The GenAI Factory

How Generative Intelligence Redefines Smart Manufacturing

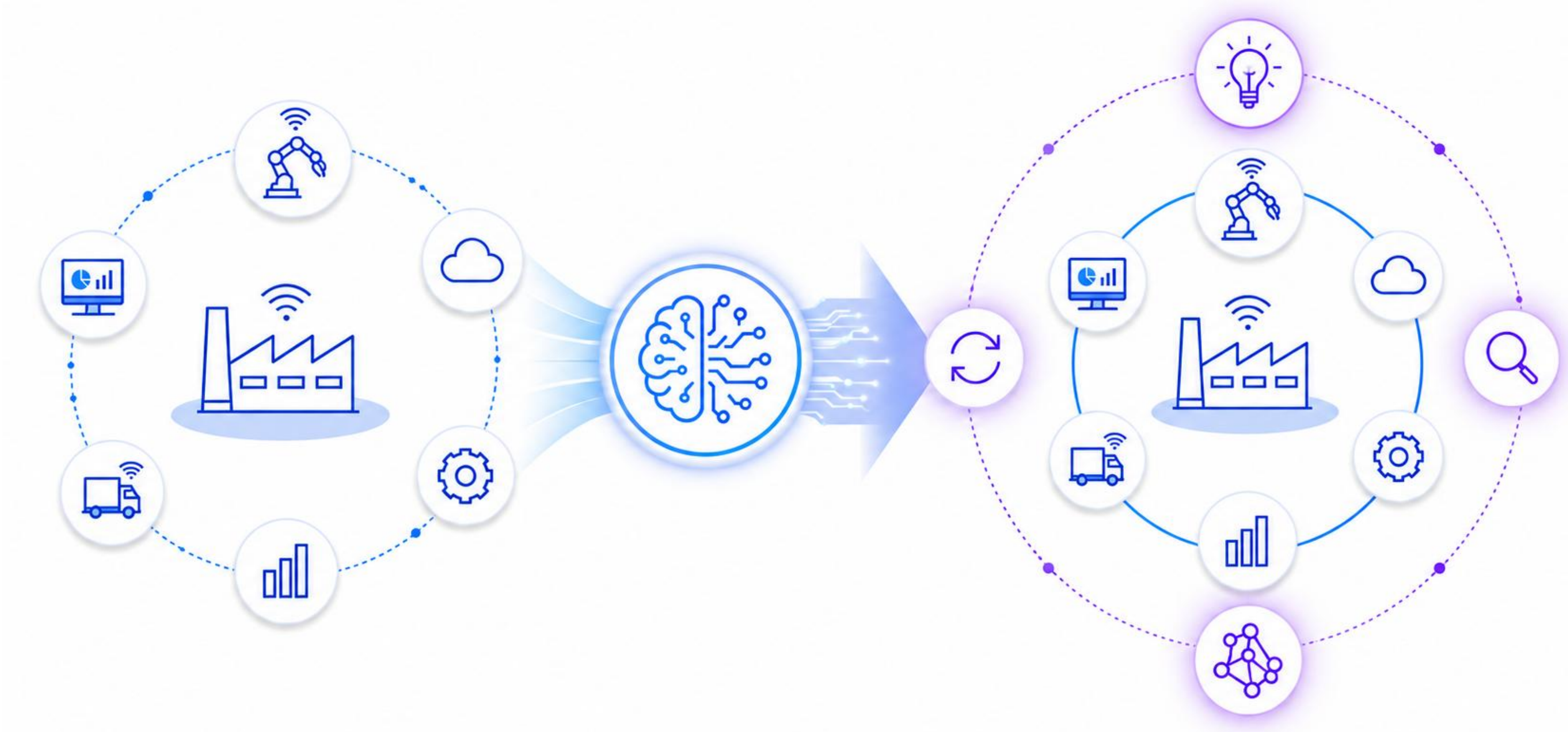
José Barata

Full Professor, NOVA School of Science and Technology ·
Leading Researcher , RICS — Robotic & Industrial Complex Systems, UNINOVA-CTS

jab@uninova.pt · rics-uninova.com

Industry 4.0 connected factories.

Generative AI allows them to imagine, explore, explain and adapt.



This is not an incremental improvement. **It is a paradigm shift** — *and this talk will defend that claim with evidence, not adjectives.*

Why ‘paradigm shift’ is the correct term

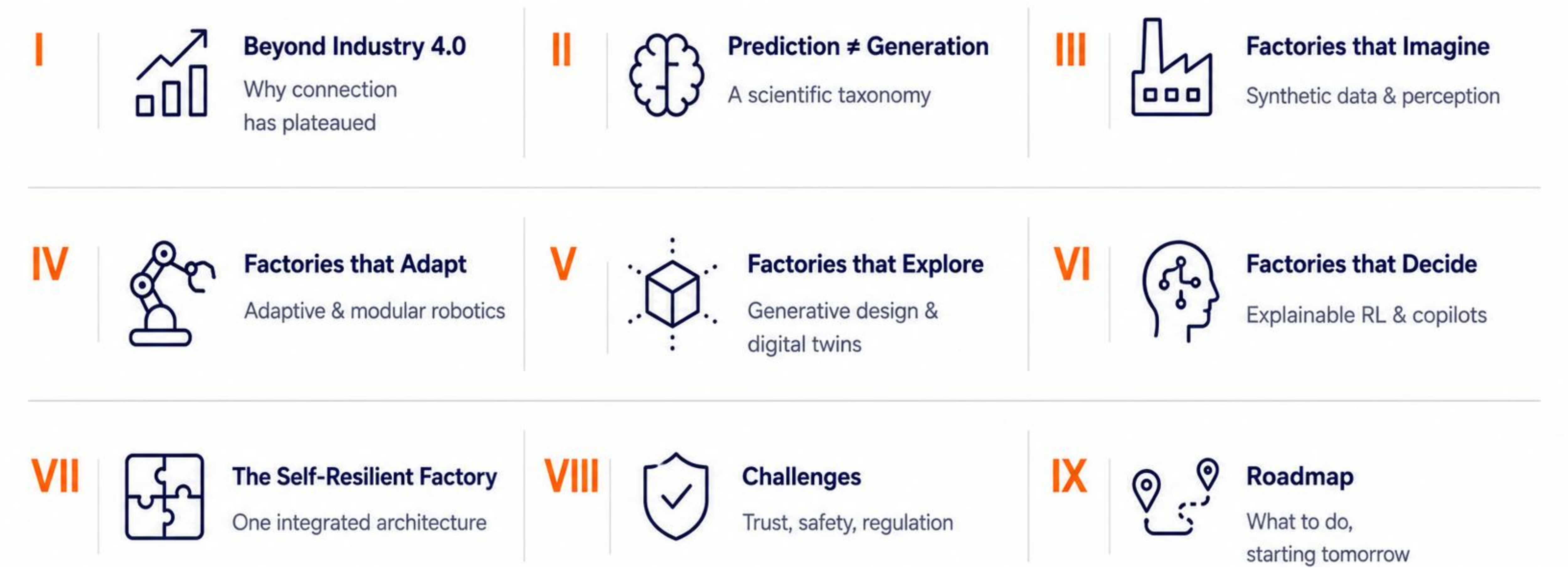
A paradigm shift is not ‘better technology.’
It is a change that propagates all the way down to how an industry thinks.

Technology changes
↓
Engineering methodology changes
↓
Operator role changes
↓
PARADIGM SHIFT → **Knowledge representation changes**
↓
Decision-making changes
↓
Business model changes

Every technology claiming to be a ‘paradigm shift’ should be able to complete this chain.
Most AI-in-manufacturing pitches stop at line one.

Nine acts, one argument

Each act answers one question. Together, they build the case for a new industrial paradigm.



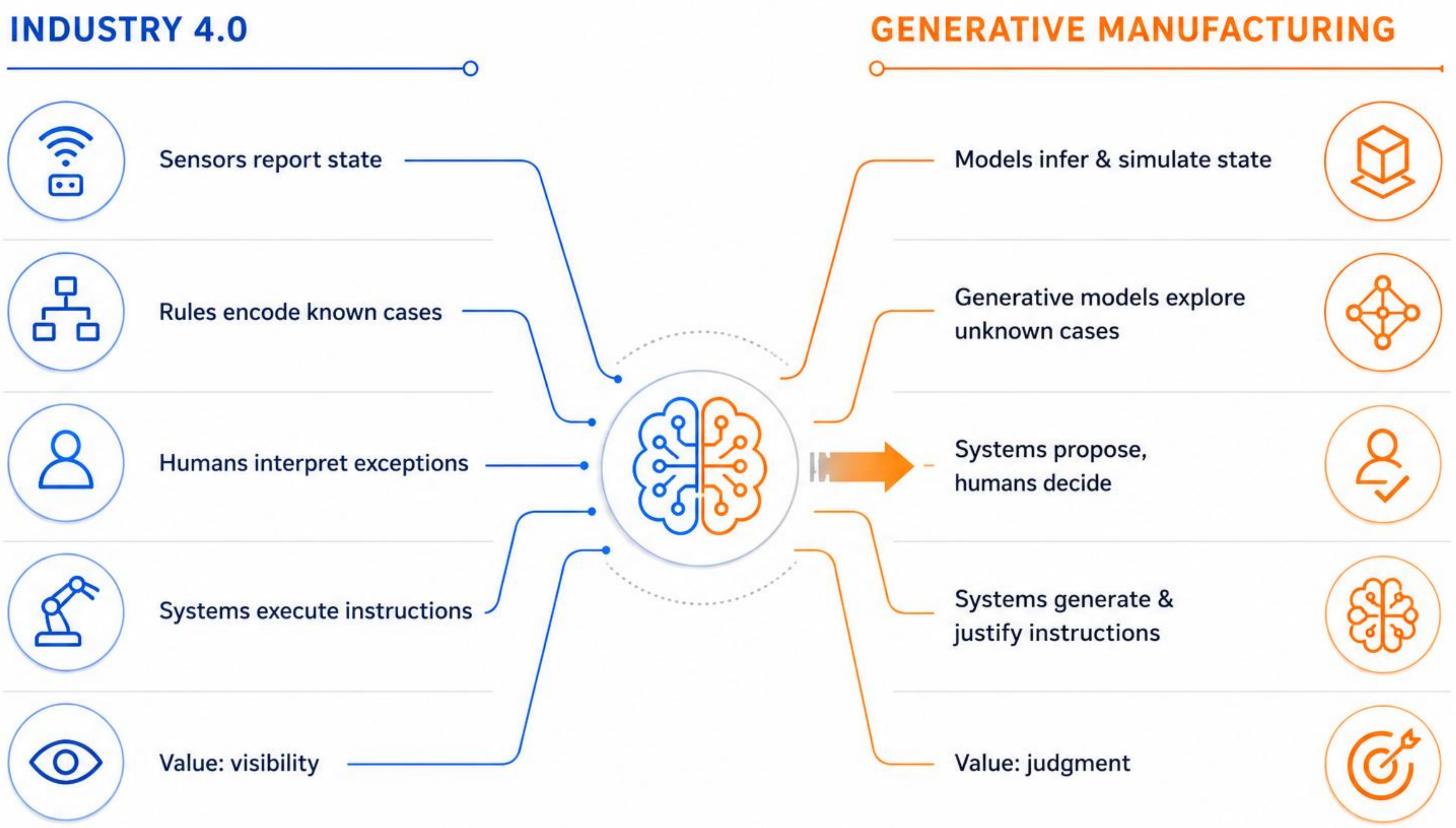
Beyond Industry 4.0

ACT I

Why the current paradigm has reached its limits

Connection is not Cognition

Industry 4.0's contribution was infrastructural. Generative AI's contribution is cognitive. These are different categories of value.



More sensors do not mean more understanding

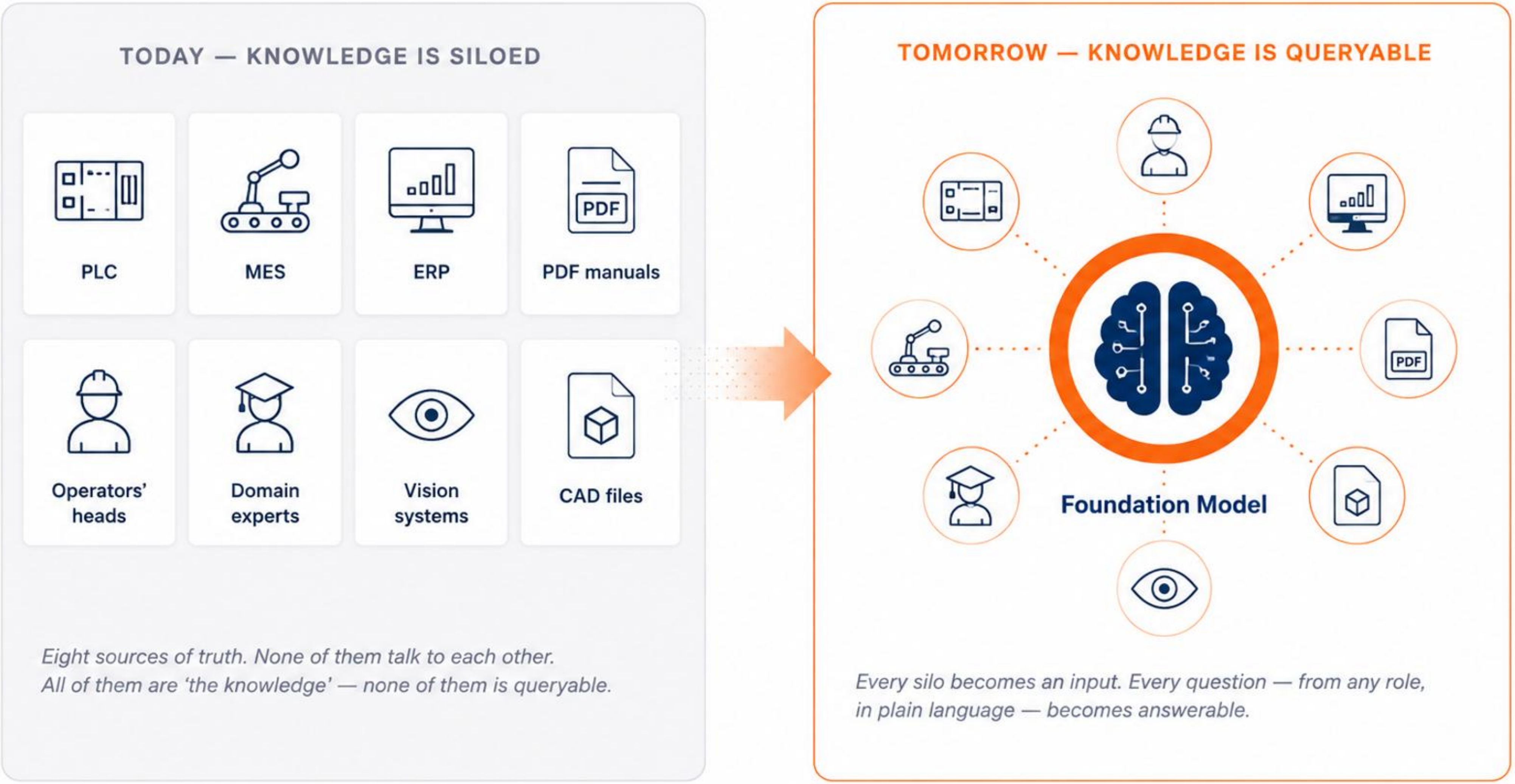
Volkswagen Autoeuropa — robotic body-shop line, fully connected, fully instrumented.

A decade into Industry 4.0, most plants are saturated with sensors.

More data did not translate linearly into better decisions; it translated into more dashboards for humans to interpret.

The bottleneck moved from <u>data acquisition</u> to <u>data interpretation</u>, and <u>interpretation is precisely what generative models are built for</u>.

Where industrial knowledge actually lives



Prediction is not Generation

ACT II

The scientific difference between predictive and generative AI

Where generative models sit in the AI landscape

This talk is about this half.

ARTIFICIAL INTELLIGENCE

MACHINE LEARNING

Discriminative / Predictive Models

Learn a decision boundary.

Answer: “what class / value?”

e.g. classification, regression, forecasting

Generative Models

Learn the underlying data distribution.

Answer: “what could plausibly exist?”

e.g. synthesis, design exploration, simulation

Inside the generative branch

FOUNDATION MODELS

Large models pre-trained on broad data,
adapted to specific industrial tasks.

LLMs

Language & reasoning
over logs, specs,
sensor summaries.

01

Diffusion Models

Iterative denoising —
image & signal synthesis,
rare-defect generation.

02

GANs

Adversarial generation —
realistic synthetic samples
for data-scarce classes.

03

Predictive AI vs. Generative AI

Two approaches. Different mindset. Different impact.

DIMENSION	PREDICTIVE / DISCRIMINATIVE AI	GENERATIVE AI
Learns	$P(y \mid x)$ — a decision boundary	$P(x)$ — the data distribution itself
Typical output	42.7 A label, a value, a forecast	A sample, an artifact, a scenario
Objective	Minimize prediction error	Maximize likelihood / sample fidelity
Example task	"Will this weld fail?"	"Show me weld failures we haven't seen"
Manufacturing use	Predictive maintenance, QA classification	Synthetic data, generative design, digital twins

VS.

Factories that Imagine

ACT III

Multimodal fusion, synthetic data, and industrial perception

Multimodal fusion for industrial inspection

No single sensor tells the whole story. Fusion architectures combine modalities the way an expert inspector combines sight, sound and experience.

Vision
Surface, geometry, alignment

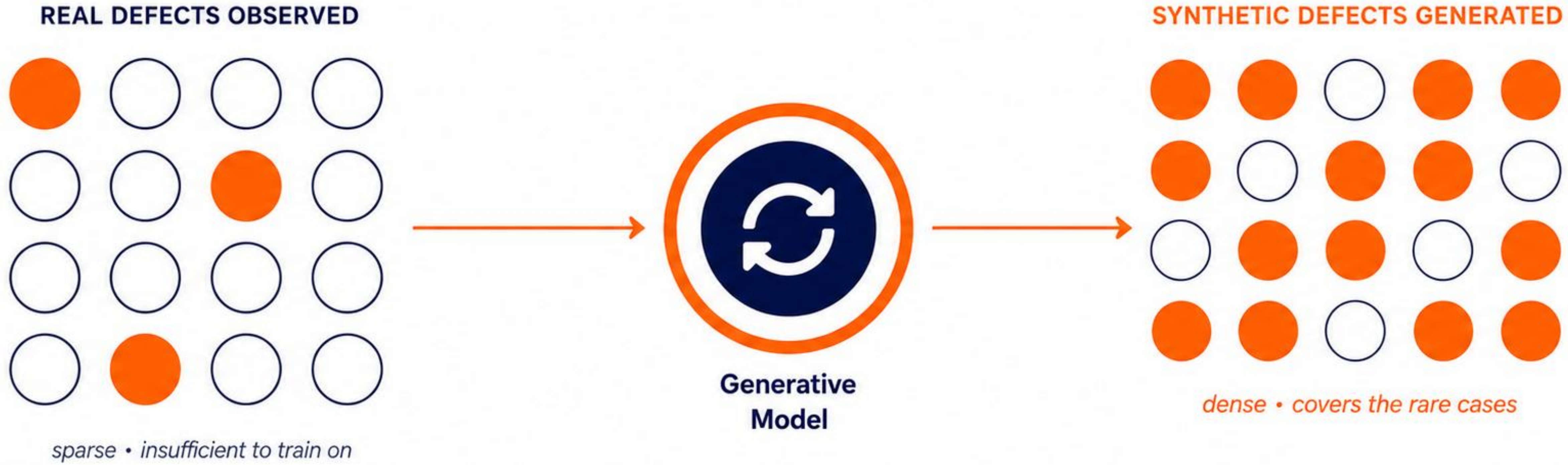
Acoustic
Sound signature, vibration

Sensor / X-ray
Internal structure, process data

Fused Representation (Foundation Model)
Cross-modal understanding captures complementary signals into a unified, rich representation.

Decision + Explanation
Actionable decision with interpretable reasoning across all modalities.

Synthesizing the defects you have not seen yet



Early defect detection at industrial scale

18,000+

vehicles' worth of measurement data used to train dimensional-defect detection models across a multistage production line.

Extending this with generative synthesis means the detector is no longer limited to the failures the line happened to produce.

Framing — Tailgate Assembly — Tailgate Alignment — Finish Line

Early prediction of defects at each stage avoids propagation downstream — the standard argument for shifting inspection left.

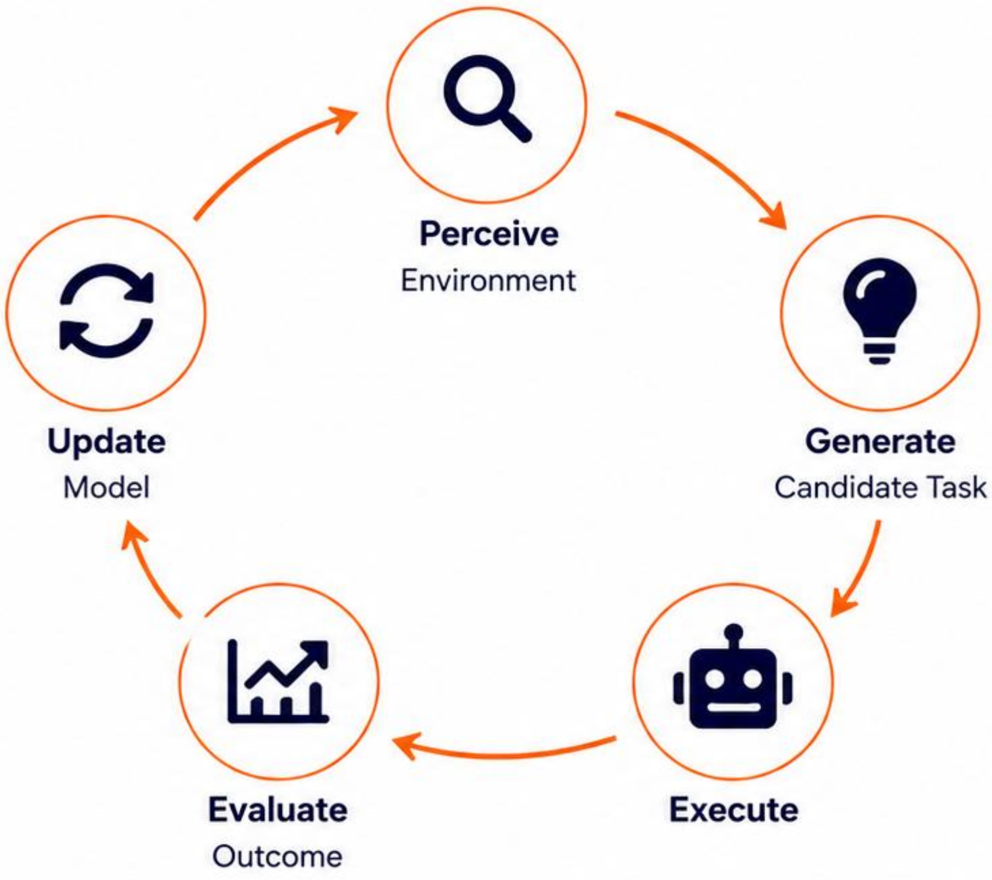
Multistage inspection line, Volkswagen Autoeuropa.

Factories that Adapt

ACT IV

Adaptive robotics, modular platforms, and generated behaviour

A closed loop of perceive, generate, act, evaluate



GrAIpe: perceive vine health ⟶ generate a targeted treatment plan ⟶ apply ⟶ evaluate outcome.

Smarf Farm: perceive fruit & canopy ⟶ generate a spray or harvest action ⟶ execute ⟶ evaluate yield impact.

Reconfiguration instead of replacement

MODULAR BY DESIGN
Swap tools.
Not machines.

MULTI-TASK PLATFORM
Spray. Harvest.
Sense. Model.

3D FIELD UNDERSTANDING
Sees and models the world in 3D.

RECONFIGURE IN MINUTES
Adapt to the task, not the other way around.

RICS modular unit for precision spraying — **greenhouse deployment, Smart Farm 4.0.**

Swap the end-effector, not the machine.

A tracked mobile platform with an interchangeable arm and sensor-guided dosing — the same chassis sprays, harvests, and models the field in 3D depending on the task it's given.

This is the physical precondition for the GenAI Factory:
a platform that can be reconfigured has something to point a generative planner at.

ONE PLATFORM. MANY MISSIONS.

SPRAY	HARVEST	SENSE	MODEL
Targeted application	Precision picking	Monitor & detect	3D mapping & analytics

From programmed motion to generated skill

This keynote spends a lot of time on models. It should — because the same generative techniques are now reaching the robot's own behaviour, not just its perception.

Tracked mobile platform, modular arm — RICS Smart Farm 4.0.

Classical robotics

An engineer programs each motion primitive by hand, for each task, on each platform.

Generative robotics

A model generates candidate motion policies from a goal and a world model — then refines them in simulation before ever touching hardware.

Sim-to-real transfer

Generative simulation produces the variety of conditions a robot needs to generalize — terrain, lighting, payload — without months of field trials.

Factories that Explore

ACT V

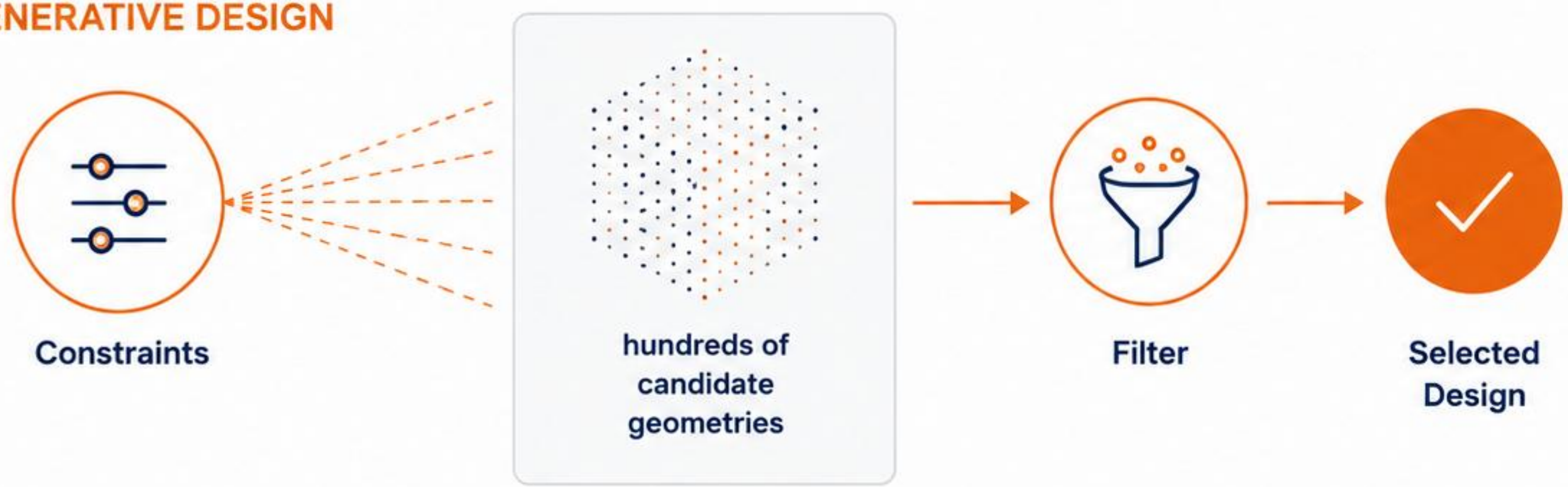
Generative design, digital twins, and simulated futures

From one design to a validated search space

Generative design flips the workflow: instead of an engineer drafting one solution, the system proposes many, and physics does the filtering.

TRADITIONAL DESIGN

GENERATIVE DESIGN

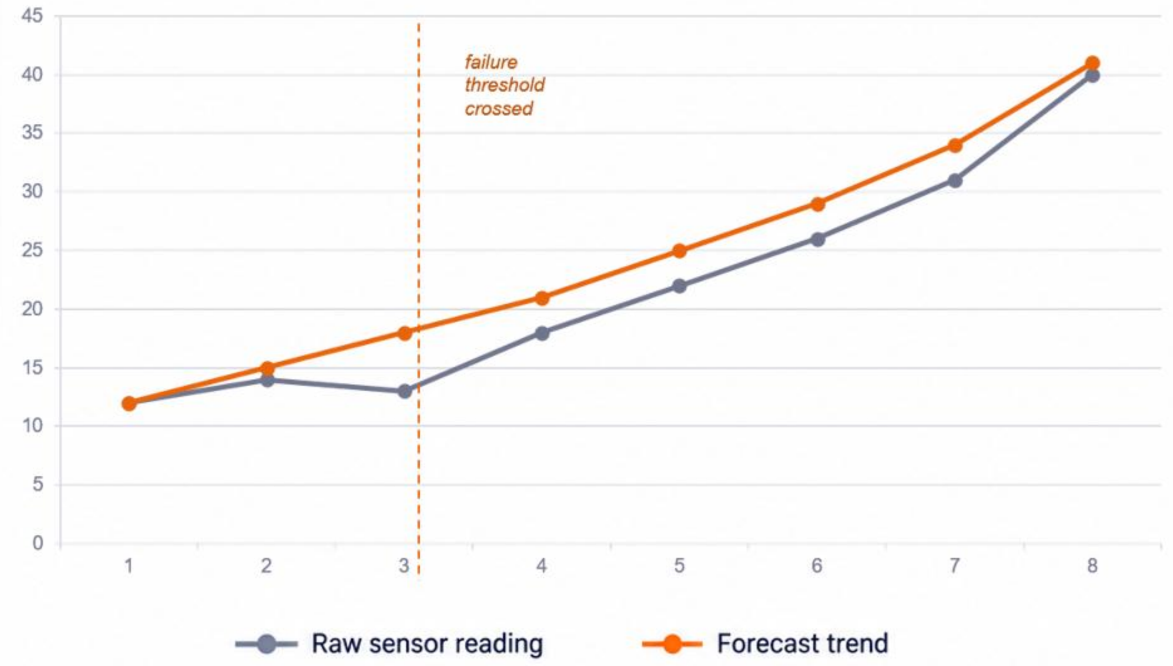


> *Human-in-the-loop selection: engineers filter the generated set against practical and aesthetic constraints the model was never given — judgment stays human.*

From forecasting failure to generating the fix

Illustrative degradation pattern, consistent with the RICS clamp-monitoring case.
Forecast models flag drift before it becomes failure.

DEGRADATION PATTERN OVER TIME



Forecasting tells you when a clamp will fail.

The generative next step: a design-space search over geometry and material, constrained by the exact wear pattern the forecast found — proposing the fix, not just the diagnosis.

Welding clamp, automotive side-member.

EARLY WARNING
Detect drift before failure occurs.

ACTIONABLE INSIGHT
Know when to intervene, not just what happened.

GENERATE THE FIX
Search and propose optimal designs within constraints.

PREVENT, DON'T REACT
Turn future failures into future uptime.

Factories that Decide

ACT VI

Explainable reinforcement learning and industrial copilots

A decision trace an operator can audit

Every decision is traceable — from what the system saw, to what it decided, to why it took that action.

State

Sensor + prediction snapshot

→

Policy

RL model selects an action

→

Action

Reschedule, reroute, hold

→

Explanation

Traced back to the triggering signal

Feedback loop — every explained decision retains the next one

> *Batch order changed because Line 2's failure risk was rising and Line 4 had slack — reordering avoided a predicted stoppage."*

Cognitive assistants for operators and engineers

For decades, expertise on a shop floor lived in one veteran engineer's head.
This is what it looks like when that tacit knowledge becomes queryable.

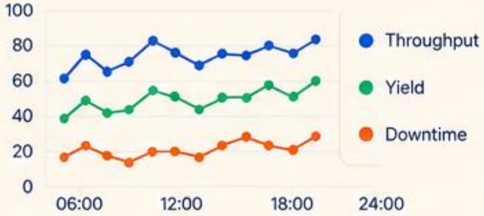
Sensors
Real-time machine signals and readings

Maintenance Logs
Work orders, failures, repairs, notes

Work Order	Status	Type	Updated
WO-88421	Closed	Repair	May 7, 09:14
WO-88420	Closed	Failure	May 6, 16:33
WO-88419	In Progress	Repair	May 6, 11:02

Vision Models
Image and video inspections

Process History
Batches, settings, events, outcomes

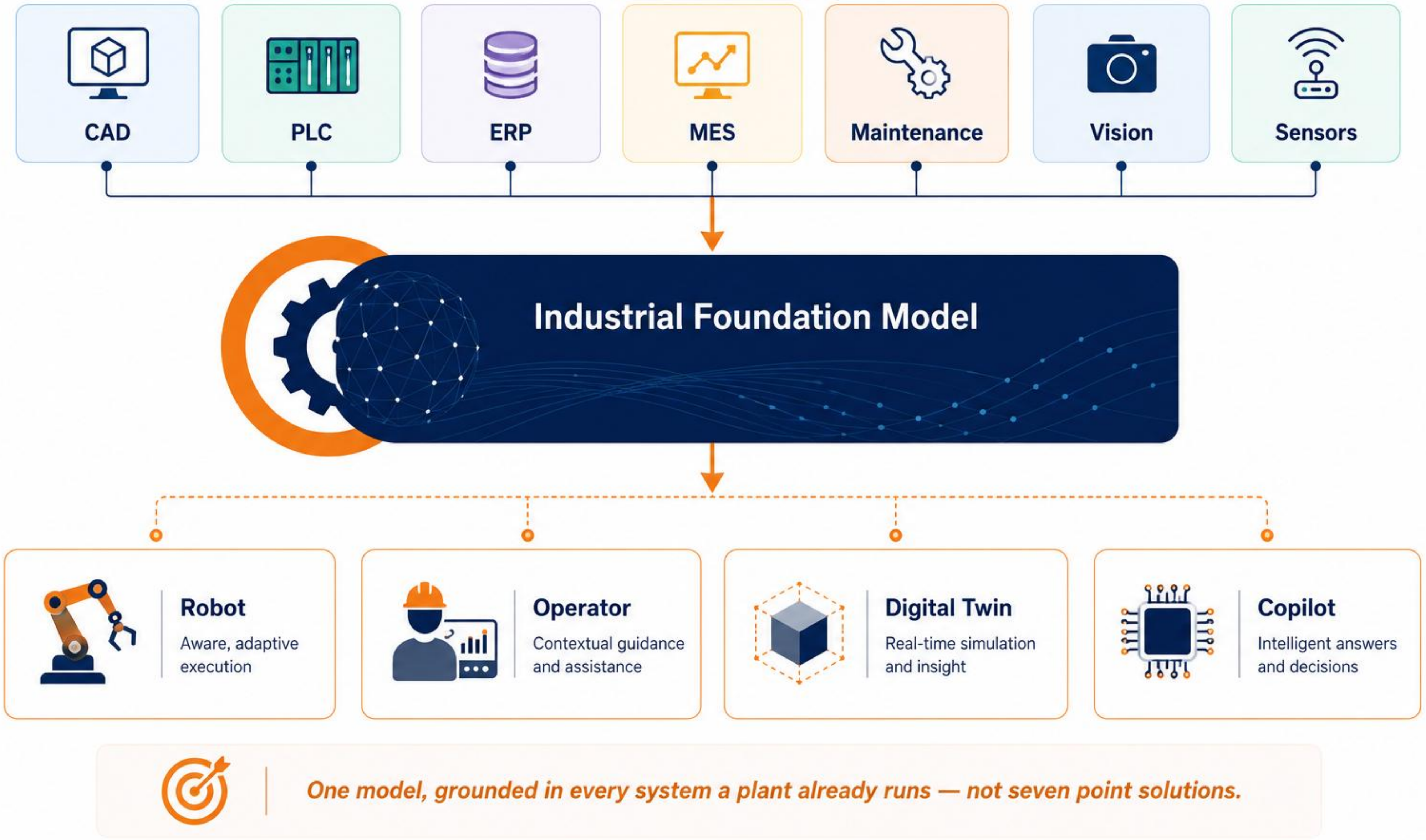


LLM Copilot
Understands context. Connects the dots.
Gives answers you can trust.

Why did Line 3 flag part #4471? → a plain-language answer, grounded in the same models already running.

The copilot proposes a person with context the model doesn't have — **cost, customer, safety margin — decides.**
This is a near-term direction, not a shipped product.

The Industrial Foundation Model



The Foundation Model Stack

Reasoning

Multi-step inference over industrial context, not single-shot lookup

Memory

Persistent state across shifts, machines, and maintenance history

Knowledge Graph

Structured relationships between assets, parts, and processes

World Model

An internal simulator of how the plant behaves, not just describes it

MCP / Tool Calling

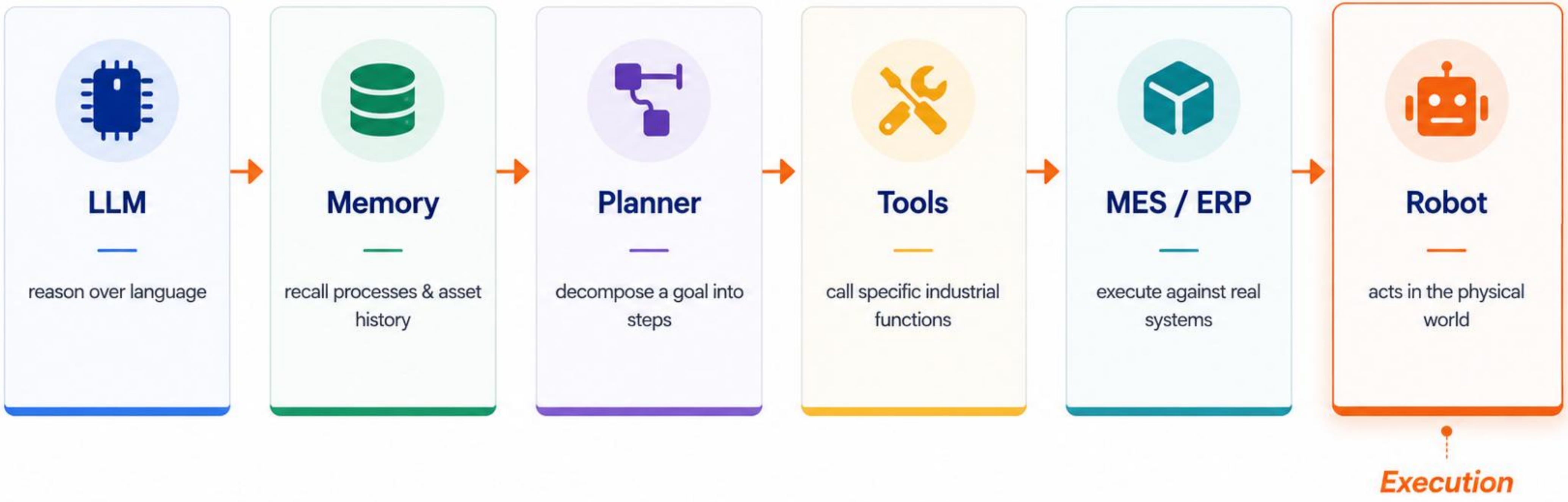
A standard interface for invoking real industrial functions

Multimodal Grounding

Vision, sensor, and language fused into one context window

This is the actual state of the art in 2026 — not a future prediction.
What's still emerging is doing all six reliably, together, inside a single industrial deployment.

From LLMs to Industrial Agents



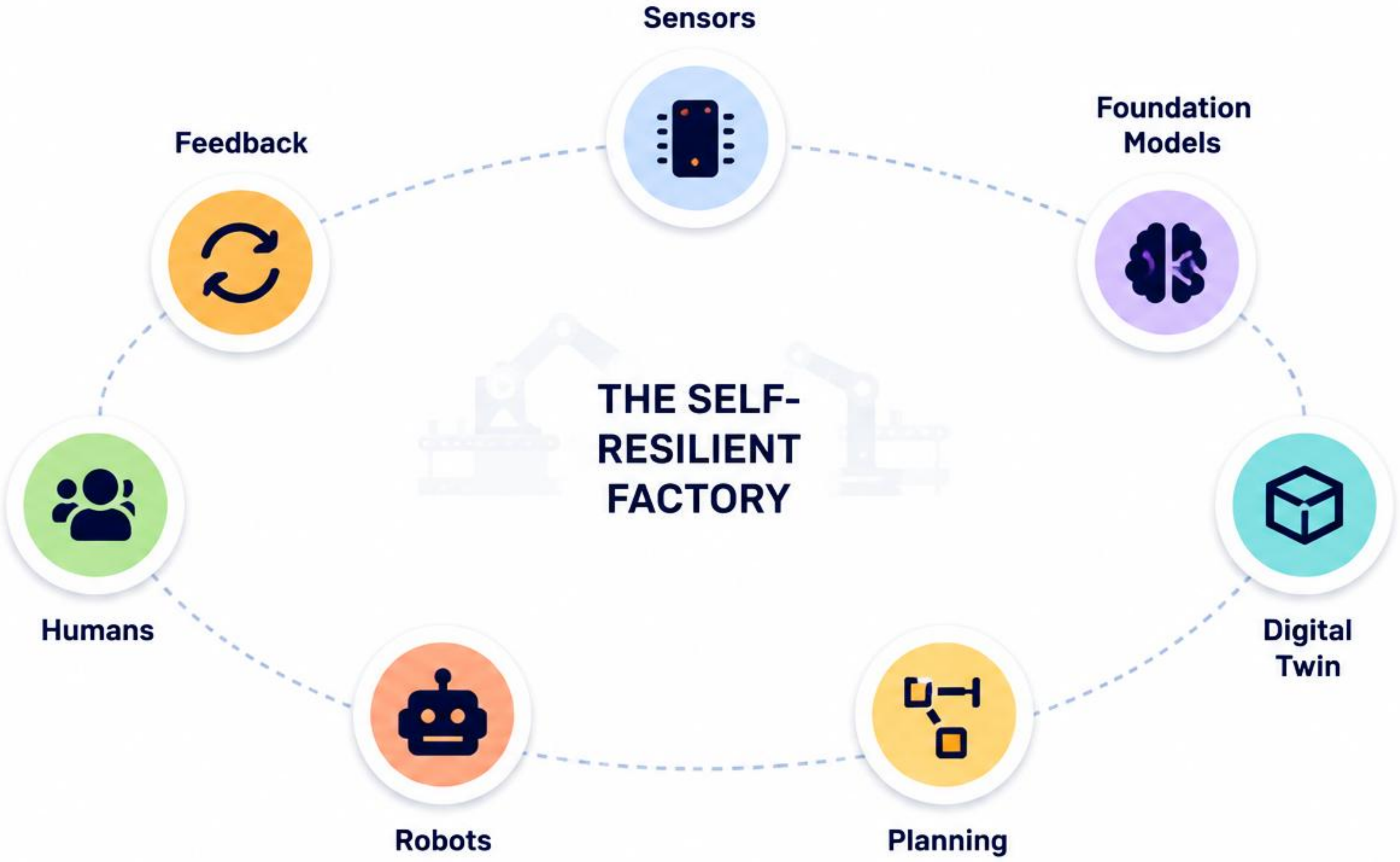
This is an extremely current reseaoch direction — and the difference between a chatbot and an agent is exactly this chain: **reasoning that terminates in a physical or transactional action**, not just a sentence.

The Self-Resilient Factory

ACT VII

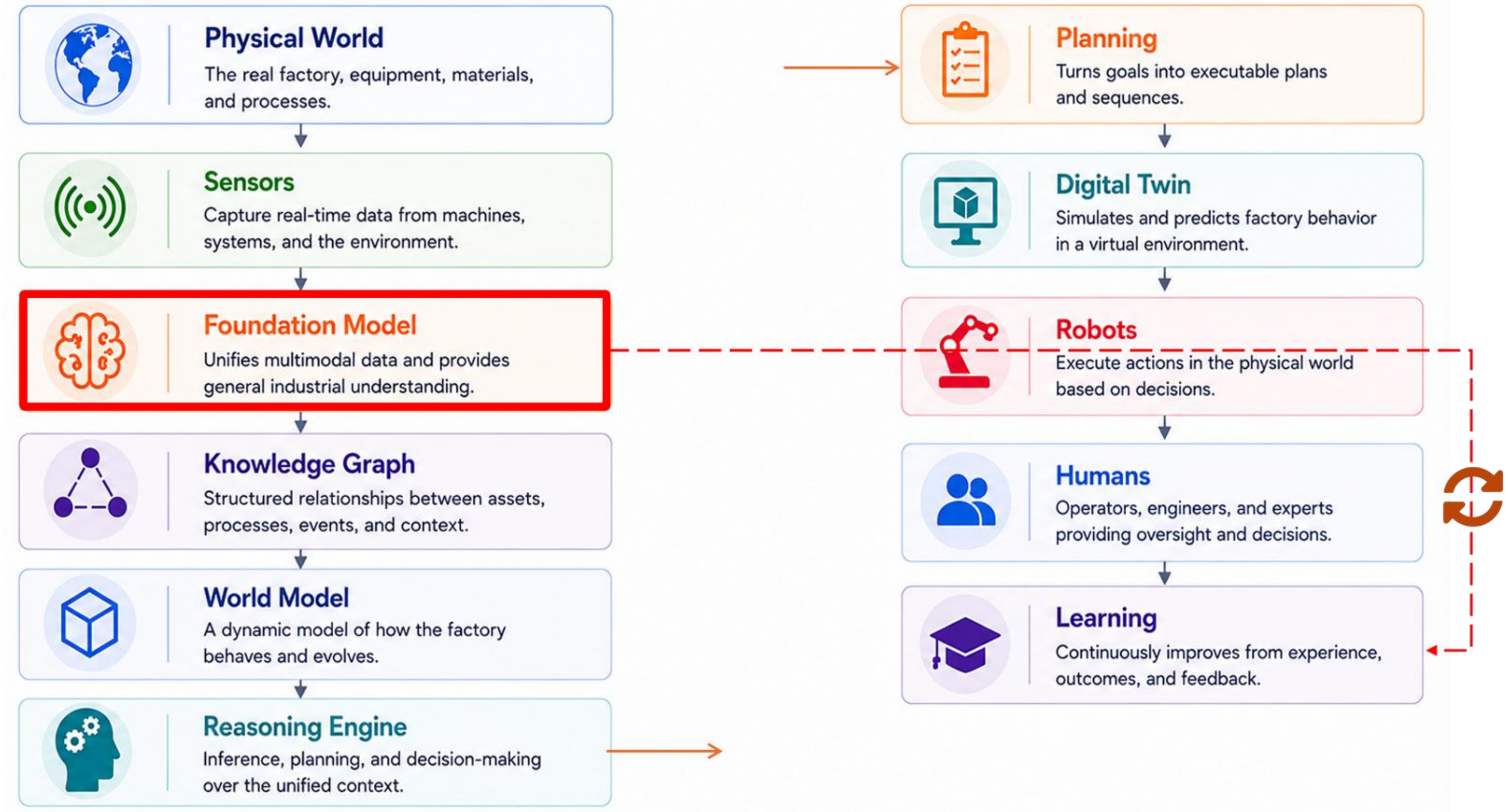
One architecture, integrating everything shown so far

THE INTEGRATED ARCHITECTURE — STEP 1



Seven components most factories already own in some form.

The full architecture, layer by layer



Challenge

ACT VIII

Scientific honesty requires naming what is still unresolved.

Challenges

Trust & Verification

Design-in-the-loop runtime trust must become standard engineering practice, not a research demo.

Explainability at Scale

Generative and RL-based decisions must remain auditable — not just accurate.

The EU AI Act

High-risk obligations for AI embedded in industrial machinery carry a real compliance deadline.

Cybersecurity

Generative pipelines widen the attack surface — synthetic data can be poisoned, models can be probed.

Functional Safety

Certifying a system that proposes its own actions is a different problem than certifying a fixed one.

Sustainability

Generative optimization must target energy and material efficiency as first-class objectives.

The Human Role

Cognitive assistance changes what operators and engineers do — training must keep pace.

Four predictions, stated plainly

Prediction is reaching diminishing returns.

Foundation Models will become industrial operating systems.

Digital Twins will evolve into World Models.

Industrial software will become agentic.

Roadmap

ACT IX

Roadmap

TODAY

Prediction

Sensors report. Models forecast. Humans decide.

TOMORROW

Generation

Systems imagine, explore, and propose — humans supervise.

2035

Self-Resilient Factory

Knowledge flows continuously. The factory reasons about itself.

When factories can imagine...

who designs?

When factories can explain...

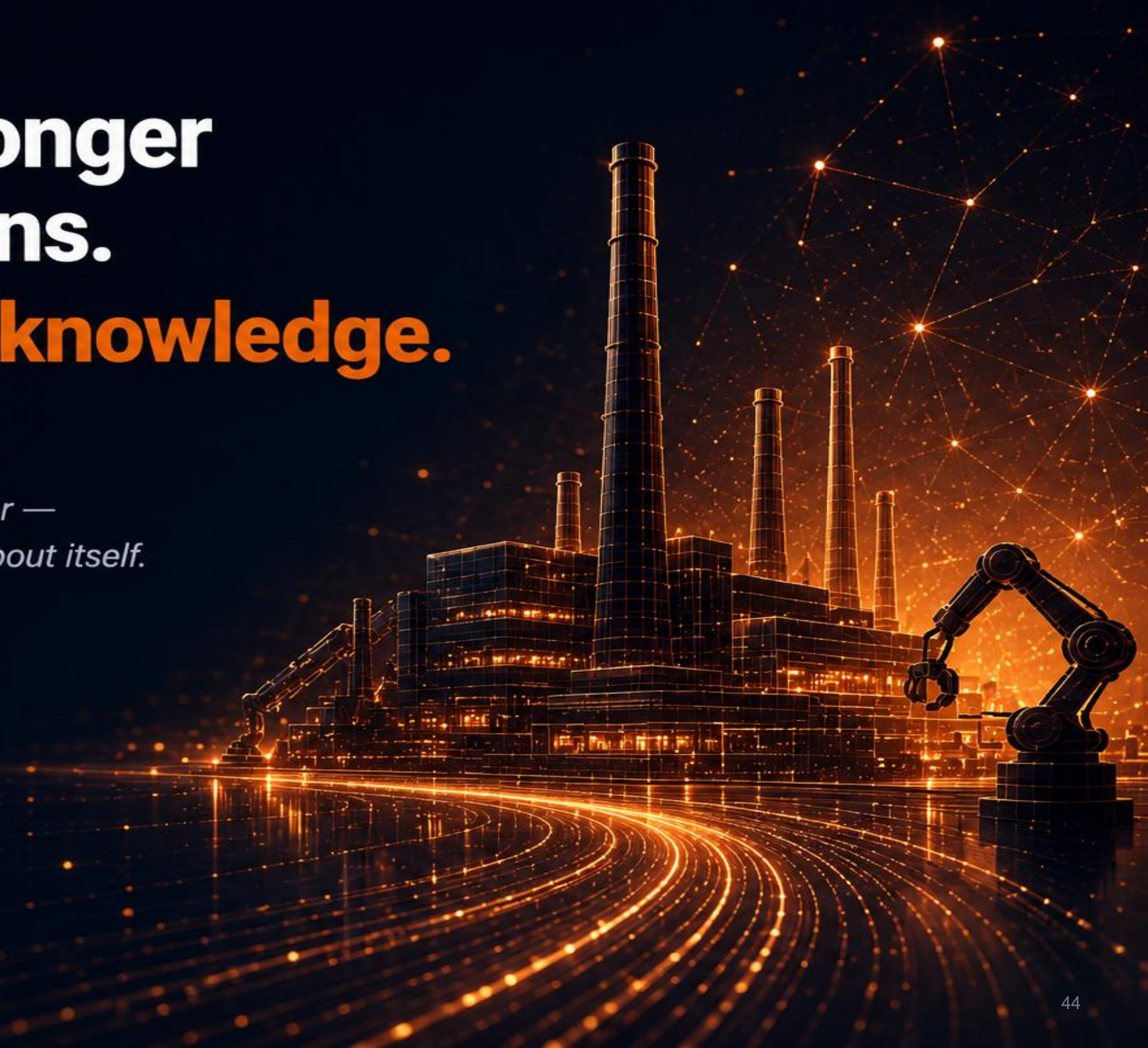
who decides?

When factories can generate knowledge...

what becomes the role of engineers?

Factories will no longer execute instructions. They will generate knowledge.

That is the paradigm shift this talk has argued for —
not a faster factory, but a factory that reasons about itself.



Jose Barata

jab@uninova.pt • rics-uninova.com