

How Better Evals Can Bring Abundance through Accelerated Scientific Discovery

Scott Clark, Founder & CEO

[DISTRIBUTIONAL]



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.**

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

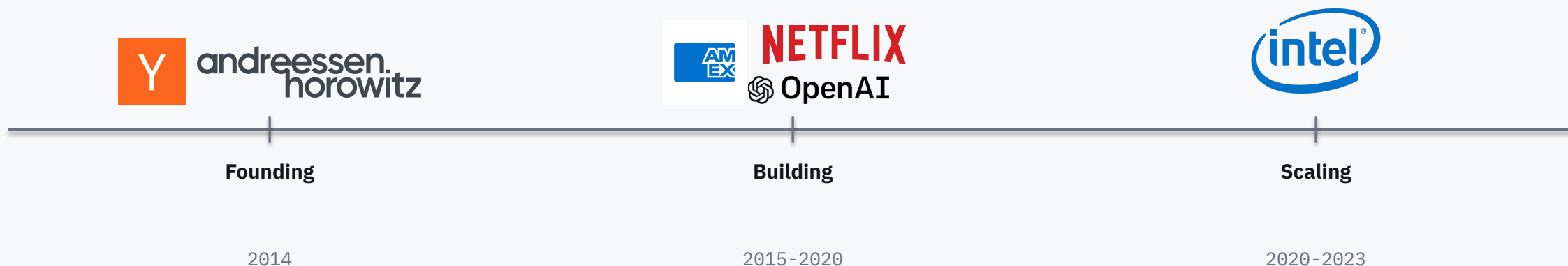
**Whatever the field,
the final step was always parameter tuning.**



Methods for efficient optimization given an eval

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 *any eval* you tell it to

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

We are still a long way from

/goal solve stable fusion containment

Have a problem? Start a startup!

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

distributional



Pre-production AI Testing

2023-2025

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 aren't monitoring for**

distributional

[DISTRIBUTIONAL]

Pre-production AI Testing

Post-production Agent Analytics

2023-2025

2025-2026

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

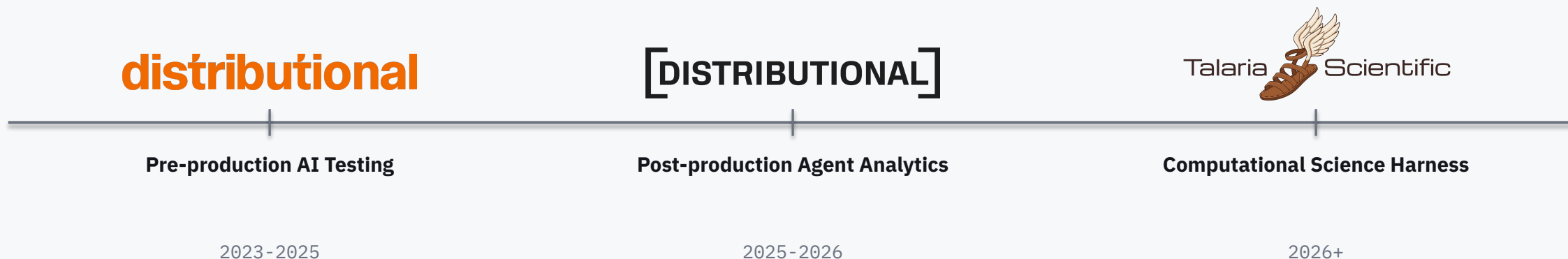
2025-2026

A feature, not a product.

(let alone a business or a startup)

Have a problem? Solve the version you care about

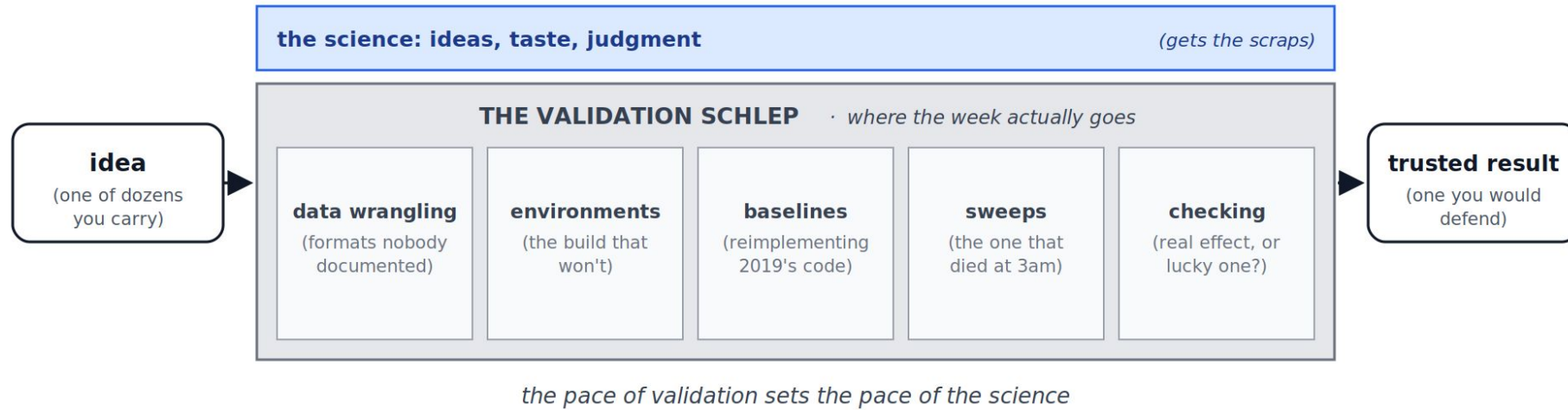
A multi-agent, multi-FM harness for computational science research acceleration.



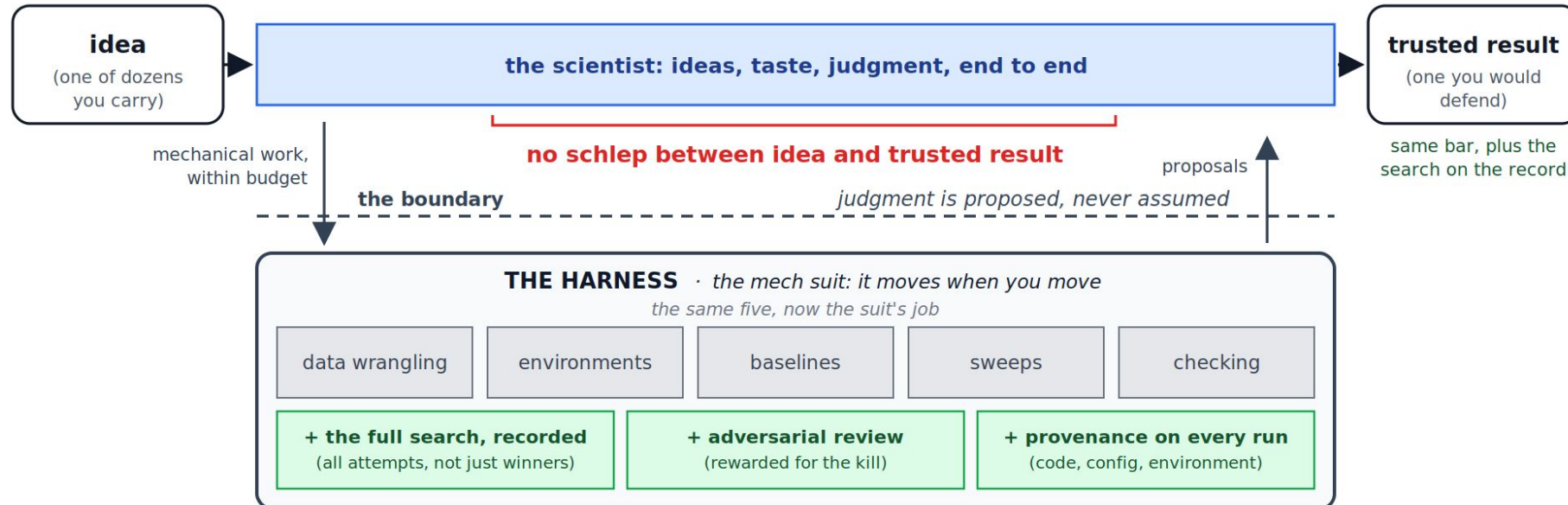
Using domain specific evals and guardrails rooted in physics and math to ensure alignment

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



Thank you

Researching computational
physics or applied math?

Building world models?



Private Beta Signups

Open Source Launch
NeurIPS'26

Always free for open science



talariasci.com



scottclark.io