You are what you measure.

...AI made me rethink something and I don't like that

Presented by David Evans

A Vibe History of Information

The Church

Only we may know things. You may not. You may, however, be really impressed by us. That is allowed.

- Sacred texts in languages only we know
- Very big buildings to show how great we are
- Not a priest? Take a hike. Give us money on your way out.

The Birth of the University

Opposition to Dogma

Universities were founded in direct opposition to the Church's dogmatic control of information.. The Church gatekept knowledge through:

- Restricted access to sacred texts
- Monopoly on interpretation
- Sole power of credentialing,

Universities Opened Up

Empirical Inquiry

Establishing truth through observation and logic rather than institutional authority or received wisdom.

Peer Review

Knowledge validation through a community of scholars rather than a single ecclesiastical hierarchy.

Uh oh, here comes the printing press analogy!

…but I respect you all too much…

The Birth of Big Tech

Infinite Grad School

Universities created intense, focused environments based on learning and benchmarking. Learn/Test repeat, working hard and playing hard. This pattern Tech-transferred its way into the private sector.

We kept some traits too:

- Restricted access to the club
- Monopoly as a natural outcome of greatness
- Money Money Money

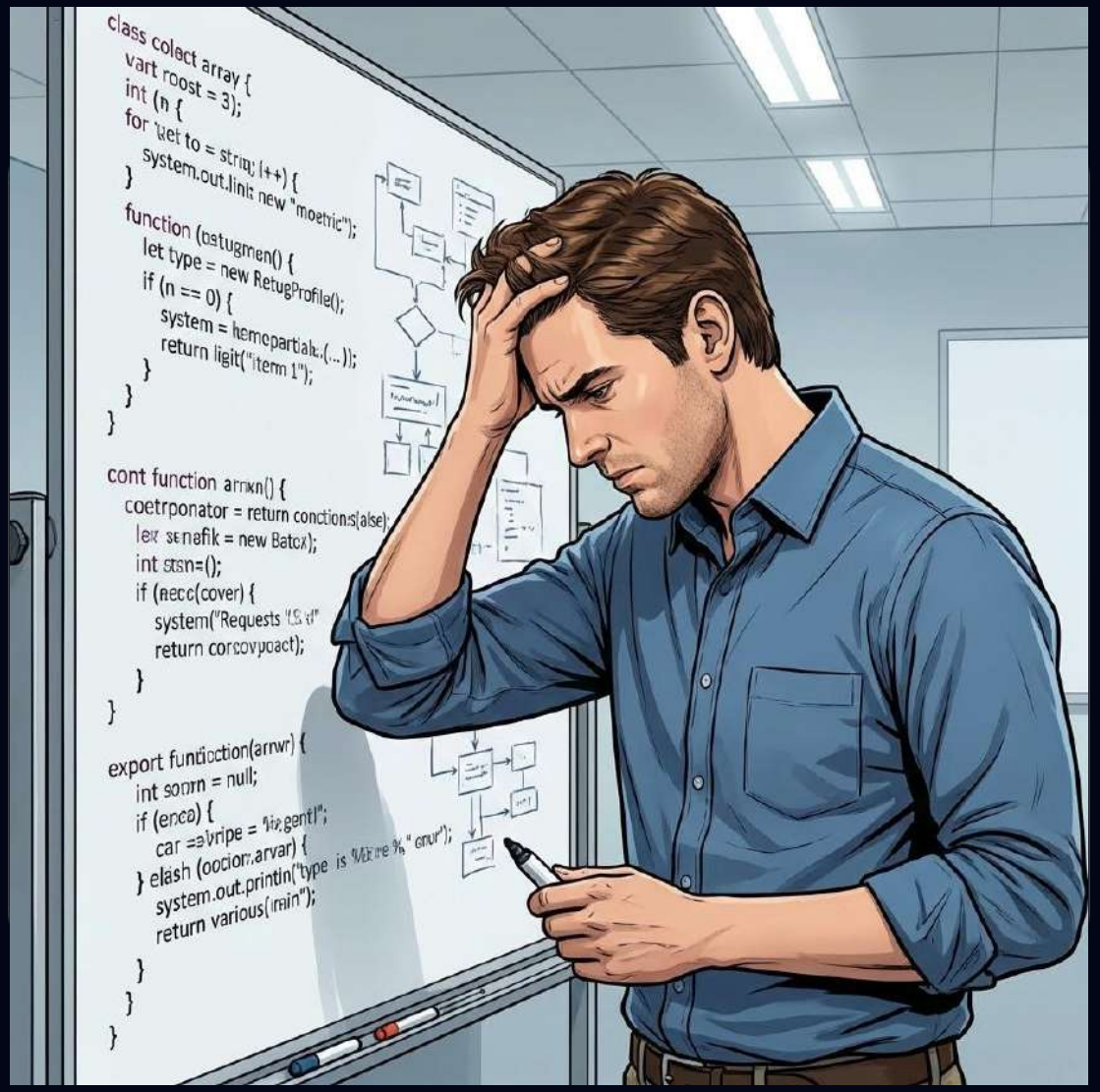
Big Tech Copied Universities

Can you even Code, Bro?

Big tech moved very quickly toward the idea that only the rarified few were good and smart enough to join the ranks.

Then AI came along and really ruined our main tool for this…

but I'll get to that.



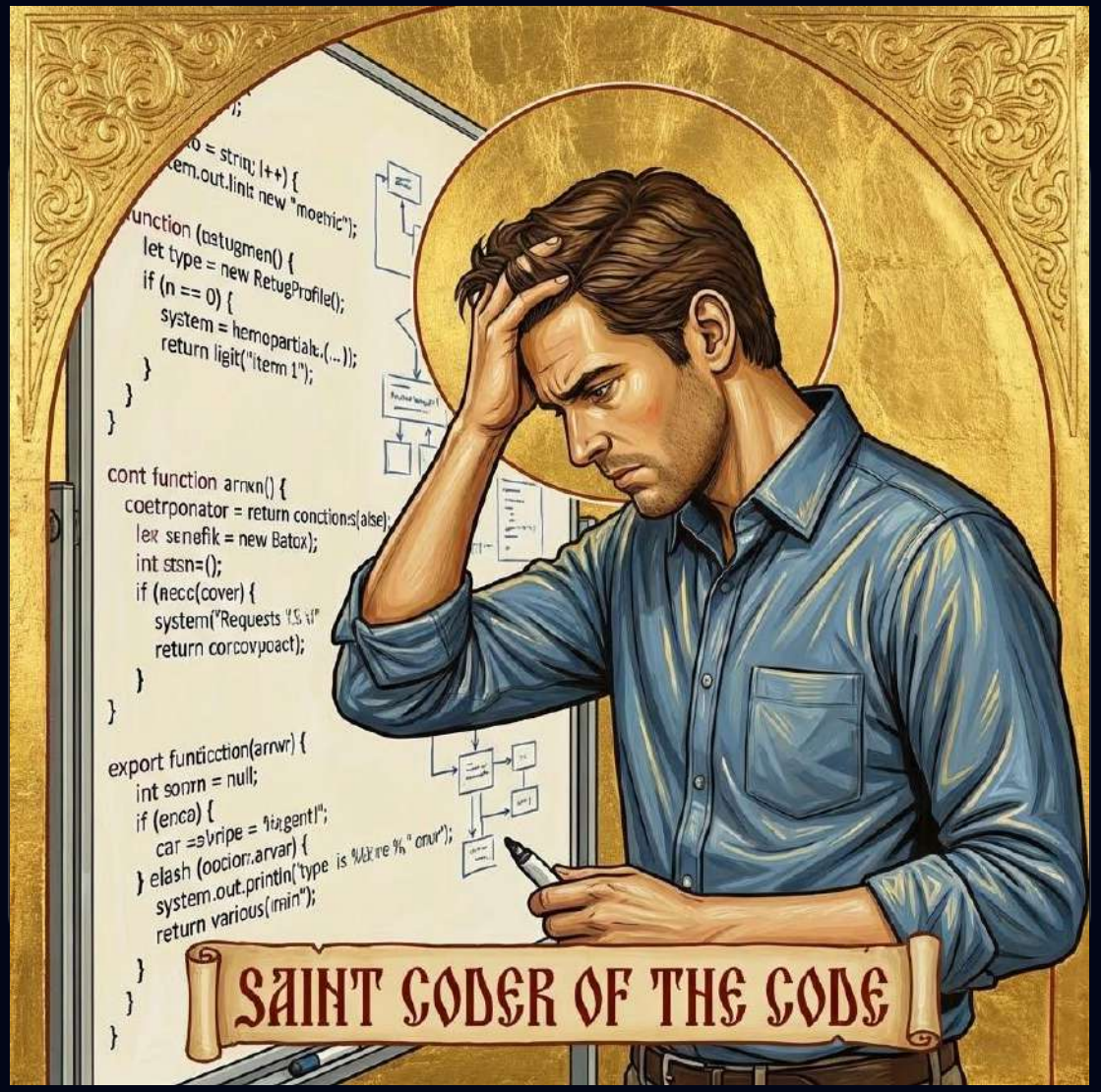
Big Tech Copied Universities

Can you even Code, Bro?

Big tech moved very quickly toward the idea that only the rarified few were good and smart enough to join the ranks.

Then AI came along and really ruined our main tool for this…

but I'll get to that.



WHO AM I

I do Computer for Job

I have spent a quarter-century inside the machine. I've watched tech waves promise democratization and really just get us addicted to things.

I’m part of maybe the last phase of folks in the field who mostly learned theory. I went to Pitt, I’ve done startups, worked at Google on spam and intelligence in texting, and now I’m at Duolingo.

I’ve done ML or AI or whatever we call it now for a long time.

Teaching Language with AI

What it means to use an LLM to teach

I currently do some LLM backed teaching at Duolingo. I have opinions, and practical outcomes, and methods,
I am here neither to promote nor defend Duolingo - that's really not my job. This is not a significant peek behind the curtain.

It's not as simple as it seems, despite many, many, many articles to the contrary.

Back to our Story

Infinite Grad School

Technology became the money-making version of Universities. Freewheeling parties, intense coding sessions, infinite youth, and absolutely 100% non-self-aware. Tech came to rely on the University system to do the dirty work of doing first-order sorting (e.g. where did you go to school)

- Finally the Nerds were cool
- Ask for Forgiveness not Permission
- Sole power of credentialing

THE TECH PROMISES

Web

Information for everyone, everywhere, instantly.

Mobile

Total freedom from location and physical barriers.

Social

Democratizing the voice of every human on earth.

The “Complications”

Every wave delivered a dark side.

Web brought vile forums.

Mobile brought always-on burnout.

Social brought the ragebait economy.

What will AI bring? (spoiler: it's just more Ads, but that's a different talk)

AI TRAINING AND OPTIMIZATION

The Technical Choice

LLMs are trained (unsupervised) on massive corpus(es?) of text data. I won't get too deep in the details here, but they learn proximity, structure, and “facts” as a side effect of this.

AI optimizes for measurable loss functions. These are design choices made by engineers, not philosophers:

- Next-token prediction
- RLHF Human Feedback ratings
- Pattern matching probability

$$L(\theta) = -\sum_{i=1}^{n} \log P(y_i \mid x_i ; \theta)$$

Optimization for what can be measured.

TECH COMPANY METRICS

Tech companies optimize for "Stickiness" and Growth above all qualitative missions.

UNIVERSITY METRICS

Claimed Goal	Measured Metric (The Shadow Goal)
Critical Thinking	Enrollment Numbers
Wisdom	Endowment Growth
Deep Understanding	Graduation Rates/GPA
Societal Impact	US News & World Report Rankings

What you measure is what you are

How you measure matters

Where is this going?

Optimizations Fail

Nothing stays the same

The Coding Gate

LLMs took our syntax and made it free.

The question of “Can you **code** on demand?” has become completely irrelevant.

This is causing an intense freakout at all levels of the industry that we haven’t fully reckoned with.

INSTITUTIONAL REACTION

Start with No

The Repetitive Cycle

Institutions always follow the same four-step reaction to disruptive technology:

Panic → Ban → Control → Adapt

Tech is somewhere in the Control to Adapt phase here.

I suspect Education is somewhere in the Ban/Control boundary

Back to Teaching Language with AI

What it means to use an LLM to teach

LLMs are broadly optimized for the normalized distribution of all text that its trained on. A concrete example of this means that it's hard to handle "more difficulty" or "less difficulty"

There's an assumption that because "LLM can translate" that means "LLM can teach"

LLM can translate: Yes

LLM can teach: lol

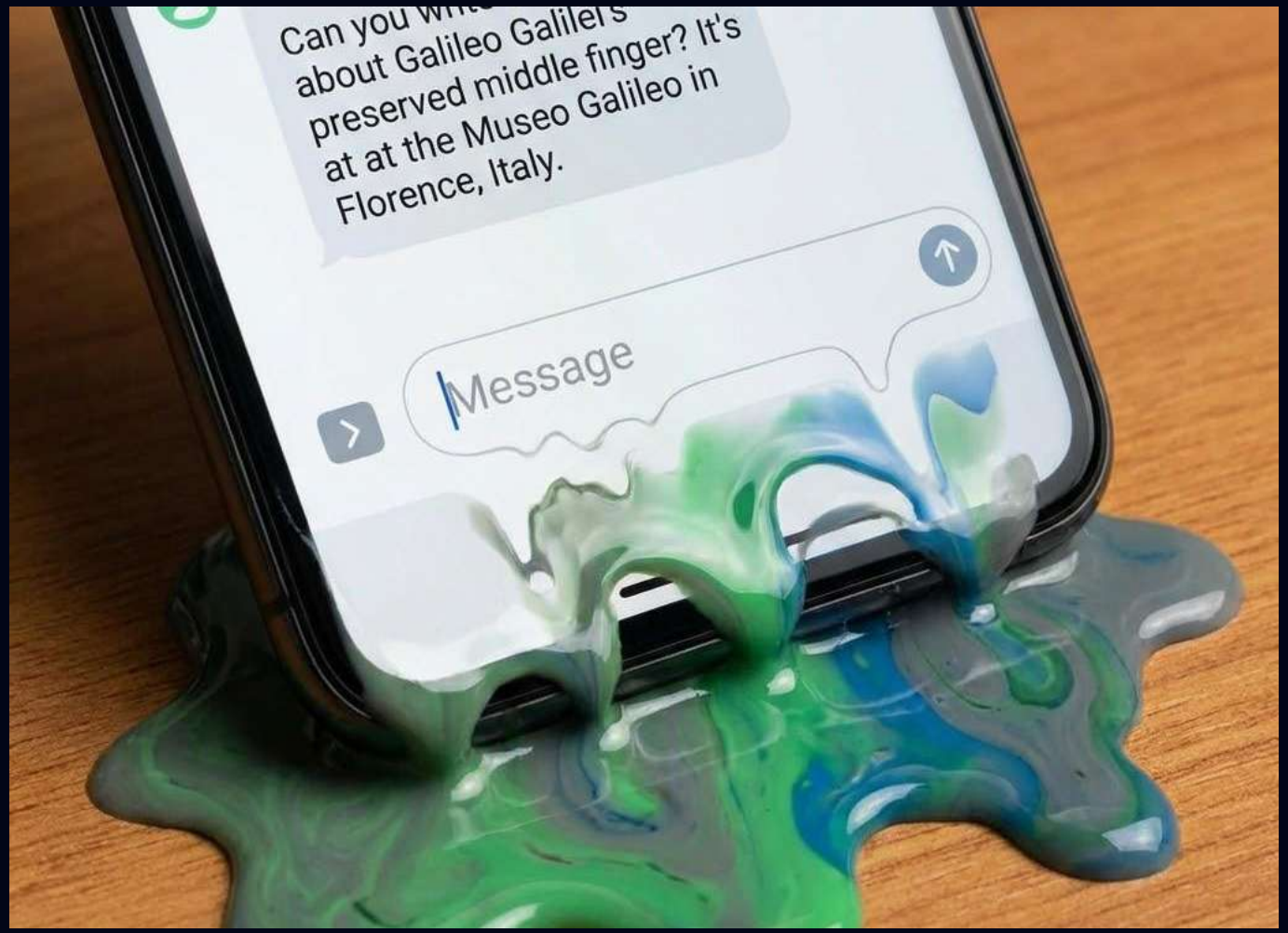
Back to Teaching Language with AI

There is an incredible amount of effort exerted by teams of experts in language learning and LLMs to Prompt, Test, Measure, Fine-Tune, etc.

Left to its own devices, the LLM will just give the user whatever they ask for. It will be the absolute mathematical average of the goal. It will not push, it will not be able to really tell if someone is understanding. It will attempt to drive right down the middle and do exactly what the user is asking.

WHAT LLMS ARE GOOD AT

- Being conversational
- Pattern matching across massive data
- Producing plausible, boilerplate text
- Writing repetitive "syntax" code



WHAT LLMS ARE BAD AT

Teaching

No Socratic method; cannot detect genuine confusion.

Curiosity

No internal stakes or genuine desire to understand.

Taste

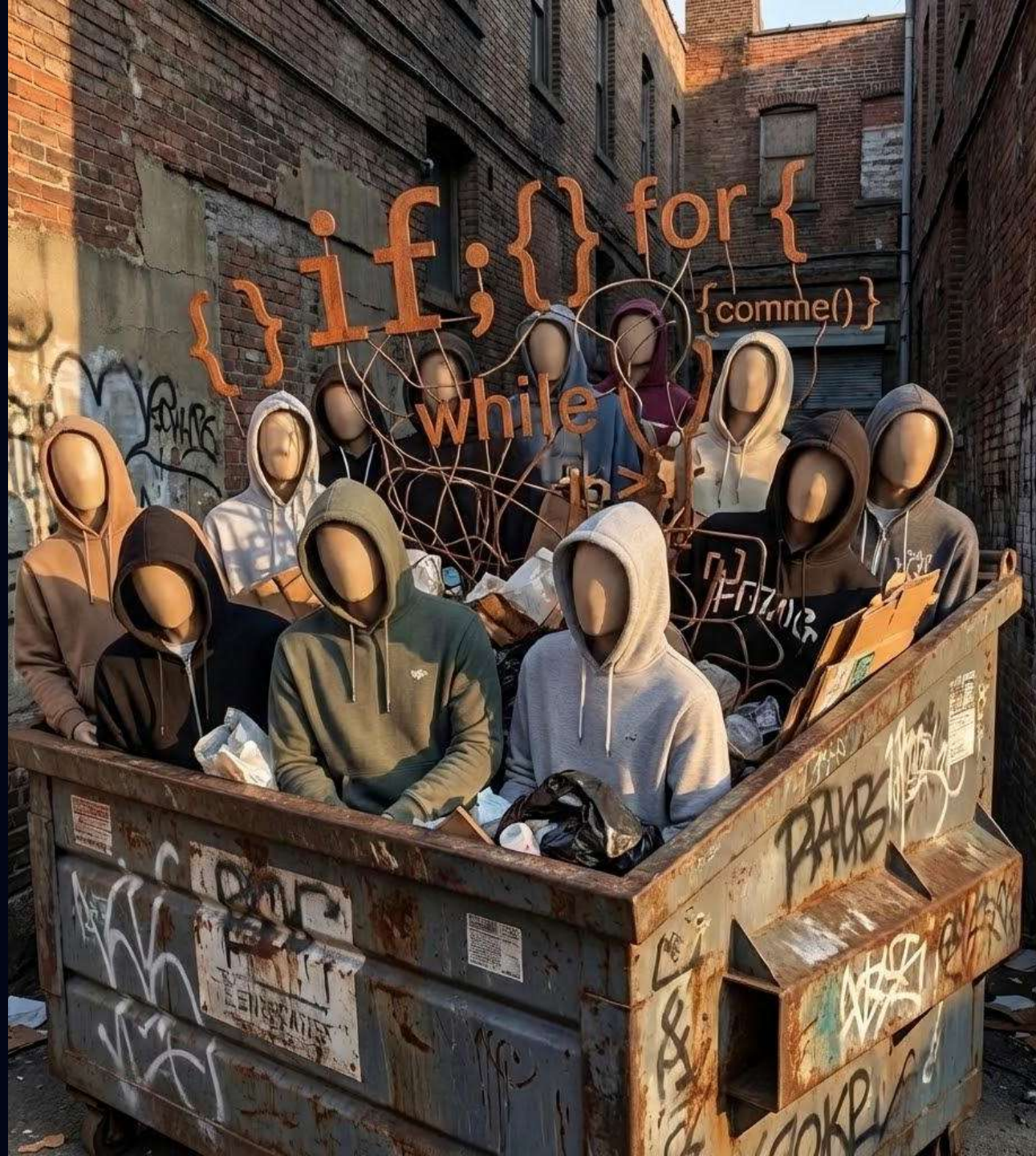
Optimizes for confidence over truth or novelty.

THE REAL RISK

The Artifacts aren't really valuable

Much of the industry is propped up on the myth of the computer genius hacker. The wunderkind who just "knows how to use the computer"

We'd been scaling the production of structured artefacts, and losing sight of problem solving while pointed at user growth and ad revenue.



AN "INTERESTING" PROBLEM

AI Capability

Can an LLM write a grammatically perfect research paper? **Yes.**

Can it write an *interesting* or *novel* one? **No.**

The Missing Ingredient

AI optimizes for **Coherence**, not Insight. It lacks the taste and discernment required for true innovation.


If AI is Extinction-Level
for 'can you code'...
Then who gets in?

"If AI can produce 'perfect' code... what does that say about what we've been optimizing for?"

Tech is staring into this mirror right now. We needed to figure out how the LLM fits into the toolchain.

We cannot ignore it.

Can someone guide the LLM well? This is important now.

We were measuring syntax, not problem-solving.

This is STILL being debated! It strikes at the identity of a lot of industry veterans.

THE NEW ASSESSMENT MODEL

Critique

"Here is AI code. What is wrong with it and why?"

Architecture

"Defend your architectural choices against AI."

Tradeoffs

"Demonstrate your process of working WITH AI."

CS Degrees are STILL IMPORTANT

there's no loss of utility for what is being taught, but the measure is perhaps broken

…but the method of teaching to the syntax won't do anyone any good on its own.


If AI is Extinction-Level
for how you teach...
What are you teaching?

"If AI can produce 'good enough' writing... what does that say about what we've been optimizing for?"

The mirror academia needs to look into

OUTPUTS VS THINKING

Assessing Outputs

Structure, correctness, demonstrated complexity.

AI Territory

Assessing Thinking

Judgment, taste, critical analysis, wisdom.

Human Territory

THE SHARED FEAR

Church

Loss of monopoly on scripture and truth.

University

Loss of monopoly on knowledge and credentials.

Tech

Loss of monopoly on syntax and complexity.

Surrender?

The Cost of being Correct is 0

When the monopoly on "Access" and "Credentials" is gone, what is left?

We’re so unconcerned with correctness that some may say we’ve entered a “post facts” era but again, this is another talk, or a few drinks…

You can just detect that someone used AI

That's an arms race for money.

You want to be in an arms race?

AI Companies Exploit human Creativity for Slop

Sure, but do you need AI for Slop?

AI is automating the artifact here, too.

They Hallucinate

I guarantee you that this is a temporary problem.

Tech is strongly incentivized to make them factual.

ASSESSING TASTE – some options

- Use AI to draft → annotate where it is shallow or wrong.
- Show your prompting strategy and iterations.
- Defend why you kept or rejected specific AI suggestions.
- Explain what the AI missed contextually or culturally.

WHAT WILL FAIL US

We will get some of this wrong.

There will be unintended consequences.

AI is happening, like the smartphone, or the stinkbug. Banning it is itself an art.

AI has no taste, no critique, no intuition. This is where the value is.


Make the Interaction
the Assessable Moment.
Stop assessing outputs. Start assessing
judgment.

Try to redesign ONE Assignment

Make it fun for **you**.

Engage your own curiosity.

Thank you for your time

and thank you to Jane and Rita for inviting me

Feel free to argue with me at either of these fine places:

Email: spaceLenny@gmail.com

LinkedIn: https://www.linkedin.com/in/davidevans16/

Some Links!

here are a few things I think are worth watching - some of which are relevant!

3Blue1Brown: Playlist on Neural Networks

Adam Curtis: Can't Get You Out of My Head

Adam Curtis: Hypernormalisation

Waldemar Januszczak: The Dark Ages