Assigning homework in a world with ChatGPT

Justin Wolfers, University of Michigan

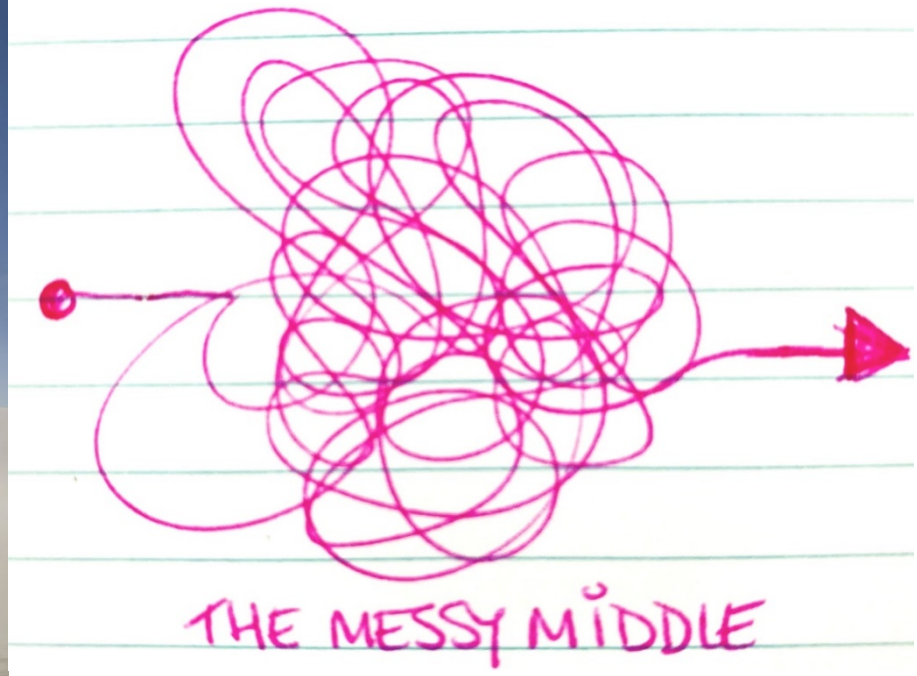
Macmillan AI webinar| Oct 11, 2023

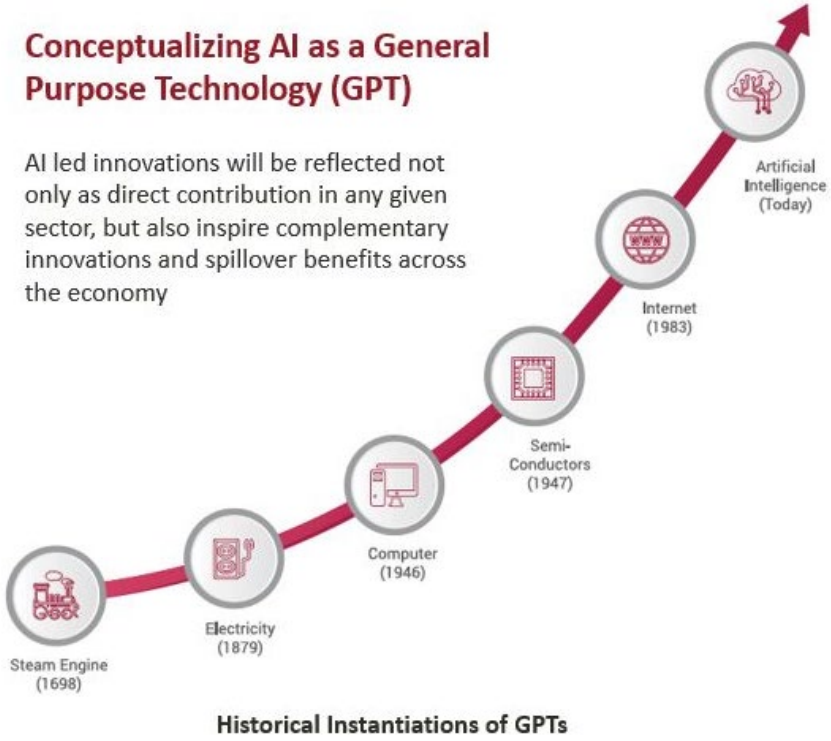
Evolution of responses to GPT

Short run
(last semester)

Medium run
(next semester and next few years)

Long run
(next decade)





Ignore the issue

Muddle through

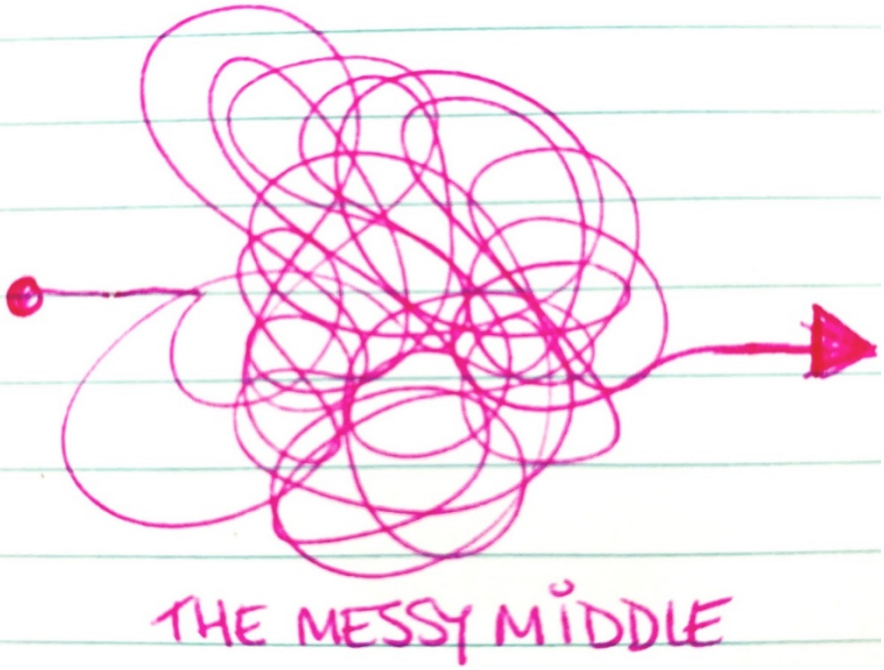
Build skills which are ***complementary*** with GPT

Focus of today's talk

Short run
(last semester)

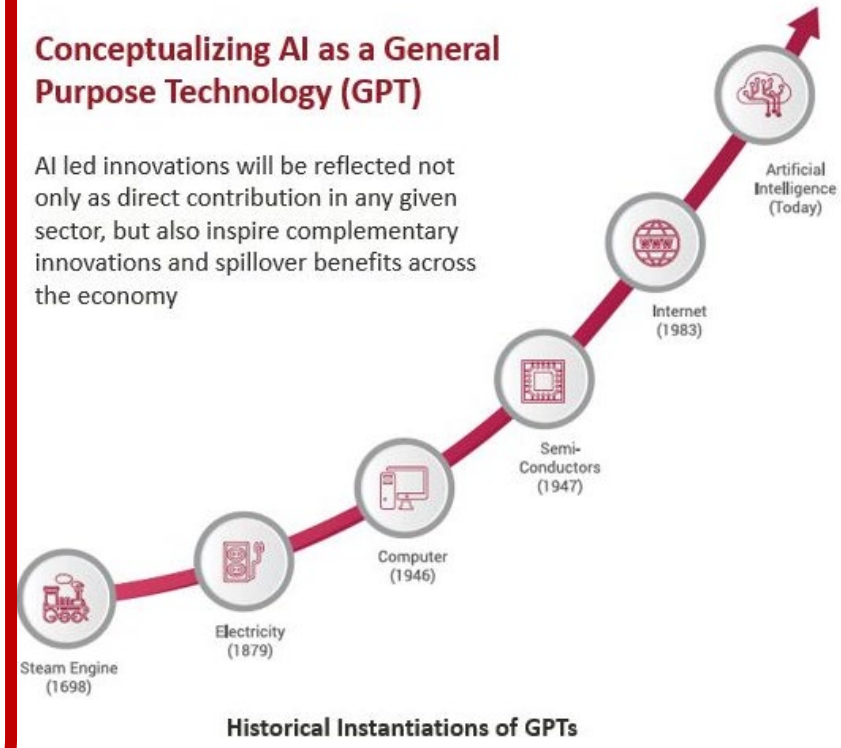
Ignore the issue

Medium run
(next semester and next few years)



Muddle through

Long run
(next decade)



Build skills which are ***complementary*** with GPT

Three motivating facts / conjectures

1. Your **students** are using Chat GPT (and its cousins)
   - And they are (or soon will be) more savvy at using it than most econ instructors

2. **Chat GPT** is better than you think
   - It is *nearly perfect* at introductory econ

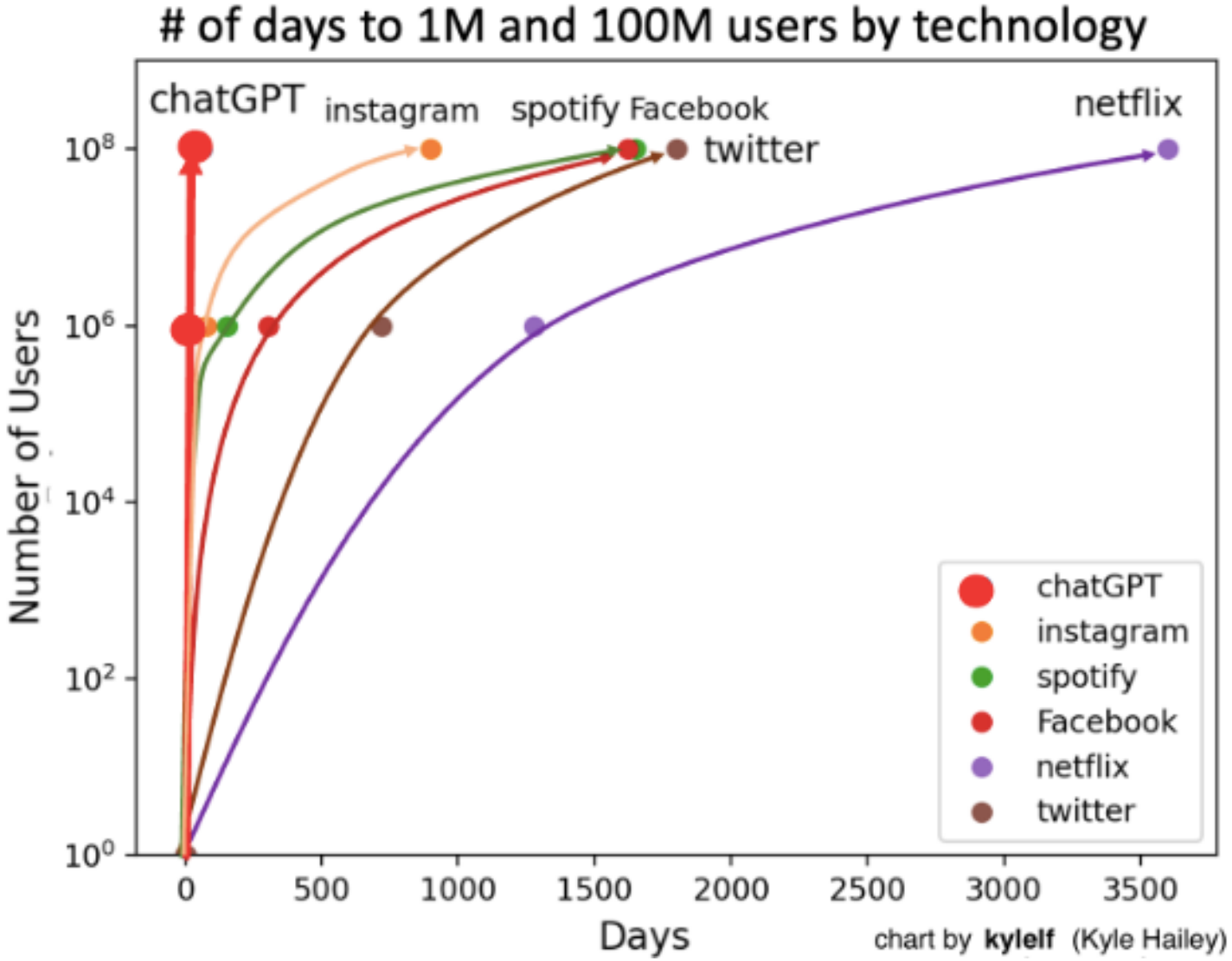
3. Your **university** is likely giving you terrible advice
   - There are no easy ways to eliminate cheating

Three motivating facts

1. **Your students are using Chat GPT**

2. Chat GPT is better than you think

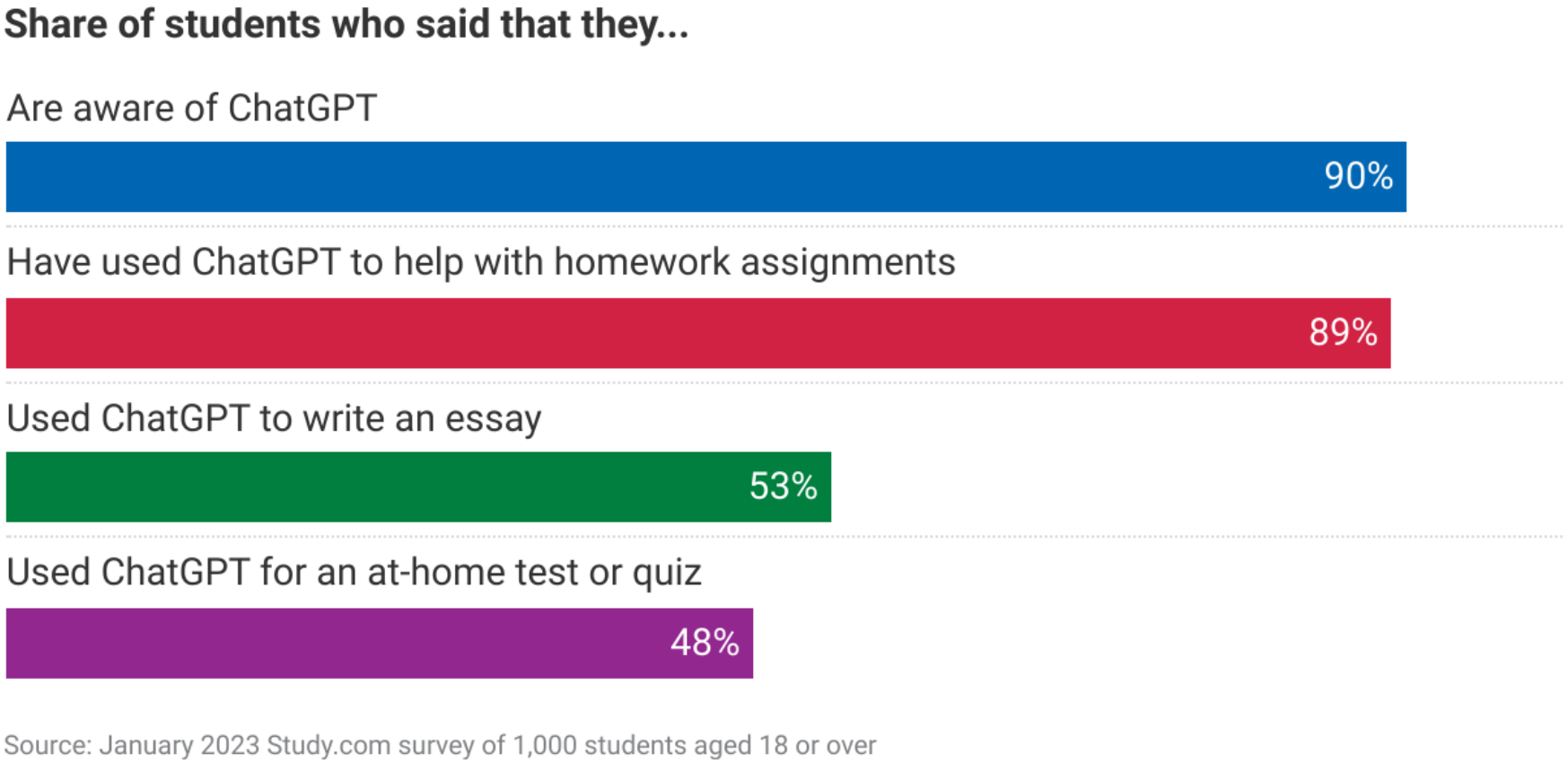
3. Your university is likely giving you terrible advice



Source: https://study.com/resources/perceptions-of-chatgpt-in-schools

Three motivating facts

1. **Your students are using Chat GPT**
2. Chat GPT is better than you think
3. Your university is likely giving you terrible advice



Three motivating facts

1. **Your students are using Chat GPT**

2. Chat GPT is better than you think

3. Your university is likely giving you terrible advice

Qualitative evidence

What Students Said About the Spring of ChatGPT

The growing adoption of AI by students is inevitable, and going into fall faculty will need to revisit their policies early and often, Ross Aikins and Albert Kuo write.

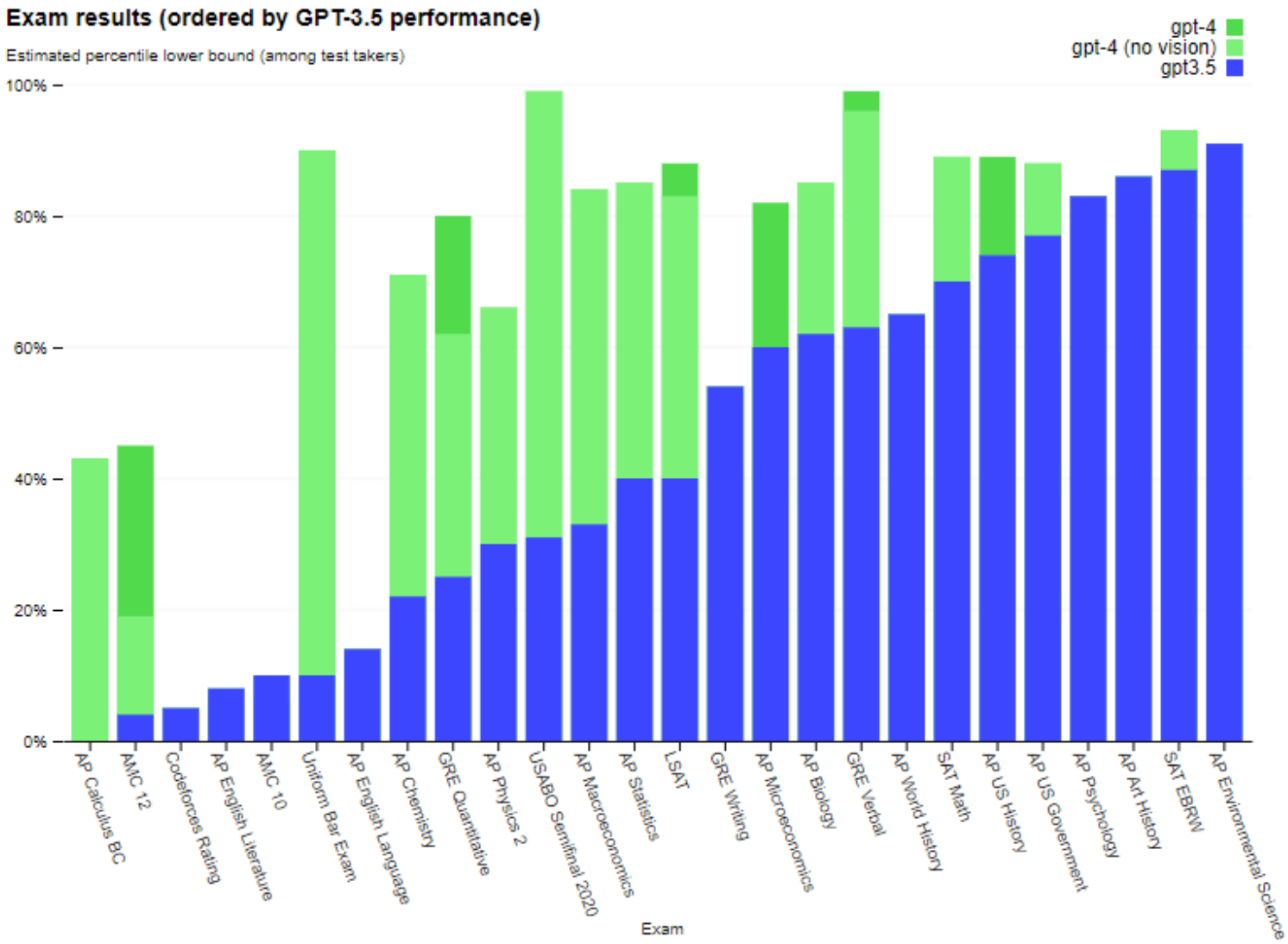
"an ever-growing number of students are turning to AI as a first resort for almost everything"

"Of the 17 students we asked if they plan to use AI again for academic work, all 17 said yes."

Source: Aikins and Kuo, 2023

Three motivating facts

1. Your students are using Chat GPT
2. **Chat GPT is better than you think**
3. Your university is likely giving you terrible advice



Source: https://openai.com/research/gpt-4

Three motivating facts

1. Your students are using Chat GPT

2. **Chat GPT is better than you think**

3. Your university is likely giving you terrible advice

A Harvard student asked her professors to grade ChatGPT's essays. It got mostly A's and B's

"It is beautifully written!"

"Well written and well articulated paper."

"Clear and vividly written."

"The writer's voice comes through very clearly."

"impressive… attention to detail"

Three motivating facts

1. Your students are using Chat GPT
2. **Chat GPT is better than you think**
3. Your university is likely giving you terrible advice

ARTIFICIAL INTELLIGENCE

M.B.A. Students vs. AI: Who Comes Up With More Innovative Ideas?

We put humans and AI to the test. The results weren't even close.

By Christian Terwiesch and Karl Ulrich

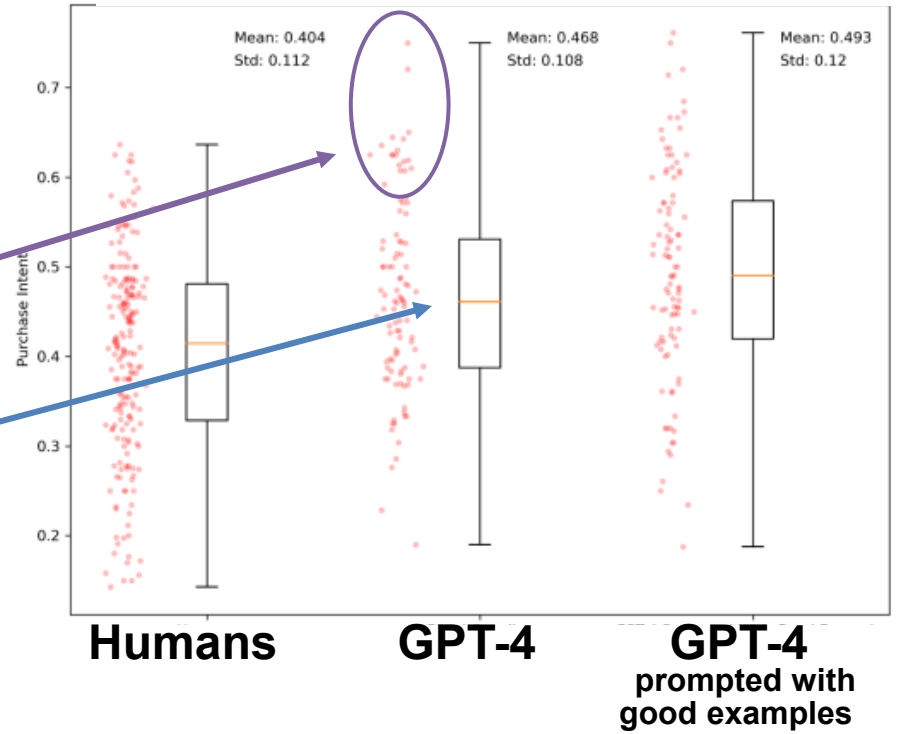
Sept. 9, 2023 9:00 am ET

Task: *Create an idea for a new physical product for the college student market that would be likely to retail for less than $50*

Purchase intent for each idea

Best ideas are better

Average idea is better



Three motivating facts

1. Your students are using Chat GPT

2. **Chat GPT is better than you think** at intro econ

3. Your university is likely giving you terrible advice

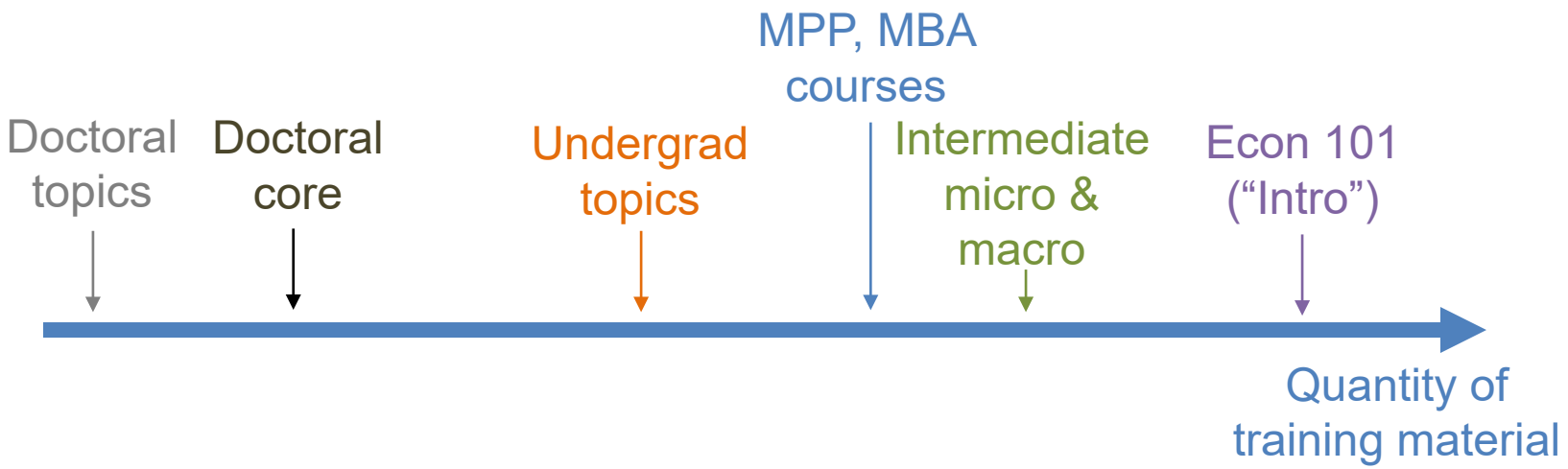
- GRE Quant exam — **80th percentile**
- Harvard Ec. 10 micro essay — **A-**
- Harvard Ec 10 macro essay — **A**
- AP Microeconomics exam — **5 / 5**
- AP Macroeconomics exam — **5 / 5**
- Mankiw test bank questions — **90% correct**

Three motivating facts

1. Your students are using Chat GPT

2. **Chat GPT is better than you think at intro econ**
   but not all econ classes

3. Your university is likely giving you terrible advice



Log scale

This is why ChatGPT is *nearly perfect* at Econ 101
…but not other subjects or levels

Three motivating facts

1. Your students are using Chat GPT
2. Chat GPT is better than you think
3. **Your university is likely giving you terrible advice**

Justin Wolfers
@JustinWolfers

I've spent much of the past few weeks plowing through the resources that universities are pushing out to faculty as a response to the rise of generative AI (ie ChatGPT).

Without calling out any institution, I'm utterly pessimistic about the responses I've seen.

Justin Wolfers
@JustinWolfers

Few universities are reading the moment correctly. Generative AI ought to lead to a wholesale reinvention of what we teach and how we teach it, not to mention how we assess it.

Instead, there's an admission that change is in the air, quickly followed by an attempt to minimize it

Justin Wolfers
@JustinWolfers

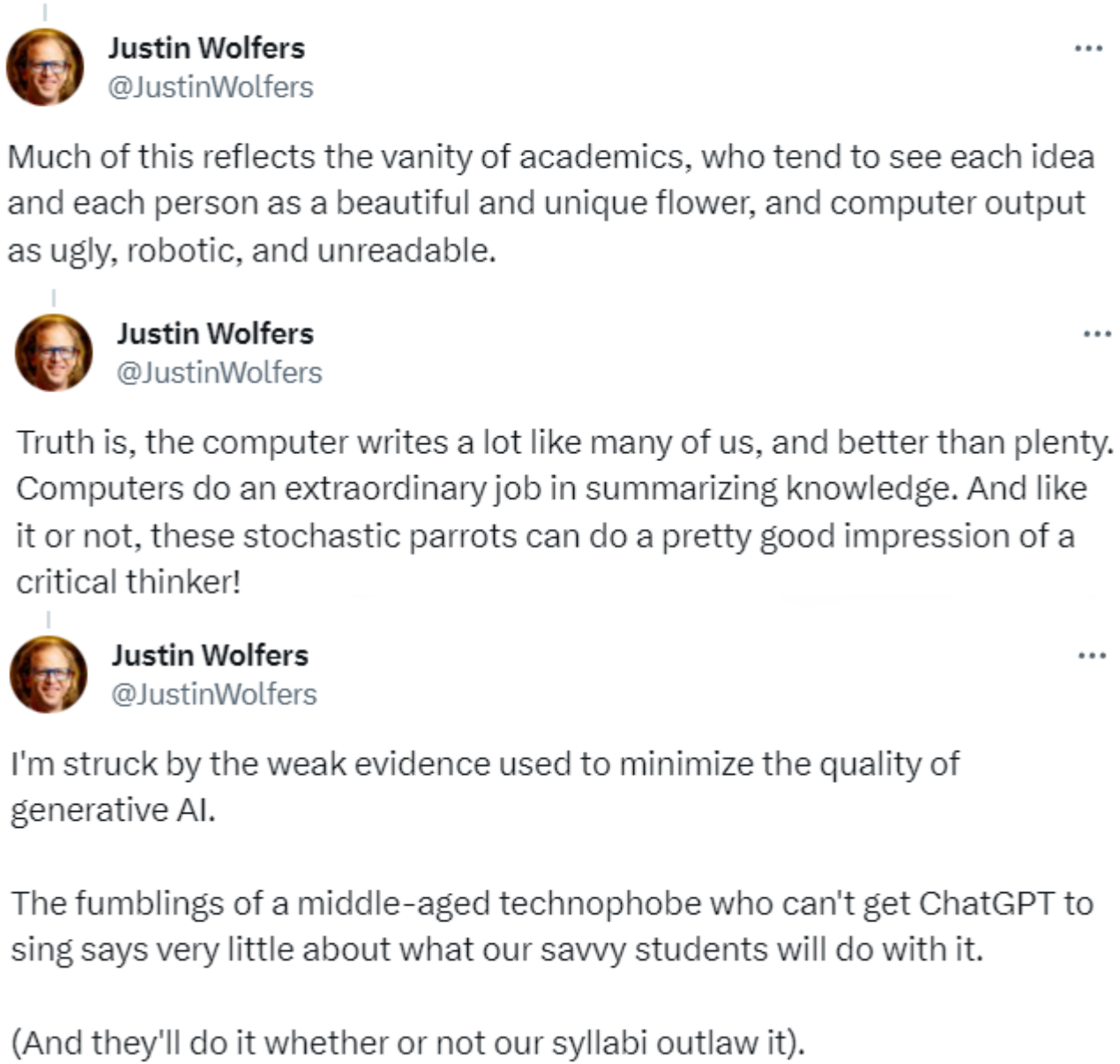
The university guidance to faculty about generative AI that I've read dwells far too much (and with far too little evidence) on the shortcomings of AI, and not enough on its extraordinary power.

Three motivating facts

1. Your students are using Chat GPT

2. Chat GPT is better than you think

3. **Your university is likely giving you terrible advice**



Justin Wolfers
@JustinWolfers

Much of this reflects the vanity of academics, who tend to see each idea and each person as a beautiful and unique flower, and computer output as ugly, robotic, and unreadable.

Justin Wolfers
@JustinWolfers

Truth is, the computer writes a lot like many of us, and better than plenty. Computers do an extraordinary job in summarizing knowledge. And like it or not, these stochastic parrots can do a pretty good impression of a critical thinker!

Justin Wolfers
@JustinWolfers

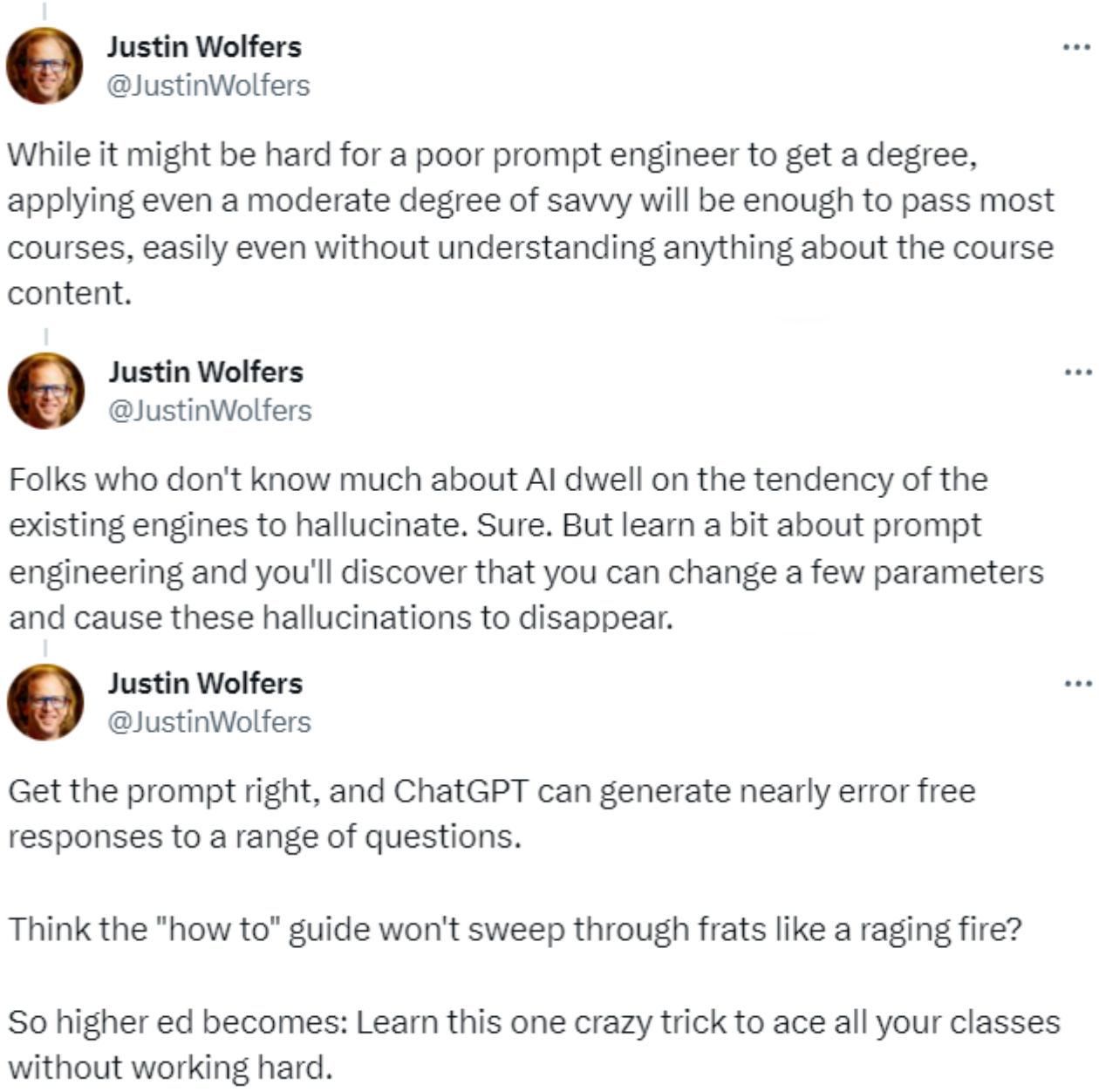
I'm struck by the weak evidence used to minimize the quality of generative AI.

The fumblings of a middle-aged technophobe who can't get ChatGPT to sing says very little about what our savvy students will do with it.

(And they'll do it whether or not our syllabi outlaw it).

Three motivating facts

1. Your students are using Chat GPT
2. Chat GPT is better than you think
3. **Your university is likely giving you terrible advice**



Justin Wolfers
@JustinWolfers

While it might be hard for a poor prompt engineer to get a degree, applying even a moderate degree of savvy will be enough to pass most courses, easily even without understanding anything about the course content.

Justin Wolfers
@JustinWolfers

Folks who don't know much about AI dwell on the tendency of the existing engines to hallucinate. Sure. But learn a bit about prompt engineering and you'll discover that you can change a few parameters and cause these hallucinations to disappear.

Justin Wolfers
@JustinWolfers

Get the prompt right, and ChatGPT can generate nearly error free responses to a range of questions.

Think the "how to" guide won't sweep through frats like a raging fire?

So higher ed becomes: Learn this one crazy trick to ace all your classes without working hard.

Three motivating facts

1. Your students are using Chat GPT
2. Chat GPT is better than you think
3. **Your university is likely giving you terrible advice**

A meta theory

Universities	Students
❑ Looking for what GPT **can not** do	❑ Looking for what GPT **can** do
❑ Confirmation bias: Search for ways in which Chat GPT *is* ineffective	❑ Problem-solving: Search for ways to *make* ChatGPT effective

Three motivating facts

1. Your students are using Chat GPT
2. Chat GPT is better than you think
3. **Your university is likely giving you terrible advice**

Universities see shortcomings	Students find solutions
Responses are formulaic	"Sharpen that essay"
Hallucinates facts (& citations)	Set temperature = 0
Makes (math) errors	Use GPT-4, not GPT 3.5
Not up to date	Use Bard not GPT
Computers lack creativity	Not true
Trained on canon of dead white men	So are our intro econ students

Prompt engineering

Poll time!

My assertion: *"Your university is likely giving you terrible advice"*

This assertion is:

A. …is (roughly) correct
B. …understates the extent of the problem
C. …overstates the extent of the problem
D. …dead wrong

Roadmap

- ☑ Three motivating facts

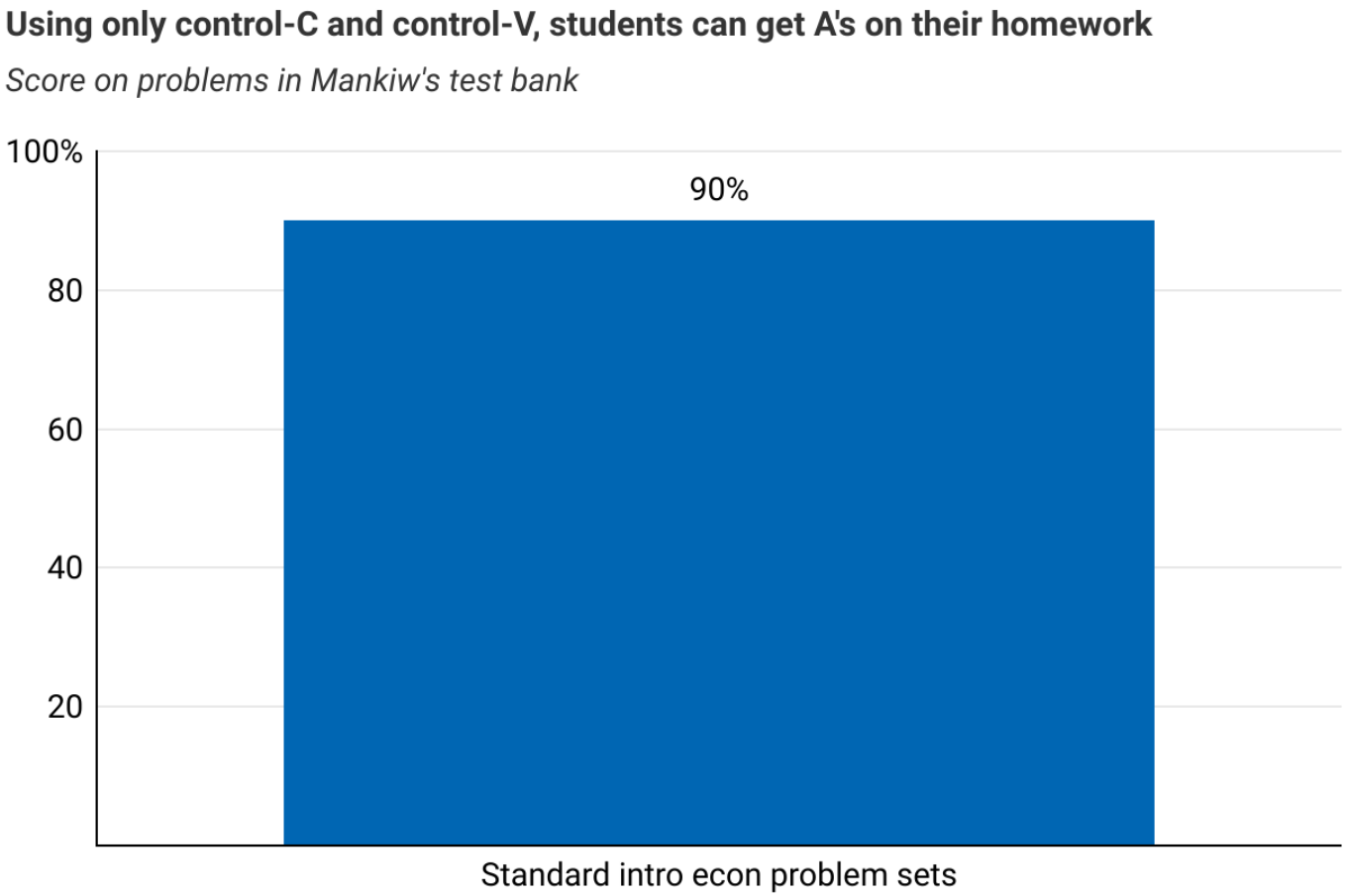
- ❑ What won't work
  - Don't ignore the budget constraint

- ❑ Chat GPT as foe: Preventing cheating
  - Lessons from the economics of crime

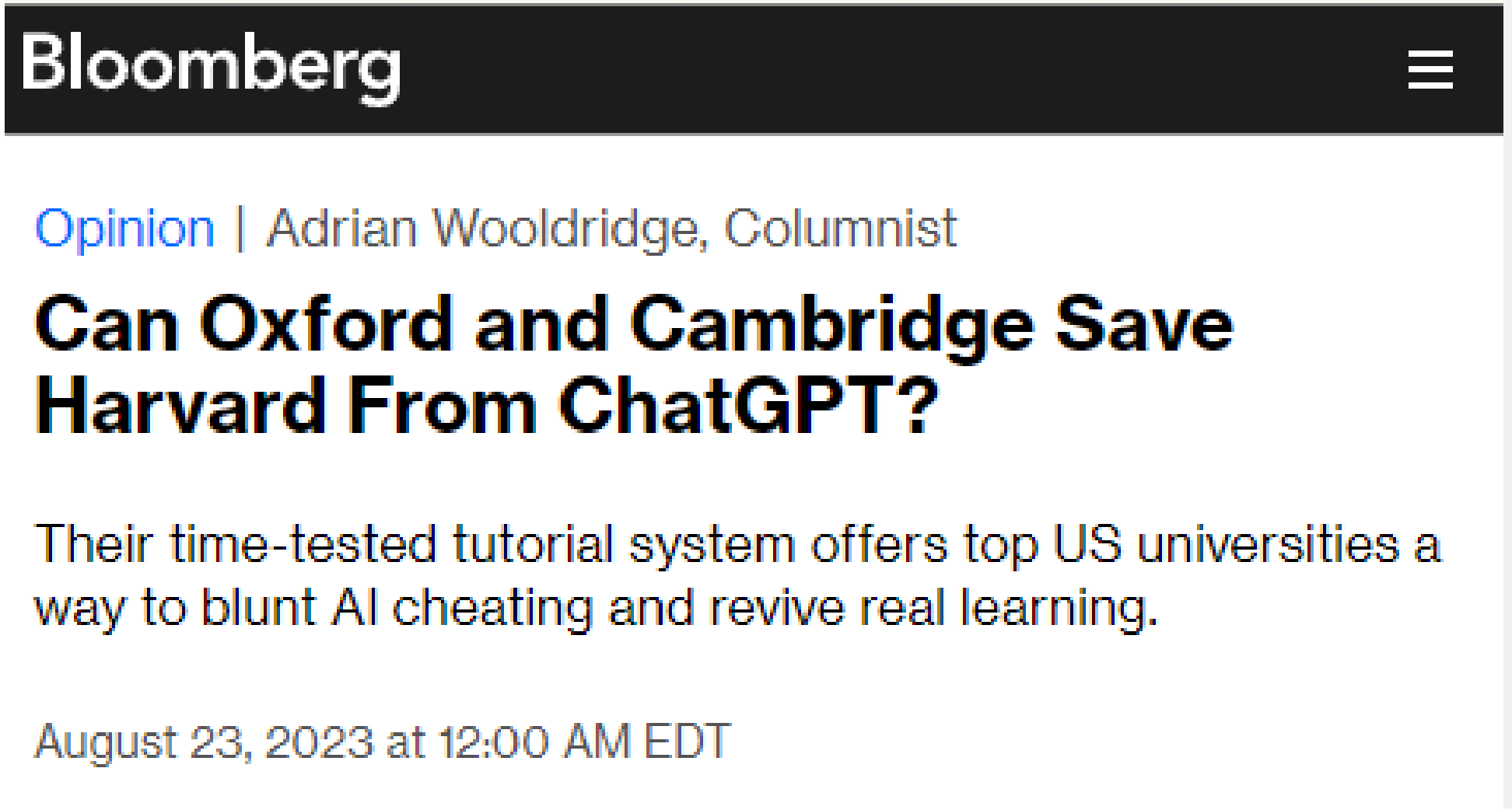
- ❑ Chat GPT as friend: Enhancing learning

What won't work: Ignoring the problem

- Assign standard homework



What won't work



Bloomberg

Opinion | Adrian Wooldridge, Columnist

Can Oxford and Cambridge Save Harvard From ChatGPT?

Their time-tested tutorial system offers top US universities a way to blunt AI cheating and revive real learning.

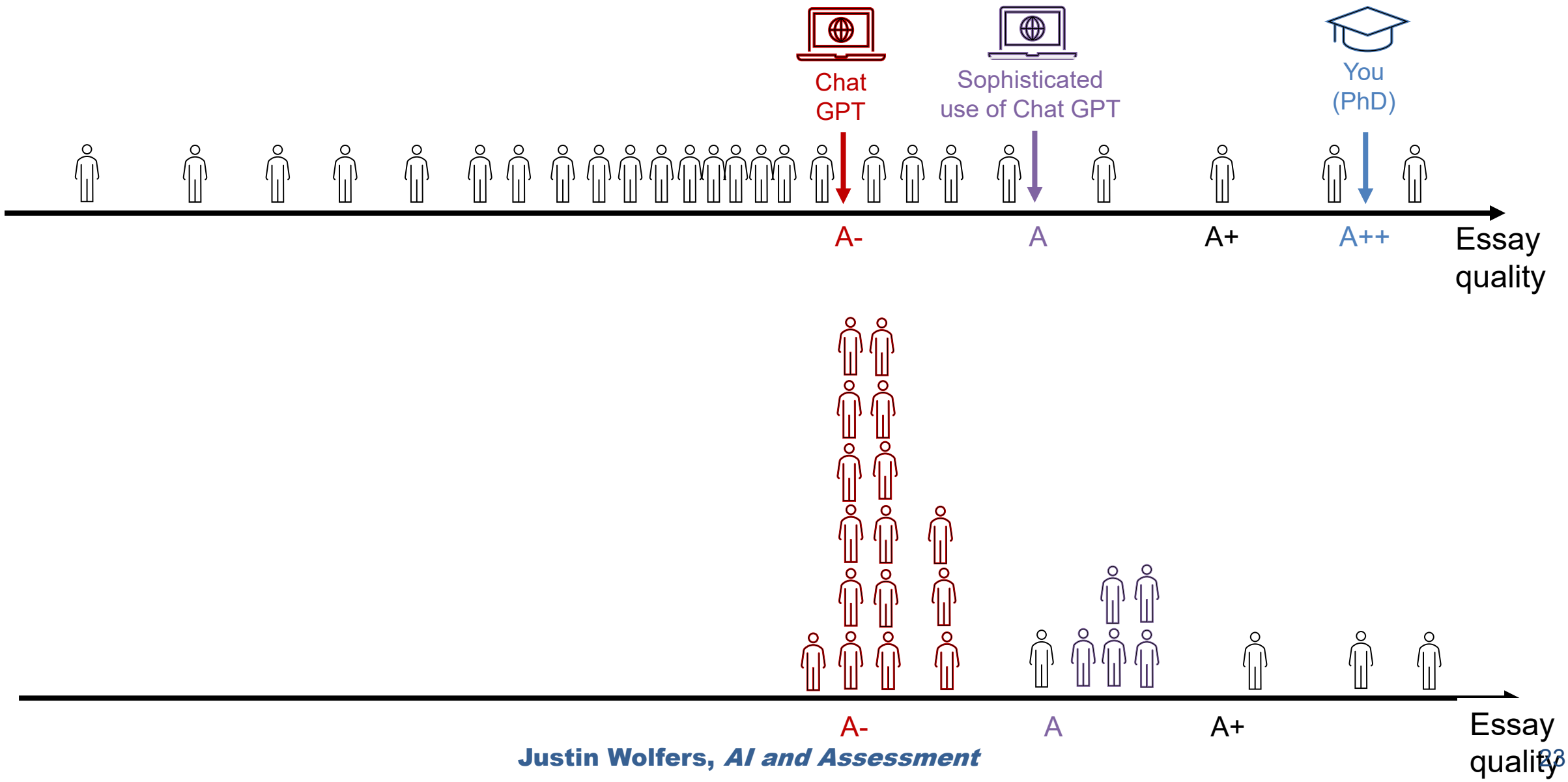
August 23, 2023 at 12:00 AM EDT

What won't work

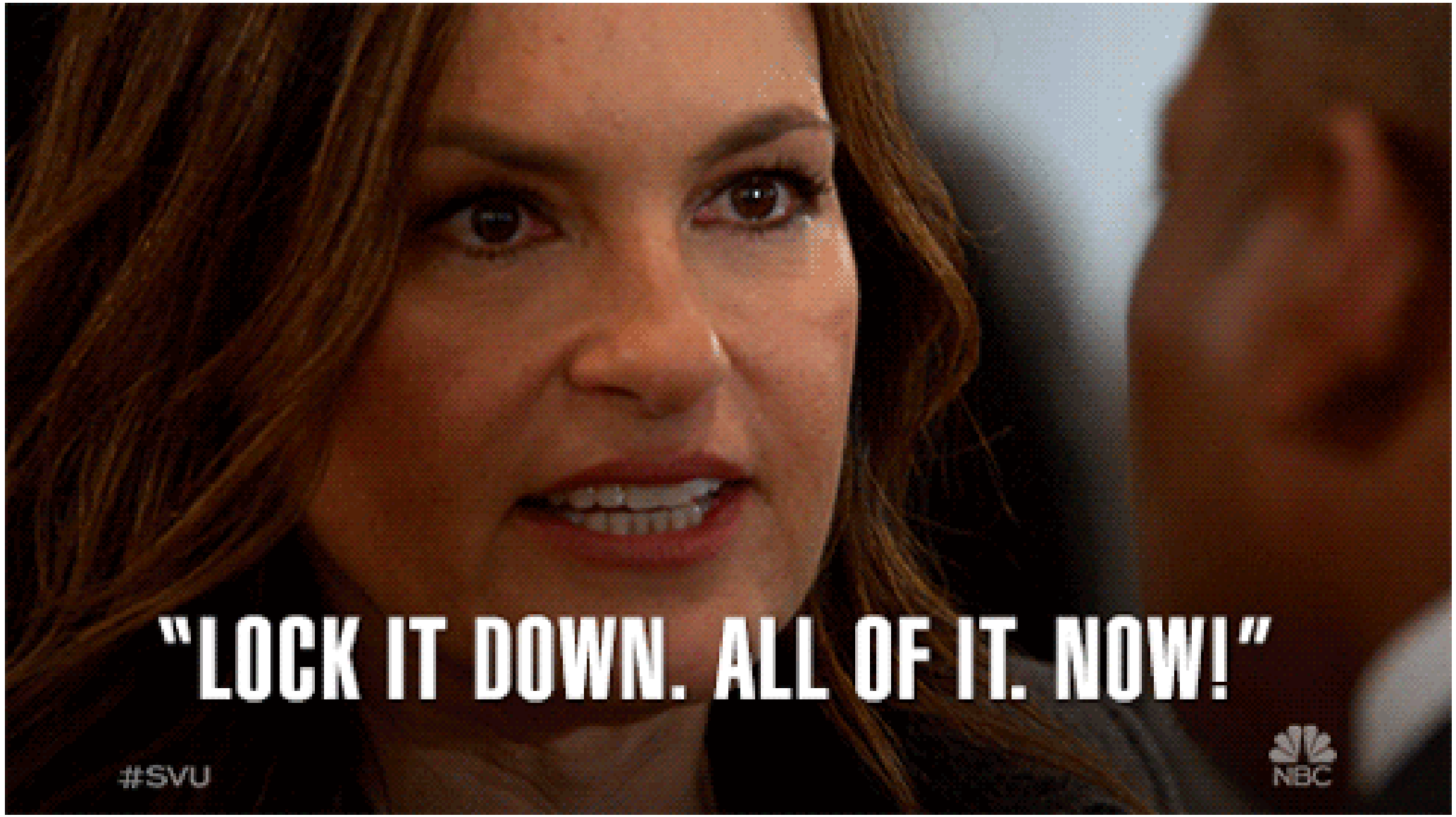
Don't Want Students to Rely on ChatGPT? Have Them Use It

It's easy to forget how little students and educators understand generative AI's flaws. Once they actually try it out, they'll see that it can't replace them.

The assessment problem with AI-aided essays



What works. But at what cost?



Roadmap

- ☑ Three motivating facts

- ❑ What won't work
  - Don't ignore the budget constraint

- ❑ Chat GPT as foe: Preventing cheating
  - Lessons from the economics of crime

- ❑ Chat GPT as friend: Enhancing learning
  - A complement that compliments

The Cost-Benefit Principle applied to tonight's homework

- Individual student: Decision to cheat (use ChatGPT) depends on:

Benefits

- Save time (and cognitive effort)
- Higher grade

Costs

- Missed opportunity to learn
- Consequences if caught

The Cost-Benefit Principle applied to tonight's homework

- Individual student: Decision to cheat (use ChatGPT) depends on:

Benefits

- Save time (and cognitive effort)
- Higher grade

Costs

- Missed opportunity to learn
- Consequences if caught

- Applying the economics of crime, a student will cheat (use ChatGPT) iff:

$$\text{Benefit} > \text{Cost}$$

$$\Delta Time + \alpha \Delta Points > p_{caught}(Punishment + Stigma) + \lambda \Delta Learning$$

The economics of crime, applied to the decision to cheat

Cheat if $\Delta Time + \alpha \Delta Points > p_{caught}(Punishment + Stigma) + \lambda \Delta Learning$

Failed strategy: Detection

Can AI-Generated Text be Reliably Detected?

Vinu Sankar Sadasivan vinu@umd.edu
Aounon Kumar aounon@umd.edu
Sriram Balasubramanian sriramb@umd.edu
Wenxiao Wang wwx@umd.edu
Soheil Feizi sfeizi@umd.edu
Department of Computer Science
University of Maryland

"Empirically, we show that paraphrasing attacks, where a light paraphraser is applied on top of the generative text model, can break a whole range of detectors, including the ones using the watermarking schemes as well as neural network-based detectors and zero-shot classifiers."

We tested a new ChatGPT-detector for teachers. It flagged an innocent student.

Five high school students helped our tech columnist test a ChatGPT detector coming from Turnitin to 2.1 million teachers. It missed enough to get someone in trouble.

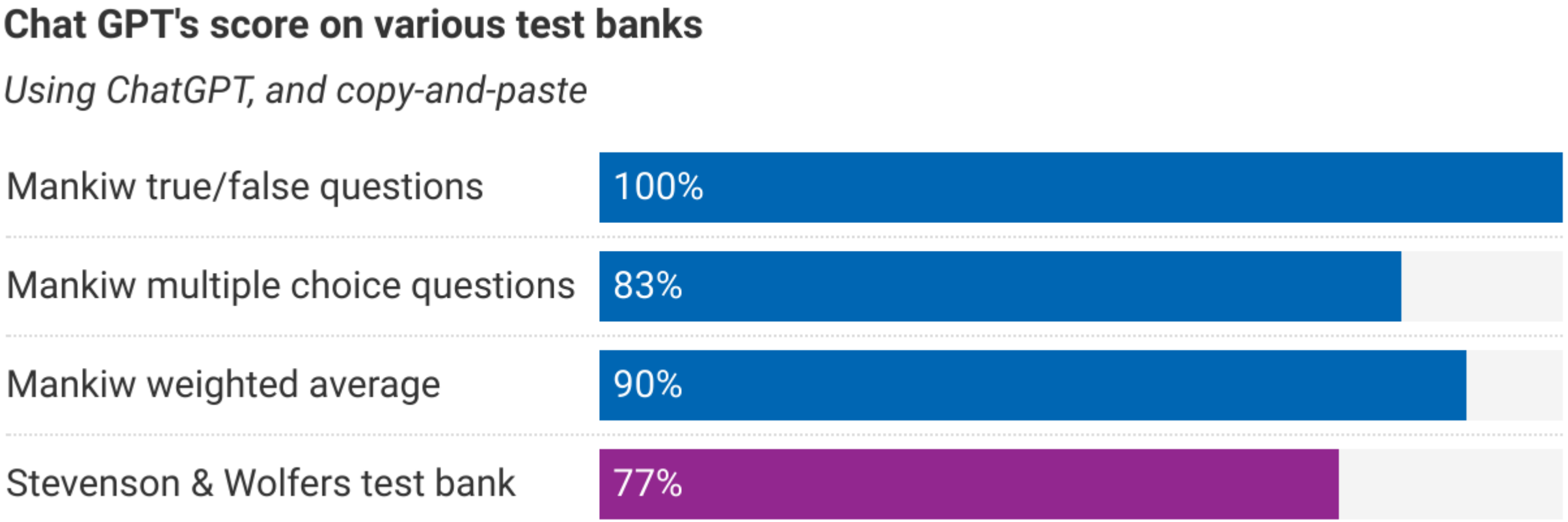
To see what's at stake, I asked Turnitin for early access to its software. Five high school students, including Goetz, volunteered to help me test it by creating 16 samples of real, AI-fabricated and mixed-source essays to run past Turnitin's detector.

The result? It got over half of them at least partly wrong. Turnitin accurately identified six of the 16 — but failed on three, including a flag on 8 percent of Goetz's original essay. And I'd give it only partial credit on the remaining seven, where it was directionally correct but misidentified some portion of ChatGPT-generated or mixed-source writing.

The economics of crime, applied to the decision to cheat

Cheat if $\Delta Time + \alpha \Delta Points > p_{caught}(Punishment + Stigma) + \lambda \Delta Learning$

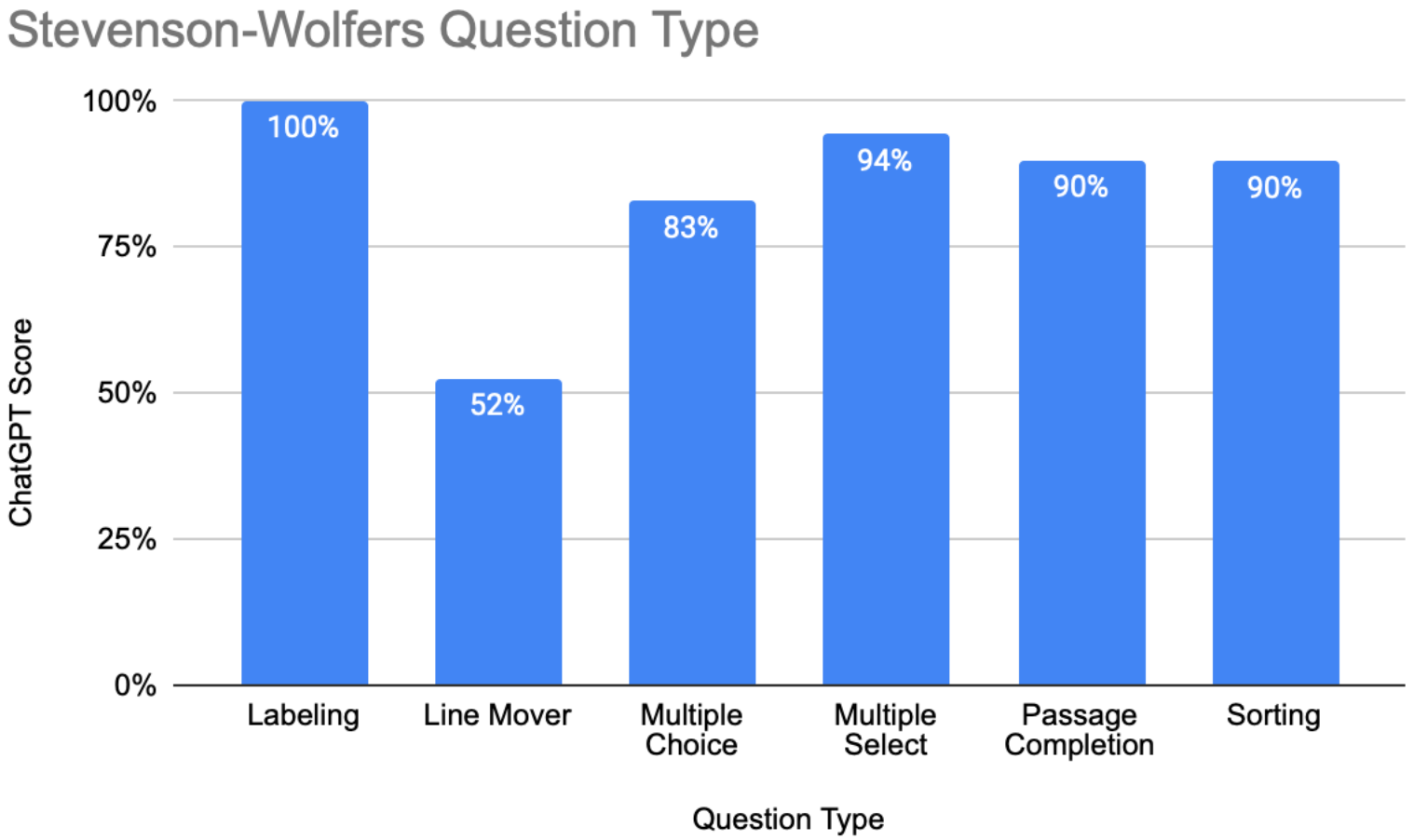
Partial strategy: Try to render LLM's less reliable



The economics of crime, applied to the decision to cheat

Cheat if $\Delta Time + \alpha \Delta Points > p_{caught}(Punishment + Stigma) + \lambda \Delta Learning$

Partial strategy: Try to render LLM's less reliable



The economics of crime, applied to the decision to cheat

Cheat if $\Delta Time + \alpha \boldsymbol{\Delta Points} > p_{caught}(Punishment + Stigma) + \lambda \Delta Learning$

Partial strategy: Try to render LLM's less reliable

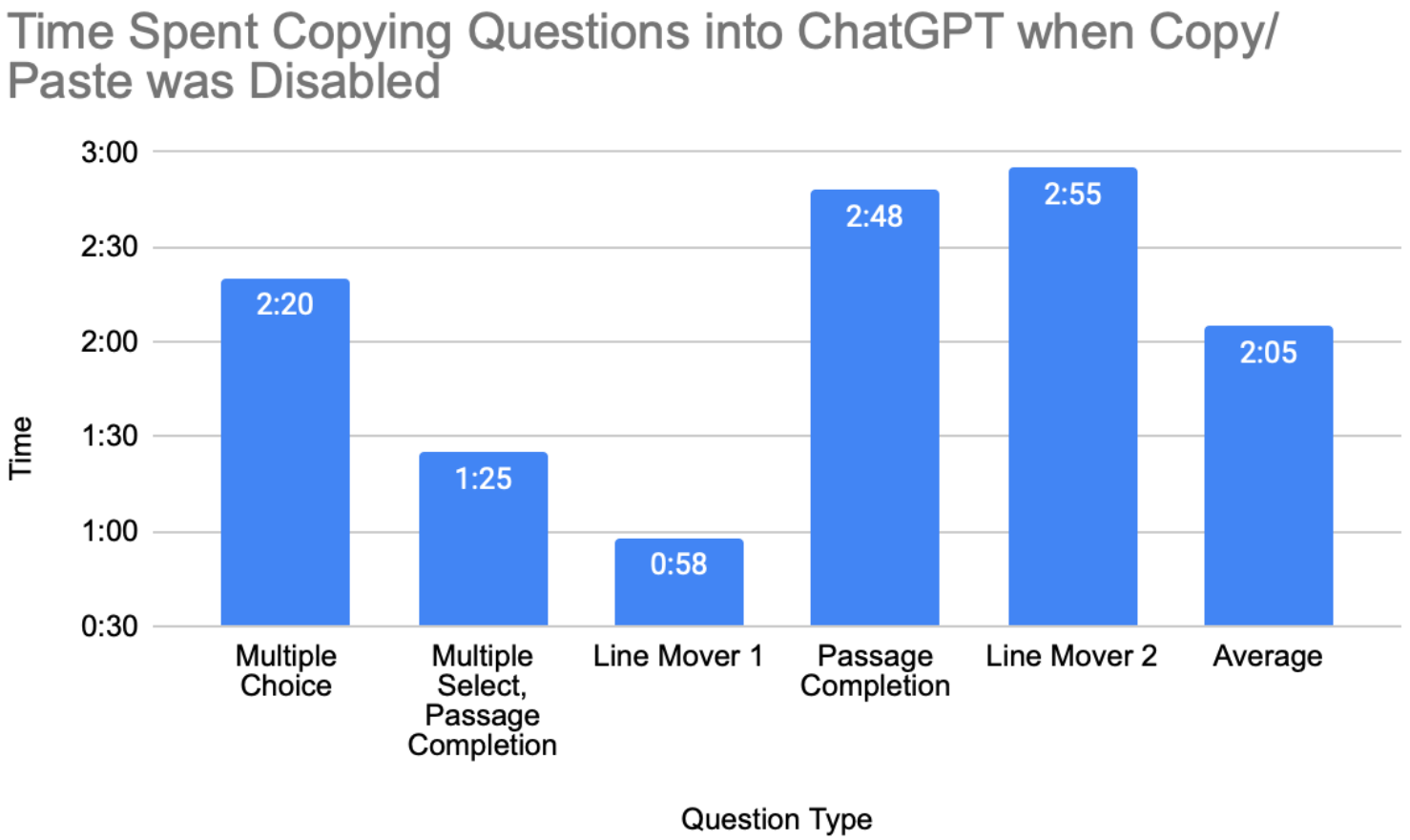
Exploit (existing) weaknesses in ChatGPT

- Use more graphs, images, and tables
- Multi-step questions that build off of embedded information in the first part of the question
- Ask for the specific values when writing "Line Mover" questions instead of just the direction of the shift.
- Ask about interdependencies between markets

The economics of crime, applied to the decision to cheat

Cheat if $\Delta Time + \alpha \Delta Points > p_{caught}(Punishment + Stigma) + \lambda \Delta Learning$

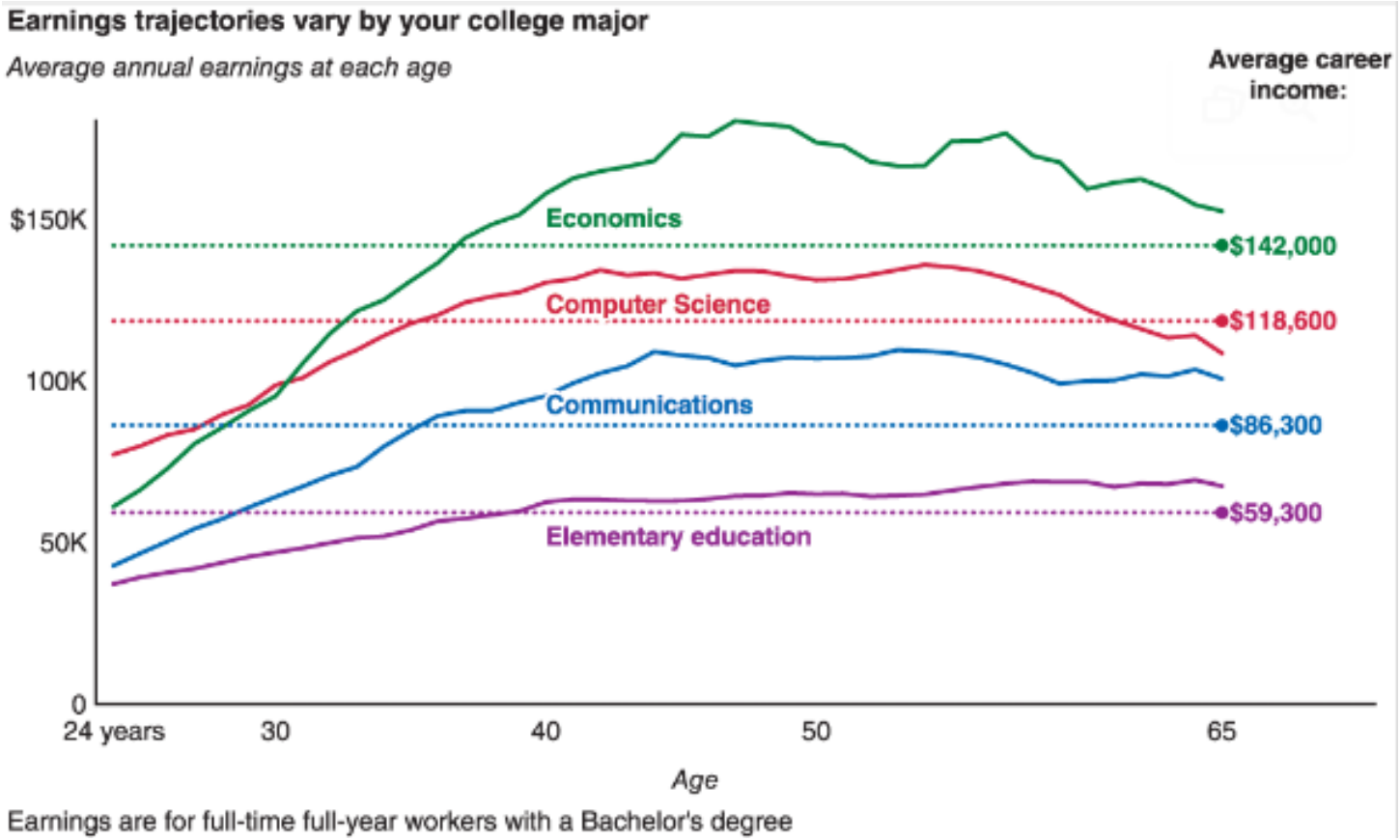
Useful strategy: Make GPT costly (in time) to use



The economics of crime, applied to the decision to cheat

Cheat if $\Delta Time + \alpha \Delta Points > p_{caught}(Punishment + Stigma) + \lambda \Delta Learning$

Useful strategy: Motivate the value of learning



The economics of crime, applied to the decision to cheat

Cheat if $\Delta Time + \alpha \Delta Points > p_{caught}(Punishment + Stigma) + \lambda \Delta Learning$

Useful strategy: Use authentic assessment

Roadmap

- ☑ Three motivating facts
- ☑ What won't work
  - Ignoring the budget constraint
- ☑ Chat GPT as foe: Preventing cheating
  - Lessons from the economics of crime
- ❑ Chat GPT as friend: Enhancing learning
  - A complement that compliments

Using AI as a complement to assessment

Introducing MAT (Macmillan Assessment Tutor)

- Socratic method: Let MAT step your students closer to the answer
  - Will not give away the answer
- Trained on your textbook
  - Uses your class concepts, definitions, and language
- Does not hallucinate
- Enhances accessibility
  - Describe the graph
  - Multilingual: “Ayuda estoy atascada”

Conclusions

- ☑ Three motivating facts
  1. Your students are using Chat GPT (and its cousins)
  2. Chat GPT is better than you think
  3. Your university is likely giving you terrible advice
- ☑ What won't work
  - Ignoring the budget constraint
- ☑ Chat GPT as foe: Preventing cheating
  - Lessons from the economics of crime
- ☑ Chat GPT as friend: Increase student learning from homework
  - A complement that compliments