

Welcome to AI Code

Why this course exists, and what you'll be able to do

Prof. Aviv Zohar

Semester A, 2026-7

AI Code @ HUJI

Why this course, and why now

AI is changing how we work

AI is changing many areas of computer science, and beyond.

Why this course, and why now

AI is changing how we work

AI is changing many areas of computer science, and beyond.

Coding: one of the most affected areas

Coding is one area where the changes are easy to see.

Why this course, and why now

AI is changing how we work

AI is changing many areas of computer science, and beyond.

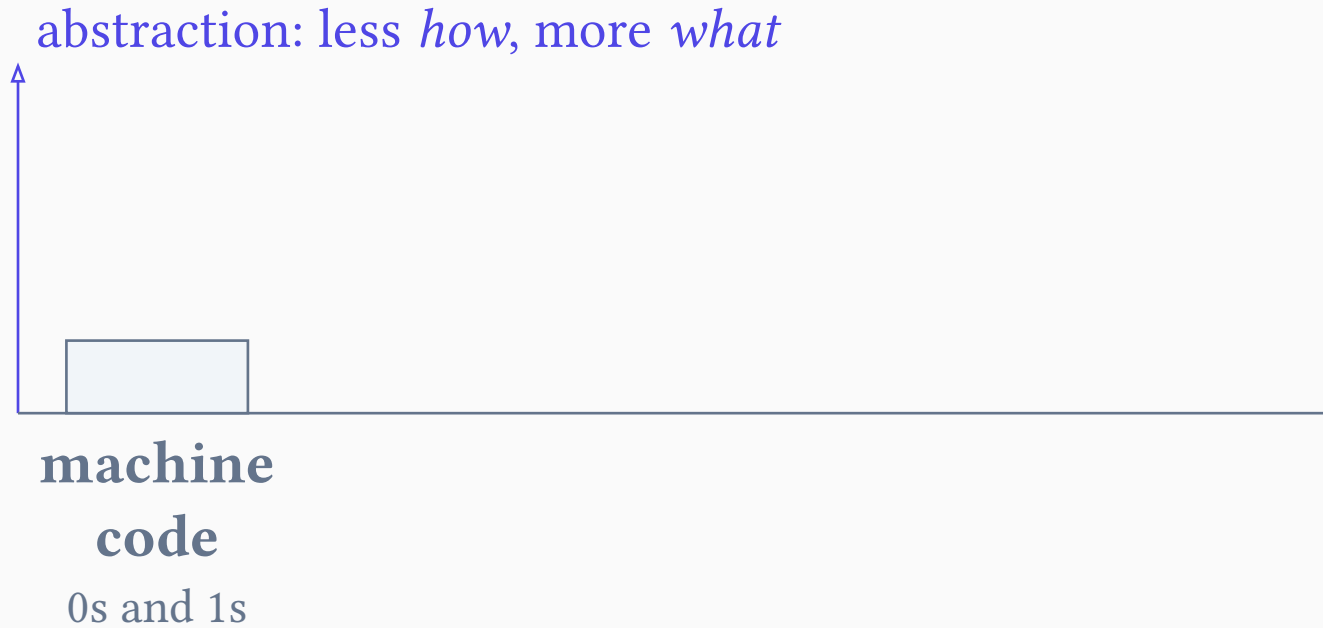
Coding: one of the most affected areas

Coding is one area where the changes are easy to see.

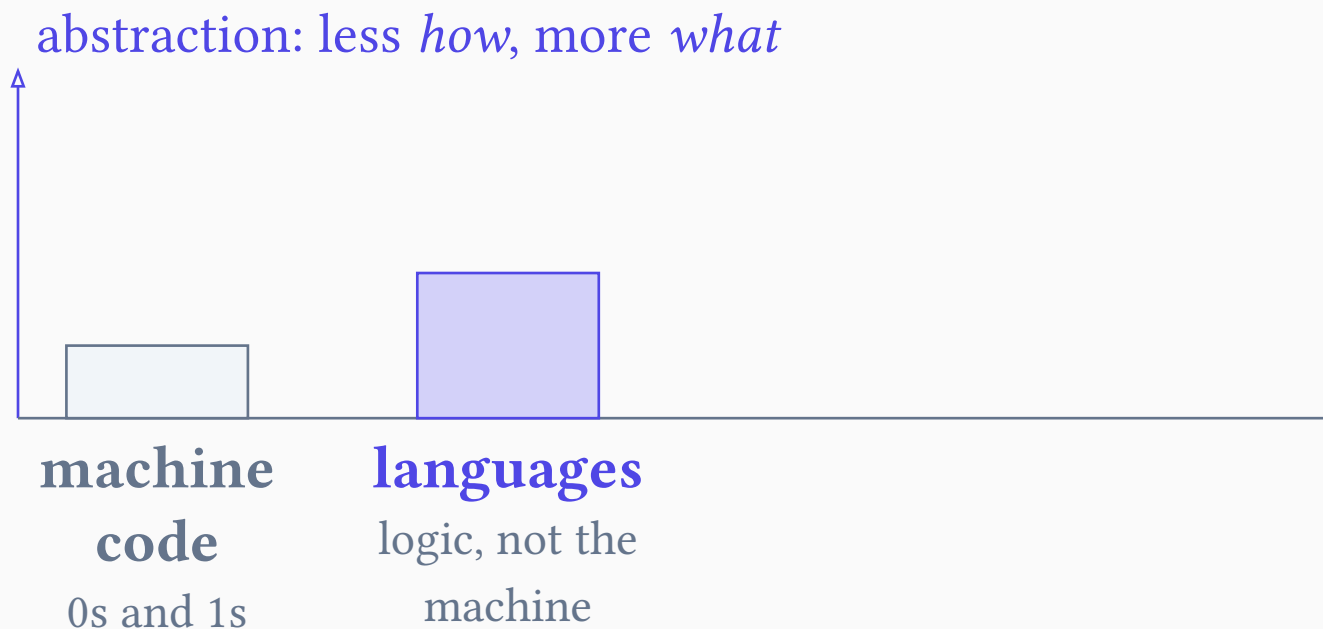
Learning this while the tools are still changing

The tools are still evolving, and best practices are not yet settled.

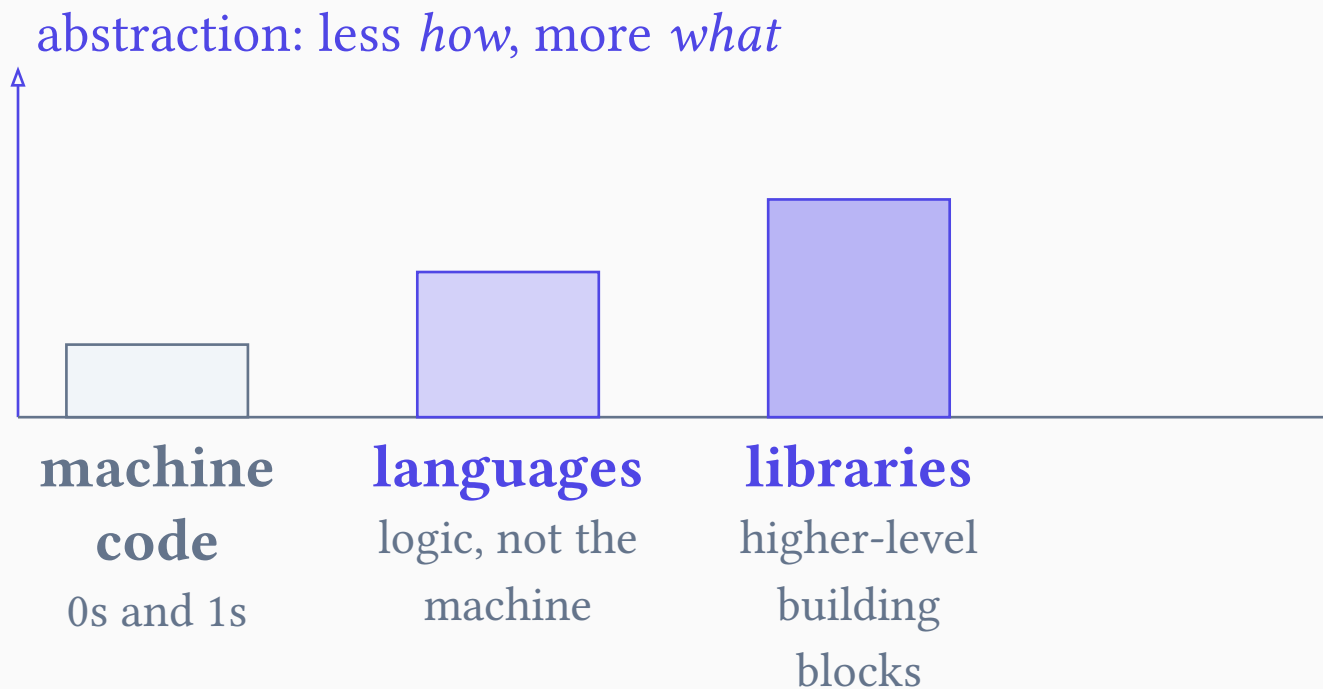
The ladder of abstraction: agents are the next rung



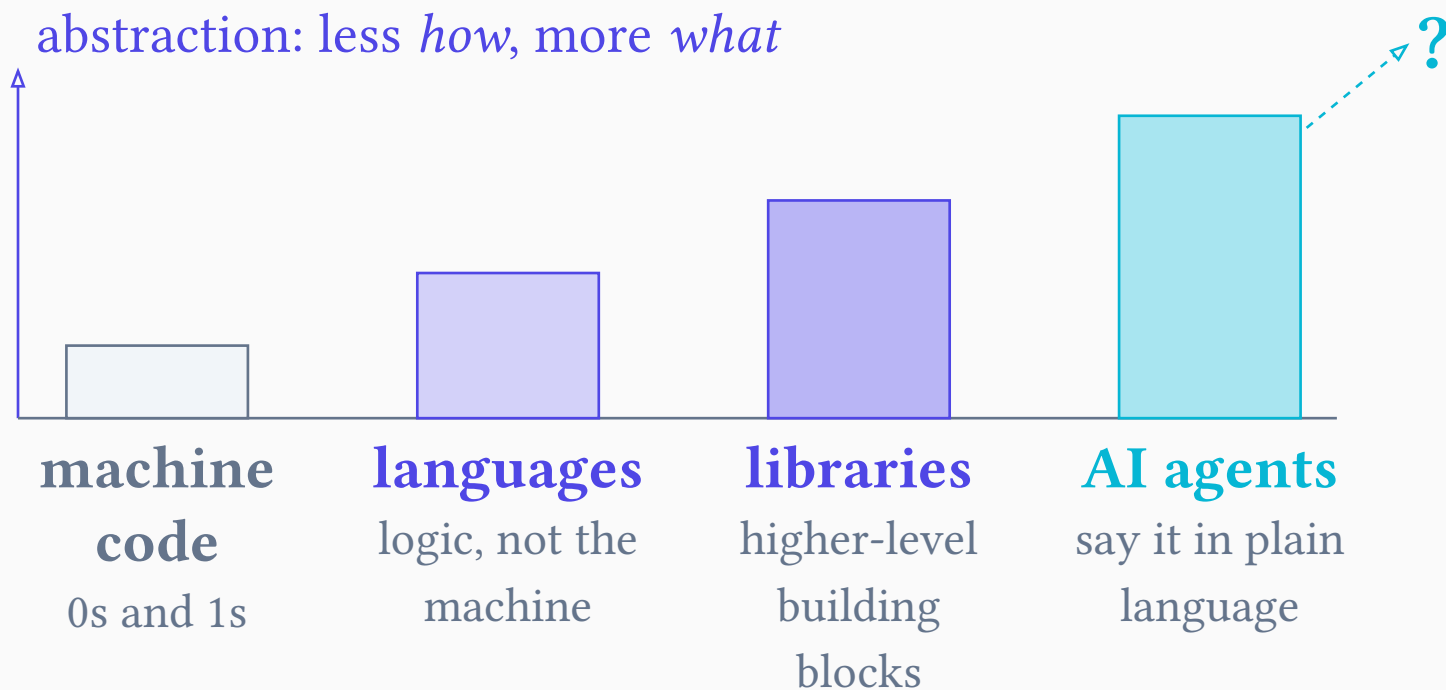
The ladder of abstraction: agents are the next rung



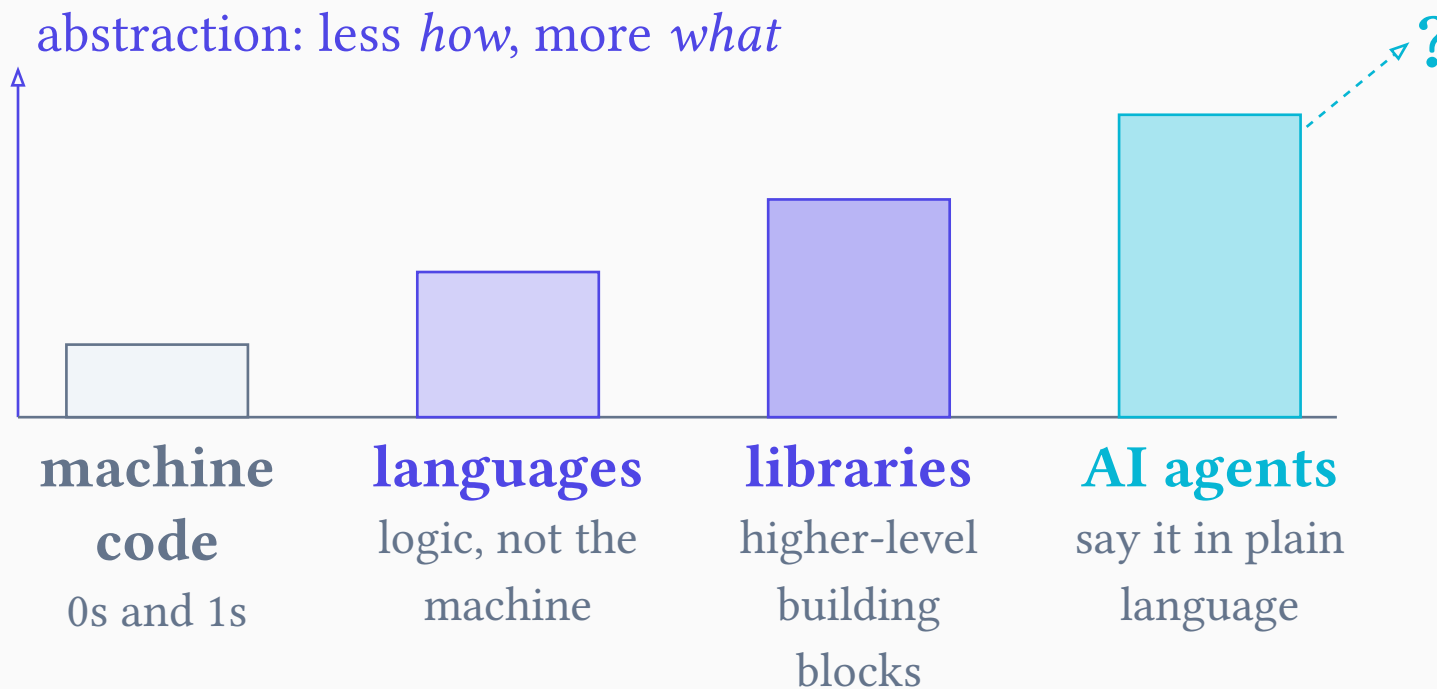
The ladder of abstraction: agents are the next rung



The ladder of abstraction: agents are the next rung



The ladder of abstraction: agents are the next rung



At every level the skill is the same: say what you want, then check you got it.

Three kinds of AI coding help

Autocomplete

finishes your line

(Copilot, tab)

Three kinds of AI coding help

most of you have used this

Autocomplete

finishes your line

(Copilot, tab)

Chat assistant

ask, copy, paste

(ChatGPT)

Three kinds of AI coding help

most of you have used this

Autocomplete

finishes your line

(Copilot, tab)

Chat assistant

ask, copy, paste

(ChatGPT)

Coding agent

reads, edits, runs

in your project

Three kinds of AI coding help

most of you have used this

Autocomplete

finishes your line

(Copilot, tab)

Chat assistant

ask, copy, paste

(ChatGPT)

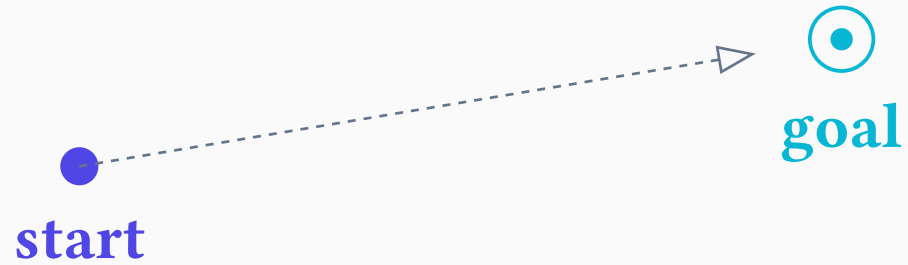
Coding agent

reads, edits, runs

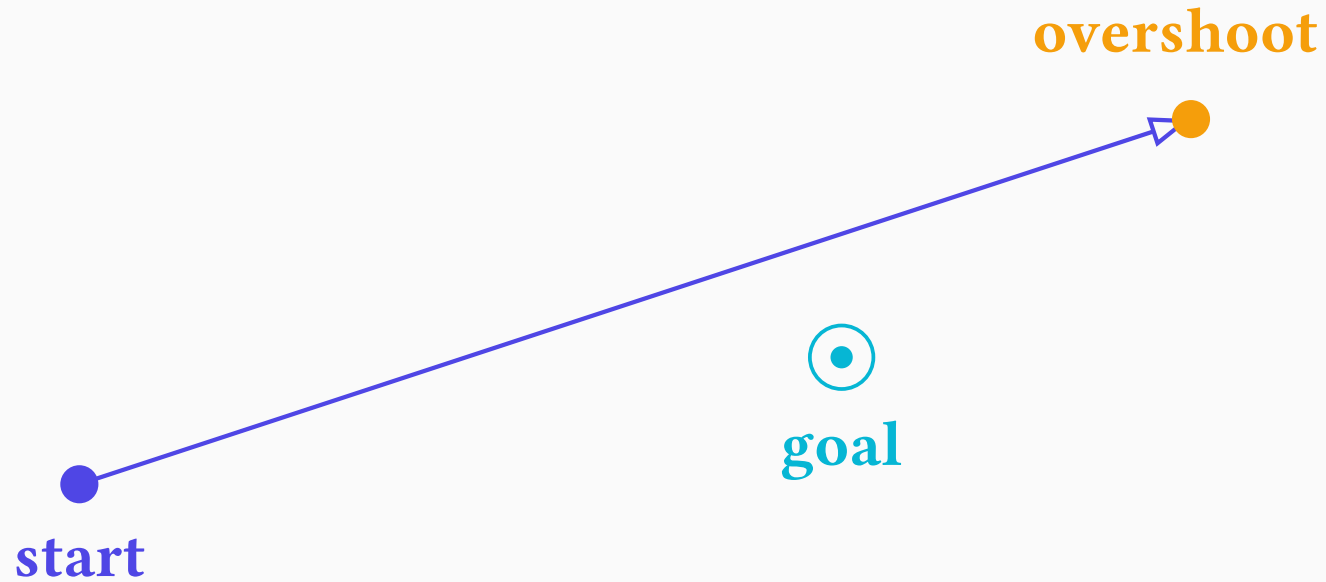
in your project

With an agent you set the task and review the work.

More autonomy, less control over each step



More autonomy, less control over each step



More autonomy, less control over each step



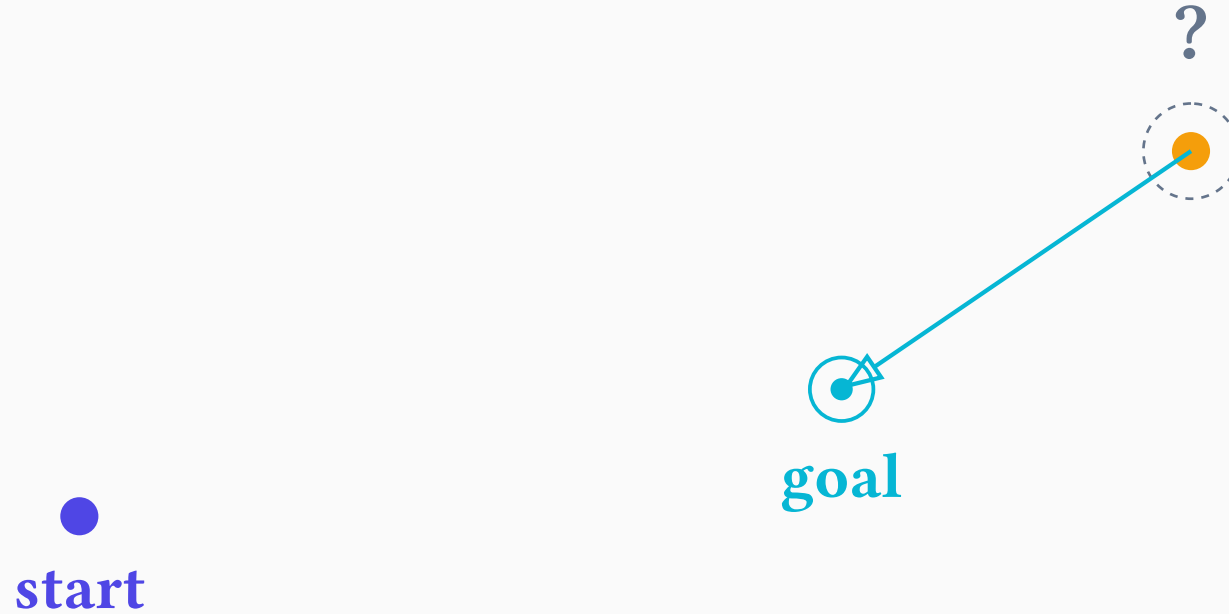
More autonomy, less control over each step

cognitive debt



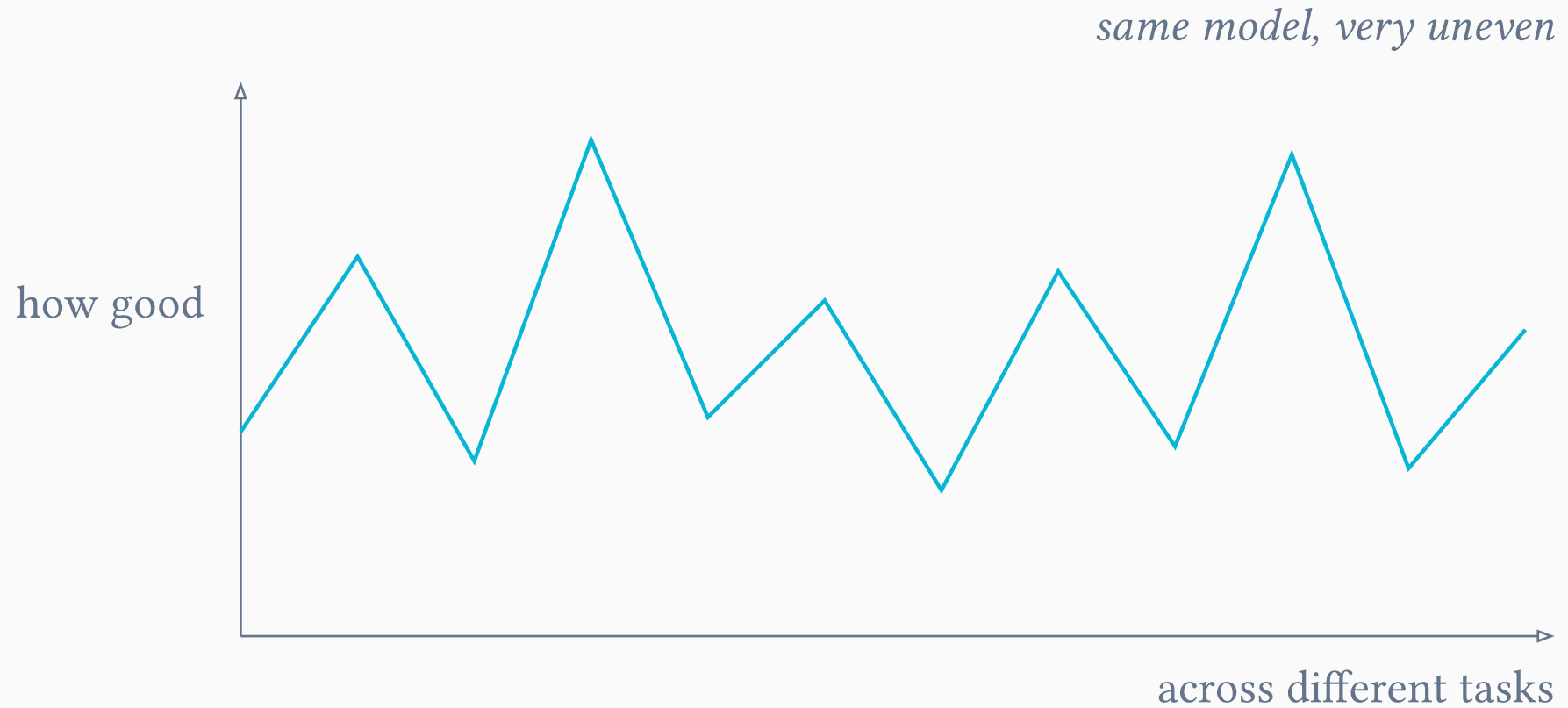
More autonomy, less control over each step

cognitive debt

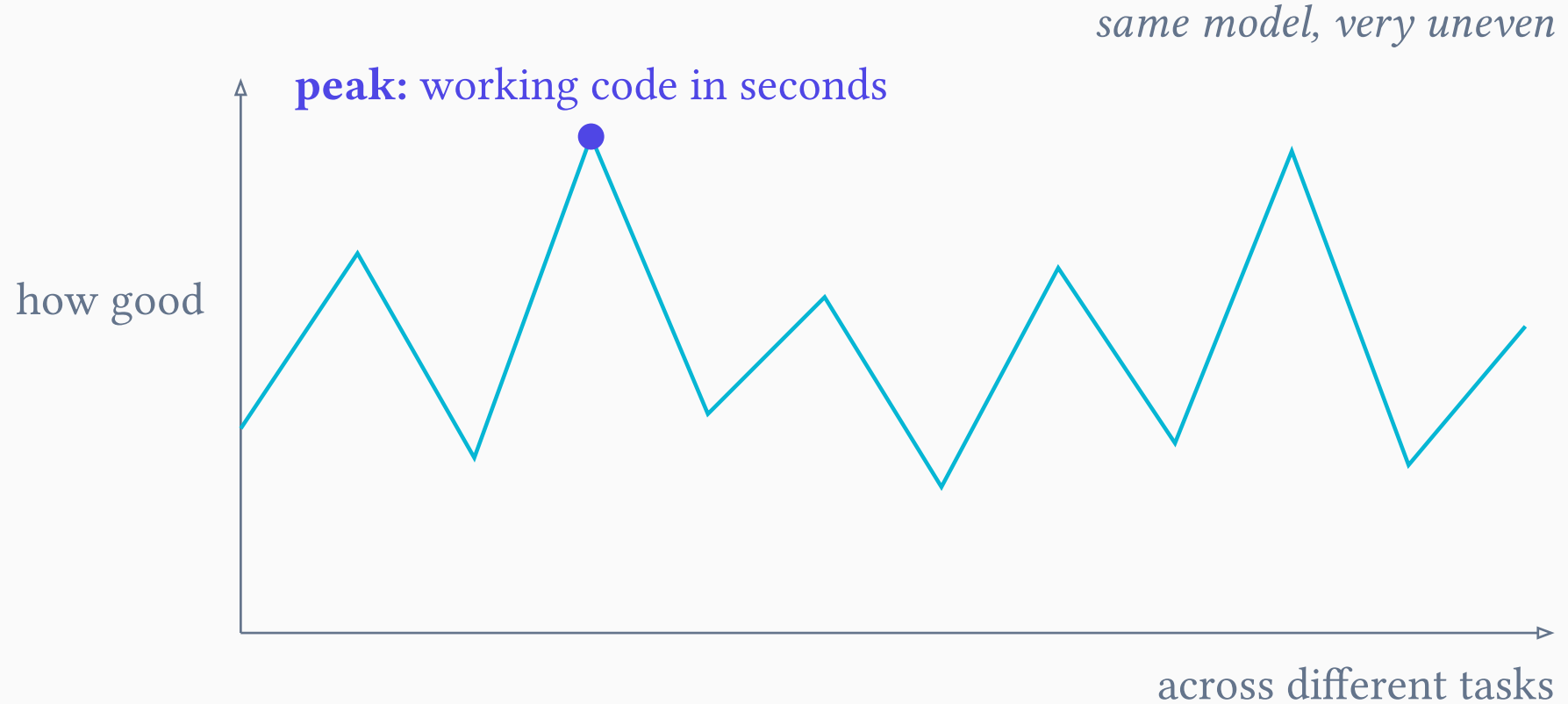


Choose how much freedom to give for each task.

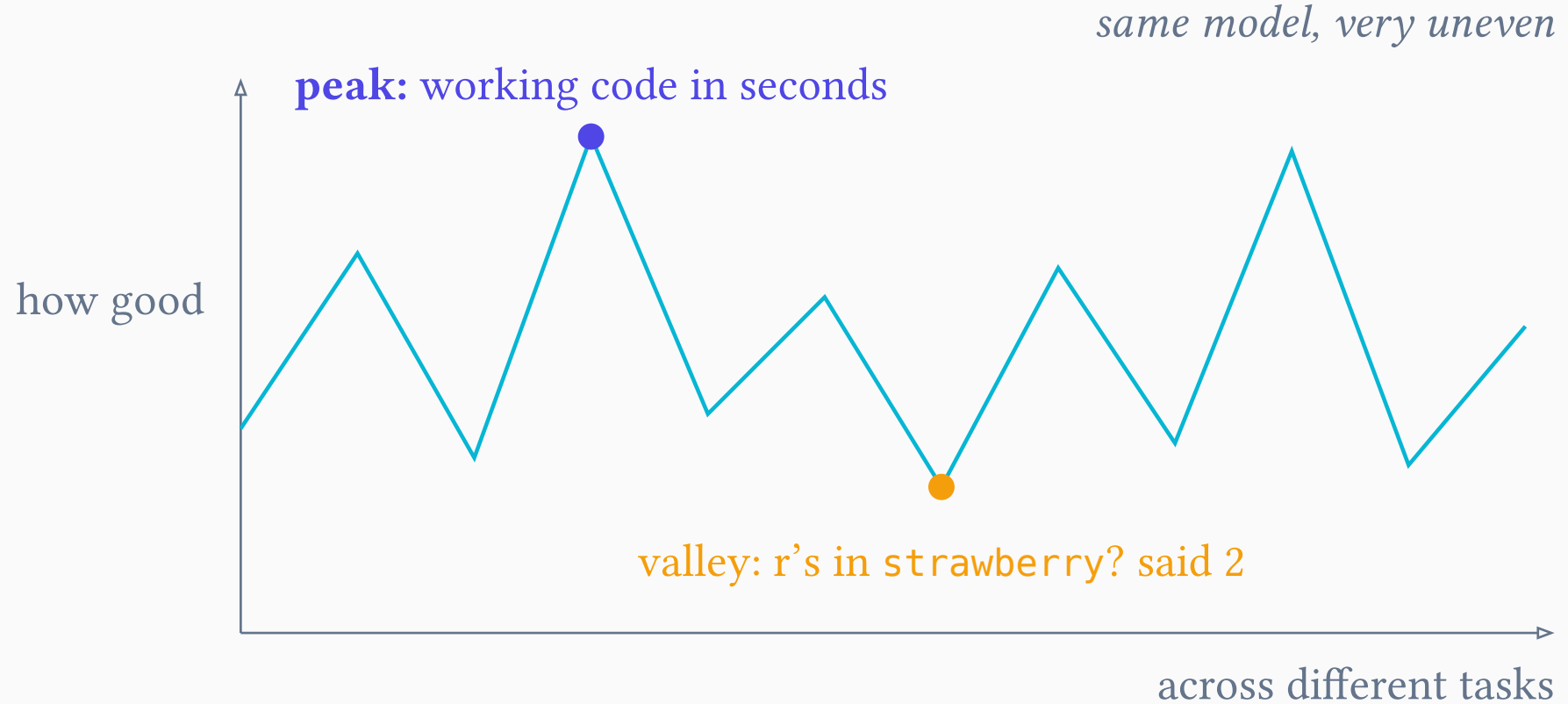
Agent ability is uneven across tasks



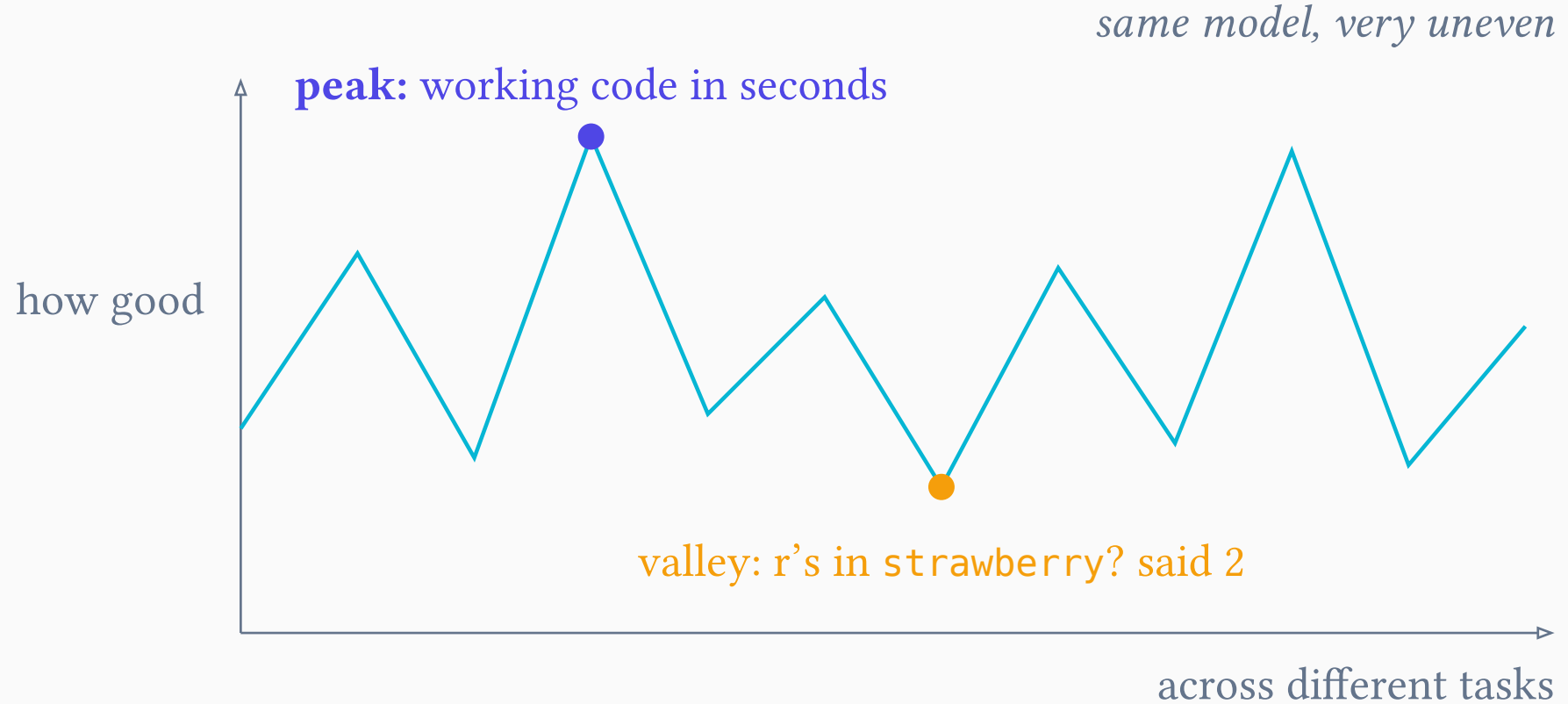
Agent ability is uneven across tasks



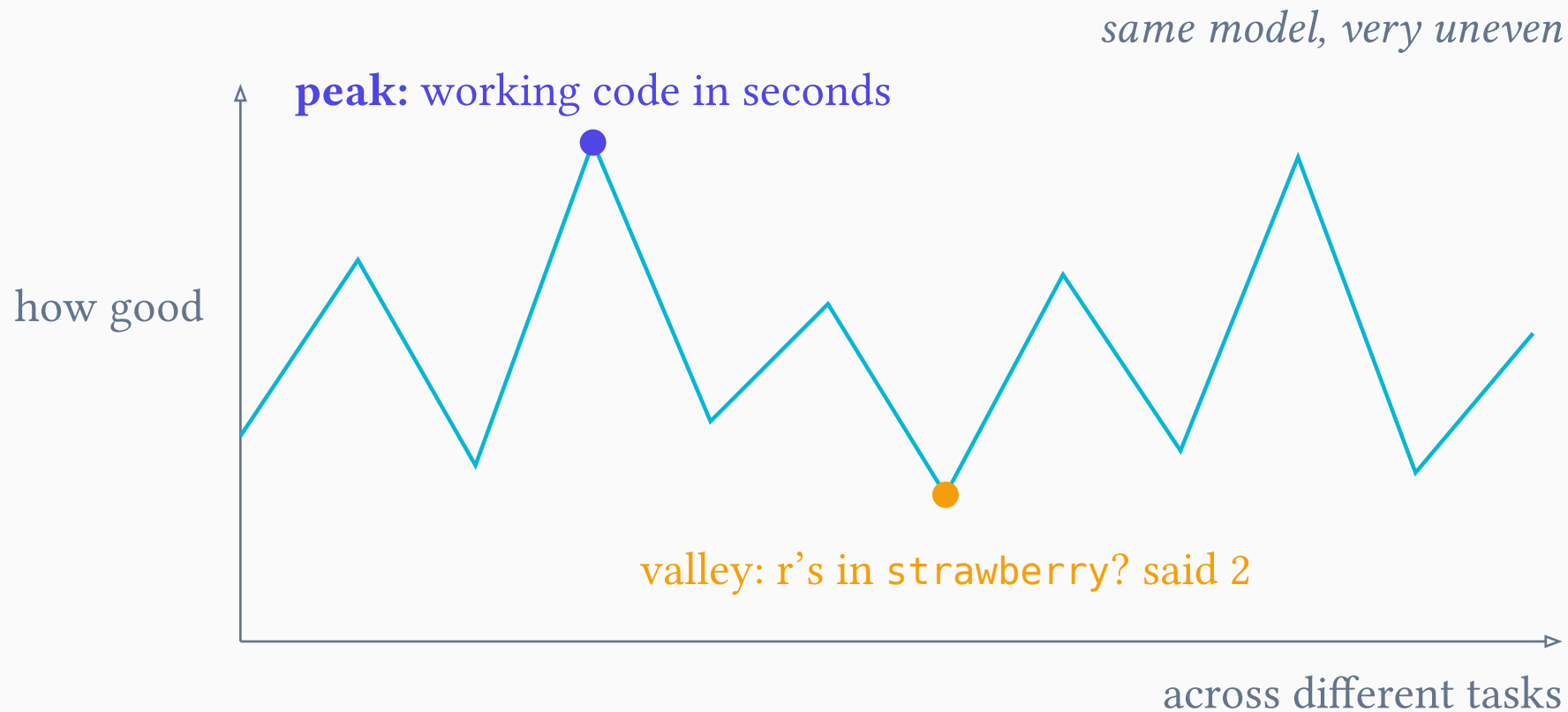
Agent ability is uneven across tasks



Agent ability is uneven across tasks



Agent ability is uneven across tasks



Learn where the agent is reliable, and where its output needs checking.

Inside an agent: a harness and a remote LLM

LLM

predicts text

runs elsewhere

Inside an agent: a harness and a remote LLM

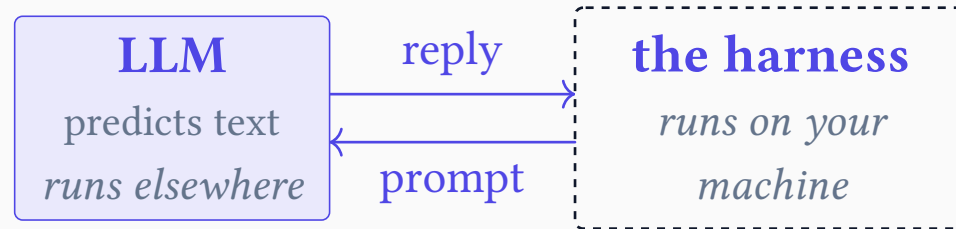
LLM
predicts text
runs elsewhere

the harness
*runs on your
machine*

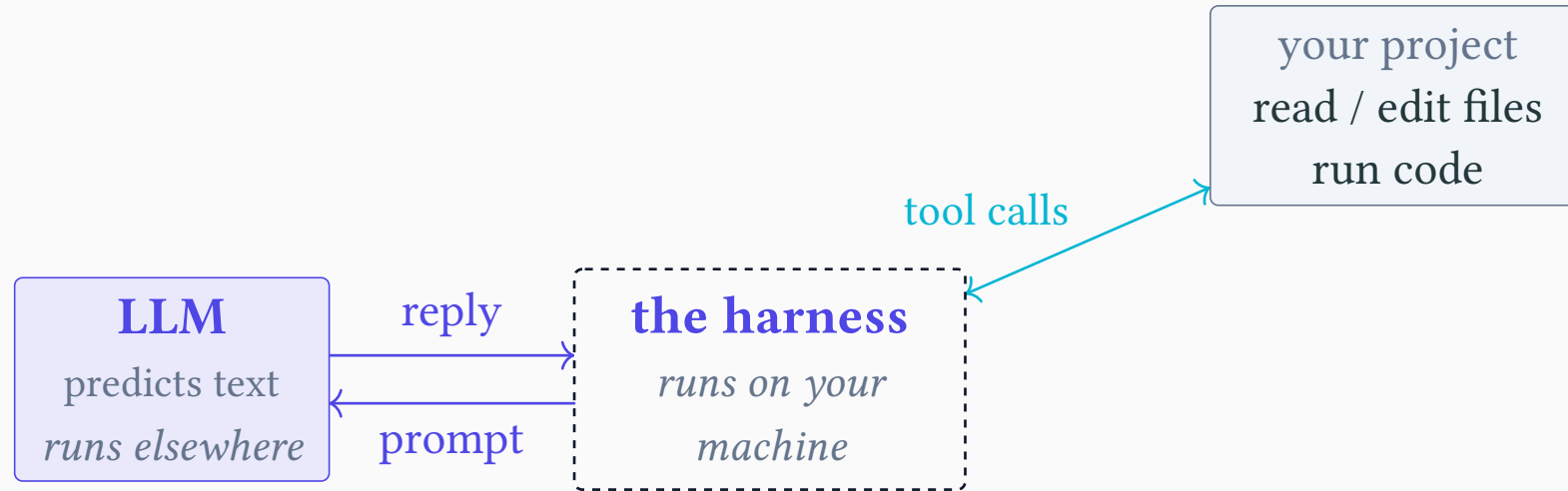
Inside an agent: a harness and a remote LLM



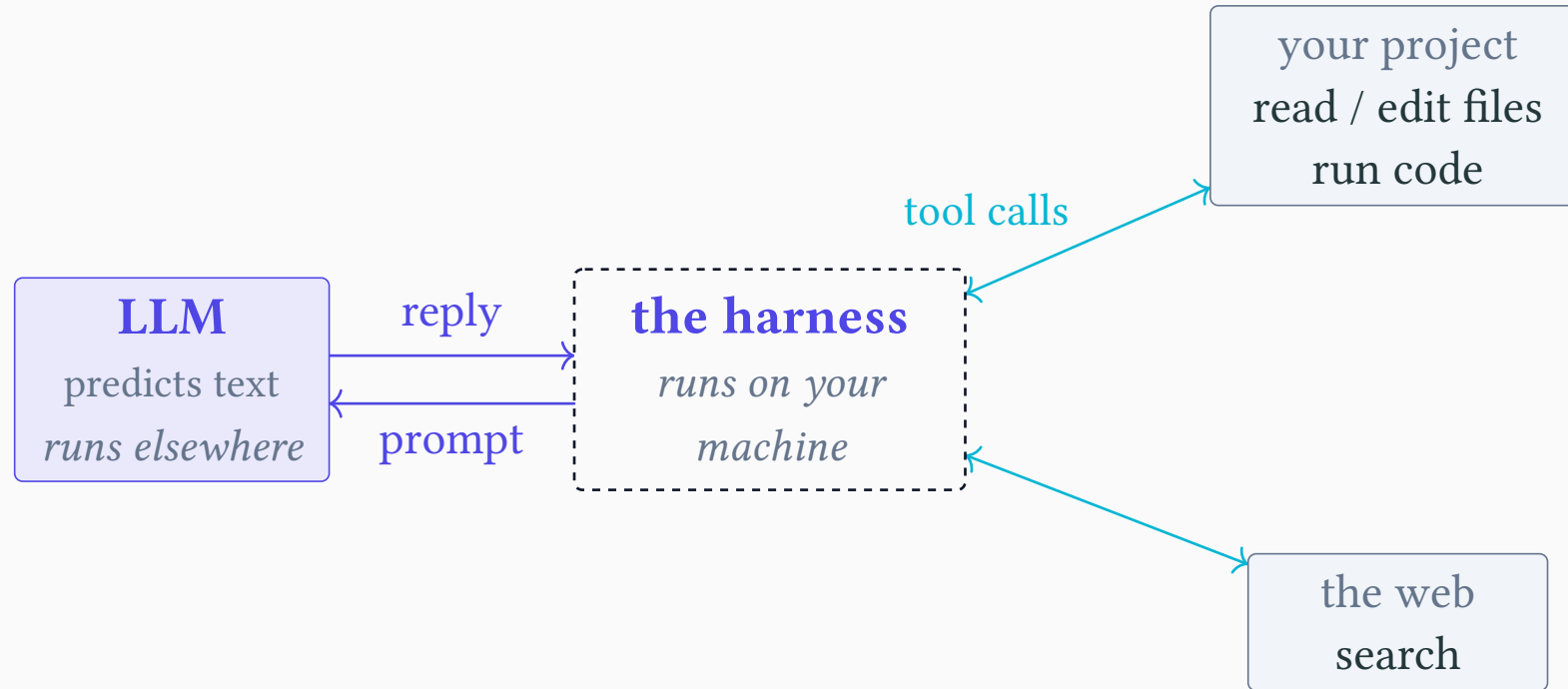
Inside an agent: a harness and a remote LLM



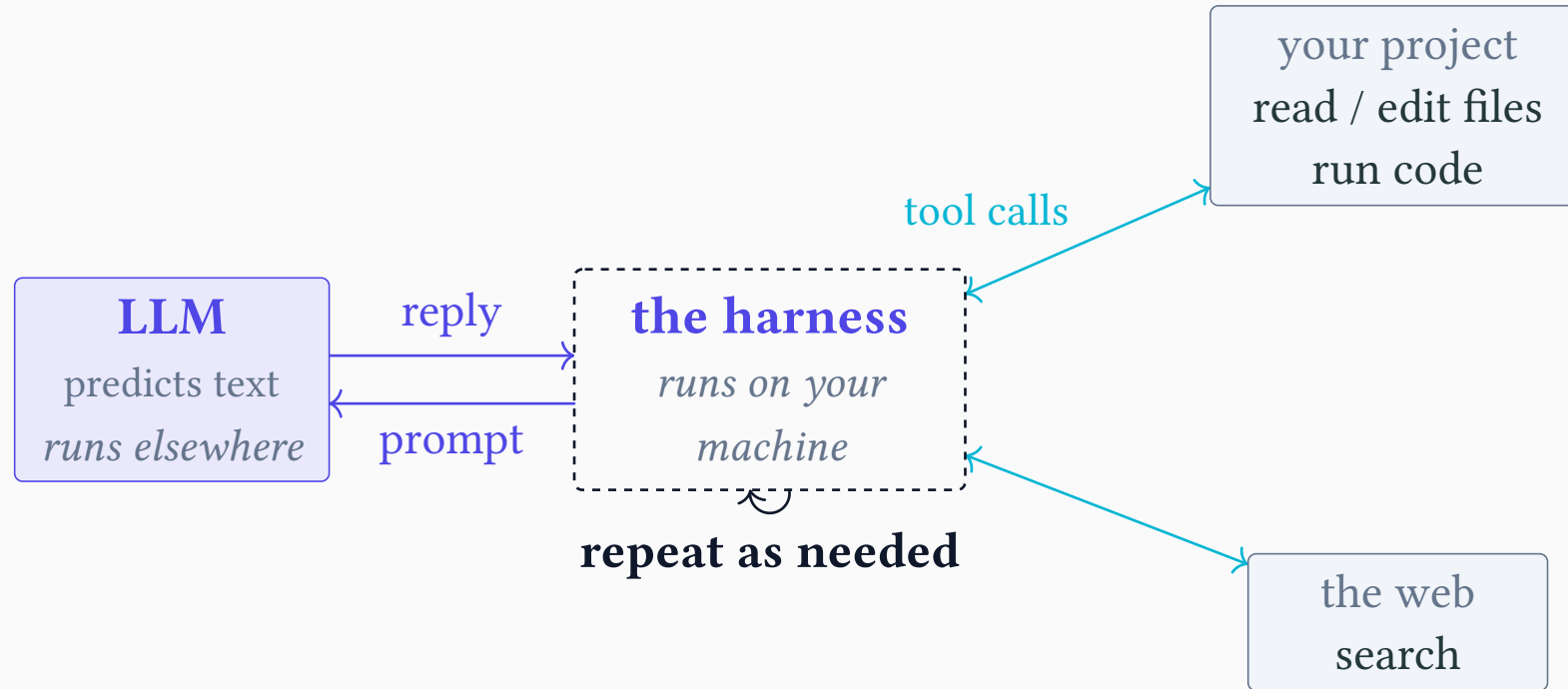
Inside an agent: a harness and a remote LLM



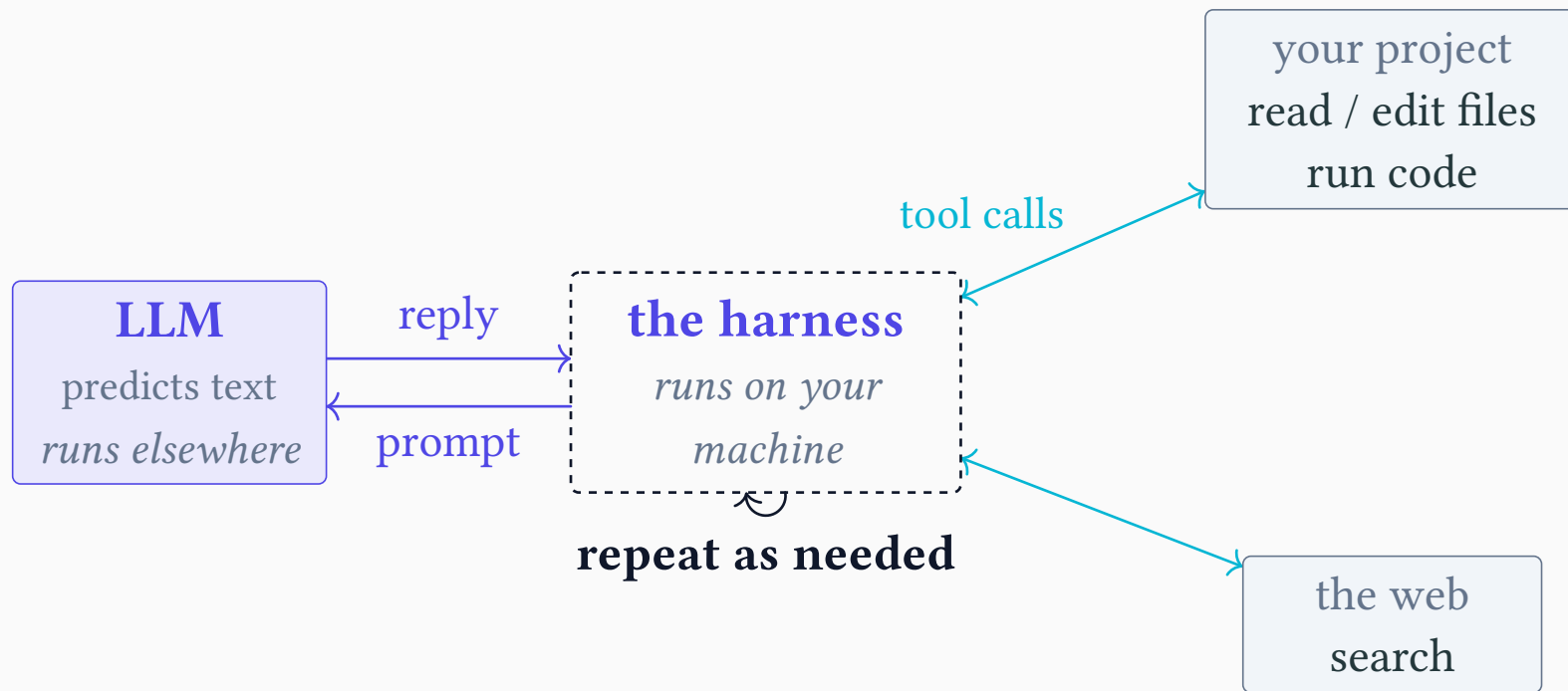
Inside an agent: a harness and a remote LLM



Inside an agent: a harness and a remote LLM



Inside an agent: a harness and a remote LLM



We'll study how the LLM and the harness work, so you can tell what an agent can and cannot do.

Agents meet real software work

This course is hands-on. You will work with these agents on real software.

Agents meet real software work

This course is hands-on. You will work with these agents on real software.

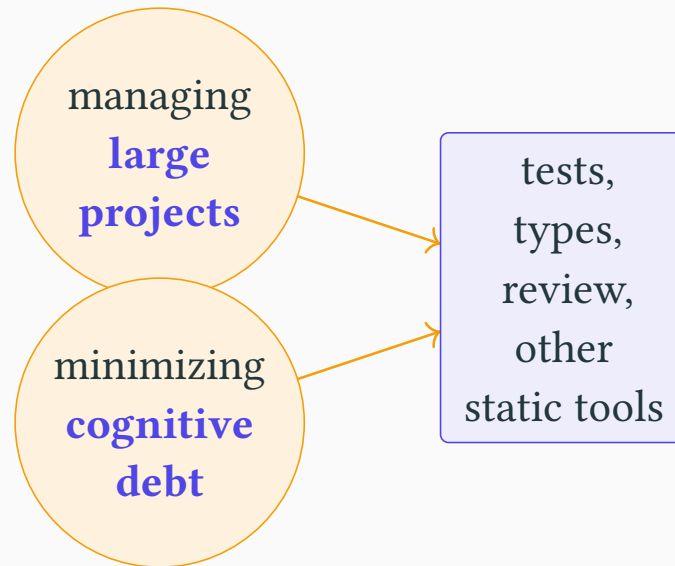
A diagram consisting of two overlapping circles. The top circle is light orange with a thin orange border and contains the text 'managing large projects'. The bottom circle is also light orange with a thin orange border and contains the text 'minimizing cognitive debt'. The circles overlap in the middle.

managing
large
projects

minimizing
cognitive
debt

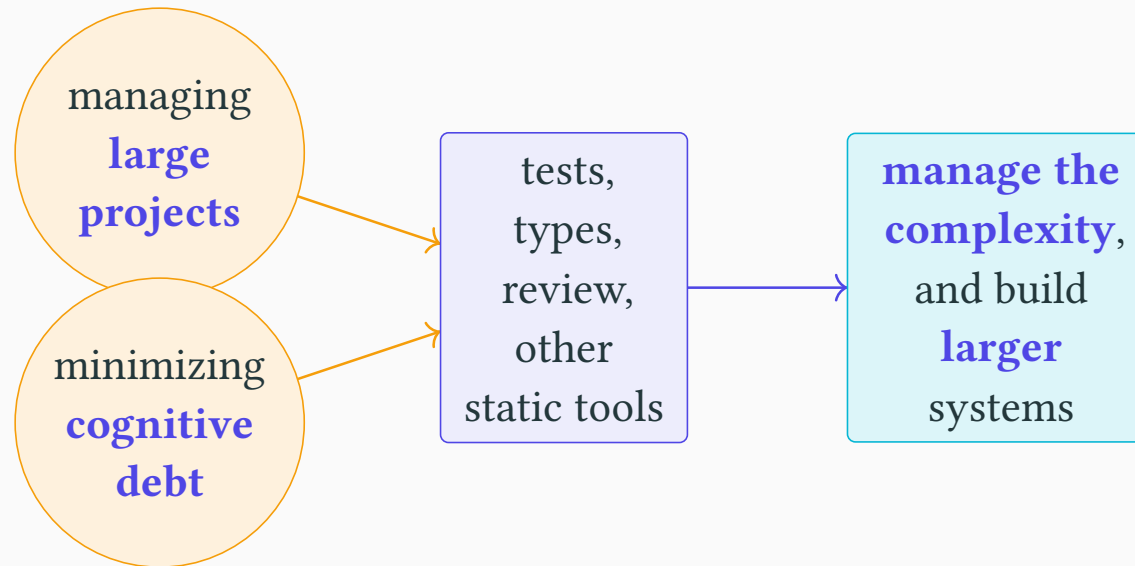
Agents meet real software work

This course is hands-on. You will work with these agents on real software.



Agents meet real software work

This course is hands-on. You will work with these agents on real software.



The risk: it runs on your machine

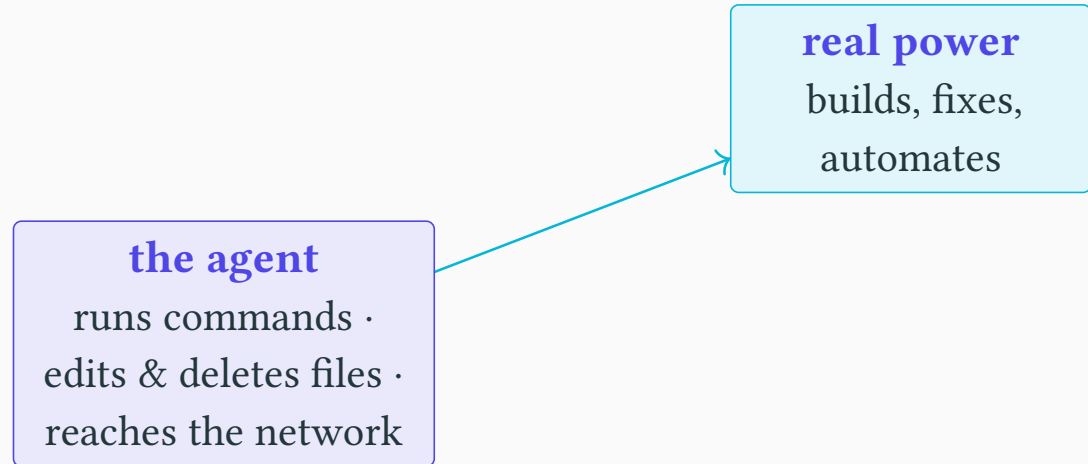
An agent can run **commands** on your machine. Mistakes can cause damage.

the agent

- runs commands ·
- edits & deletes files ·
- reaches the network

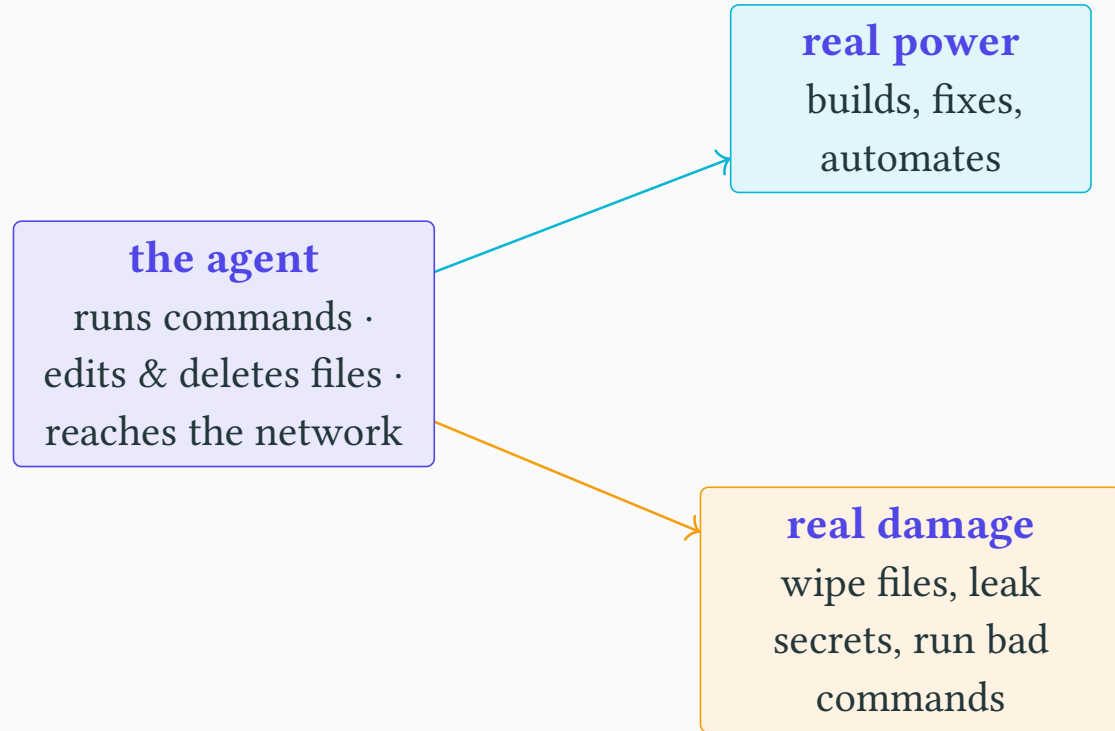
The risk: it runs on your machine

An agent can run **commands** on your machine. Mistakes can cause damage.



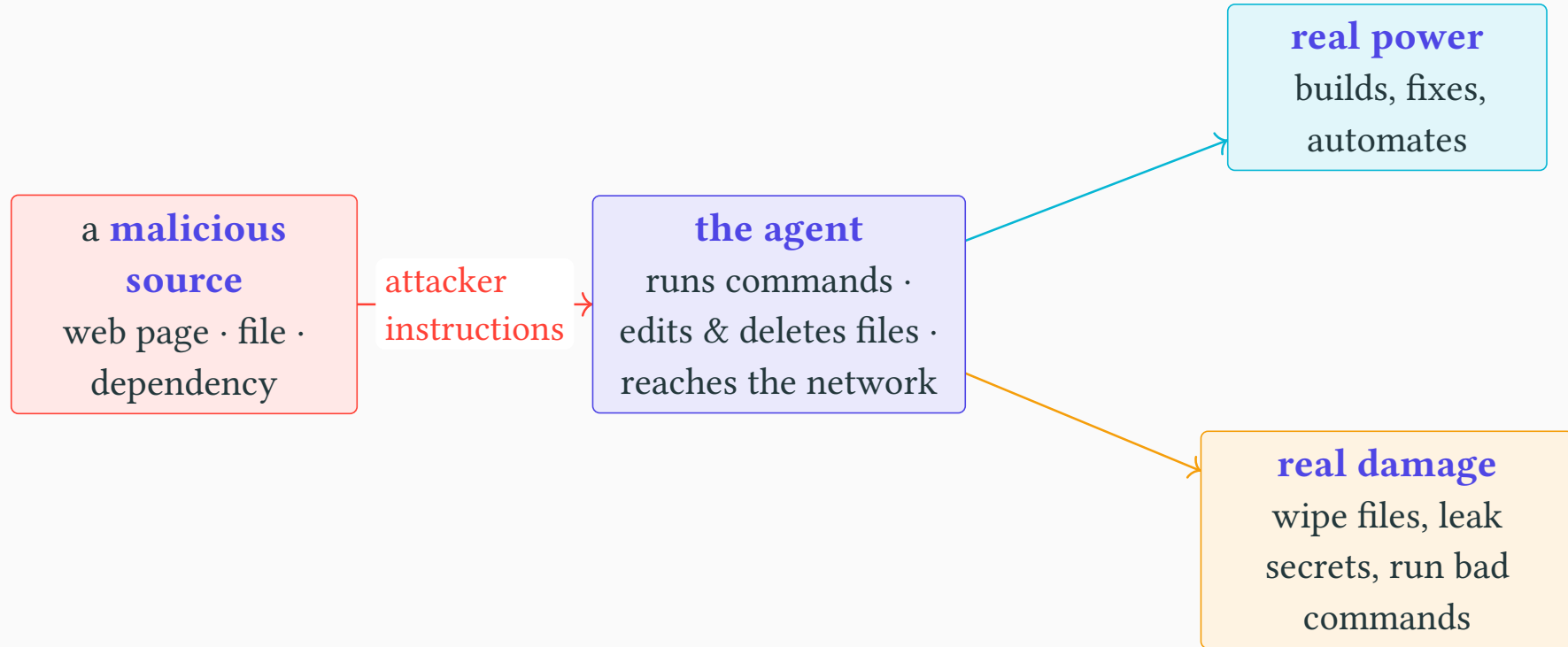
The risk: it runs on your machine

An agent can run **commands** on your machine. Mistakes can cause damage.



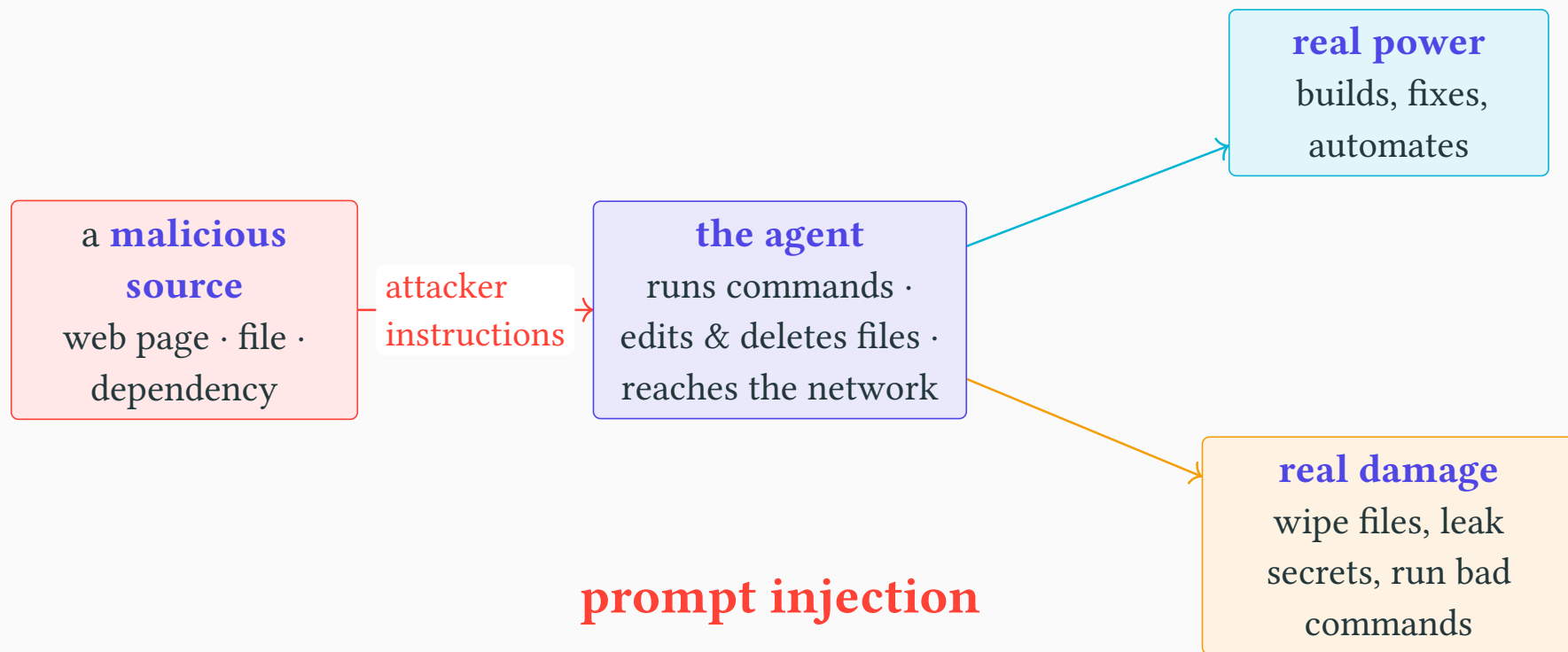
The risk: it runs on your machine

An agent can run **commands** on your machine. Mistakes can cause damage.



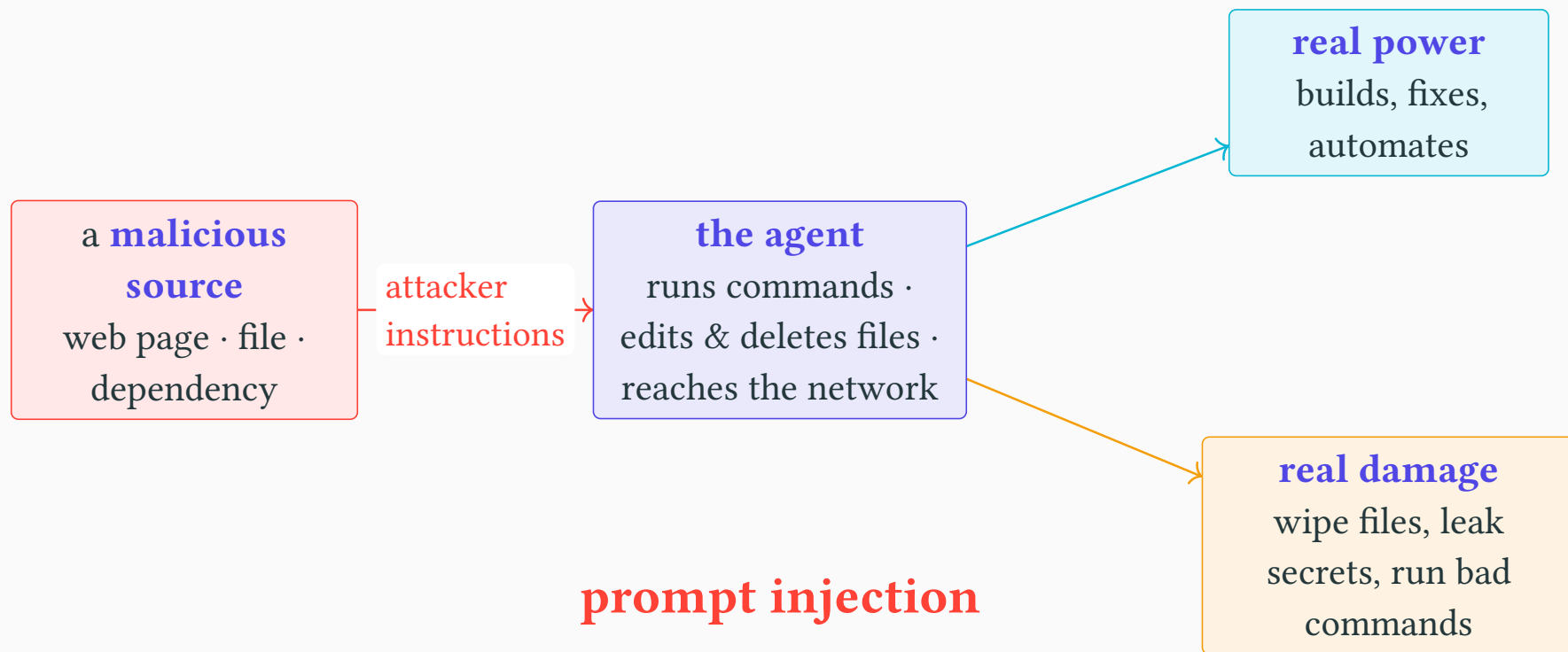
The risk: it runs on your machine

An agent can run **commands** on your machine. Mistakes can cause damage.



The risk: it runs on your machine

An agent can run **commands** on your machine. Mistakes can cause damage.



We need to limit the harm from mistakes and attacks.

What we'll cover

- 1 **Directing agents** through real, non-trivial tasks

What we'll cover

- 1 **Directing agents** through real, non-trivial tasks
- 2 **Managing context** CLAUDE.md, skills, subagents

What we'll cover

- 1 **Directing agents** through real, non-trivial tasks
- 2 **Managing context** CLAUDE.md, skills, subagents
- 3 **Making output reliable** tests, types, linters, CI

What we'll cover

- 1 **Directing agents** through real, non-trivial tasks
- 2 **Managing context** CLAUDE.md, skills, subagents
- 3 **Making output reliable** tests, types, linters, CI
- 4 **Sandboxing & security** containers, prompt injection

What we'll cover

- 1 **Directing agents** through real, non-trivial tasks
- 2 **Managing context** CLAUDE.md, skills, subagents
- 3 **Making output reliable** tests, types, linters, CI
- 4 **Sandboxing & security** containers, prompt injection
- 5 **Orchestrating agents** several on one problem at once

What we'll cover

- 1 **Directing agents** through real, non-trivial tasks
- 2 **Managing context** CLAUDE.md, skills, subagents
- 3 **Making output reliable** tests, types, linters, CI
- 4 **Sandboxing & security** containers, prompt injection
- 5 **Orchestrating agents** several on one problem at once

Throughout, we measure whether the agent actually helps.

Next: how the course runs, the weekly progression, grading, and what you need to set up.