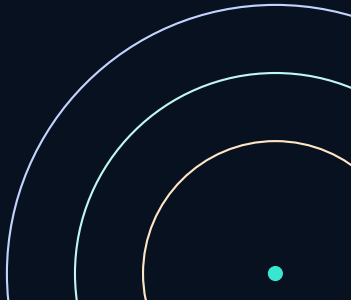
TECHNOLOGY CONFERENCE KEYNOTE

Beyond the Demo

Engineering AI systems that earn trust

SYSTEMS TRACK / 2026



THE INFLECTION POINT

The demo is no longer the hard part.

The hard part is knowing what the system will do after the applause, under real load, for a real person.

Prototype — **Production behavior**

AI reshapes software

CONVENTIONAL SOFTWARE

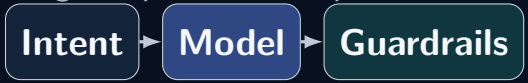
Specification becomes behavior

We debug code-to-output paths against an explicit contract.

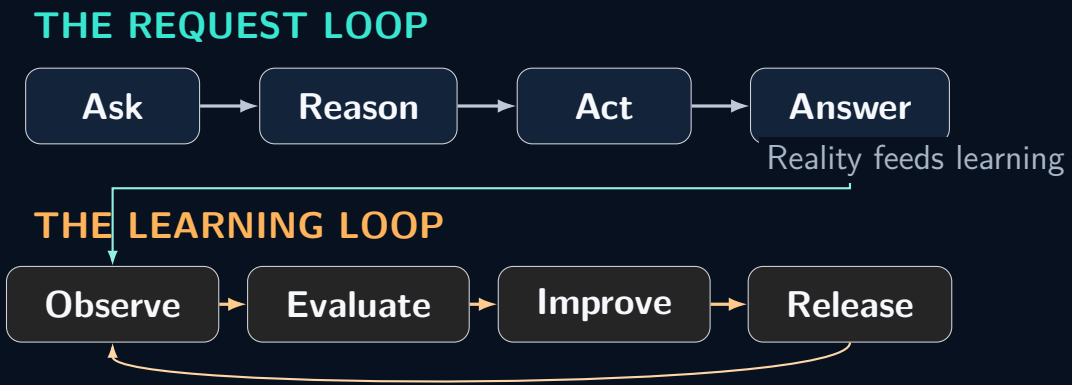
AI-NATIVE SOFTWARE

Intent constrains behavior

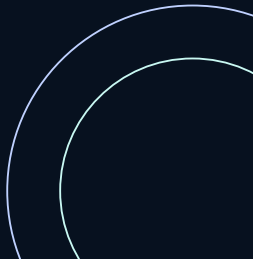
We engineer boundaries around a range of plausible outputs.



Products run two loops



Ambiguity is the first outage.



Make intent executable

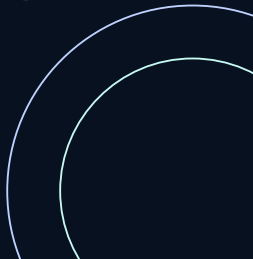
VAGUE REQUEST

“Make the assistant helpful.”

EXECUTABLE CONTRACT

Task	Answer account questions using approved records.
Bounds	Never change data; never infer missing identity.
Evidence	Cite the record and preserve the audit trail.
Escalate	Ask for a human when confidence or authority is low.

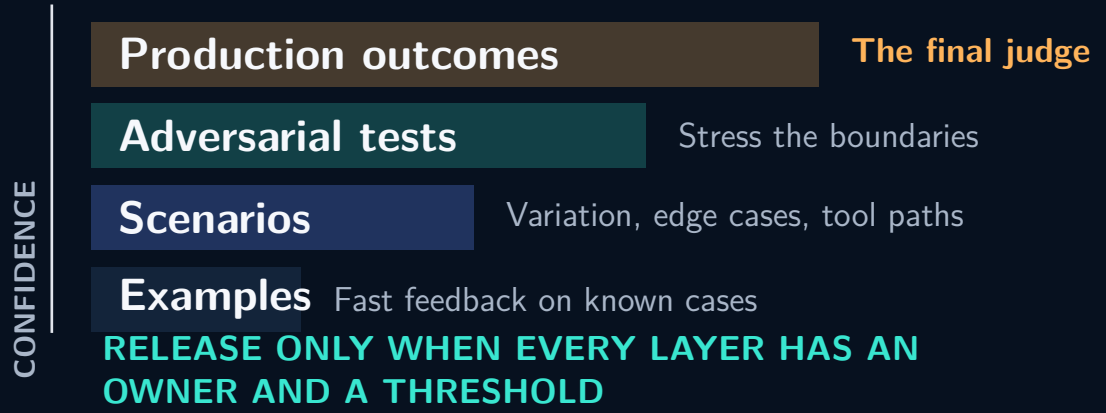
Trust is a property of the whole system.



Four signals reveal trust

- **Correctness** — Did it complete the intended task?
- **Grounding** — Can the answer be traced to evidence?
- **Control** — Did it respect authority and limits?
- **Recovery** — Can a bad outcome be repaired?

Evaluate real behavior



RELEASE ONLY WHEN EVERY LAYER HAS AN OWNER AND A THRESHOLD

Production must explain

A USEFUL TRACE ANSWERS

1. Which context shaped the answer?
2. Which decision spent time, money, or authority?
3. Where did uncertainty enter the path?
4. Can the team reproduce the failure?

Give failure a budget

ONE OPERATING EQUATION

$$\text{risk burn} = \frac{\text{weighted failures}}{\text{allowed failures}}$$

Weight errors by consequence, not by how easy they are to count.

WHEN THE BUDGET BURNS

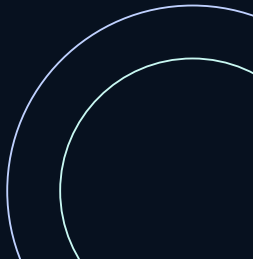
Slow down the rollout

Narrow authority

Escalate review

Freeze the release

A human should have authority, not homework.

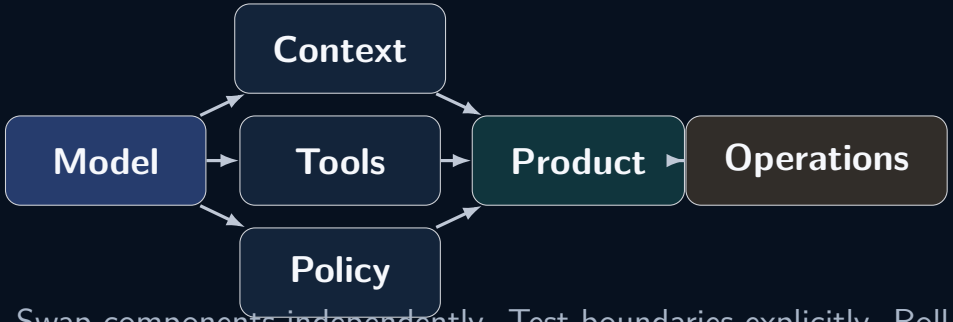


Human intervention

Auto	Confirm	Approve	Human-only
Low impact, reversible, observable.	System proposes; person confirms.	Evidence appears before commitment.	Irreversible or sensitive actions.

Match the checkpoint to the cost of being wrong.

Reduce the blast radius



Swap components independently. Test boundaries explicitly. Roll back without rebuilding the world.

Ship one thin vertical slice

ONE USER ▪ ONE JOB ▪ ONE SAFE ACTION

- **Observe the current work** — Capture real inputs
- **Define the contract** — State success and limits
- **Instrument before scale** — Trace decisions and outcomes
- **Expand only with evidence** — Add scope, then autonomy

The operating loop

release-loop.py

```
contract = define(task, bounds, evidence, escalation)
candidate = system.run(request, contract)
trace = observe(candidate)
score = evaluate(trace, contract)
if score.within_budget():
    release(candidate)
else:
    narrow_authority()
    learn(trace)
```

Six weeks to evidence

The milestone is not a launch. It is a decision backed by evidence.

Questions before scale

01 What behavior are we willing to own?

02 What evidence would stop this release?

03 Who can intervene, and how quickly?

04 What will production teach us next?

THE STANDARD FOR WHAT COMES NEXT

Ship evidence, not confidence.

Define the contract. Instrument the path.
Earn autonomy one verified outcome at a time.

BEYOND THE DEMO