

Accelerating Science with AI: Surfing the Frontier Between the Bitter Lesson and No Free Lunch

Scott Clark, Founder & CEO

[DISTRIBUTIONAL]



Three Takeaways

- 1. AI x Science can lead to abundance**
- 2. Researchers will always have a role**
- 3. Better tools can accelerate this**

What abundance with AI actually means

new scientific discoveries

new materials & more efficient energy

advances in medicine, aerospace & more

living in a scifi future

The real promise of abundance through AI

**Improving quality of life,
not just replacing repetitive work.**

The pieces are finally falling into place

**We can now meaningfully accelerate
science with AI.**

The Bitter Lesson meets No Free Lunch

Scale beats cleverness

but

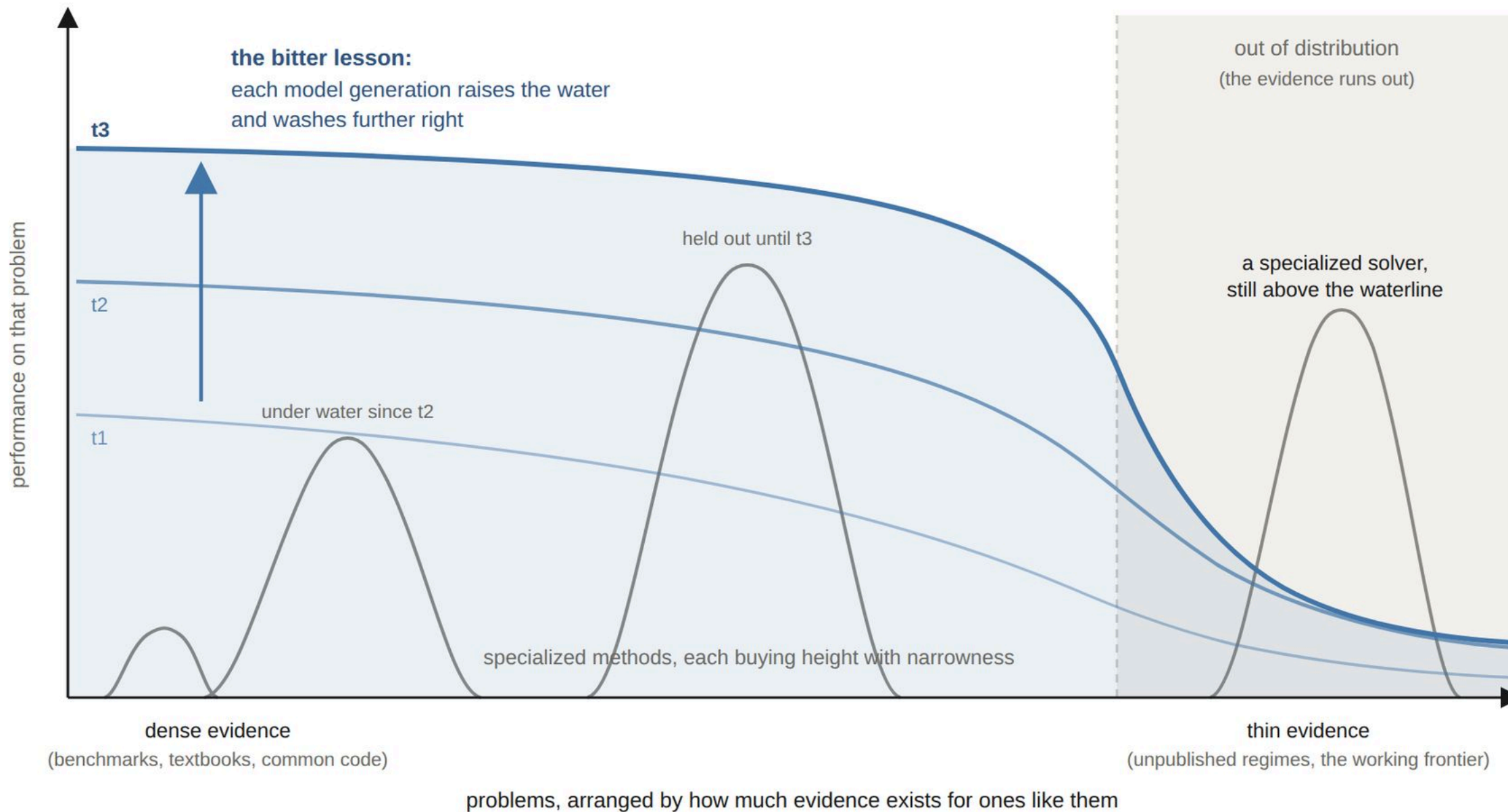
Specificity beats generality

Rich Sutton, The Bitter Lesson, 2019

Wolpert and Macready, No Free Lunch Theorems for Optimization, 1997

The bitter lesson, drawn as a rising wave

a conceptual sketch, not measured data



But there is hope

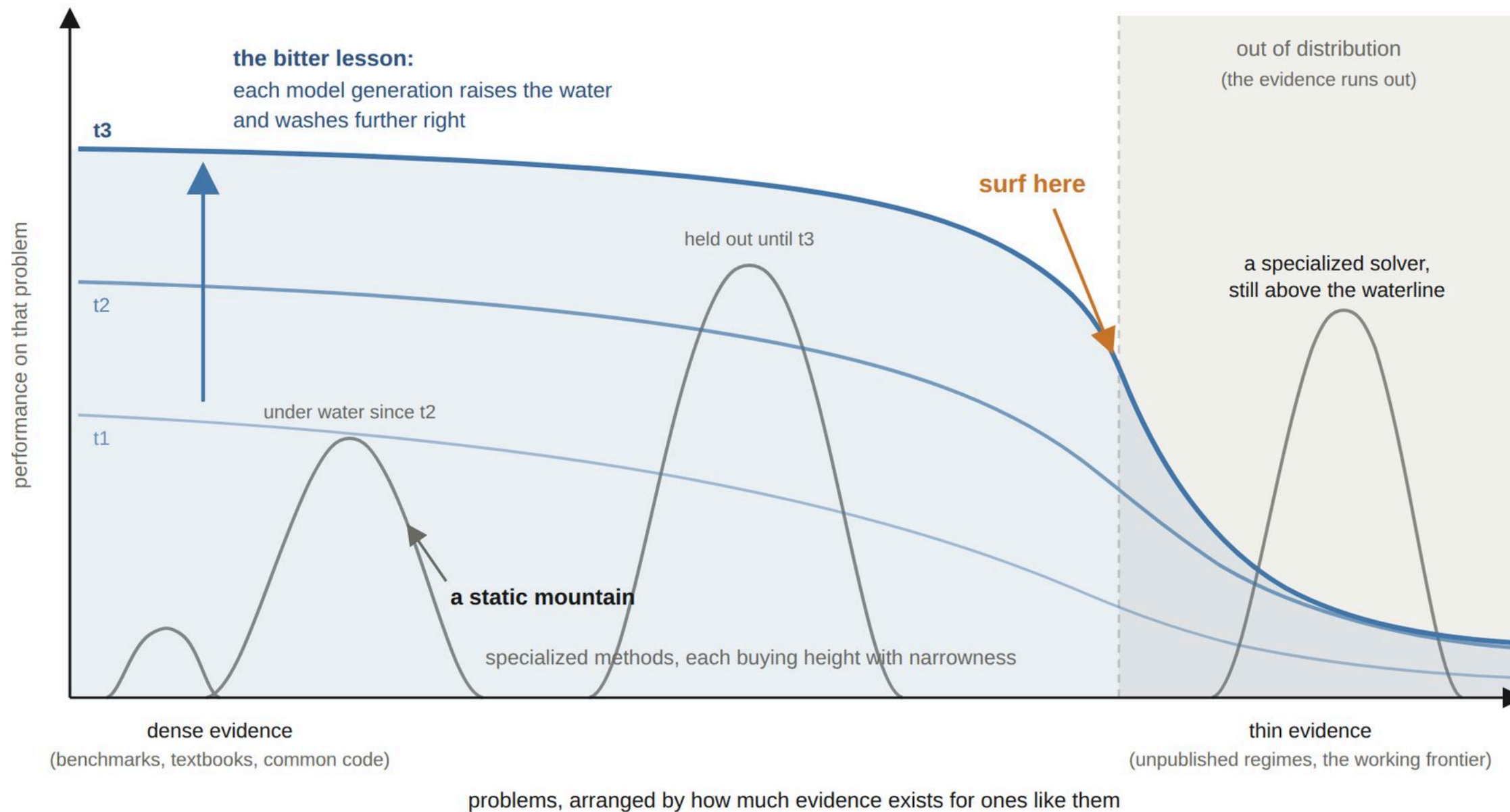
There will always be a frontier

Surf the wave.

**Don't climb a static mountain
that is about to be washed over.**

The bitter lesson, drawn as a rising wave

a conceptual sketch, not measured data



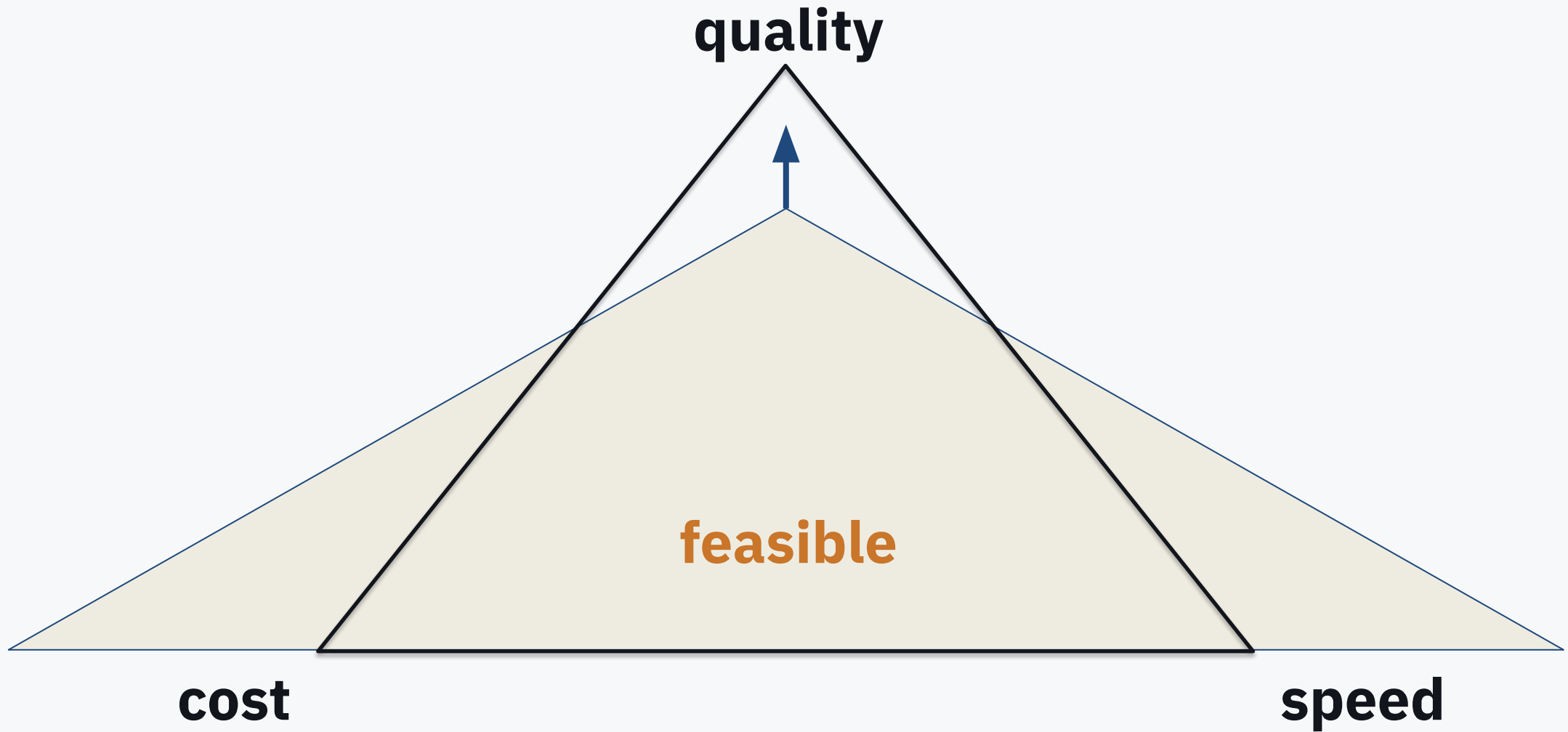
We are still a long way from

/goal solve stable fusion containment

AI is changing the game

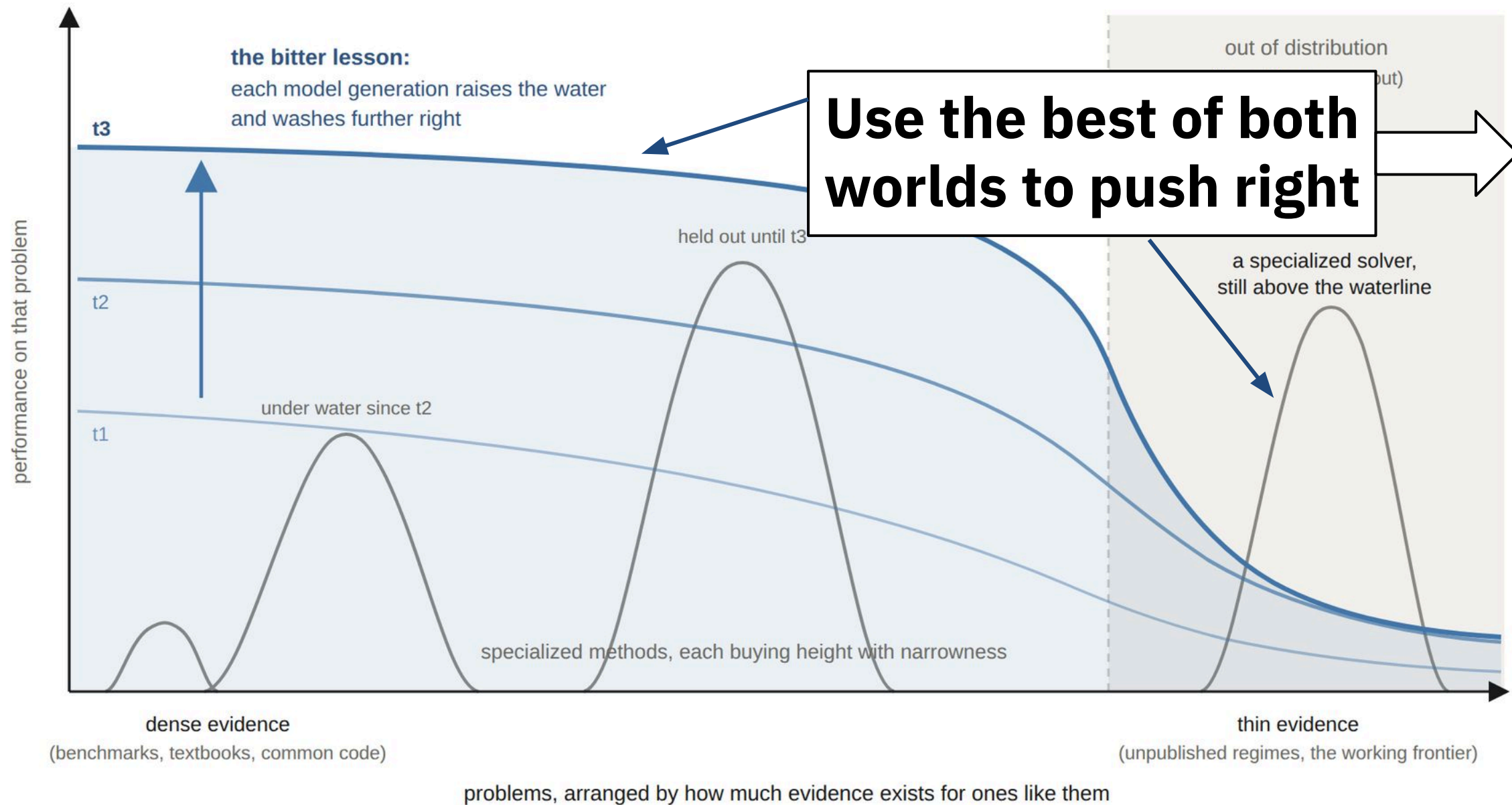
**Things that were previously impractical
are now achievable**

AI now lets you do more, faster, with increasing quality



The bitter lesson, drawn as a rising wave

a conceptual sketch, not measured data



A culmination of my life's work

**I have been trying to solve this problem
for the last 20 years of my career
with varying degrees of success.**

Grad school: Solving the same problem over and over

**Whatever the field,
the final step was always parameter tuning.**

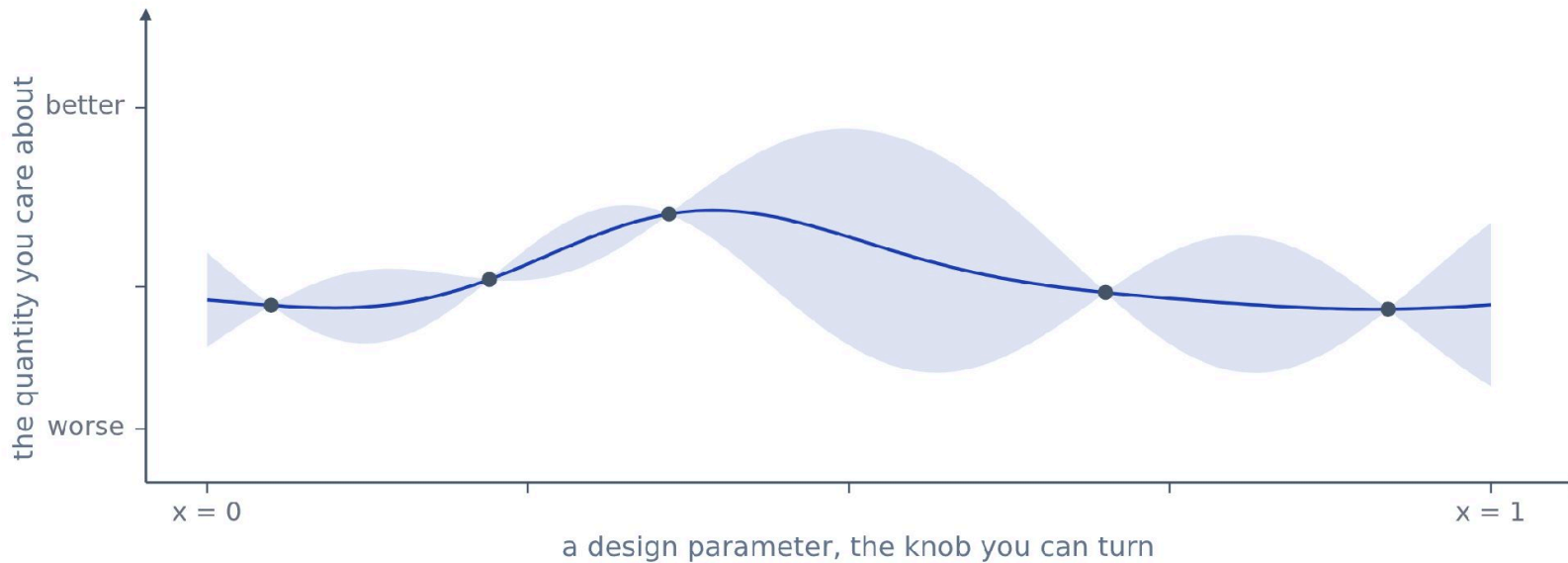


Methods for efficient optimization given an objective

Bayesian optimization: a few evaluations are all you have

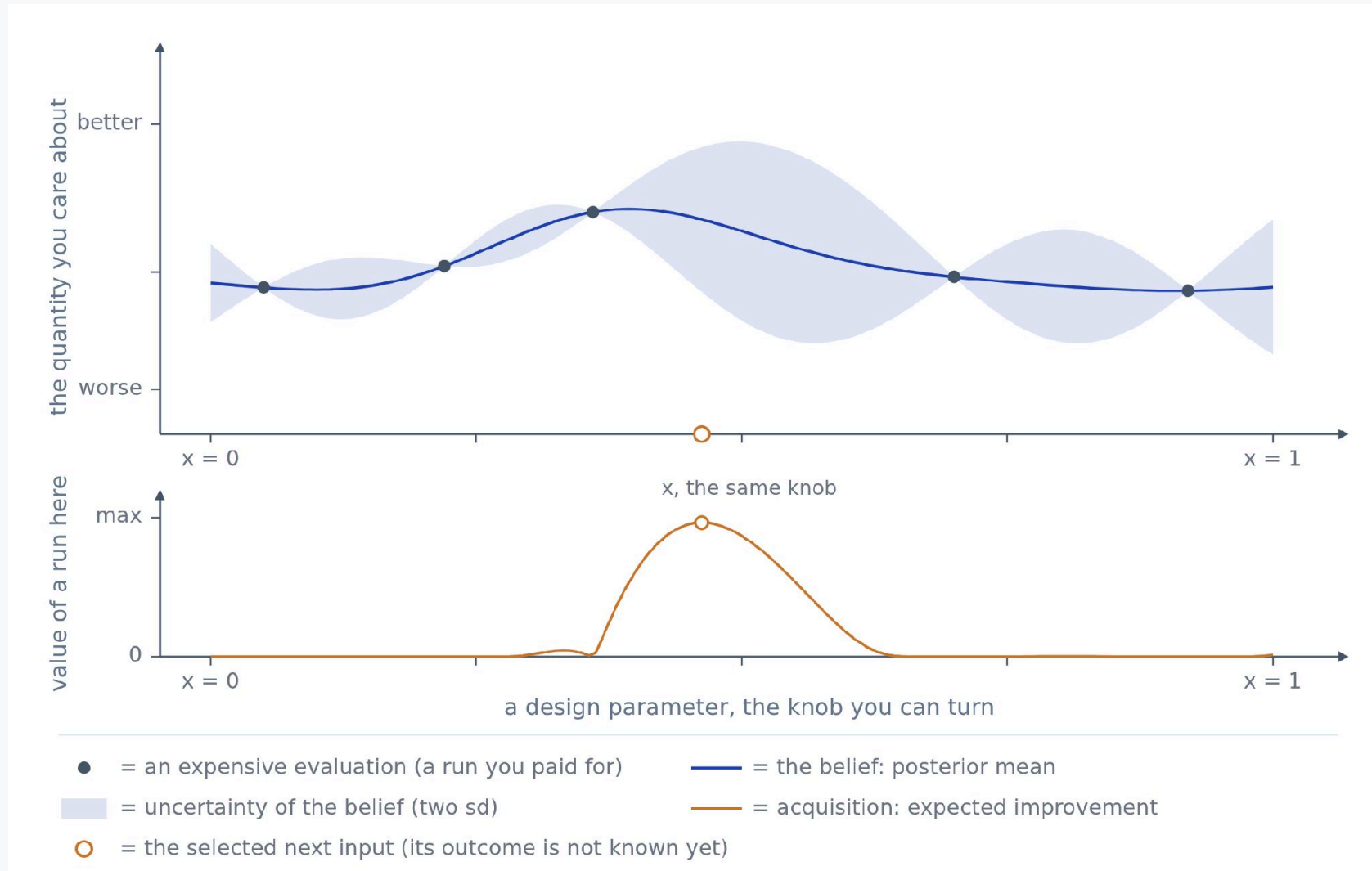


A Gaussian process turns them into a belief about what you have not run yet

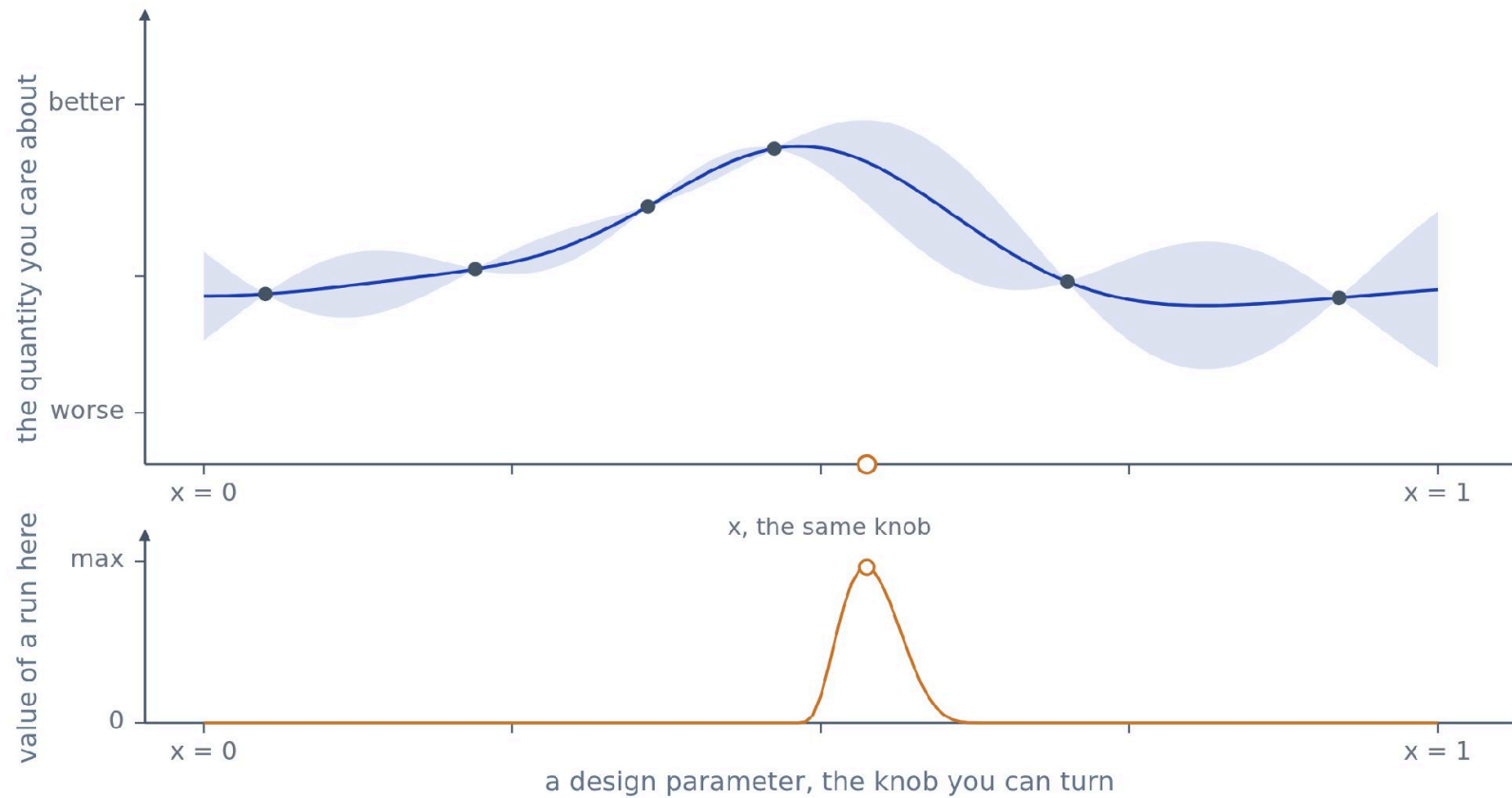


- = an expensive evaluation (a run you paid for)
- = the belief: posterior mean
- = uncertainty of the belief (two sd)

An acquisition function values the next evaluation



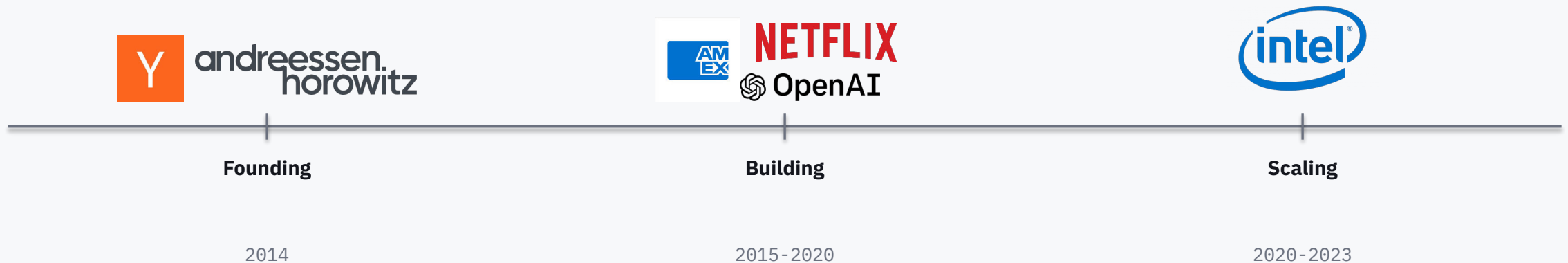
Measure, update, repeat



- = an expensive evaluation (a run you paid for)
- = the belief: posterior mean
- = uncertainty of the belief (two sd)
- = acquisition: expected improvement
- = the selected next input (its outcome is not known yet)

SigOpt: Optimize everything

Bayesian Optimization as a Service



Optimization at scale for problems that mattered

Learning I was only solving half the problem

The optimizer worked

...except when it didn't.

It overfit, or optimized the “wrong” thing

The problem with optimizers

The *best* thing about a black box optimizer is that it will optimize *anything* you tell it to

The *worst* part about a black box optimizer is that it will optimize *exactly* what you tell it to

Optimization was only half the problem

You need to be able to trust the numbers

Have a problem? Start a startup!

High dimensional Bayesian statistical tests for
chaotic and non-stationary AI systems / agents
It failed.

distributional

Pre-production AI Testing

2023-2025

- 1. People didn't know what to test
(or have the data before deployment)**
- 2. They couldn't understand tests well enough to act**
- 3. They shipped to prod anyway**

Have a problem? Pivot!

Post-production analytics on agent traces to find the evals that you are monitoring for

It failed.

distributional

[DISTRIBUTIONAL]

Pre-production AI Testing

Post-production Agent Analytics

2023-2025

2025-2026

A feature, not a product.

(let alone a business or a startup)

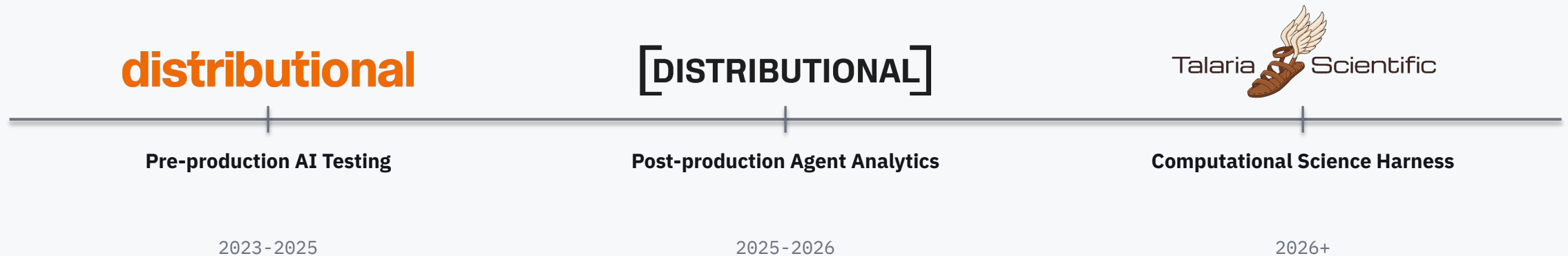
But through failure comes wisdom

Knowing what to measure matters

Trusting what you measure matters

Have a problem? Solve the version you care about

A multi-agent, multi-FM harness for computational science research acceleration.



Using domain specific guardrails rooted in physics and simulation to ensure trustworthiness

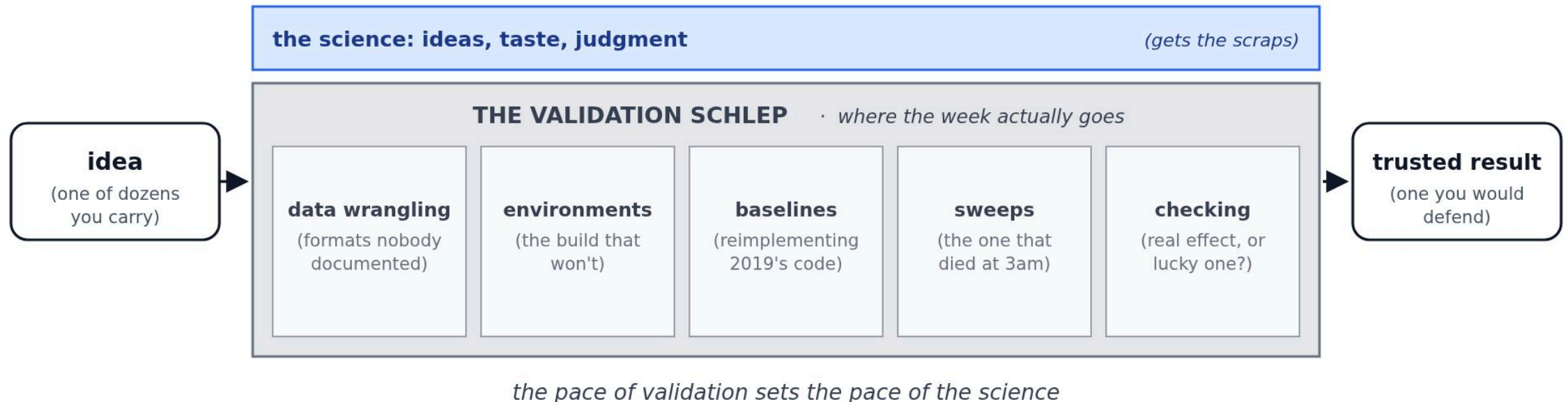
The pace of science is set by validation

**Ideas move science forward,
but testing them is expensive in researcher
time.**

Where the week actually goes

FROM IDEA TO TRUSTED RESULT: WHERE THE WEEK GOES

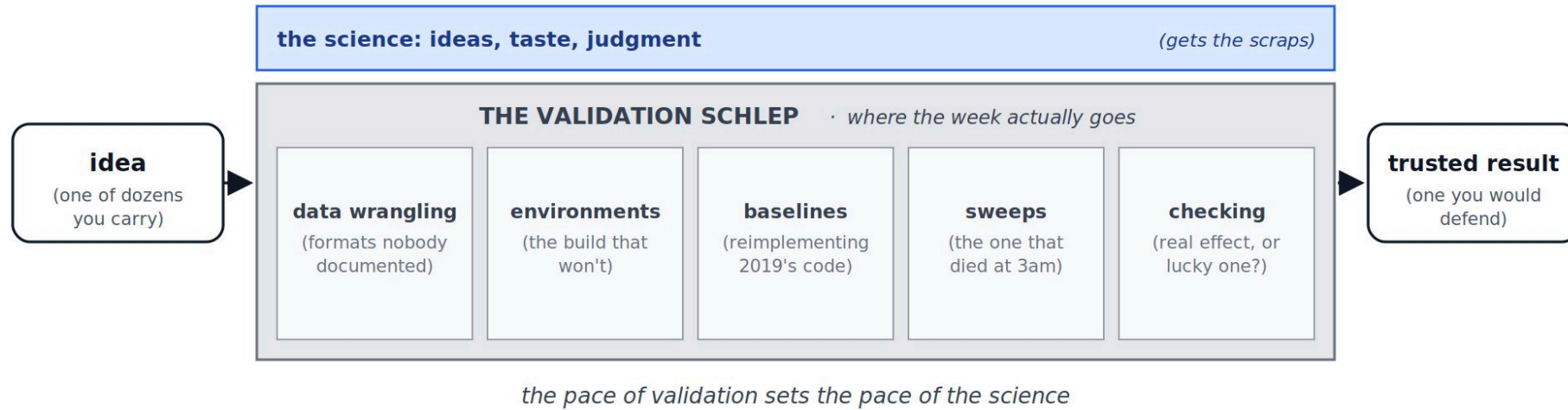
1 • TODAY the scientist walks the whole span by hand



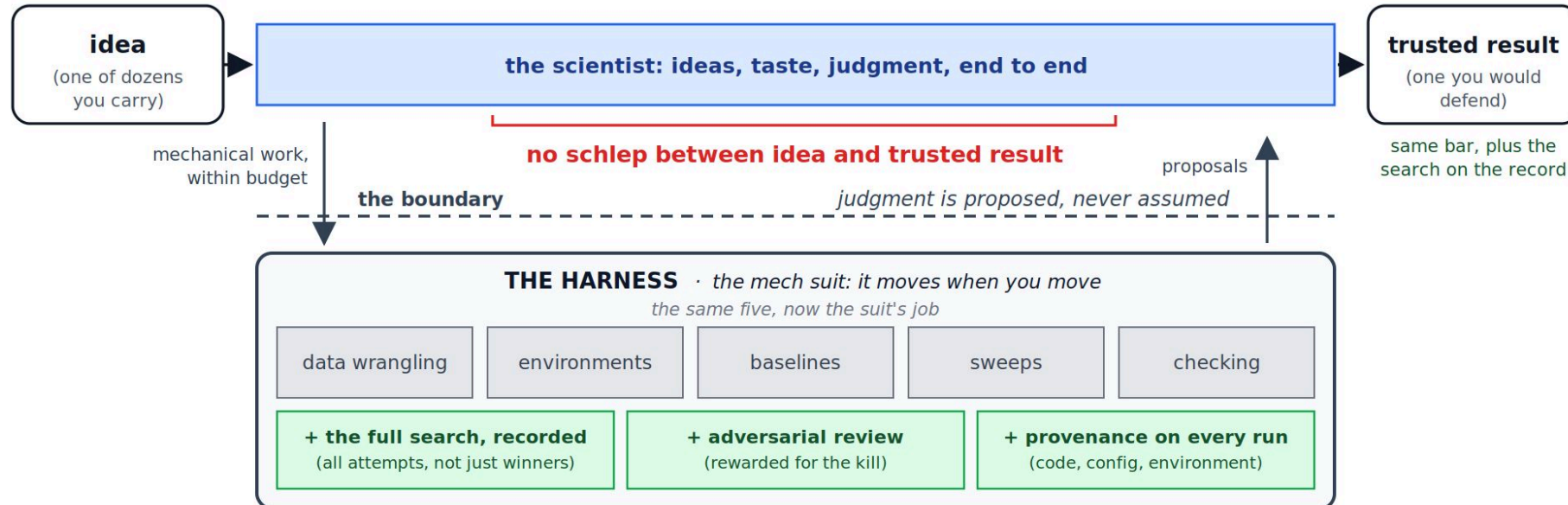
AI is now good at the mechanical middle of research

FROM IDEA TO TRUSTED RESULT: WHERE THE WEEK GOES

1 • TODAY the scientist walks the whole span by hand



2 • IN THE SUIT the same span, worn as equipment

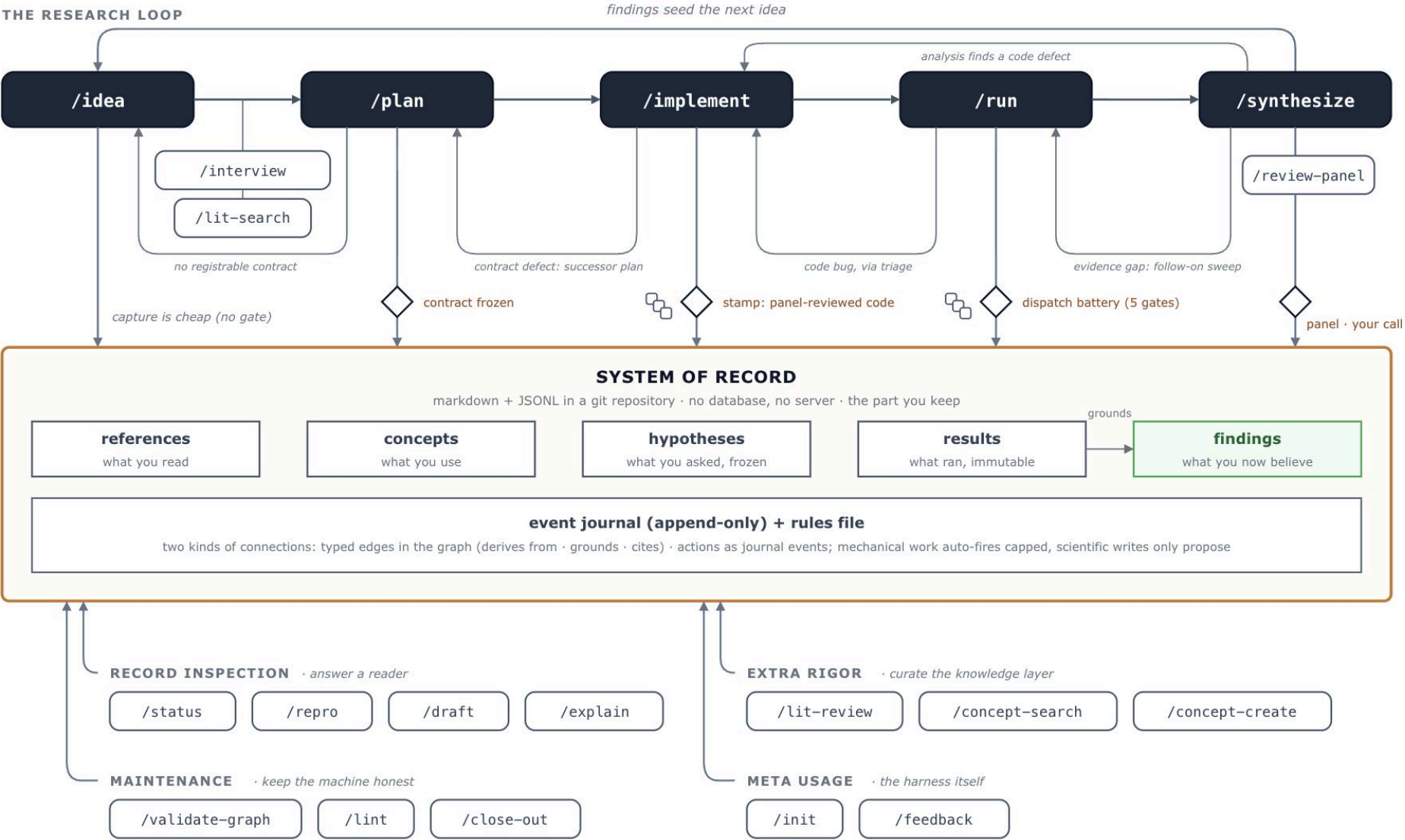


A computational science “what-if” machine

- **Start new research from scratch**
- **Respond to reviewers**
- **Convert old academic software**
- **Extend current papers**

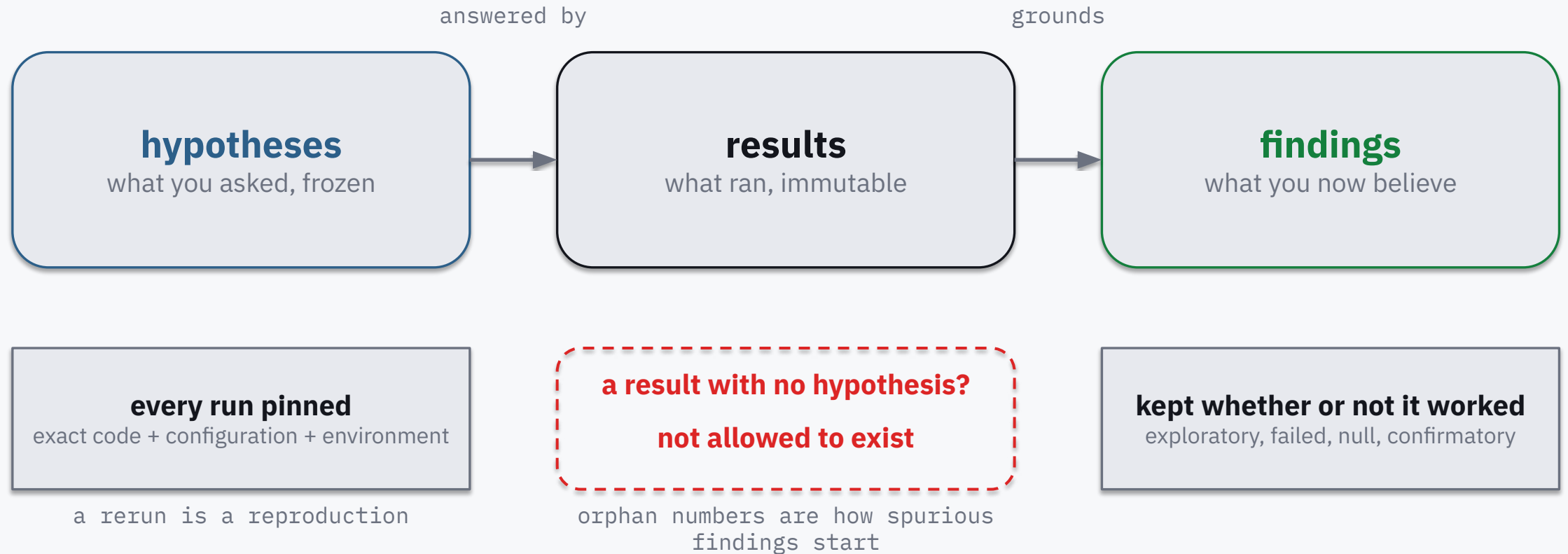
THE TALARIA ARCHITECTURE

THE RESEARCH LOOP



- = loop stage (the five actions)
- = the record
- = supporting action
- = record artifact
- = gate (names what it refuses)
- = /review-panel rides this deposit
- = earned belief (findings)
- = kickback / return (label names the trigger)

Receipts: every idea, run, result, and decision, with its provenance



the lab notebook, inverted: a record the work keeps of itself

A mech suit instead of a replacement

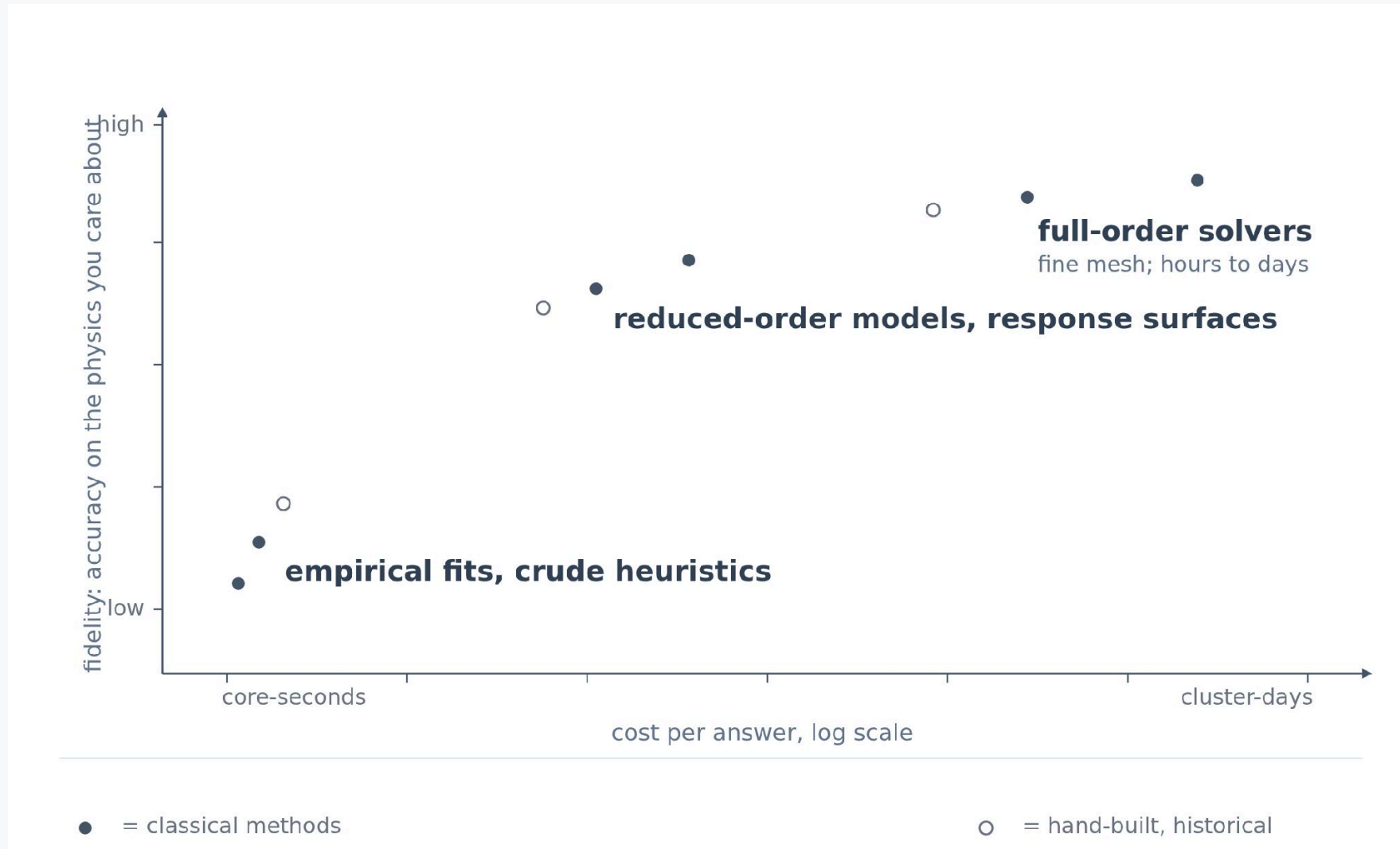
So what can we do with it?

Every field that simulates the physical world keeps a portfolio of models from the full-fidelity solvers it trusts to the cheap surrogates it can afford

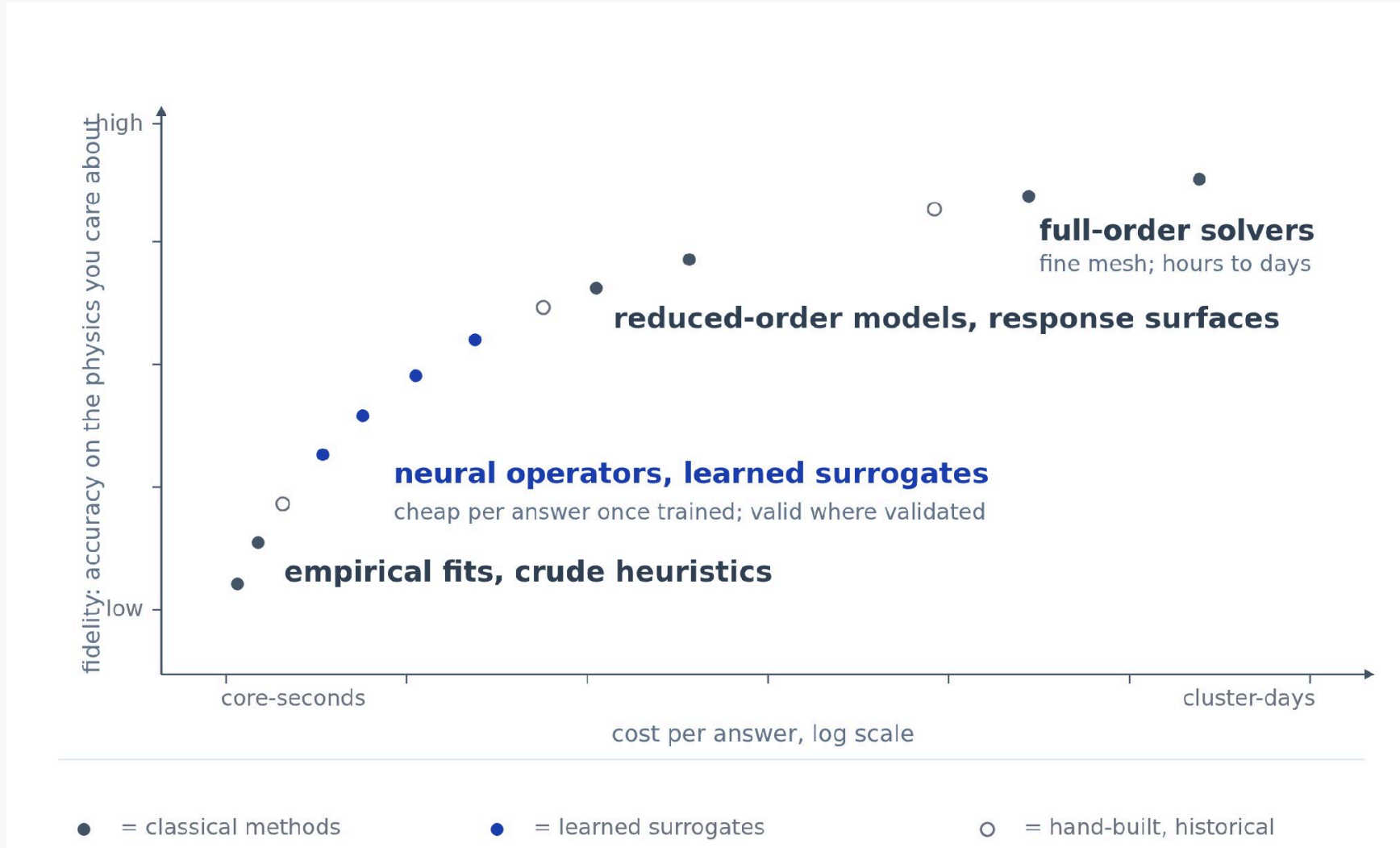
surrogate: any cheap model standing in for an expensive one

fidelity: how faithfully it reproduces the metric you care about, in the regime you care about

The portfolio has always been sparse



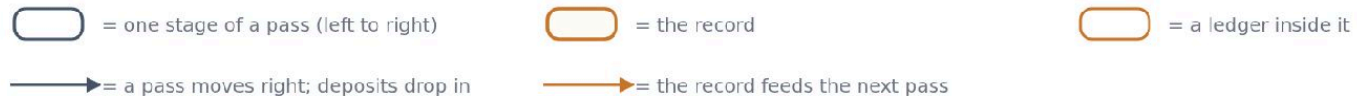
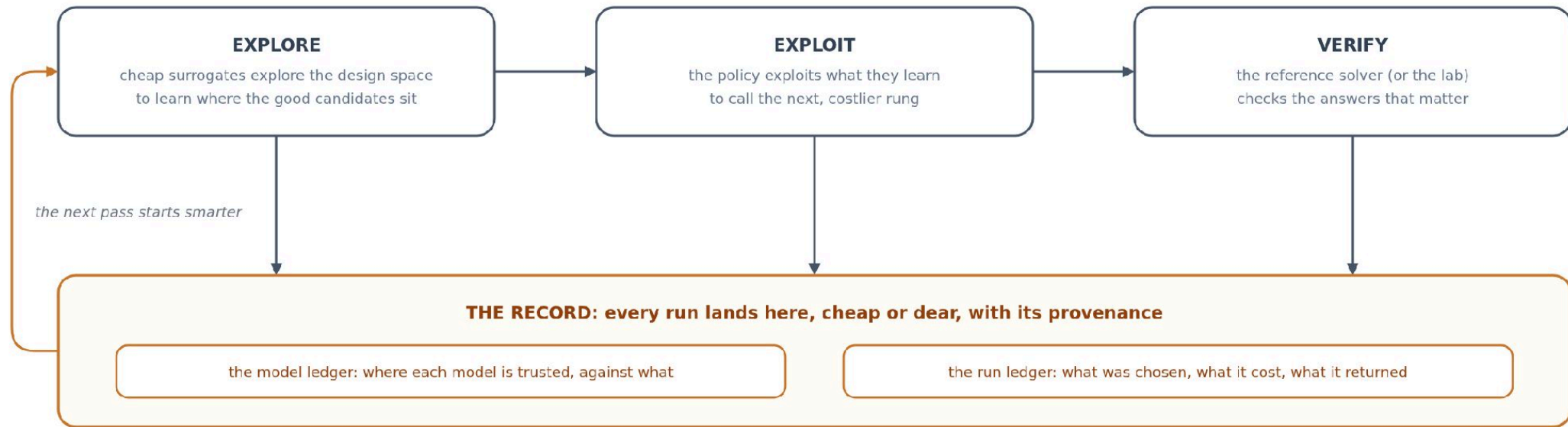
Learned surrogates sit at points the classical stack could not cheaply reach



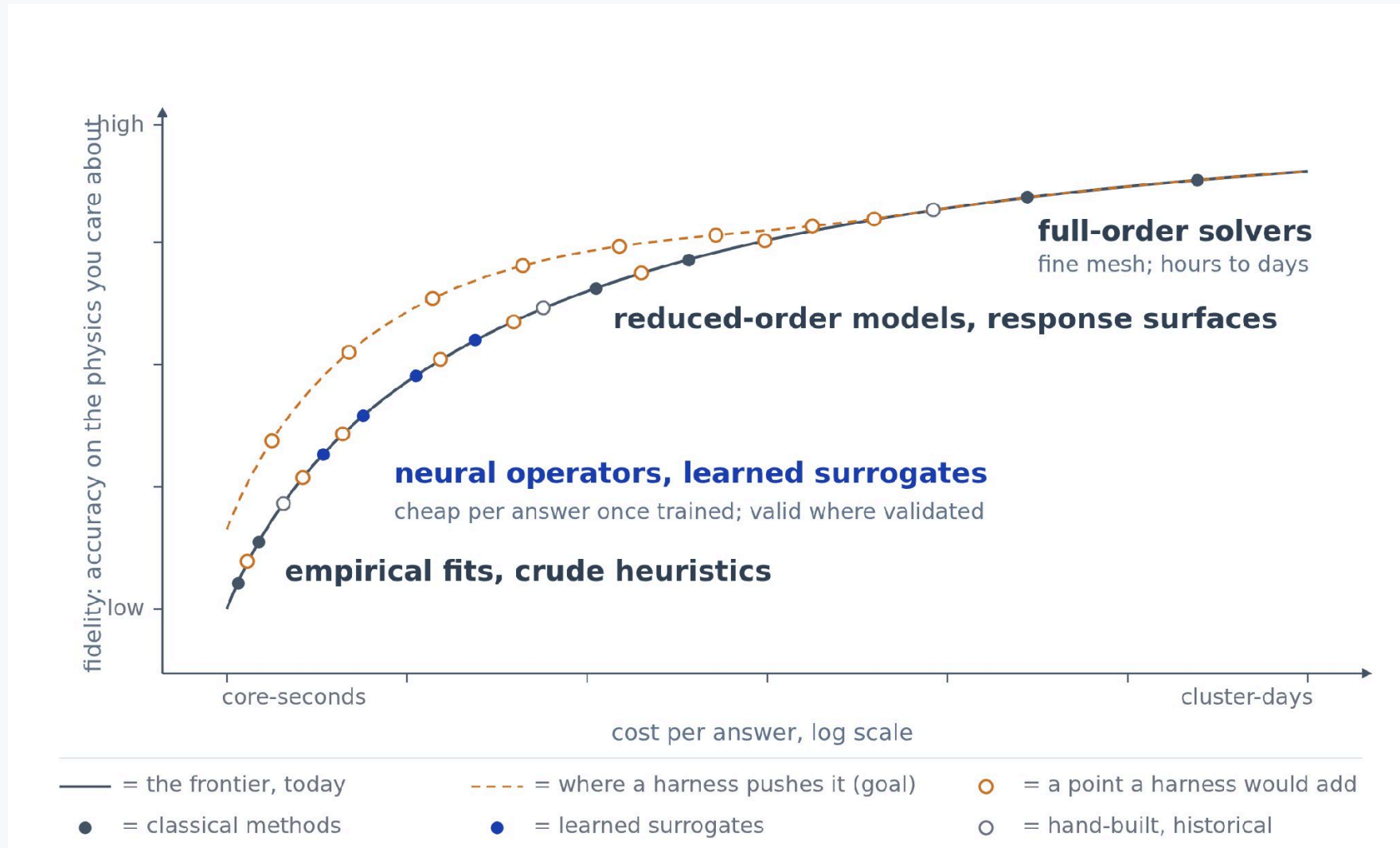
a neural operator learns the solution map across instances; once trained, an answer is one pass through the network

Validation travels with every switch

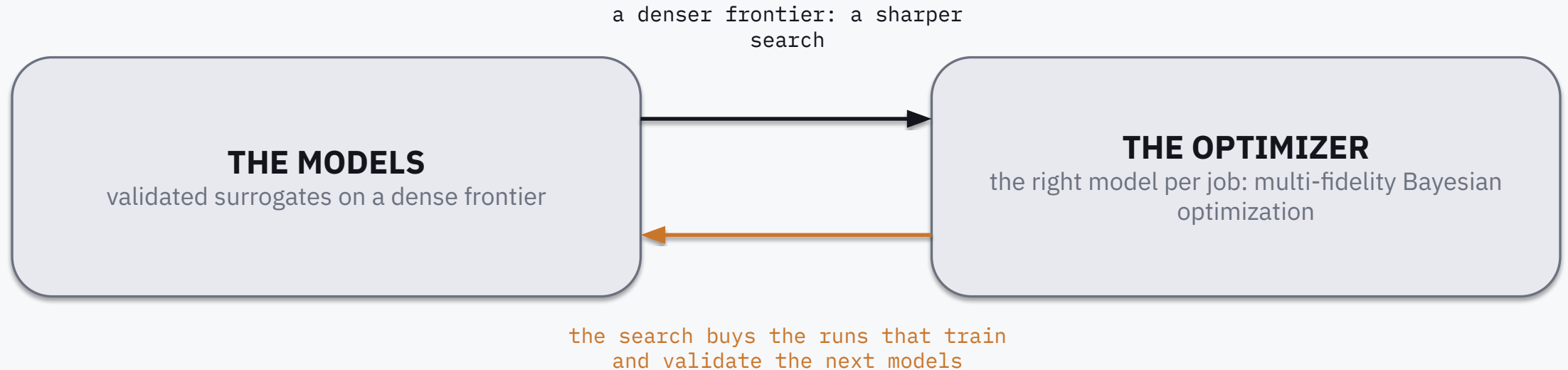
THE MULTI-FIDELITY LOOP



A harness that manufactures validated surrogates cheaply makes the frontier dense



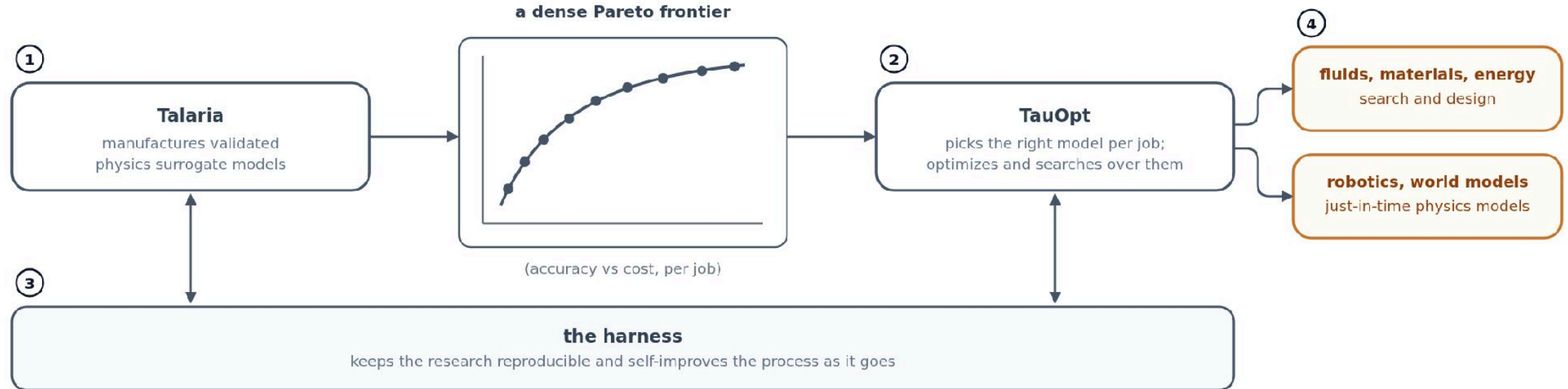
The models and the search improve each other



the policy itself can be learned: over surrogates, acquisition functions, and fidelities

Manufacture the models, densify the frontier, pick the right one per job, keep the record

THE STRATEGY, IN FOUR PARTS



= a product (what we build)



= the Pareto frontier



= a validated model on it



= the harness



= the outputs



= strategy part (see the list)



= feeds



= the harness uses and improves it

Search over surrogates toward what you build next

- **materials discovery**
- **energy**
- **physical AI**

the classic multi-fidelity design search, with a dense set of validated models to search over parameter *and* model space

Built in the open, free for open science

Private beta today

Public, open source, and always free for open scientific work

Public launch at NeurIPS (December 2026)

Thank you

**Researching computational
physics or applied math?**

Running physics simulations?



Private Beta Signups

**Open Source Launch
NeurIPS'26**

Always free for open science



talariasci.com



scottclark.io

Backup slides

for the Q&A

Isn't this multi-fidelity Bayesian optimization?

Yes. Krige 1951: gold-grade statistics (kriging, named later) . Kushner 1964: the first sequential rule . Mockus, 1970s: Bayesian global optimization . Sacks, Welch, Mitchell and Wynn 1989: computer experiments . Jones, Schonlau and Welch 1998: EGO . Kennedy and O'Hagan 2000: multi-fidelity modeling; 2001: calibration . Forrester, Sobester and Keane 2007: multi-fidelity optimization . Snoek, Larochelle and Adams 2012: practical Bayesian optimization for machine learning . Poloczek, Wang and Frazier 2017: multi-information-source optimization

The research agenda here is upstream: automate the construction and validation of a task-specific portfolio, and keep the evidence behind each choice.

PINNs, neural operators, classical solvers: how they sit on the plot

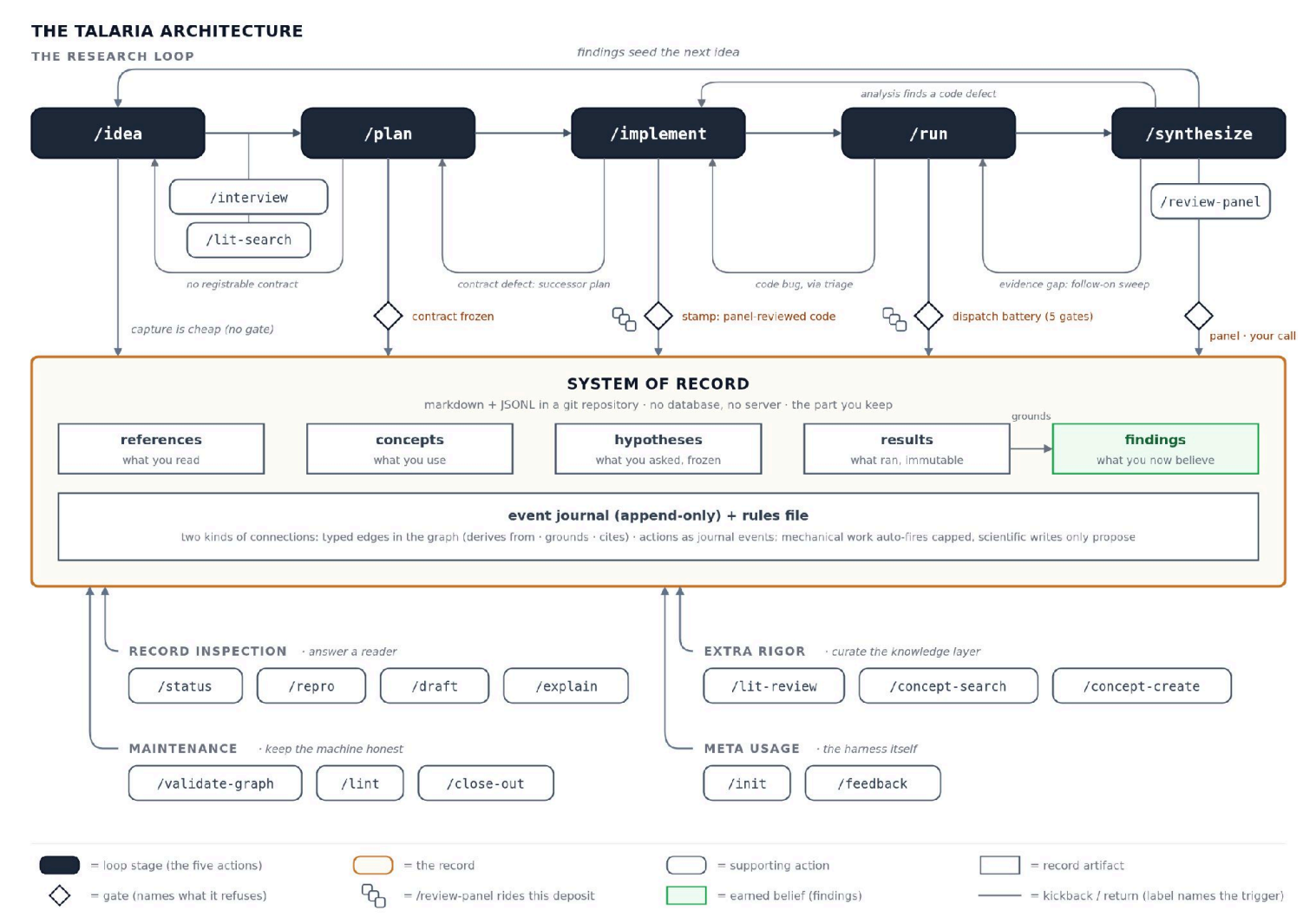
PINN: a network fit to one PDE instance, with the PDE residual plus boundary and initial conditions as the loss

neural operator: a network that learns the solution operator across instances (FNO, DeepONet)

classical solver: the reference; its error is controlled by analysis under stated assumptions, then verified by convergence studies and validated against experiment

on the frontier plot each is a point with a measured error and a cost, valid where validated

What the harness actually does: a research loop on a system of record



Who is the scientist in the loop?

judgment is proposed, never assumed

what was asked, what was checked, what was believed: all on the record

Where uncertainty quantification lives

beliefs about each model's error and cost, updated online: that is uncertainty quantification

run the loop backward and it is an inverse problem: parameters that explain observations

model error has a classical home: Kennedy and O'Hagan's calibration and discrepancy framework (2001)

Where the reference solver is unaffordable

validate at the reachable edge of the regime of use, and carry the belief outward with its uncertainty

trust never extends past measured error: the trust region is the deliverable, not the model

where it matters, pay for the reference run; that is what it matters means

Where the physics goes

into the architecture: structure-preserving operators and compatible discretizations (open research, and this college's lane)

into the loss: residual penalties per domain, PINN-style

into the harness: the invariants become validation checks on every run, so a surrogate that breaks a conservation law never reaches the shelf

How to get access

private beta now: research groups, admitted lab by lab

open source at NeurIPS, December 2026; free for open scientific work, forever

talariasci.com . scott@talariasci.com