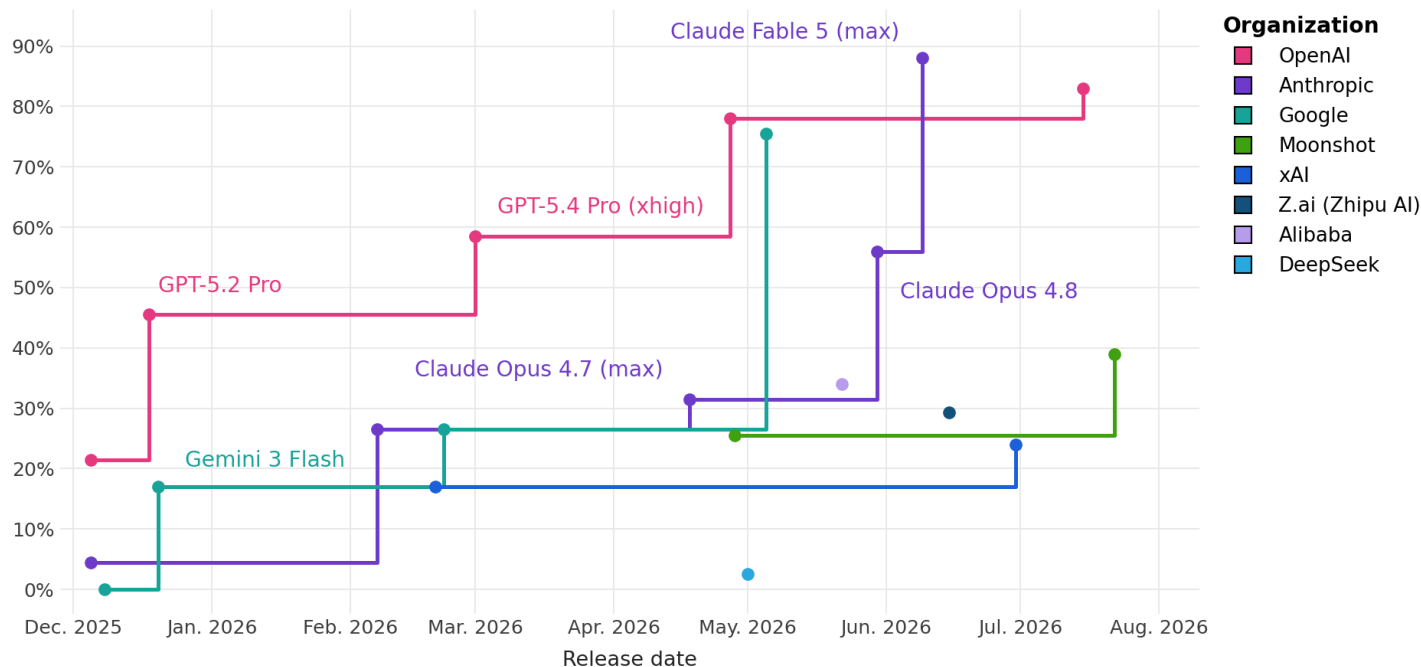
Data and chart design: Epoch AI (CC BY) · epoch.ai · through early June 2026

- Pink staircase on top all spring: 5.2 Pro → 5.4 Pro → 5.5. Our June advice: ChatGPT Pro
- Claude: best writer and coder, not a math contender. Gemini Pro: in between, free in AI Studio

Fable 5: the purple spike

AI performance on the most challenging expert-level mathematics problems

FrontierMath Tier 4 accuracy



Data and chart design: Epoch AI (CC BY) · epoch.ai · as of July 22, 2026

- **88 on Tier 4, the highest point ever. Deep intuitions; born agentic (Cowork, Claude Code)**
- The catch: scarcity. Offline Jun 12 to Jul 1; four access regimes in five weeks; API at twice the Sol price

GPT-5.6: Sol on the web, an agent army in Codex

5.6 Pro in the browser

The 5.5 Pro successor for heavy chat thinking. Tier 4: 83

Sol Ultra in ChatGPT Work

Up to 64 cooperative subagents; the setup behind the cycle double cover proof

Limits and price

Far more generous limits than Claude; API at roughly half the Fable price

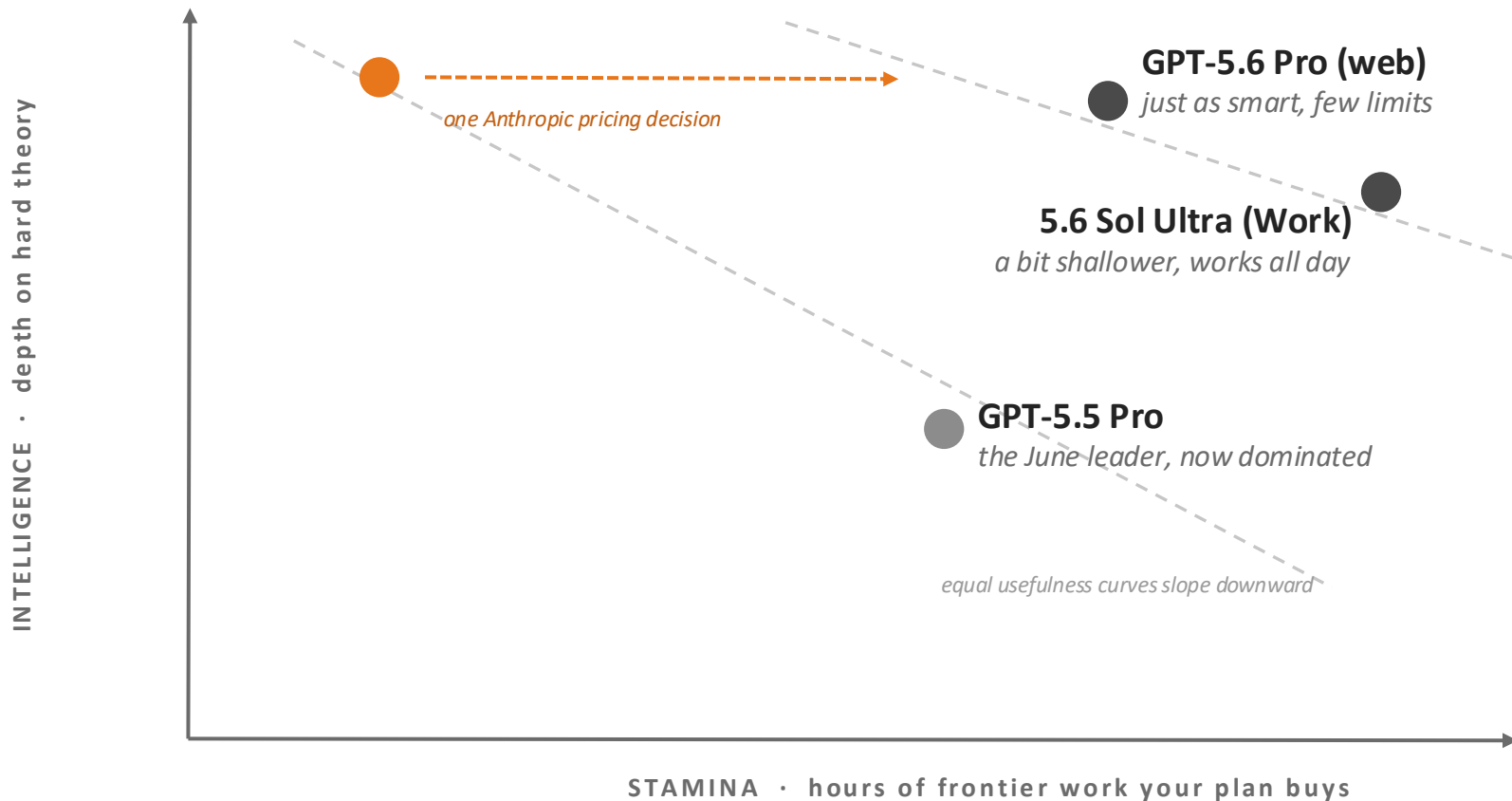
Four weeks after Fable. This pace is the new normal

openai.com/index/gpt-5-6

Intelligence and stamina are substitutes

Fable 5

brilliant, rationed



A slightly shallower model that grinds all day beats a genius you can afford for twenty minutes

Pick 5.6 today, narrowly

- Intelligence: Fable $\approx$ 5.6 Pro, both above the ChatGPT Work flavor of Sol Ultra. Depth of intuition: Fable maybe a touch ahead (speculative)
- **In our experiments, same conjectures cracked, same ones failed. But Fable fails better: more enlightening dead ends**
- Stamina decides $\rightarrow$ 5.6. We switch back if Anthropic makes limits much more generous for Fable

Access matters for theory: top tier Pro at max effort, the 20 dollar tier, and free chat are three different experiences. Buy the best tier you can; always set maximum effort. Agentic environments add the fourth dimension: stamina

PRO BUDGET

5.6 Pro on the web + Sol Ultra
in ChatGPT Work

20 DOLLARS

OpenAI Plus: Sol Ultra in
ChatGPT Work included

FREE

Gemini Pro in AI Studio; chats
train Google models

The browser is fine. The agents earned a try

A MUST

Empirical work: lots of data and code. Agentic environments are how that work is done now

VERY USEFUL

A theory project folder: LaTeX, related literature, referee reports, notes. Read in context, long multi step jobs

NEW: EVEN WITH AN EMPTY FOLDER

Sol Ultra stamina alone turns ChatGPT Work (ex Codex) into a heavy thinking engine. The CDC proof: an agent run, not a chat

Browser still fine for brainstorming. June: you do not need agents. July: give one a week



[Paul Goldsmith-Pinkham](#)

Claude Code for Applied Economists, Markus Academy mini series



[Ben Golub](#)

Modern AI for Economics Research: An Overview of Tools, Markus Academy

Five parts, built around broad questions

- 1 **Core uses of AI for economic theory** the jobs these tools do in the theory workflow
- 2 **Can models be creative?** why frontier math results matter for our expectations
- 3 **Which model, and when to run it** intelligence, stamina, tiers, browser versus agents
- 4 **Prompting and practices** prompting basics, then agents, LaTeX, restarts, proofreading
- 5 **Conclusions and tools** takeaways, writing, voice, meeting memory

AI for Economic Theorists and Mathematicians

Part 4/5: Prompting

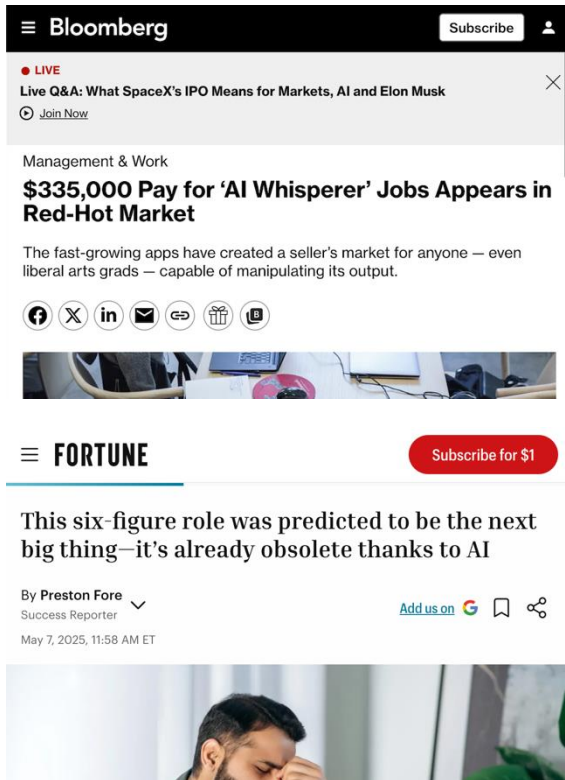
Pietro Ortoleva
& Fedor Sandomirskiy

Princeton University

Five parts, built around broad questions

- 1 **Core uses of AI for economic theory** the jobs these tools do in the theory workflow
- 2 **Can models be creative?** why frontier math results matter for our expectations
- 3 **Which model, and when to run it** intelligence, stamina, tiers, browser versus agents
- 4 **Prompting and practices** prompting basics, then agents, LaTeX, restarts, proofreading
- 5 **Conclusions and tools** takeaways, writing, voice, meeting memory

Is prompt engineering still a thing?



Bloomberg 2023: 335,000 dollars for AI whisperers.
Fortune 2025: already obsolete

- Incantation era over: strong models handle short, vague prompts
- Old view: prompting as magic wording, role play, fixed formulas
- **What survives: precise task specification, context, output format, quality criteria**
- New: in agent land, prompts become management documents

The weaker the model, the more prompting still matters

The minimal prompting template

Task

What exactly should the model do: prove, disprove, map, simplify, rewrite, check?

Context

Which definitions, files, notation, examples, and constraints matter?

Output

What form should the answer take: proof skeleton, table, referee report, code, Markdown?

Quality criteria

What counts as success: exact verification, smallest counterexample, no new notation?

Incentives

Should it help, attack, verify, repair, or adjudicate?

Let the model write the prompt

Models reward detailed prompts; writing them is slow. And it is hard to beat an LLM at writing prompts for LLMs

Write your short, lazy prompt. Never run it as is. First send:

“Optimize the prompt below for GPT-5.6 Pro (or Fable 5, or Gemini 3.1 Pro). Keep its intent, and add whatever it needs to run as is: missing details, relevant context from the attached files, output format, and quality criteria. The result must be fully self contained. Return only the optimized prompt.”

- 1 Saves time**
- 2 Catches misunderstandings**
- 3 Surfaces quality controls**

Did our Theorem 1 really need regularity?

Theorem 1 assumes a regular equilibrium. Hunch: redundant. The lazy prompt:

"Theorem 1 in the attached paper is proved under an equilibrium regularity assumption. I am interested in whether this assumption is redundant. Try proving the statement without the regularity assumption, or construct a counterexample. If a counterexample exists, come up with alternative assumptions under which the statement is still plausible."

Deliberately lazy: two sentences, no definitions, just the attached PDF. Never run it as is: we expanded it first

Gemini: regularity cannot be dropped; boring alternatives. 5.5 Pro: creative ones, including a trick we had missed when writing the paper

Extreme Equilibria: The Benefits of Correlation *

Kirill Rudov[†] Fedor Sandomirskiy[‡] Leeat Yariv[§]

May 1, 2026

Abstract. Correlated equilibria arise naturally when agents communicate or rely on intermediaries such as recommendation systems. We study when a given Nash equilibrium can be improved within the set of correlated equilibria for general objectives. Our key insight is a detail-free criterion: any Nash equilibrium with three or more randomizing agents is generically improvable. We refine this insight to specific classes of games and objectives, including Pareto and utilitarian welfare, and provide constructive methods to obtain improvements. Our findings underscore the ubiquity of improvable Nash equilibria and the crucial role of correlation in enhancing strategic outcomes.

Rudov, Sandomirskiy, Yariv (2026)



Kirill Rudov



Leeat Yariv

One task per prompt. Mind your incentives

ONE TASK PER PROMPT

Do not bundle prove / disprove / repair
→ separate prompts, run in parallel

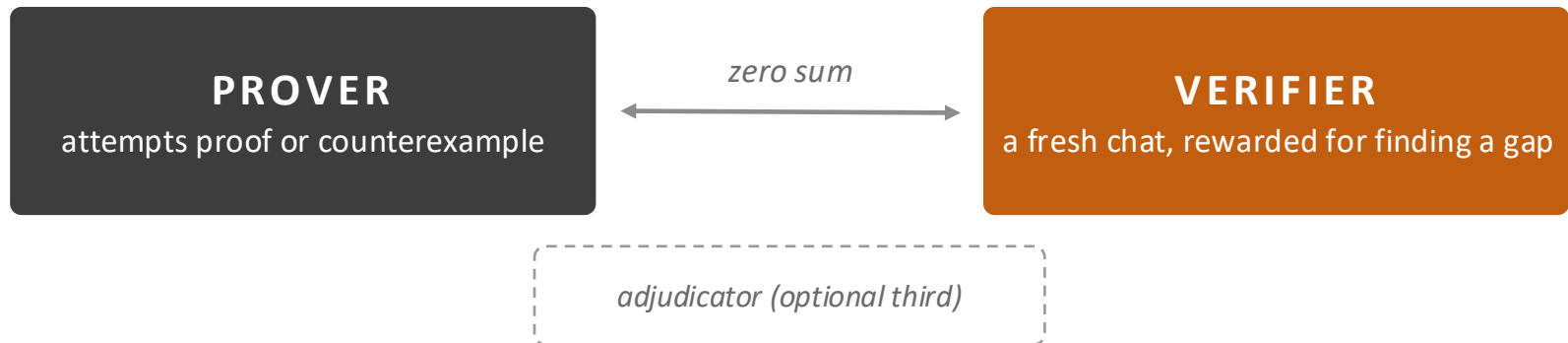
On weaker models, splitting = the difference between an answer and mush

INCENTIVES MATTER

Asking for a proof reveals the verdict you want → sycophancy rises exactly when your guard is down

Top models now admit failure more often; smaller ones still flatter

Prover versus verifier: a zero sum game



- 1 **Run the prover** and do not read the long output yet
- 2 **Hand the output to the verifier** instructed to hunt for the weakest step
- 3 **If they disagree** adjudicate with a third instance, or feed the gap back to the prover

Only a cleared output earns your eyes

Format rule: LaTeX in, never PDF

WHY PDF IS BAD

A PDF is a layout object that merely looks like text. The model wastes effort reconstructing notation, line breaks, references, and equations

WHAT TO DO INSTEAD

Your paper → paste the LaTeX or give the source folder

No source → one conversion run to Markdown, then use the Markdown in the real task

Know when to restart

THREADS ROT

Long conversations dilute attention,
and early wrong turns keep steering

**In long theory projects this quietly
degrades every answer**

THE RESTART PROMPT

“Summarize our conversation for a fresh chat: the problem, working assumptions, results so far, dead ends to avoid. Fully self contained Markdown.”

Restart habit: Ben Golub



Obvious blunder or a loop → do not argue. Restart with a clean summary

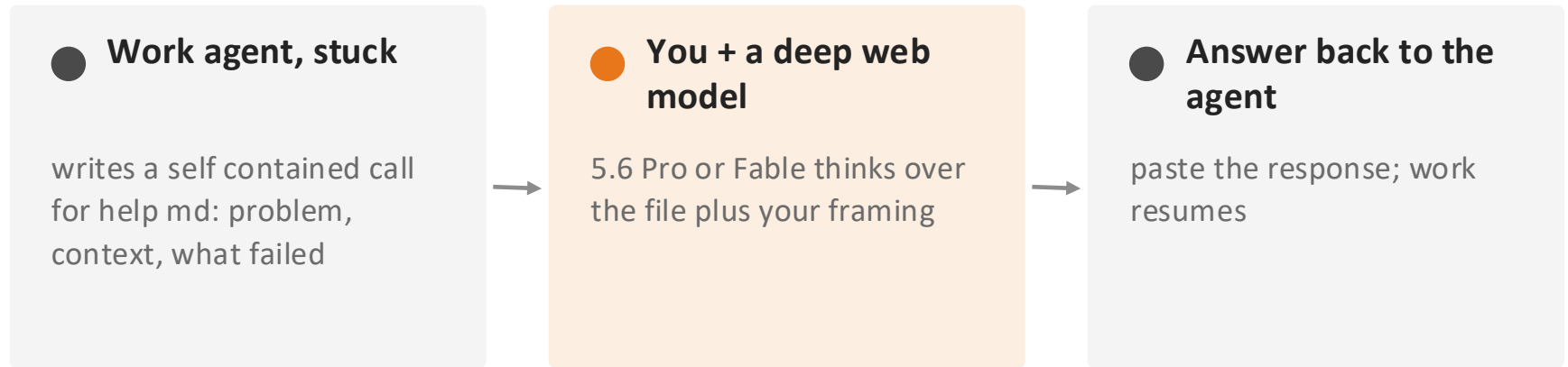
A multi agent workflow for hard targets

“Spend at least 8 hours on this before even thinking of returning or giving up.”

“Use up to 64 concurrent agents. Do not use a fixed assignment such as N agents for strategy X .”

“Keep several incompatible proof routes alive through multiple rounds. Do not tell most agents the currently favored approach.”

Chats and agents are complements (advanced trick)



- The browser model is still deeper at some heavy thinking → use both, they are complements

AI proofreading catches real gaps

- Not just typos: inconsistencies, and actual holes in proofs
- **Refine (by Ben Golub):** has caught subtle, fixable gaps in our proofs several times. Use a dedicated tool for the final pass
- Same space: coarse.ink, reviewer3.com
- Along the way: DIY passes on the subscription you already pay for → recipe next

Proofreading, in three steps

1 Short initial prompt naming the venue

“Proofread the following paper prepared for Econometrica.”

2 Expand it into a detailed proofreading prompt. Not on a top tier model? Ask for separate prompts: presentation, notation, consistency, typos, proofs; or per section

*“Improve the following proofreading prompt for [model] used in [webchat / agentic environment].
Optional: produce separate prompts focusing on presentation, notation, consistency...”*

3 Or hand the whole job to an agentic environment. ChatGPT Work or Cowork: comprehensive proofreading from one prompt, even with a weaker model

FINE PRINT

LaTeX in, never PDF

Run repeatedly; verify before fixing

Easy fixes → one Markdown file: an agent applies, you review the diff

Five parts, built around broad questions

- 1 **Core uses of AI for economic theory** the jobs these tools do in the theory workflow
- 2 **Can models be creative?** why frontier math results matter for our expectations
- 3 **Which model, and when to run it** intelligence, stamina, tiers, browser versus agents
- 4 **Prompting and practices** prompting basics, then agents, LaTeX, restarts, proofreading
- 5 **Conclusions and tools** takeaways, writing, voice, meeting memory

AI for Economic Theorists and Mathematicians

Part 5/5: Other AI Tools & Main Takeaways

Pietro Ortoleva
& Fedor Sandomirskiy

Princeton University

Five parts, built around broad questions

- 1 **Core uses of AI for economic theory** the jobs these tools do in the theory workflow
- 2 **Can models be creative?** why frontier math results matter for our expectations
- 3 **Which model, and when to run it** intelligence, stamina, tiers, browser versus agents
- 4 **Prompting and practices** prompting basics, then agents, LaTeX, restarts, proofreading
- 5 **Other AI Tools & Main Takeaways** takeaways, writing, voice, meeting memory

The writing setup we actually use

- **Overleaf:** unbeatable real time coauthoring. Weakness: AI
- **VS Code + LaTeX Workshop + Copilot:** completes your sentence and your equation
- **The bridge:** Overleaf Workshop plugin = a live Overleaf session inside VS Code.
Setup: 15 minutes. Why we prefer VS Code to Cursor
- **Prism (OpenAI LaTeX workspace):** free, improving; not yet worth migrating a live project

Voice input, and a memory for meetings

VOICE INPUT

Dictation with AI cleanup beats typing: emails, prompts, even paper paragraphs

Wispr Flow: what we use (subscription)

Pipit: free, good. Apple dictation: no cleanup, short window

AI NOTE TAKING

Granola: AI notepad for calls and in person chats. Queryable summaries; transcripts kept. Good free tier

Closes the loop: coauthor conversation
→ transcript → brainstorming prompt

Tell people you are recording

Five parts, take-home messages

- 1 Core uses of AI for economic theory** model set-up, proofs, literature
- 2 Can models be creative?** why frontier math results matter for our expectations
- 3 Which model, and when to run it** intelligence, stamina, tiers, browser versus agents
- 4 Prompting and practices** prompting basics, then agents, LaTeX, restarts, proofreading
- 5 Other AI Tools & Main Takeaways**

UNTIL VERSION 4

Takeaways

- Treat the model like a brilliant colleague: tolerate the dead ends, harvest the flashes of genuine intelligence
- Use it to sketch models, brainstorm proof ideas, check and extend results, improve papers
- Top tier models at max effort make a real difference for theory
- Prompting subtleties matter: LaTeX not PDF, name the audience, expand your prompts
- The landscape changes fast: try to keep up, and most importantly, experiment with workflows to see what is relevant for your work