Teaching with AI

My experience leveraging AI for teaching in the last three years
& RCT results across 19 universities

Sébastien Martin
Associate Professor of Operations · Northwestern Kellogg
Chicago Booth · November 16, 2026

INTRODUCTION

How I relate to AI

Research

- Reinforcement learning
- AI & optimization
- Robotaxis (Waymo)
- AI in education
- Long-term AI memory

Teaching

- AI Foundations for Managers
- Leveraging AI in education at Kellogg

Industry

- Board member, NYSE: ESAB, focusing on AI strategy

INTRODUCTION

Key questions for AI in education

What should we teach?

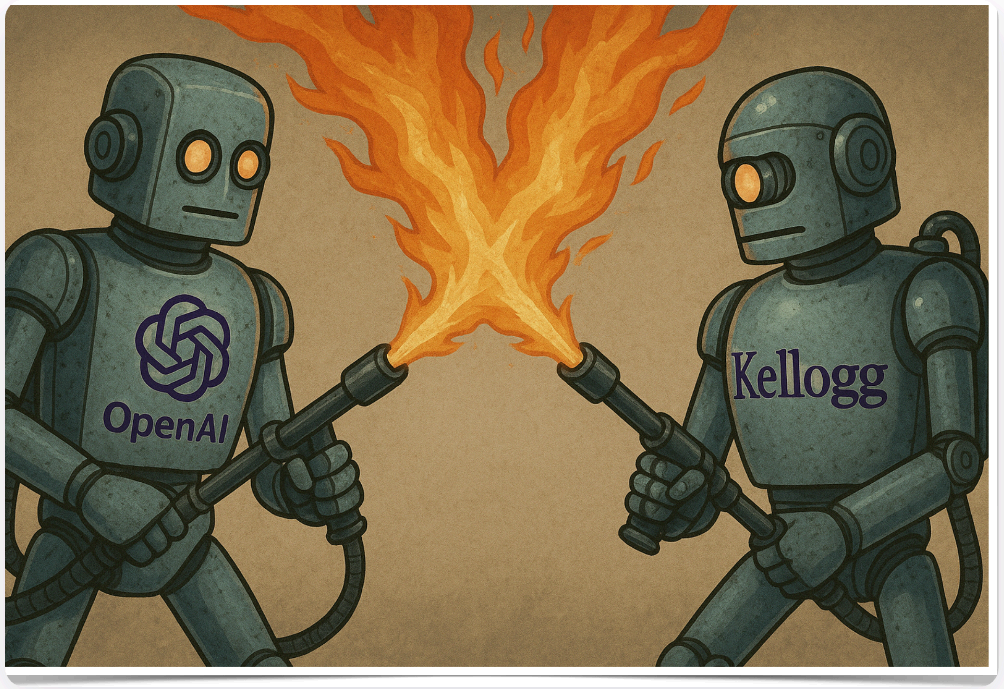
How should we teach?

Damage control, or fighting fire with fire?



Damage control

The new technology hurts our old way of doing things, and we try to work around it.



Build something new

Or we use that same technology to do things we could never do before.

INTRODUCTION

For today

PART 1
SHOWCASE

- Teaching with an AI TA
- AI case studies
- Instructor empowerment
- Teaching AI

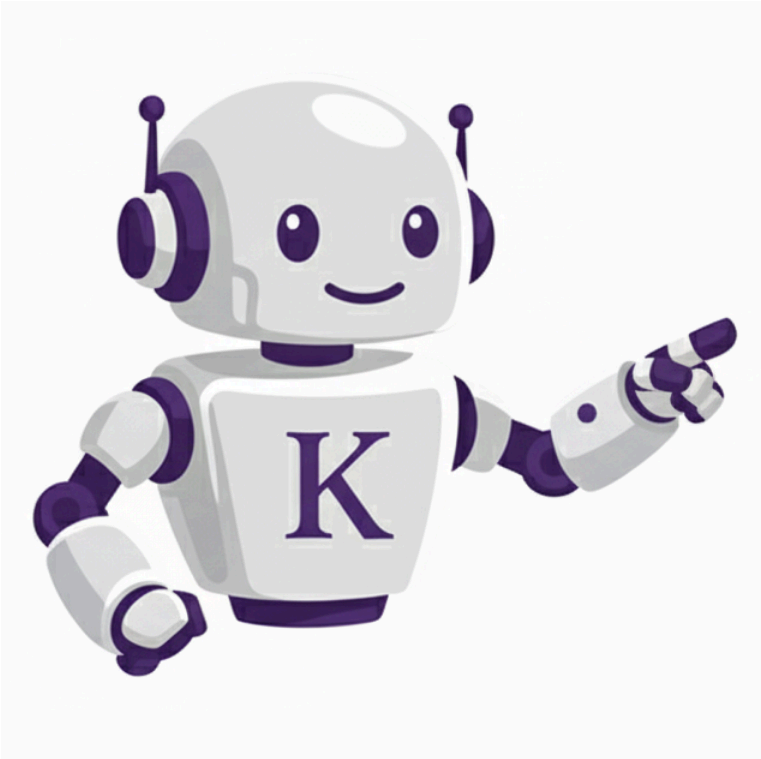
PART 2
RESEARCH

- Does it work? A large experiment

Teaching with an AI TA

LIVE DEMO

Kai, Kellogg's AI teaching assistant



Our teaching assistant

- For students' questions about the course
- For interactive homework
- In my classes since 2024

Some thoughts

The right incentive systems

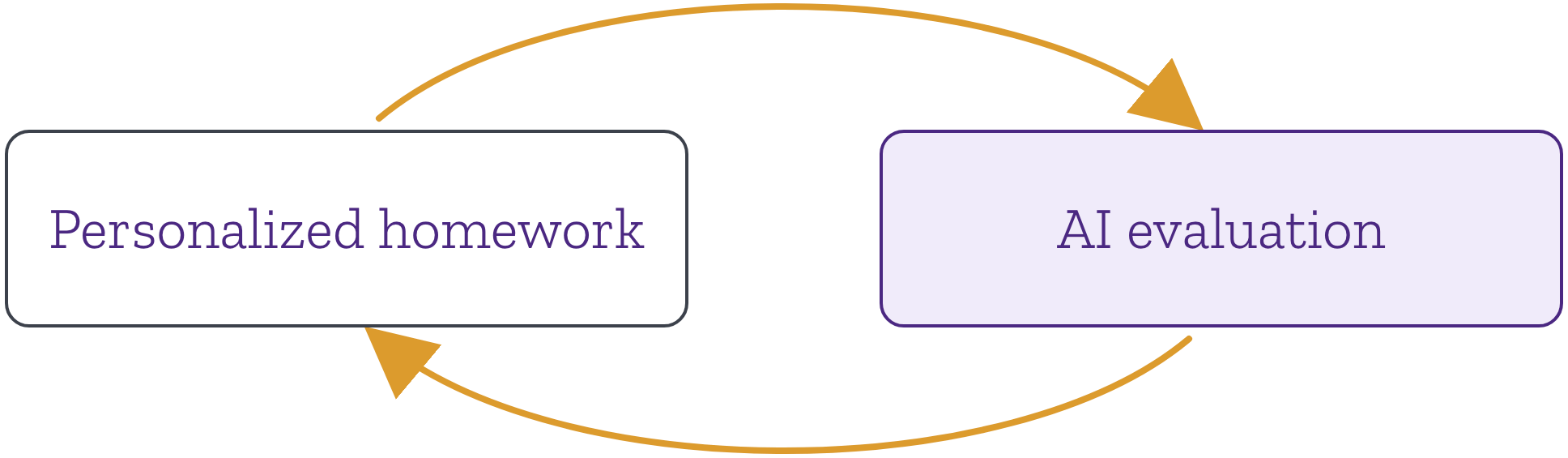
The instructor in control

The contract with your students

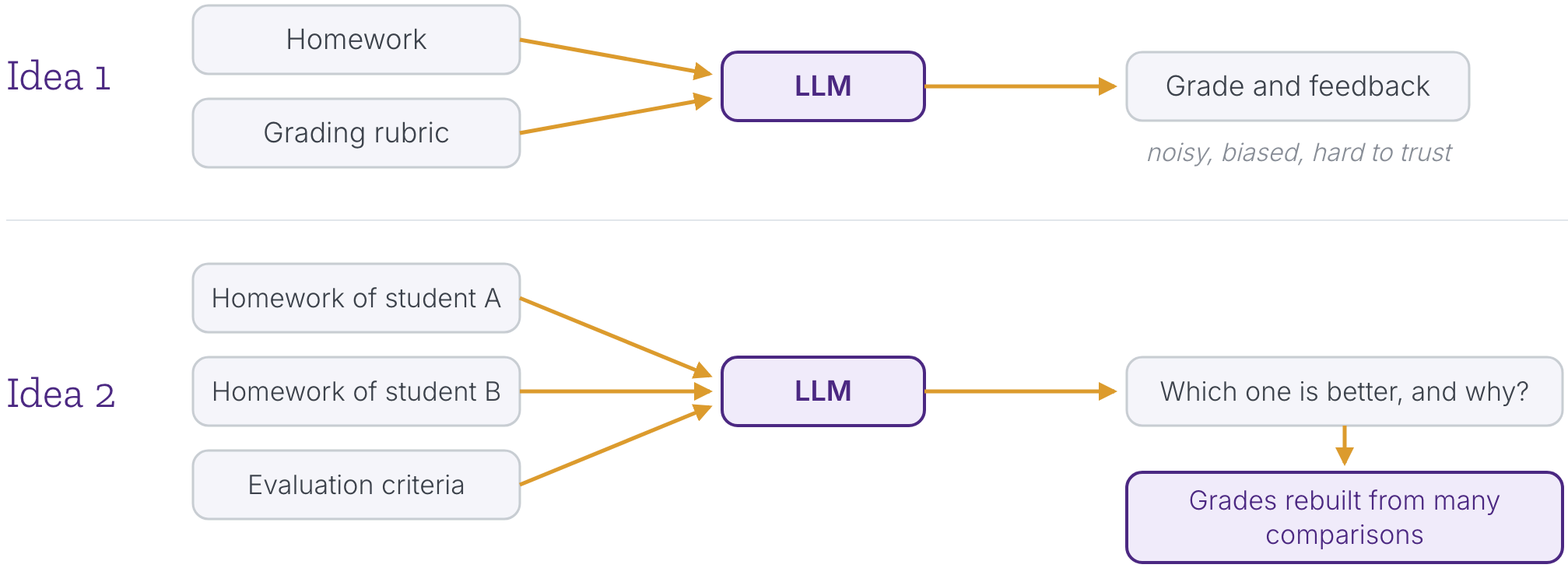
Creativity

LIVE DEMO

AI evaluation and grading



How to grade with AI?



Arnaudo, Bennouna & Martin. Interpretable Feature Discovery from Unstructured Data Using LLMs and Statistical Choice Models. Under review at ICLR 2027. arXiv 2610.38082.

AI case studies

AI CASE STUDIES

Traditional cases, in the age of AI

One document for everyone

- Static, the same for every student.
- Hard to adapt: identical information for every student and every professor.
- Hard to resist just asking ChatGPT to read it for you.
- Richer cases with software or simulations are a costly, heavy lift.



What if students could **interact with the characters** of the case?

An immersive, adaptive simulation, powered by AI, instead of a static PDF.

LIVE DEMO

Transportation Crisis at the Boston Metro School District



Superintendent



CFO



COO



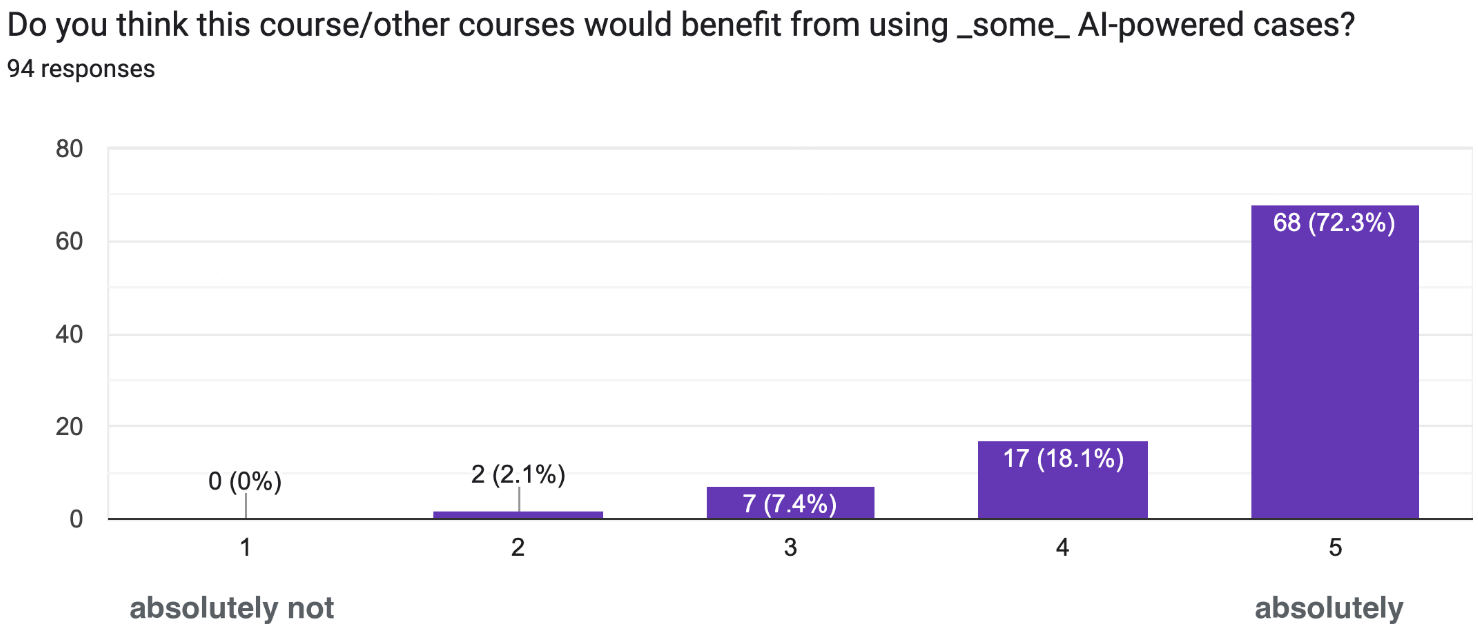
Principal



Bus driver

My final Operations case, on NoYes AI: **students interview five characters**.

Students want more AI cases



Kellogg end-of-course survey, 94 responses. 72% gave the top rating; none gave the lowest.

Student testimonies

"It felt more realistic. I worked in **consulting** before, and it really felt like the first days of a project, interviewing clients, gathering insights. It was **more engaging** too."

"The AI case gave me the taste of real projects where I need to **structure my thinking** and apply what I learned to **ask useful questions**."

"I liked that it incorporated the real-life element of **conflicting perspectives and priorities**, and that it was based on a real case (Boston public schools)."

"**Accessible anytime**, and easy to talk to."

AI CASE STUDIES

Want to try it yourself?


THE WALL STREET JOURNAL.

AI Is Teaching the Next Generation of M.B.A.s the Classic Case Study

Business students at Northwestern aren't using AI to shortcut their way through assignments. It's helping them problem-solve real strategic challenges

By Lindsay Ellis Follow

Nov. 1, 2025 12:00 pm ET

The Wall Street Journal, November 2025


Bloomberg Businessweek

AI Is Changing the Business School Case Study

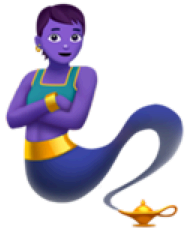
Bloomberg, September 22, 2026

The WSJ article links to a live AI case anyone can open. Both links are on the page at the end.

LIVE DEMO

What's changing with the latest AI technologies

ChatGPT OPENAI	Claude ANTHROPIC	Gemini GOOGLE	Claude Code & Cowork ANTHROPIC	Codex & Work OPENAI
OpenClaw OPEN SOURCE	Hermes NOUS RESEARCH	Muse META	Dots OPENAI	Instinct STARTUP



At this point, think about AI as **a genie**.

WHAT I LEARNED

Why I'm optimistic about AI in education

"Fighting fire with fire"

More work from the human staff!

Increasing my ability to connect with students

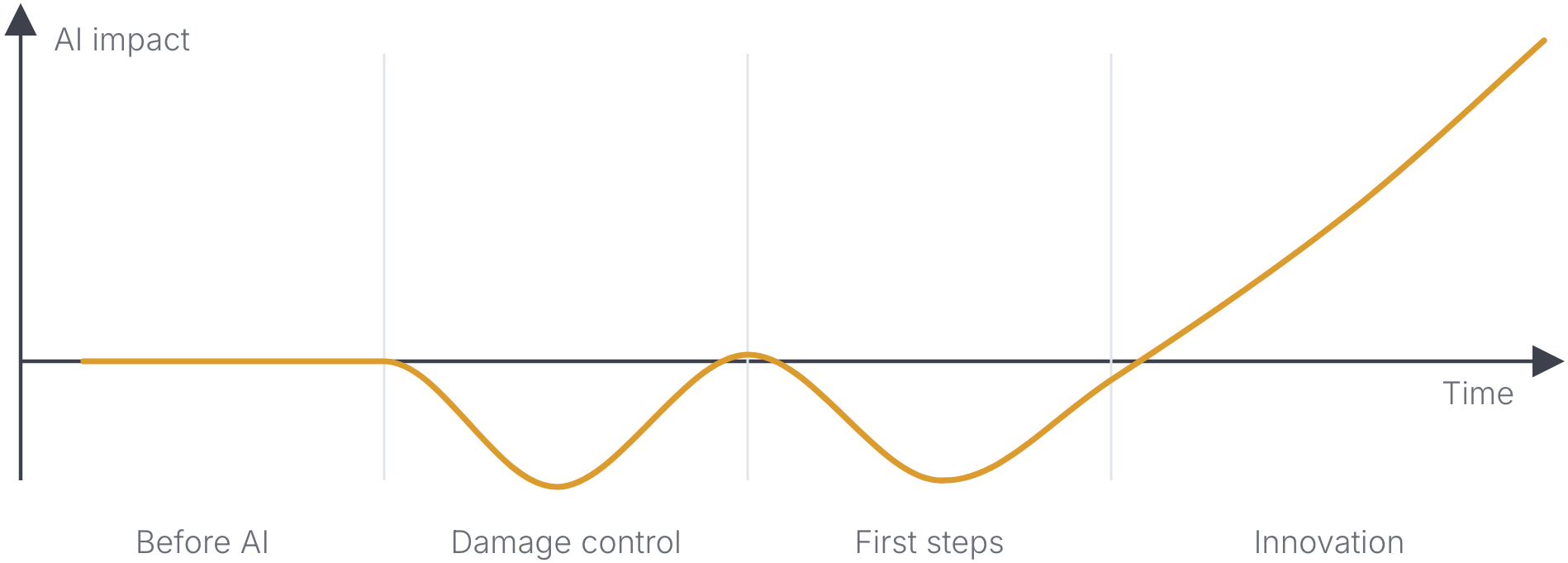
The benefits of personalization

5% AI expertise, 95% instructor expertise

The instructor's imagination is the limit!

WHAT I LEARNED

My experience with AI



Four years in, the one thing that predicts success: **how involved the instructor is.**

Teaching AI

TEACHING AI

AI Foundations for Managers: AgentOps

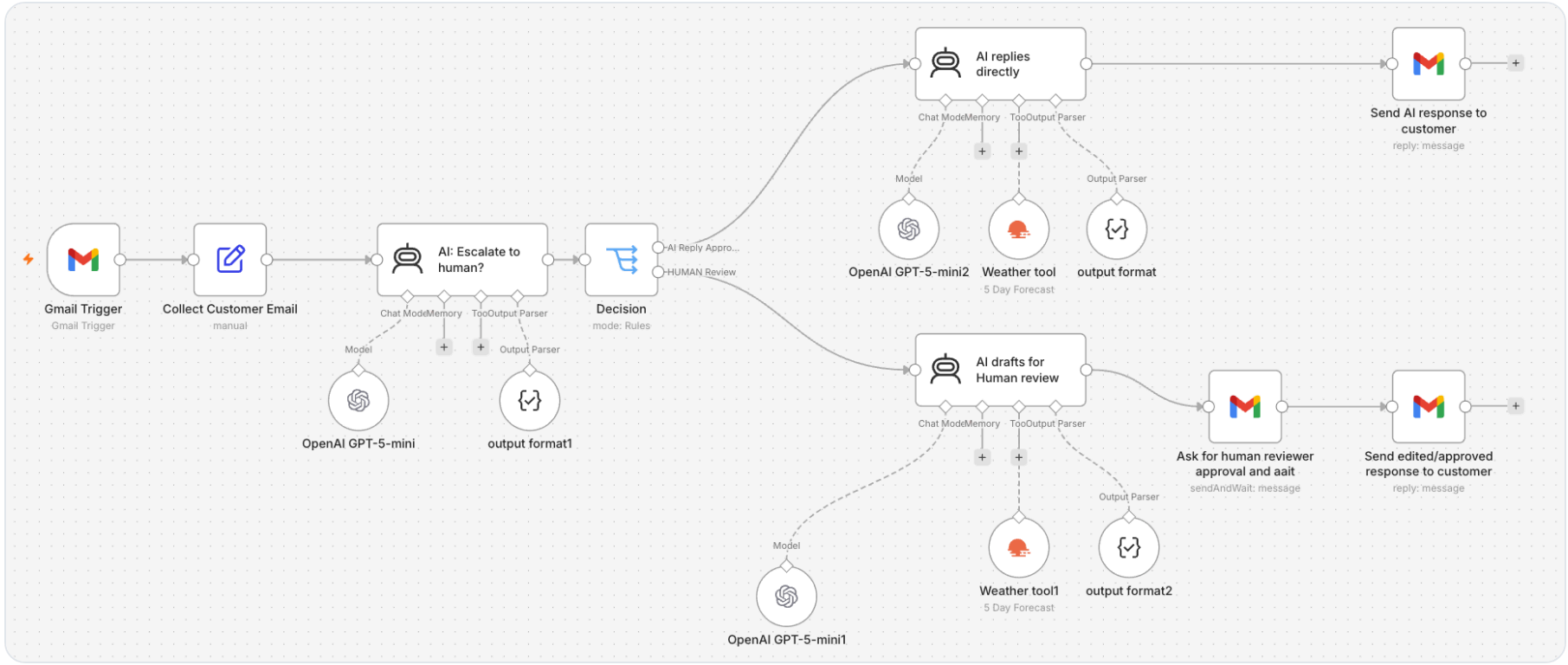
MBA elective · five weeks · no prerequisites

Build from day one → How generative AI works → Agents and the frontier → **AgentOps**

Evaluate, deploy, manage change

Project: an AI agent for a business problem you know

An AI workflow is an operations problem



Who decides, who reviews, and how do we know it works?

Instructor-designed AI improves learning in a randomized trial across 19 universities

with Jun Li, Robert Bray, Nils Van den Steen and Andrew Wu

and Xiao Ni, Arielle Anderer, Ekaterina Astashkina, Volodymyr O. Babich, Jean-Paul Baldwin, Hamsa Bastani, Christian Blanco, Oguzhan Celebi, Christopher Chen, Shi Chen, Benjamin Collier, Yao Cui, Pnina Feldman, Blair Flicker, Brian Han, Yichun Hu, Daniela Hurtado-Lange, Jane Jiang, Zhaohui Jiang, Ismail Kirci, Murat Köksalan, Zhen Lian, Hongyan Liang, Mengshi Lu, Lesley Meng, Ken Moon, Mi Kyong Newsom, Minje Park, Chris Parker, Anyan Qi, Meng Qi, Debjit Roy, Neha Sharma, Alp Sungu, Tom Tan, Necati Tereyagoglu, Hanu Tyagi, Sriram Venkataraman, Guihua Wang, Yixin Wang, Owen Wu, Zonghao Yang, Qiuping Yu, Sophie Yu, Jingwei Zhang, Wenchang Zhang and Ruihao Zhu

arXiv:2611.19099 · Submitted to Nature

THE AI RCT

What do we know about AI in education?

EARLIER TRIALS

Students receive a generic chatbot, or one AI tool designed centrally.

→ **Students learned less**

Bastani et al. 2025 · Fan et al. 2025 · Liu et al. 2026

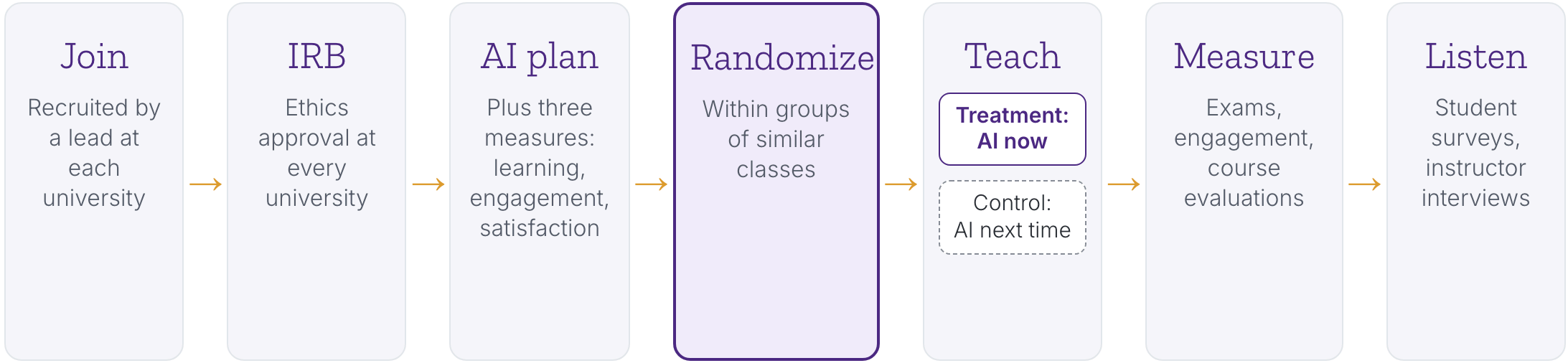
OUR QUESTION

What if **each instructor designs the AI** for their own class?

→ **Does learning improve?**

THE AI RCT

Our overall design



FOR EVERYONE, ALL ALONG

Training and tools: regular Zoom meetings · a yearly in-person workshop · Faculty Twin, an AI platform co-developed with Google

↻ Rolling entry: a new cohort each term, spring 2025 to spring 2026

THE AI RCT

What is randomized: when the class starts

	EARLIER TERMS same course, same instructor	EXPERIMENTAL TERM randomized
Treatment class starts now	No AI from the instructor	**The instructor's AI**
Control class waits one term	No AI from the instructor	Taught as before

A unique treatment: **each instructor designs the AI** for their own class

Not one tool handed to every student. Training and tools are provided to all.

The consortium: 19 universities, 55 classes

University	S25	F25	S26	Class-terms	Enrollment
Carnegie Mellon University	●	·	●	4	412
Cornell University	●	●	●	7	897
Georgetown University	·	·	●	1	100
Indian Institute of Management Ahmedabad	●	·	·	1	157
Indiana University	●	·	●	2	138
Northwestern University	·	·	●	1	131
The Ohio State University	·	●	●	7	852
Purdue University	·	●	●	3	170
Southern Methodist University	●	●	●	5	441
Stevens Institute of Technology	·	●	●	2	50
The University of Hong Kong	·	●	·	2	155
University of Illinois Urbana-Champaign	●	●	●	5	594
University of Michigan	●	●	●	7	835
University of Pennsylvania	●	·	●	5	611
University of South Carolina	●	●	●	4	424
The University of Texas at Dallas	●	●	●	4	318
University of Virginia	·	·	●	3	342
University of Washington	·	●	·	1	156
Yale University	·	●	●	3	417
Total, 19 universities				**67**	**7,200**

19
universities

42
instructors

55
classes
one course, one instructor

7,200
student enrollments

What instructors built: 33 tools, 7 functions

Course navigation	25	A concierge for the syllabus, deadlines and course policies
Socratic coaching	19	A practice-problem tutor that withholds the final answer
Study tutoring	18	A practice-exam generator with new real-world scenarios
Stakeholder role-play	8	A skeptical investment analyst who challenges the student's model
Simulation	8	A negotiation game against an AI counterpart
Reflection	6	A career mentor that ties the course to each student's goals
AI literacy	5	An error hunt: compare the agent's work with your own reasoning

One square = one of the 30 treated classes using AI for this function.

The average class combines 3.0 functions.

Who is in the analysis

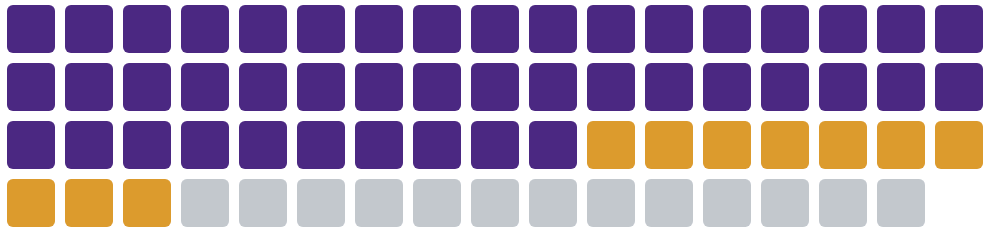
To join

- Has **taught the course before**
- Has **not built AI into it yet**
- No other major change planned

Not pre-registered

- The AI-proof exam filter
- The class-level analysis
- OLS: the pre-registered estimators give the same answer

67 CLASS-TERMS IN THE TRIAL

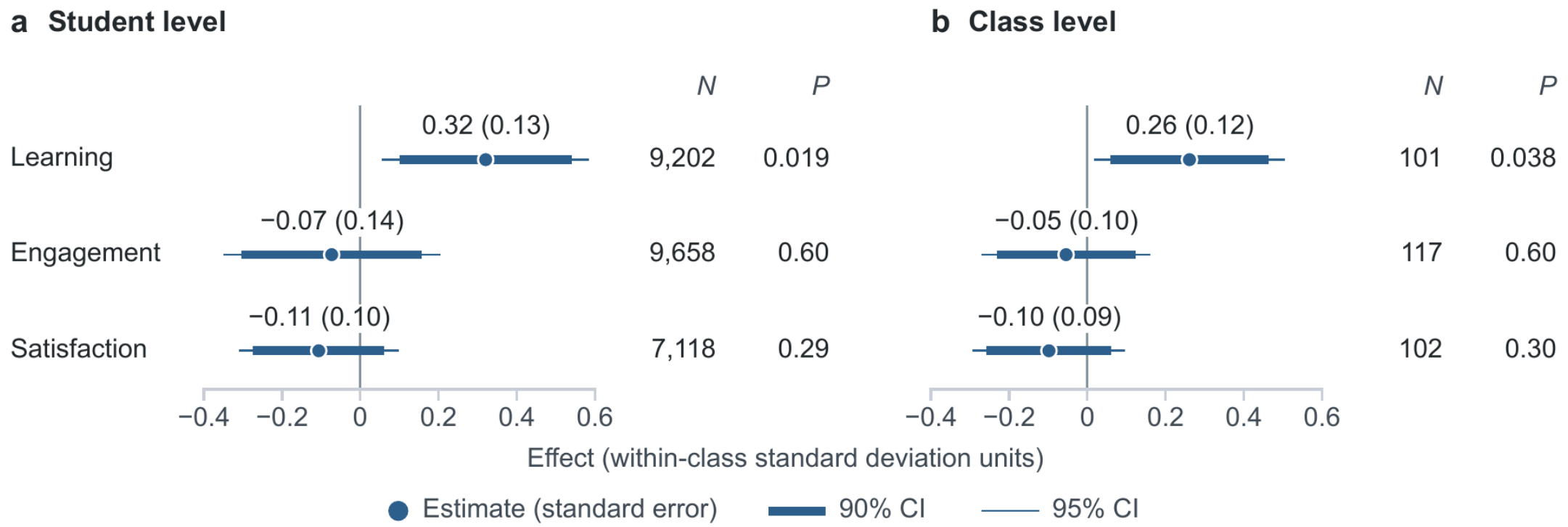


44 in the learning analysis
4,768 student enrollments

10 dropped: the exam was not AI-proof
only proctored, in-person exams count

13 dropped for other reasons
course or assessment changed (6), instructor changed (4), no AI implemented (2), not randomized by mistake (1)

Learning went up by 0.32 standard deviations



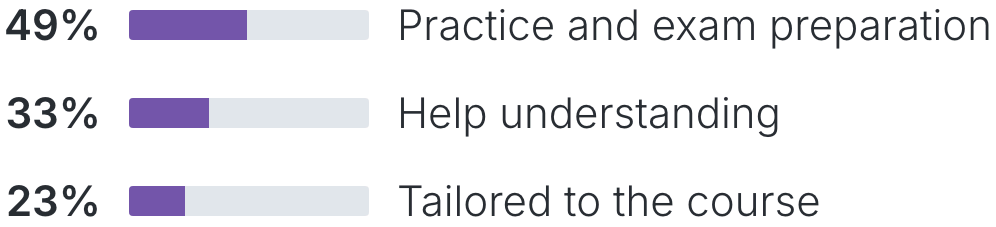
A **statistically significant increase in exam scores**, on proctored, in-person exams.

Student level: the students who agreed to share their data. Class level: every enrolled student, from class averages.

What students and instructors told us

Students

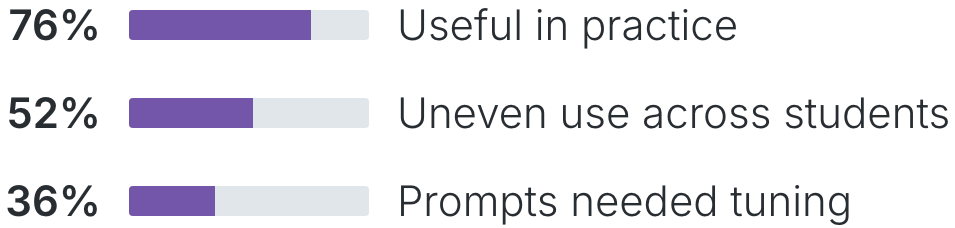
What they liked about the AI · 1,414 open answers



> *"It allowed me to have endless problems."*
>
> A student, end-of-course survey

Instructors

What happened in the classroom · 25 interviews



> *"Students don't use it unless you give bonus points or tie it to a particular class requirement."*
>
> Professor 3

THE AI RCT

Takeaway: put the instructor in charge

AI improved learning when **each instructor designed it** for their own classroom.

WHAT WE CANNOT SAY YET

- Whether it holds beyond applied STEM
- Whether the gain lasts
- Which designs work best

We are growing. **Join us!**

Send me an email if you want to take part:
sebastien.martin@kellogg.northwestern.edu

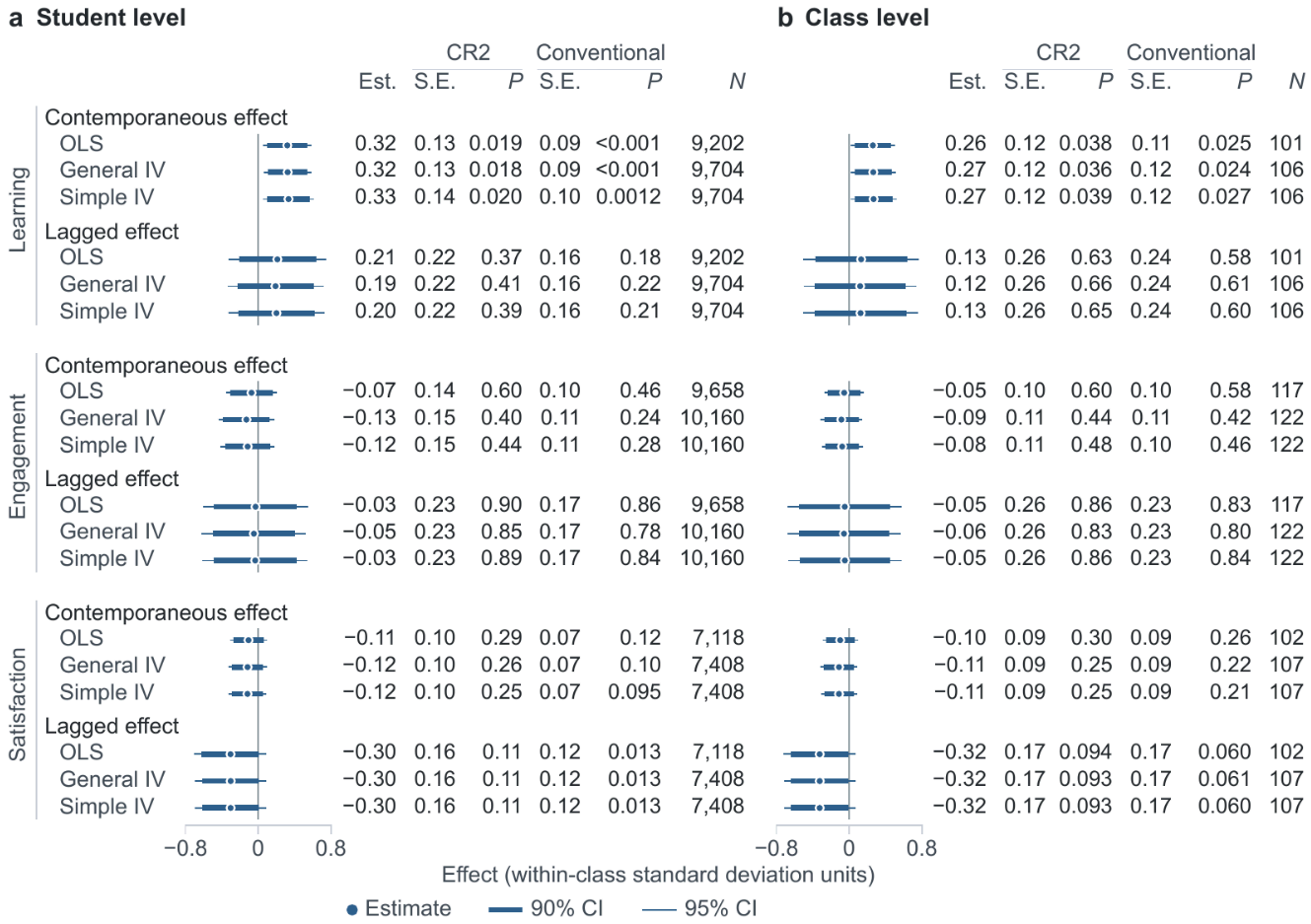
Thank you



The slides, Kai, and all the links:

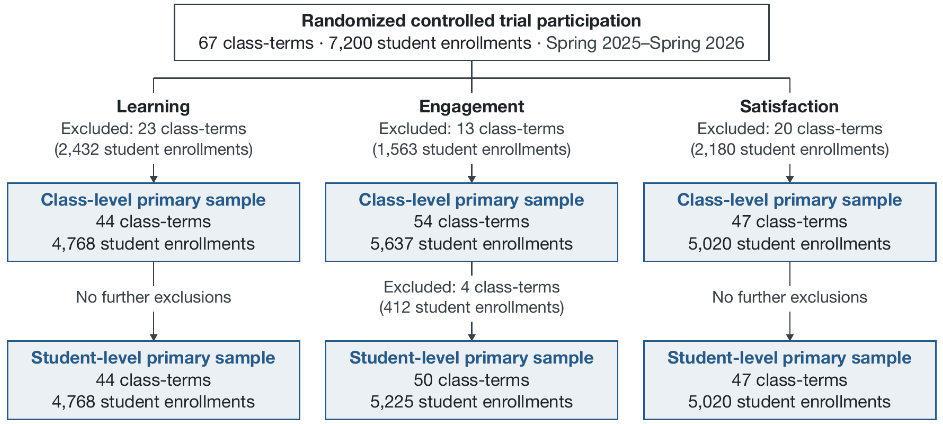
sebastienmartin.info/p/booth.html

All specifications



Sample flowchart

a Selection of primary analysis samples



b Exclusions from primary analyses

Entries are class-terms (student enrollments).

Primary documented reason	Learning	Engagement	Satisfaction
Before class-level analysis			
Major curriculum or assessment change	6 (972)	6 (972)	6 (972)
Assessments not AI proof	10 (869)	0 (0)	0 (0)
Instructor change	4 (383)	4 (383)	4 (383)
No or minimal AI implementation	2 (186)	2 (186)	2 (186)
Missed randomization due to course misclassification	1 (22)	1 (22)	0 (0)
University course evaluation change	0 (0)	0 (0)	8 (639)
Total excluded before class-level analysis	**23 (2,432)**	**13 (1,563)**	**20 (2,180)**
Additional exclusions before student-level analysis			
Unavailable engagement data	0 (0)	4 (412)	0 (0)
Total additional exclusions	**0 (0)**	**4 (412)**	**0 (0)**

Class-term = one term of a course–instructor pair. Each excluded class-term is listed under one reason; the same class-term can appear in several outcome columns.