



LEAN CULTURE EVENT

From Lab to Market:

35 Years of Closing the AI Gap

Why AI Pilot Projects Fail — What to Do Before You Start One

Evan Steeg — AI Researcher, Biotech Entrepreneur & Innovation Strategist

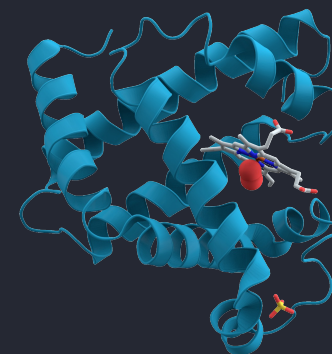
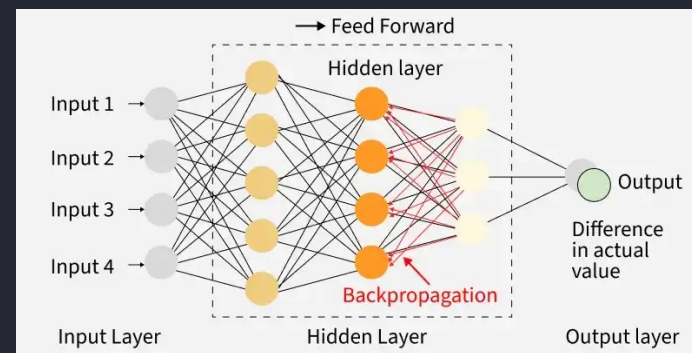
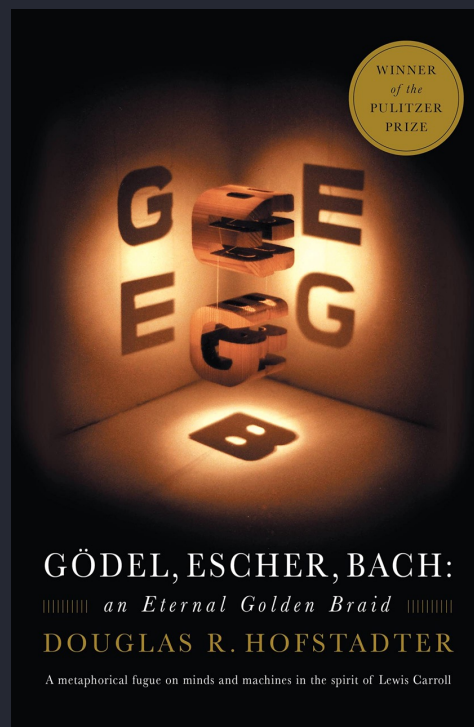
Hosted by Theresa S. and Sean M.



Where It Started for Me, in 1982

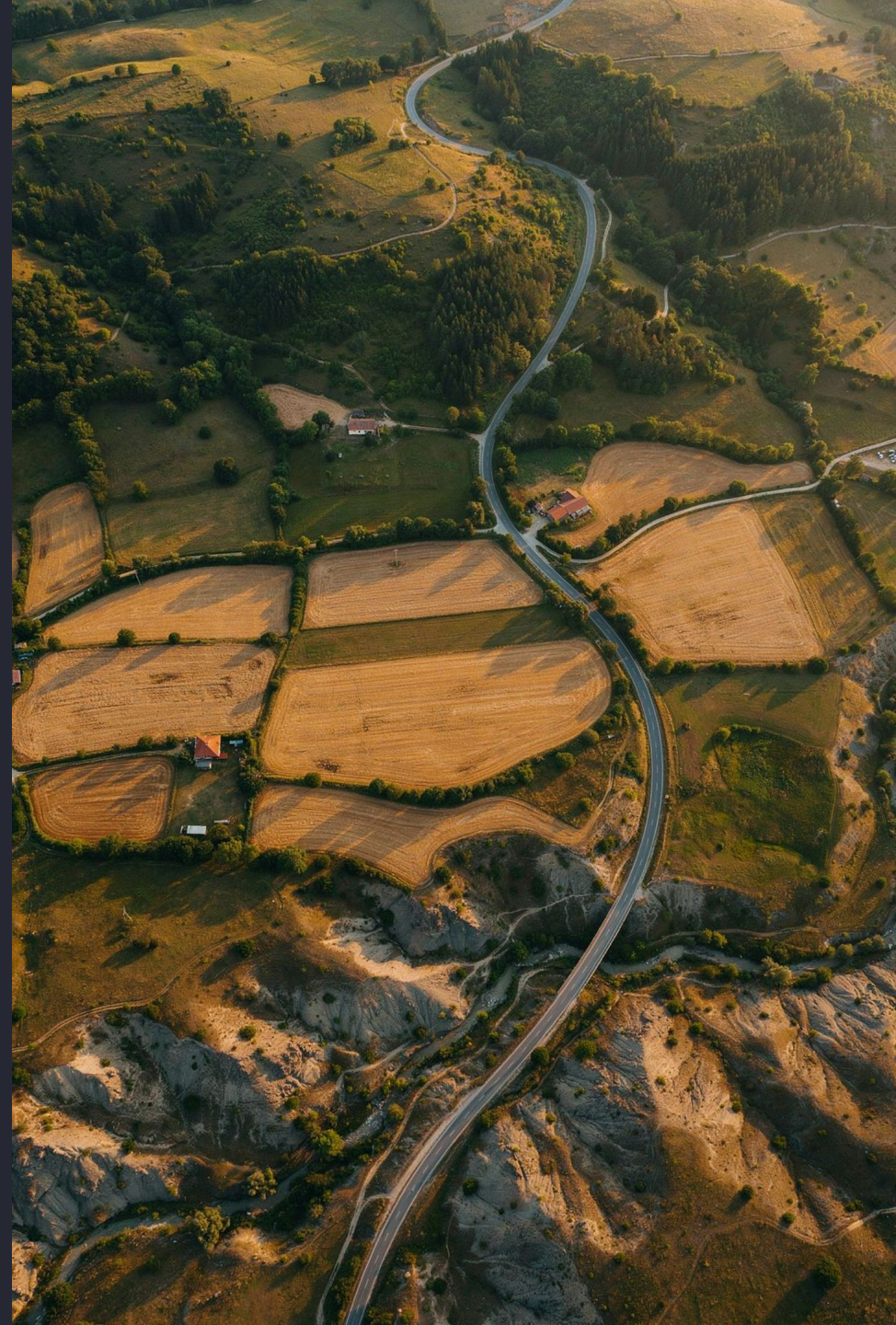
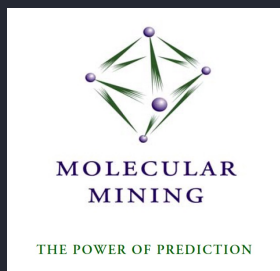
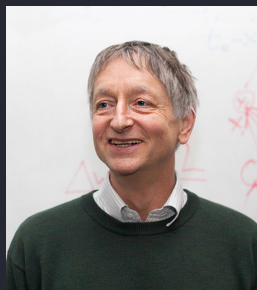
I read *Gödel, Escher, Bach* on vacation as an undergraduate and switched majors because of it. That eventually took me to the University of Toronto, specifically to get into Geoffrey Hinton's lab — long before "Nobel laureate" was attached to his name.

There, I decided to apply ML to molecular structure prediction.



35 Years in Academia, Startups, Government, Consulting... One Recurring Question:

Why don't lab results survive contact with
the real world?



Why Biology Cracked AI & Vice Versa

It didn't work in 1988. (I tried!) AlphaFold cracked it, spectacularly, in 2021.

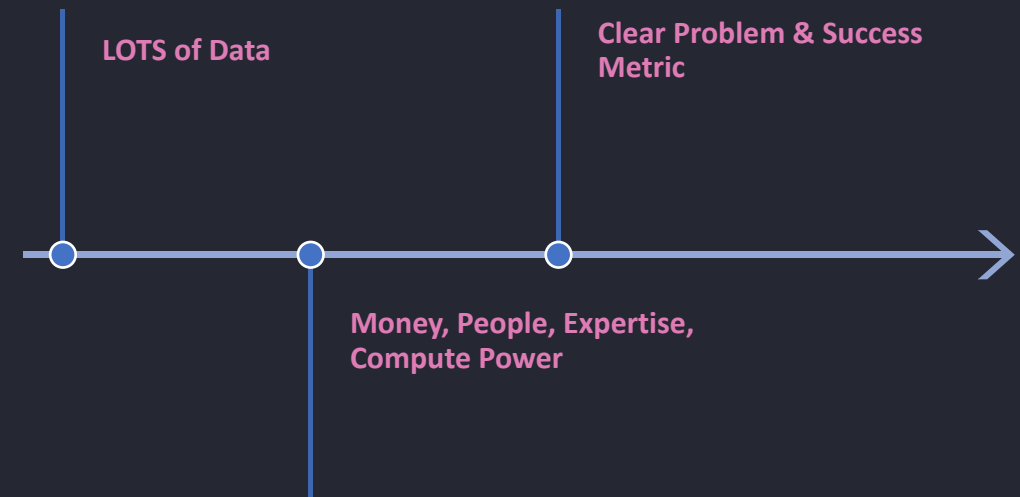
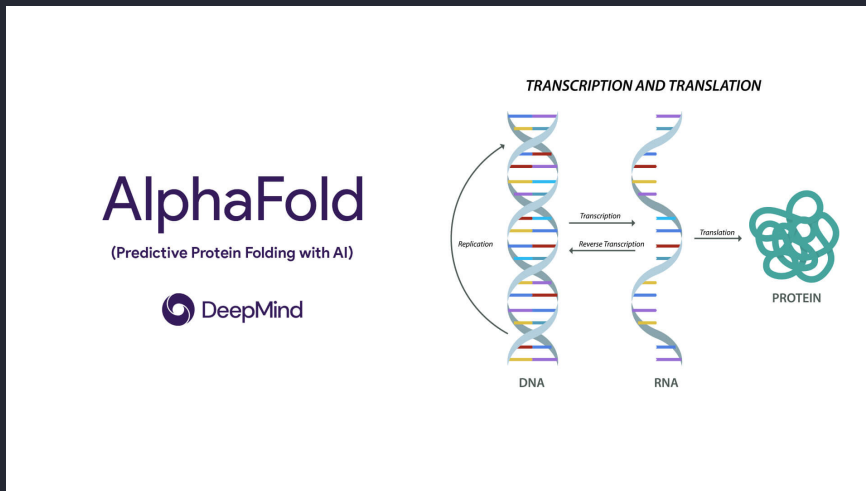
Three reasons: 20+ years of public structural data in the Protein Data Bank, a precise problem with an unambiguous benchmark (CASP — no proxy metrics), and an open-source release that was as strategically significant as the breakthrough itself.

Fast-forward to September 2026:

Anthropic's biology lab set Claude loose as a swarm of roughly 950 AI agents. In 21 hours, surveying 1.9 billion protein clusters, it surfaced a previously unknown enzyme system with CRISPR-like repeats. Anthropic's own caveat: they don't yet know what it does.

Biology had the formula: Lots of data, lots of money, lots of talent. (Finance has the same formula).

That's why "AI transforms everything equally and immediately" is the wrong story.





Startups: The Moment We're In

Building with AI has never been more accessible. The raw materials of a production-grade AI system are now free, open, and within reach of a two-person team on a bootstrap budget.

HuggingFace

1M+ models and 250k+ datasets —
free to use, ready to deploy

GitHub

Complete production AI stacks from
Meta, Mistral, Google, and thousands
of open contributors

Claude Code, Codex, Cursor, etc....

Go from idea to working prototype in a single afternoon



So why is building a **successful** AI startup still hard?

Speed Is Not Strategy

AI dev tools answer **how to build** — they don't answer **what to build** or **why it will win**.
You can have a demo by Friday. So can your competitor.

Product-Market Fit

Still requires talking to real customers.
No AI tool shortcuts genuine discovery.

Demo ≠ Product

Security, edge cases, UX, and user trust
separate a demo from a deployed,
defensible product.

The Operating Model

How does AI change who you hire, how you price, how you market, and what you
can actually do at scale? That's the real question.



Real-World AI Failure Story:

IBM Watson for Oncology

IBM Watson for Oncology failed to revolutionize cancer treatment, offering valuable implementation lessons.



Ambitious Technology

Developed to analyze medical literature and recommend cancer treatments to oncologists.



Implementation Challenges

Low hospital adoption rates with instances of unsafe or inappropriate treatment recommendations.



Lessons Learned

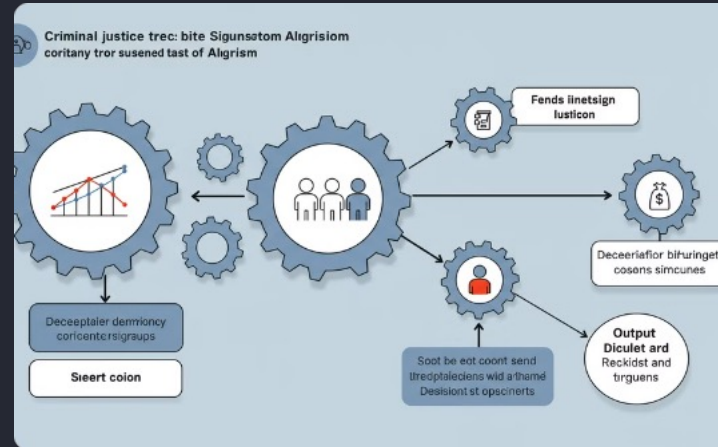
Demonstrates the critical need for rigorous testing and expert collaboration in healthcare AI.

AI Failure: COMPAS Recidivism Prediction



Ambitious Technology

COMPAS was designed to predict criminal reoffending probability, with its assessments influencing judicial sentencing decisions.



Implementation Challenges

A 2016 ProPublica study found the algorithm falsely flagged Black defendants as future criminals at twice the rate of white defendants, revealing significant racial bias. And it performed no better than humans in terms of predictive accuracy.



Lessons Learned

This case underscores concerns about algorithmic bias and accountability when deploying AI in high-stakes contexts like criminal justice.

Why AI Pilots Fail

Wrong Problem

AI applied because the tool is exciting, not because the problem demands it.

Data Isn't Ready

Garbage in, garbage out, at enterprise scale and speed.

No Change Management

Tools deployed, adoption never happens.

No Measurement, Or Measuring the Wrong Things

Nobody defined success before launch, and most teams measure the wrong thing: the model scores well internally on proxy data, but the customer cares about something else entirely.

Governance Friction

Teams route around slow approval processes; or give up in frustration.

Scaling Too Fast

Pilot conditions don't survive contact with the real world rollout.

The antidote: the right problem, the right requirements, the right consultation, the right follow-through. That is, an AI STRATEGY.

Practical Roadmap to AI Strategy: Define Goals, Plan, Budget, Responsibilities

The same seven steps apply whether you're a 5-person shop or a 500-person plant — just different scale of effort at each stage.



The AI Readiness Framework — Key Ideas

Assess readiness before launching a major initiative, not after it stalls.

Data Readiness

Is it clean, accessible, and itself governed?

Infrastructure Readiness

Do your systems support secure AI integration?

People Readiness

Do staff have the skills and the judgment to use or to override AI when needed?

Process Readiness

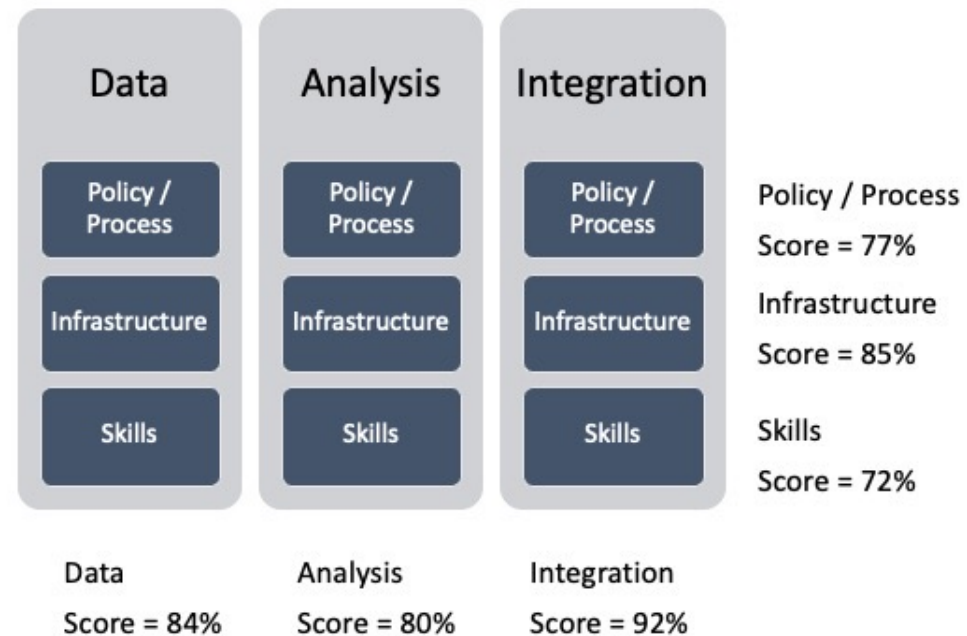
Are workflows built to act on AI output, with clear accountability?

Governance Readiness

Are policies and review mechanisms in place and understood?

The AI Readiness Framework — A Real Example from UAE

An AI Readiness Assessment (AIRA) for a Large Enterprise



How to Measure Real AI Value

ROI is rarely a single number. Think across four categories — and set baseline metrics **before** deployment, not after. Measure at 30, 90, and 180 days. Include qualitative signals alongside quantitative ones.

<u>Category</u>	<u>What to Measure</u>
Efficiency	Hours saved, error rates reduced, cycle times shortened, manual steps eliminated
Revenue	New products or services enabled, faster time-to-market, improved conversion rates
Risk Reduction	Compliance incidents avoided, fraud detected, decision quality improved, audit findings resolved
Strategic Optionality	Capabilities built that enable future moves — platforms, datasets, and skills that compound over time

- ☐ Include qualitative signals: staff confidence levels, client satisfaction, and board comfort with AI topics. These leading indicators often predict quantitative outcomes before the numbers move.

Where Does Your Organization Sit? AI Maturity Index

(Note also: Lean, Kaizen, Six Sigma, ISO)

1

A — Exploring

Ad hoc tool use, no policy, limited organizational awareness

2

B — Experimenting

Pilots underway; some informal guidelines beginning to form

3

C — Operationalizing

Approved use cases, governance in place, staff training underway

4

D — Scaling

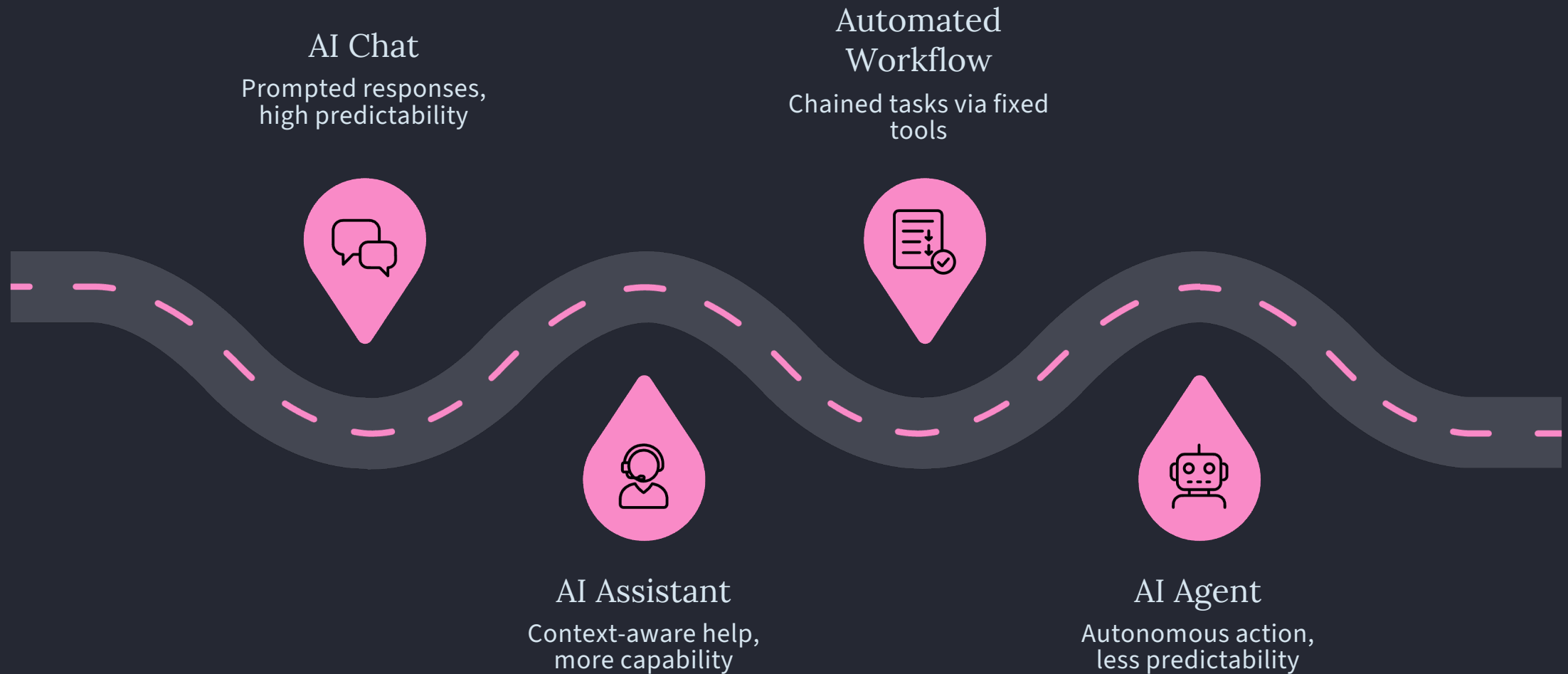
AI embedded in core workflows, outcomes measured and continuously improved

Productivity: Individual & Immediate vs. Organizational

AI generates many possibilities.
But real-world validation takes time and human effort.



AI Use is Evolving: From Getting Advice to Managing a Brilliant, Unpredictable “Employee”





Well-Harnessed AI Agents & Workflows

The Reliability Problem

Real-world AI agents fail to correctly follow instructions roughly 25% of the time in independent testing.

The Solution

You don't fix a flaky agent with a better prompt — you wrap it in a harness.

- fixed tool registry
- swappable models
- deliberate context management
- hard guardrails on tool calls and retries
- logged agent loop
- verification against the trace, not against what the model claims happened

*"While the whole world was having a big old party, a few outsiders and weirdos saw what no one else could... These outsiders saw the giant lie at the heart of the economy, and they saw it by doing something the rest of the suckers never thought to do: **They looked.**"*

- The Big Short

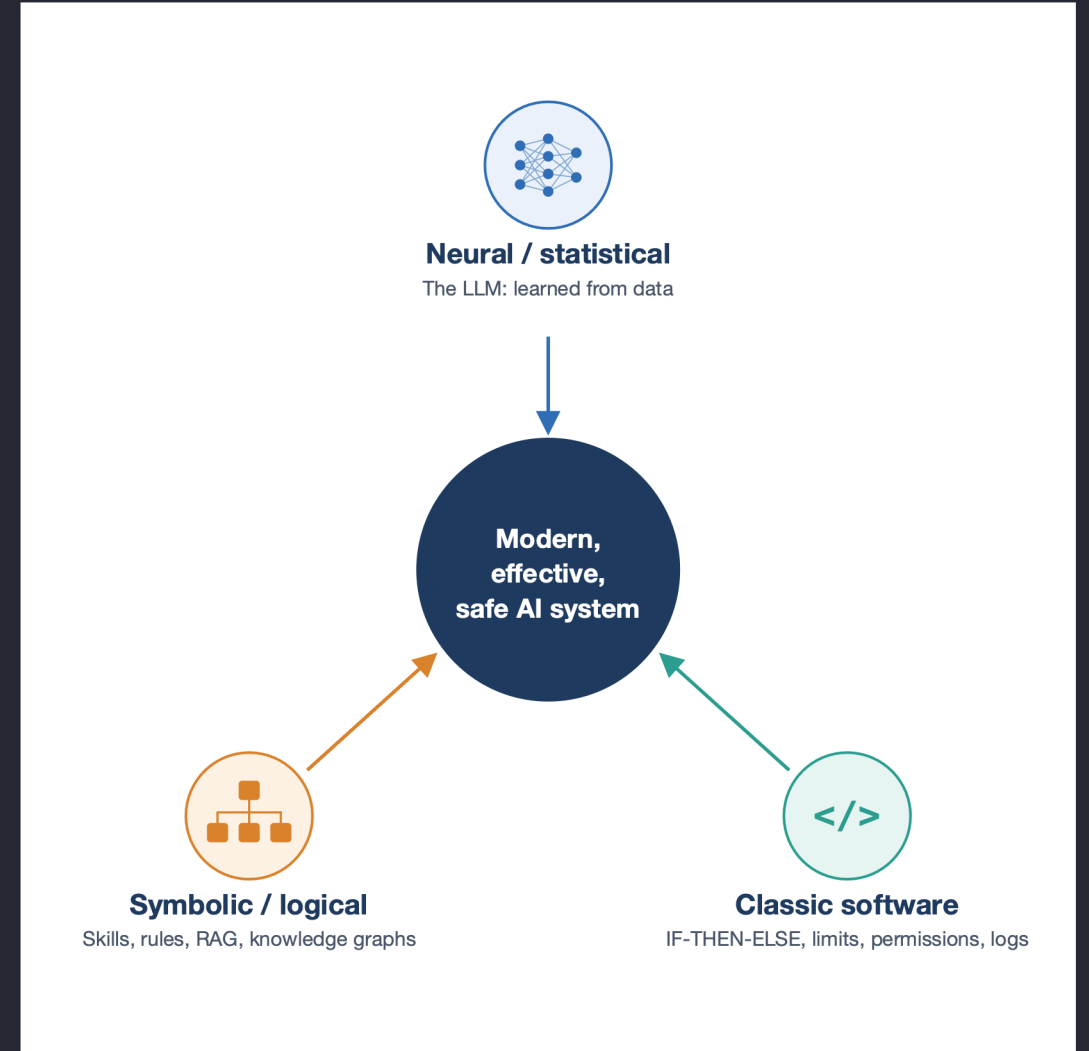
Managing — AI or anything else — is largely about
PAYING ATTENTION.

Coming Full Circle: Neural AND Symbolic AND Non-AI

The three old camps from the 1990s AI wars — neural networks, symbolic rules, classic deterministic software — aren't competing anymore.

They're all alive, together, inside a well-harnessed LLM-based AI / agentic system.

This is the Way.



Thank You.

Questions?

evansteeg1@gmail.com · www.linkedin.com/in/evansteeg/