A LECTURE

What Are Universities For When AI Can Think?

Higher education, cognitive offloading, and the return of judgment

Herman Cappelen · University of Hong Kong

OVERARCHING THESIS

The thesis I will defend.

The question is not whether AI can think, or write, or argue better than you.

The question is:

“What do I, on reflection, take to be true, good, or mine?”

That step — the taking-it-as-mine — is not a piece of cognition the AI has failed to do well.

It is structurally not something anyone else can do for you.

OPENING PUZZLE

We are asking the wrong first question.

If we begin with plagiarism, classroom rules, or assessment design,
we miss the deeper institutional question underneath.

Why do we have universities at all —
and what are they for, now that machines can perform many cognitive tasks?

OPENING PUZZLE

Universities have always had plural aims.

Professional training

Doctors, lawyers, engineers, clergy.

Formation of persons

Character, articulacy, taste, reflection.

Elite reproduction

A network for the already-advantaged.

Research

Discovery and validation of new knowledge.

Knowledge transmission

Preservation of texts, methods, traditions.

Social mobility

A ladder out of inherited circumstance.

Citizenship

Deliberation, judgment, public reason.

National development

Workforce and capacity-building.

*Three of these face direct AI pressure — **work, self-development, and citizenship.** Part 1 focuses there.*

OPENING PUZZLE

The development-and-workforce model

China after 1999

~50%

jump in university admissions in 1999 alone
followed by 15 years of rapid growth

After 1999, China expanded higher education on a scale unprecedented in modern history.

One account calls it the largest influx of university-educated workers into a labor market — ever.

It is the clearest case of treating mass higher education as engine of cognitive labor for a modernizing economy.

The AI disruption

If a major reason for mass higher education was that modern economies required large numbers of people with advanced cognitive skills…

…then AI raises the possibility that many such skills can be partly offloaded to machines.

That does not show universities are obsolete. But it forces them to explain what they are now for.

The argument, in three moves.

I **Why universities?**

Three classical justifications — and how AI puts pressure on each.

II **Kant and AI.**

The romantic answer: just teach them to think for themselves. Why it fails.

III **The return of judgment.**

What still cannot be settled by any authority — and what universities can form instead.

PART ONE

Why universities?

Three classical justifications under pressure.

The familiar answers — and their AI pressure.

Aim	University promise	AI pressure
Work	Train people for cognitively demanding jobs.	AI may perform or assist many such tasks.
Self-development	Make people more reflective, articulate, capable.	Hard to measure — students may not pay if job payoff weakens.
Citizenship	Form people able to deliberate and judge.	AI and social media may degrade public reason as much as support it.

The central dilemma.

For five decades the dominant frame has been ***human capital***: education as investment in skills, productivity, employability, and growth. The university became central because cognition became economically valuable.

IF universities sell themselves as job training,

AI may undercut the need for some of what they teach.

IF they sell themselves as self-development or citizenship,

they owe students and society an honest account of what that means and how success is measured.

Part 1 focuses on the human-capital answer. Self-development and citizenship return in Parts 2 and 3.

China: the human-capital model, made explicit.

We saw the scale earlier. The interesting thing is what the scale was *for*
The expansion was driven by an explicit national bet: universities exist to produce the cognitive labor a modernizing economy needs.

MBA programs and computer engineering grew first and fastest. Humanities and social sciences were treated as less directly “useful” and squeezed.

Many countries have tried the same bet — South Korea, Singapore, Ireland, the Gulf states, parts of the United States and the UK. China is just the most explicit recent case.

In each case, the human-capital answer to “what are universities for?” is treated as the obvious one — and the obviousness is itself the thing worth examining.

The new question

If AI can do more of the cognitive work for which universities trained students,
what cognitive skills remain educationally fundamental?

So what counts as foundational?

If AI absorbs much of the cognitive work, universities have to name what is left — the capacities that remain educationally fundamental. The usual list:

Critical sense	Asking questions	Logical reasoning
Understanding (not output)	Curiosity	Creativity
Moral & political judgment	Evaluating testimony	Knowing when not to delegate

Universities answer the AI challenge by naming these as foundational. ***That answer needs scrutiny — because what counts as foundational has shifted before.***

PART 1B

What counts as foundational?

Every cognitive technology — writing, print, the internet — has reorganized what counts as foundational. AI is the latest such technology, not the first.

Avoid the easy story.

“People always panic about new technology, and they are always wrong.”

That is too quick.

The better point is harder:

a skill can move from being an internal human requirement to being an externalized support — and education then reorganizes around the new ecology.

PLATO ON WRITING

“This invention will produce forgetfulness in the souls of those who learn it.”

— Socrates, in Plato’s Phaedrus (c. 370 BCE)

Memory looked foundational from his standpoint.
Writing looked like a technological shortcut that would hollow out the mind.

Plato's argument: memory is foundational.

1 **Writing is external.**

It places the marks of thought outside the soul, on a surface anyone can carry, copy, or lose.

2 **External marks displace internal practice.**

Once we trust them, we no longer exercise memory "from within ourselves." The faculty atrophies.

3 **Writing creates the appearance of wisdom without its substance.**

Readers receive much information but little teaching — they will seem to know much while knowing nothing.

Why this is not a silly argument.

In an oral culture, memory was not a private hobby.

It was the medium of law, religion, philosophy, and politics.

A trained mind held entire poems, proofs, genealogies, and arguments — and could revise them in real dialectic.

To outsource that to ink and papyrus really did dismantle a form of cognition we no longer have access to.

What Plato sees that we forget.

Cognition has architecture.

New media don’t just add tools; they redistribute which faculties get exercised and which atrophy.

Externalization has costs.

Something is genuinely lost when an internal capacity migrates to a surface — and that loss may be invisible to those who never had the capacity.

“Seems wise” ≠ “is wise.”

A culture of fluent retrieval can mistake access for understanding. The diagnosis lands on us as much as on Athens.

The lesson: what is foundational gets reorganized.

A skill can move from being an internal human requirement

to being an externalized support.

Education then reorganizes around the new ecology — and what once looked foundational becomes infrastructure.

Hold this thought. The same question now faces us about AI — and we are no better placed than Plato to see what we are losing.

Print: from scarcity to abundance.

Manuscript culture

Memorization of scarce texts.

Oral disputation.

Close apprenticeship to a master.

Knowledge as something held in the body of a community.

Print culture

Navigating abundance.

Comparing sources.

Producing new syntheses.

Knowledge as something accessible — and contestable — at scale.

Internet: from recall to navigation.

The complaint was familiar:

“Students no longer know things — they just look them up.”

But the serious educational question shifted:

Can they tell what is worth looking up, what to trust, and how to integrate information into understanding?

AI does not merely retrieve information.

It generates plausible reasoning, prose, code, images, plans, explanations.

> *Do "critical thinking," "creativity," and "understanding" remain foundational in the old way — or will they too be partially externalized?*

A thought experiment: two programmers.

PROGRAMMER A

Understands code.

Years of training in languages, systems, and architecture.

Can read what she writes, justify it, and see why it might be wrong.

PROGRAMMER B

Vibe-codes with a tool as reliable as A's output.

A second model audits the first; a third catches the second's mistakes.

Faster, cheaper, and — measured by output — better.

B does not understand the code, cannot think critically about it, and cannot tell when the whole stack is quietly wrong about something new.

Who do you hire? And what does your answer commit you to?

If you pick B, you are saying: understanding is not foundational — reliable output is.

Then most of what a computer-science degree teaches becomes ornamental: nice to have, not necessary to do the job.

If you pick A, you owe an account of what understanding is for — beyond the code that gets shipped.

Maybe: judgment when the tools fail, responsibility when they ship something dangerous, the capacity to ask whether the system should exist at all.

The same question, generalized, is the question of this talk.

PART TWO

Kant and AI

The traditional answer to “what is foundational?” — the capacity to think for oneself.

The traditional answer: think for yourself.

There is a long-standing answer to what is foundational — older than the AI debate, and older than the university itself. The Enlightenment names it:

“Enlightenment is emergence from immaturity — the inability to use one’s own understanding without another’s guidance.”

— Kant, “What Is Enlightenment?” (1784)

On this view, what is foundational is autonomous reasoning — *and AI now appears as a new guardian, ready to relieve the student of it.*

A familiar university response

Students need AI-free spaces, because they must first acquire the underlying capacities that make critical AI use possible.

- Writing without AI
- Reading without summaries
- Solving problems unaided

The objection: maybe the ideal was always an illusion.

Why think "thinking for yourself" is foundational at all?

No one has ever reasoned alone.

Real reasoning runs on instruments, experts, institutions, textbooks, peer review, search engines, and conversation. Modern rationality has always been distributed cognition.

We do not expect citizens to re-run vaccine trials, inspect bridges, audit aircraft software, or test weapons systems themselves — and we are right not to.

So "think for yourself" — if it means "think unaided" — is a romantic 19th-century picture that never described how reasoning actually works. Maybe AI is simply revealing this.

If AI joins the epistemic infrastructure…

Refusing AI may be like refusing:

Writing	Print	Calculators	Search engines	Statistical software

It may preserve a romantic picture of cognition at the expense of actual capability.

The objection, sharpened: thinking for yourself can be harmful.

The strong form of the objection is not just that “think for yourself” is unrealistic.

It is that thinking for yourself is often worse than letting the right authority think for you.

Parents reasoning for themselves about vaccines kill their children.

Patients reasoning for themselves about cancer skip treatment.

Citizens reasoning for themselves about epidemiology fill hospitals.

On this view, an AI that gently routes the agent away from her own untrained judgment is not stealing her autonomy. It is saving her from a worse version of it — and often, saving other people from her.

“Think for yourself” has always been a slogan, not a method.

It does not tell you what to read, whom to trust, or which arguments deserve weight.

Left alone with it, students do not become autonomous.

They become confident — and wrong, in their own private way.

The romantic picture of unaided cognition is not just nostalgic. It is a way of training people to mistake their own first impressions for reasons.

Where we are: the easy answers are gone.

We cannot just say:

— *“Universities should teach you to think for yourself.”*

— *“They should make you critical.”*

— *“They should form autonomous reasoners.”*

These slogans are too simple. We have just seen why:

no one has ever reasoned alone; thinking for yourself can be worse than deferring; and modern rationality is already distributed across people, instruments, and institutions — AI is just the next member of that list.

So the question of Part 3 is not whether to defer, but ***when*** — and to ***what*** — and which questions are even ***the kind*** **that any authority — human or AI — can settle for you.**

PART THREE

The return of judgment

Some questions are not the kind authority can settle for you — and a few have no settled answer at all.

What is foundational: participation in a cognitive infrastructure.

The Enlightenment was right about what matters and wrong about what it requires.

What is foundational is not unaided cognition.

It is participation in a shared cognitive infrastructure — books, archives, teachers, peer review, instruments, conversation, and now AI.

Call the capacity that participation requires ***judgment***: deciding when to defer, when to resist, when an answer is the kind of thing authority can settle, and when it is not.
Judgment is what survives every move from writing to print to internet to AI — because every cognitive technology raises the question of when to defer to it, and none of them answers it.

But why can’t the AI do that for you?

A natural reply:

“If judgment is just deciding when to defer and when to resist, surely an AI can do that too — maybe better than I can.”

It cannot. Or rather: it cannot do the part that matters.

AI can recommend “defer here, resist there.” But a recommendation is not yet a belief, a commitment, or an action. It becomes one only when someone takes it up — or refuses to.

AI can be a participant in this infrastructure. It cannot be the final node — the place where input becomes belief, commitment, action.

The final node has to be a person.

Questions where expertise helps but does not settle.

01 What is a good and virtuous life?

02 What is the best performance of the Goldberg Variations?

03 What is a just distribution of goods?

04 What is the best form of collective decision-making?

05 What is the distinction between the human and the synthetic?

06 Is there a universal morality, or is everything relative?

07 How should we weigh climate against welfare?

08 What should be the punishment for involuntary manslaughter?

09 What is the correct definition of gender?

Two ways a question can resist authority.

A — JUDGMENT-RELATIVE

There is no answer independent of the agent.

What the right answer is depends on the person whose life, taste, or commitments are in question.

Aesthetic, hedonic, vocational, and many practical questions live here.

B — TRUTH WITHOUT ACCESS

There may be an objective answer.

But humans have no reliable procedure for finding it — no priesthood, no algorithm, no model.

Deep moral, political, and religious questions live here.

In neither case can authority — human or machine — relieve the agent of judgment.

Case A — “Is rakfisk delicious?”

Rakfisk is fermented Norwegian trout. Some people love it. Many do not.

No expert, no panel, no model can deliver the right answer to the question whether it is delicious — for you.

The answer is constituted by the agent’s own response, formed under conditions of attention, exposure, and reflection.

And this generalizes far beyond food: what kind of work is meaningful for me, who I should love, what city I can live in, what risks I am willing to run.

Case B — truth we cannot reach.

Suppose there is an objective fact about what justice requires, what we owe future generations, or how to weigh a human life against a thousand animal lives.

Even then: we have no track record of converging on it.

Two thousand years of moral philosophy, every world religion, every political tradition — all serious, none decisive.

A more capable AI does not change this. It is trained on the same disagreement, and it inherits it. Scale does not produce moral consensus; it amplifies the disagreement.

So even in the best case for moral realism, AI cannot deliver the answer — because no one can.

A stronger reading: truth that depends on us.

Case B assumed there is a mind-independent fact, just out of reach. But for many of the hardest questions, even that is too generous.

Many serious philosophical positions — constructivism, expressivism, response-dependence, sensibility theories, Humean and Wittgensteinian views — hold that our responses are partly ***constitutive*** of the truth of moral, aesthetic, and existential claims.
On these views, “what is just?”, “what is beautiful?”, “what counts as a life well lived?” are not subject-independent. There is no fact about them apart from how creatures like us respond.

If that is right, an AI giving you the answer is not merely unreliable. It is structurally disqualified — because your response is part of what would make any answer true.

So what falls to the university?

Not delivering answers — not anymore.

Forming the cognitive responsiveness that belief-formation requires.

Reasoning can be aided — by colleagues, by texts, by AI. But the last step — taking up the result as *my* belief, *my* commitment, *my* action — cannot be aided. It has to be done by the agent.
The university's task is to form agents capable of that final step — across the full range of questions where it is the step that matters.

Where the final step has to be yours.

Take the questions only you can answer for yourself:

What is delicious to me. What is beautiful. What I owe my parents. What counts as a life well lived. What I, on reflection, believe about justice.

An AI can lay out the considerations, simulate objections, model consequences, sharpen the alternatives. That is real help, and it is not nothing.

But the question is not “what is the answer?” It is “what do ***I***, on reflection, take to be true, good, or mine?”
That step — the taking-it-as-mine — is not a piece of cognition the AI has failed to do well. It is structurally not something anyone else can do for you.

When AI hands you a finished answer to a question of this kind, the answer is not yet a belief. It becomes one only when you take it up — or refuse to.

THE ARGUMENT IN ONE PAGE

1. Universities were justified by producing scarce cognitive labor.
2. AI makes much of that labor cheap, raising the question of what universities are for.
3. What counts as a foundational skill is reorganized by every cognitive technology — writing, print, the internet, now AI.
4. The romantic “think for yourself” response collapses: autonomy is not autarky, and unstructured anti-AI norms produce fragmentation, not independence.
5. But many central questions — judgment-relative ones, and morally objective but inaccessible ones — still cannot be settled by any authority, AI or otherwise.
6. The university’s task is to form agents who can carry such questions: discriminate them, sit with them, and act under them.

Why does it matter what I, on reflection, endorse?

One might ask: who cares what I personally take to be true, good, or mine?

It matters — because it is your life.

A life lived on borrowed endorsements — from a teacher, a culture, an algorithm — is, in a precise sense, not yet yours.

The beliefs and commitments you act on shape what you do, who you become, and what you can answer for.

No one — no expert, no model — can take that step for you. If they did, it would not be your life they were living.

And the same holds for a “we.”

The point scales up.

It matters what a group, a community, a church, a corporation, a nation, a government takes — on reflection — to be true, good, or its own.

A collective that runs on outsourced endorsements — inherited, marketed, machine-generated — is not yet acting as a collective. It is being moved.

And here too: the act of endorsement is something no outside party can perform on the group’s behalf.

Self-government — personal and political — begins exactly where delegation ends.

CLOSING THOUGHT

AI can think, write, and argue. Often better than we can.

But there is one question it cannot answer for you:

"What do I, on reflection, take to be true, good, or mine?"

The taking-it-as-mine is not a piece of cognition the AI has failed at.
It is structurally not something anyone — or anything — else can do for you.

That is what the university is now for: to form people who can take that step.

Thank you.