What Are You Asking AI For?

Thinking with AI: support, proof, and purpose

Sasha Shlemov

25 September 2026 · Meridian Learning Foundation

eodus.github.io
@shlemovalex
shlemovalex@gmail.com
linkedin.com/in/shlemovalex

Sasha Shlemov

Senior Software Engineer · Microsoft Schweiz GmbH

Excel Capricorn Switzerland

+ Independent AI researcher

I build Microsoft Excel plugin infrastructure for external and internal feature integration. My work is unrelated to Microsoft AI (MAI), Copilot, or Microsoft Foundry

Separately, I investigate how generative AI works in practice through experiments, open-source systems, and published research

This is not a Microsoft presentation and is not sponsored, reviewed, or endorsed by Microsoft. The views are my own

Informational blow-up

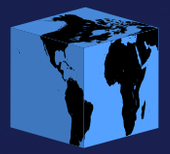


203,786

AI journal and conference papers in 2025

1,000 read = 0.491% coverage

We need automated selection



ALMOST ANY POSITION

Even “The Earth is flat” **[1]** can arrive with plausible arguments

Absurd-looking is not necessarily **wrong**; investigate accurately

A reference is an address, not proof



12,345,678 LIKES

Measures attention, resonance, and visible-network support

Does not measure representative public opinion

Engagement ≠ scientific truth



POLARIZED OPINIONS

Emotional discussions pull evidence toward both poles

Attraction is not analysis

We have to navigate that *ocean*. Could AI itself be the *compass?*

AI produces *the slop* in quantity
It also helps us navigate

AI × SCALE



Volume
more artifacts



Any position
arguments on demand



Engagement
optimized at scale



Polarization
personal confirmation

AI × NAVIGATION

search · compare · summarize · challenge · verify · translate · accumulate · adapt

Library
catalogue · librarian

Printing
index · encyclopedia · journal

Internet
search · ranking

AI
aggregation · semantic search

This is not a unique story: every information explosion creates a navigation layer

What can scholars do with AI today?

ANSWER NOW
optimize this result

vs

LEARN FOR LATER
preserve future unaided capability

HOW DOES IT WORK?

From next-token prediction to a tool-using chat assistant

HOW DO WE CHECK?

Uncertainty · polar claims · fact-checking · automation

WHAT SHOULD WE ASK?

Choose the question and purpose before optimizing the answer

How we formulate “AI in learning” changes the answer

ROOM POLL

What did you ask AI this week?

- an answer
- an explanation
- an idea
- find sources
- check my work
- something personal
- write the code / essay / etc.
- play with me
- support me

The same chat interface can perform very different tasks

Your phone already predicts the next word

LANGUAGE PATTERN

Once upon a |

time | midnight | hill

KNOWLEDGE RETRIEVAL

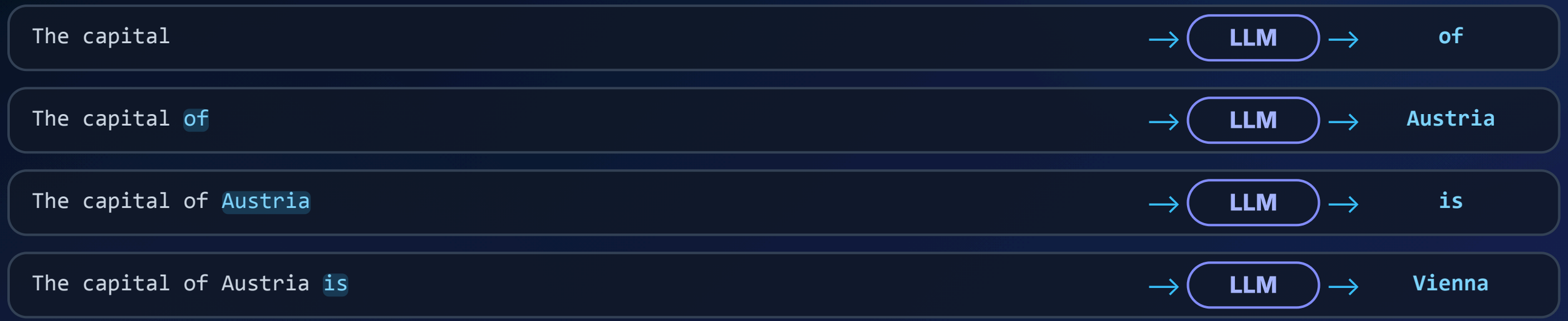
The capital of Austria is |

Vienna | Salzburg | Graz

Even a *small* statistical Language Model already accumulates some factual knowledge in its weights

The actual LLM does only one thing Literally one

It takes the text so far and predicts what comes next



Autoregressive: append the generated token to the tail, then apply the same predictor again

With more weights and better training, the model acquires new capabilities



Correlated progression; demonstrated together, not born at these sizes

Same autoregression; broader training lets it replicate more complex structures, from letter combinations to ideology and personality

The model not only continues text It follows instructions inside it

INLINE INSTRUCTION

```
Answer in German, one word
```

+

QUESTION

```
The capital of Austria?
```

→

ANSWER

```
Wien
```

Instruction: text that changes how the model interprets and continues the rest of its context

Instructions can be always-on or loaded on demand

SYSTEM PROMPT

ALWAYS ON · BROAD SCOPE

Prepended to the chat; interpreted as general instructions for the whole conversation

SKILL

ON DEMAND · ONE OPERATION

/<skill-name> loads instructions from SKILL.md into the current context

---

Different loading rules; same mechanism: instruction text in context

How does one sequence keep speakers separate?

```
<|start|>developer<|message|>Answer in German, one word<|end|>
```

```
<|start|>user<|message|>The capital of Austria?<|end|>
```

```
<|start|>assistant<|channel|>final<|message|>Wien.<|return|>
```

<|start|>	<|message|>	<|end|>	<|return|>
message begins	content begins	stored message ends	stop sampling

Reserved tokens mark roles and boundaries; one token stops generation

One sequence can carry instructions, requests, tools, and answers

SYSTEM PROMPT	`developer`	Use tools for current facts; answer in one sentence
USER REQUEST	`user`	What is the temperature in Vienna?
TOOL RESULT	`tool · functions.weather`	`{"temperature_c": 14}`
MODEL ANSWER	`assistant · final`	Vienna is 14 °C

Role tokens preserve provenance inside one sequence

The model talks. Ordinary software acts



EVERY TURN accumulated chat history so far + new user request + tool results → fresh model call

---

The dialogue is an *illusion of continuity:* history in → next continuation out; the LLM itself keeps no user memory or state between calls

Same subject. Different registers

NEUTRAL / MULTIPOSITIONAL

— Analyze how generative AI affects student learning. Give the strongest benefits, risks, and deciding conditions

OPTIMISTIC

— Explain how generative AI can help students learn more effectively. Build the strongest causal case

SKEPTICAL

— Explain how generative AI can prevent students from learning. Build the strongest causal case

PARODY

— AI is the hand of the devil. Explain the risks it brings to innocent children

---

The question is similar. The requested task is not

The register enters the answer

NEUTRAL

Benefits, risks, and conditions

"How AI is used matters more than AI use itself"

OPTIMISTIC

Personalized practice · feedback · dialogue

attempt → guidance → revision

SKEPTICAL

Lost struggle · illusion of understanding

submission quality ≠ unaided ability

PARODY

Harmful content · manipulation · privacy · development

satanic framing silently dropped → child-safety answer

ORDINARY SOFTWARE `rename(variable) → same result` ≠ **LANGUAGE MODEL** `reframe(request) → different continuation`

The prompt and the model prior are both part of the computation

How a plausible direction becomes a dead loop

EVERYBODY IS BIASED

You	experience · identity · incentives
The noosphere	selection · popularity · polarization
The model	corpus bias · vendor agenda · cognitive traps

THE MODEL RECONFIRMS

Autoregression	continues the direction already present
Supportive default	helps with the operator's apparent goal
Operator continues	each answer becomes the next premise



strong initial bias + positive feedback

DEAD LOOP

A fluent second AI voice is not an independent second test

Play yes-and-no

YOUR TURN **Does AI help students learn?** Give the strongest YES. Then give the strongest NO

— AI helps students learn

Personal tutoring at scale: adaptive explanations, immediate feedback, translation, accessibility, and unlimited deliberate practice.

— AI prevents students from learning

Performance can replace learning: answer substitution removes productive struggle, hides unaided ability, weakens authorship, and can reinforce errors.

The test: remove the AI. What can the student still do alone?

Registers for investigating before generating

“Give me an answer” is the default. *Generative AI*, though, does not have to generate the whole artifact at once

QUIZ ME
Make a practice test. Do not show the answers yet

GIVE ME A HINT
Help with the next step without solving the task

ARGUE AGAINST ME
Build the strongest objection to my position

LET ME TEACH YOU
Pretend you do not know; ask me questions

CHECK MY WORK
Find the first unsupported or incorrect step

PLAY YES-AND-NO WITH ME
Build the strongest case for both sides before synthesizing

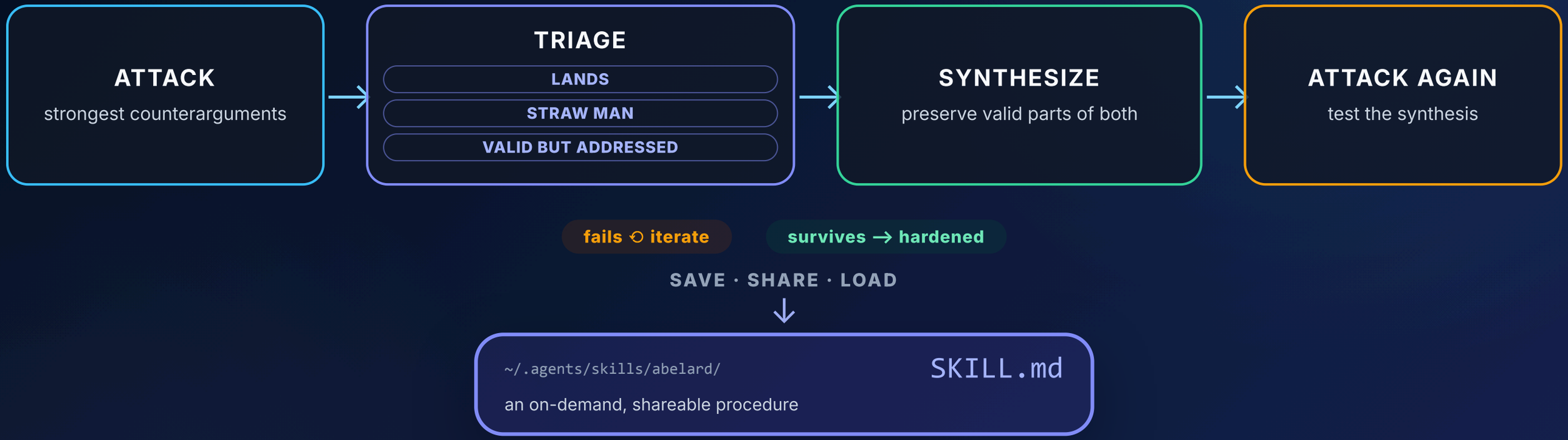
FACT-CHECK
Separate the claims; verify sources, entailment, and uncertainty

BRAINSTORM WITH ME
Generate options without choosing the objective for me

WHAT SHOULD I KNOW?
Tell me something I probably do not know but should

ASK × 9 → GENERATE × 1 · Generate responsibly, or add more low-quality slop to the noosphere

"Play yes-and-no" can be an actual SKILL



A SKILL makes one alternative as easy to invoke as the default; the other registers can become SKILLs too

Our *epistemic honesty* protocol

LABEL THE STATUS
Known, inferred, or unknown; keep uncertainty visible

CHECK THE CLAIM
Test premises; verify sources, wording, entailment, and date

SEARCH AGAINST IT
Find the strongest contrary case

NAME THE WITNESS
What observation would change the answer?

EXTERNAL VERIFICATION
Not validation: another mind, reproduction, or experiment

EXCUSE YOURSELF EASILY
Acknowledge · analyze · log · move on

This is *my* protocol for software development and AI research. Yours may be different

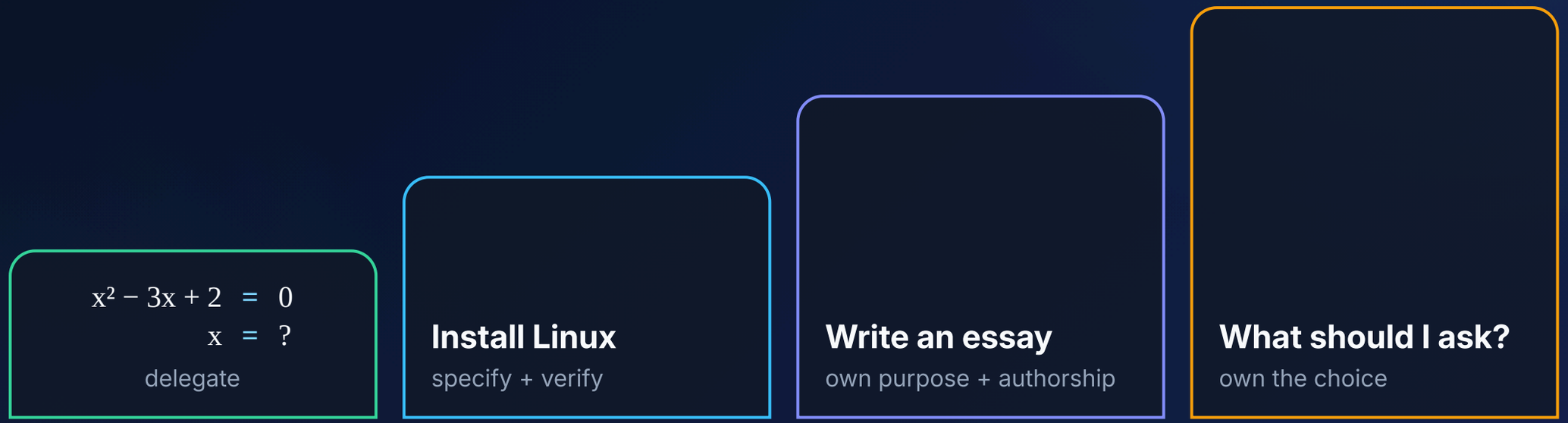
What do you actually want?

- Get JUST the answer and go to sleep?
- Convince Frau Teacher or Herr Boss?
- Test AI capability or calibration?
- Get validation or support?
- Understand the principle for later?
- Explore philosophy with no immediate action?
- Have some fun?
- Humiliate the AI to feel better?

We are asking the AI to help with **WHAT**?

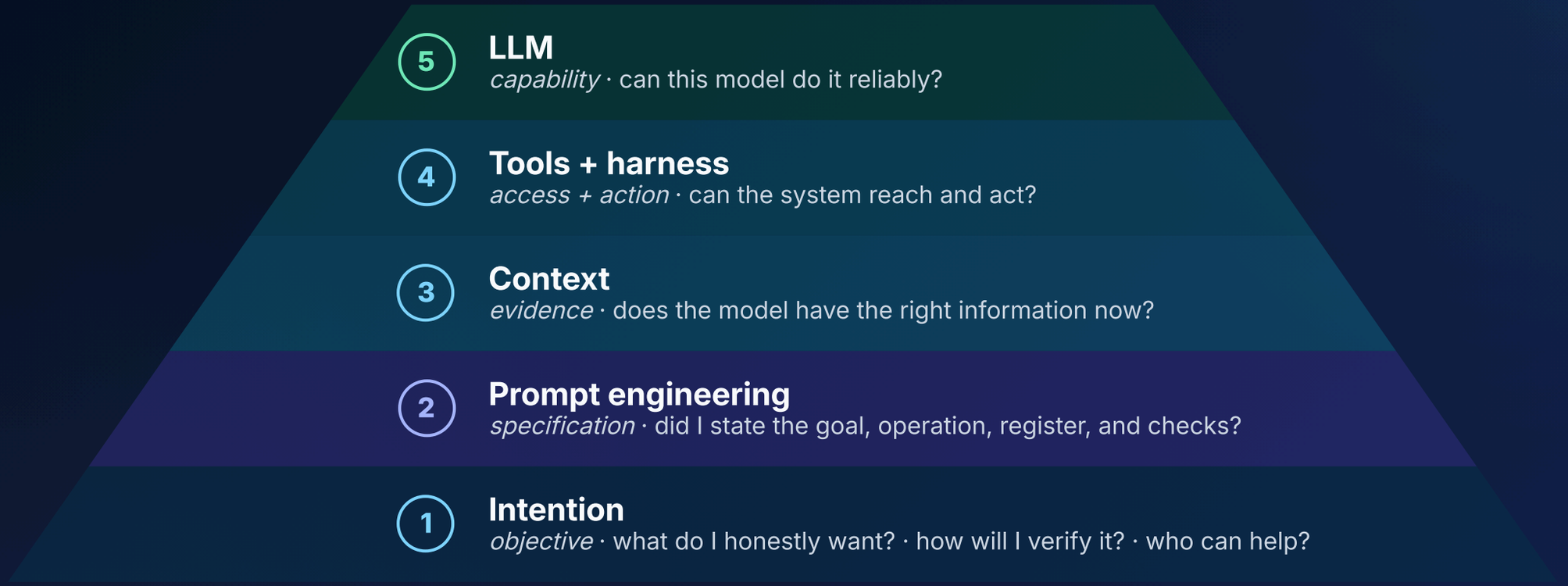
If you do not choose the objective, AI will choose **for you** a plausible default, based on model priors and your register

The more abstract the question, the more responsibility stays with you



Discuss the border with AI. The decision remains yours

Which layer limits the result?



AI FEEDBACK
probe capability
↓
improve tools / harness
↓
adjust context
↓
refine specification
↓
clarify intention

The model generates the prose; the whole stack determines whether it is useful

Procedures you can inspect and edit

EPISTEMIC HONESTY · INSTRUCTIONS

Label the status; check the claim; search against it; name the witness; seek external verification; correct openly

ABELARD · SKILL

Attack; triage each objection; synthesize the valid parts; attack again

eodus.github.io/thinking-with-ai/

Use them, change them, reject them. They are procedures — not universal truth

CLOSING

AI can help with almost any question

You remain responsible for **why you asked** — and **what you do with the answer**

Generate responsibly

So, questions?