In collaboration with Kearney

WORLD ECONOMIC FORUM

The AI-First Operating System: A Blueprint for Operating and Business Model Innovation

WHITE PAPER

JUNE 2026

Images: Getty Images

Contents

Foreword 3

Executive summary 4

Introduction 5

Block 1: Intelligence engine 9

1.1 The speed loop 10

1.2 The scale loop 12

1.3 The scope loop 14

Block 2: Adaptive AI technology stack 17

2.1 Turn data into fuel for the intelligence engine 18

2.2 Own the control layers 19

2.3 Compose a model-agnostic portfolio 20

2.4 Make context dynamic, not static 21

Block 3: Operations redesign 22

3.1 Where to allocate intelligence 24

3.2 How to redesign work 26

3.3 Building intelligence-native operations 29

3.4 Operational leverage: the new business economics of intelligence 32

Block 4: Human-AI teaming 33

4.1 Growing human-AI talent 34

4.2 Organizing for human-AI teams 36

4.3 Evolving the organization for AI-first execution 38

Block 5: New value creation 40

5.1 Designing product 41

5.2 Prioritizing trust 42

5.3 Positioning in the market 43

Design your own AI-first blueprint 45

Conclusion 47

Contributors 48

Endnotes 50

Disclaimer

This document is published by the World Economic Forum as a contribution to a project, insight area or interaction. The findings, interpretations and conclusions expressed herein are a result of a collaborative process facilitated and endorsed by the World Economic Forum but whose results do not necessarily represent the views of the World Economic Forum, nor the entirety of its Members, Partners or other stakeholders.

© 2026 World Economic Forum. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, or by any information storage and retrieval system.

Foreword

Maria Basso
Head, AI Applications and Impact, Centre for AI Excellence, World Economic Forum

Cathy Li
Head, Centre for AI Excellence; Member, Executive Committee

Michael Römer
Senior Partner & Global Lead, Activate: Digital and Analytics, Kearney

For most of business history, the enterprise has had a ceiling. There were limits on how fast enterprises could learn and how fast information could travel from the edge of an organization to the centre to inform action. Additionally, there were constraints on how rapidly decisions could be made and executed, and how far enterprises could scale without losing coherence or exhausting talent, capital or time. These were structural limits – built into the architecture of how organizations think, decide and act.

A small but growing number of enterprises are dismantling those limits. They are achieving this not by adopting AI as a tool, but by designing their entire operating logic around artificial intelligence (AI) – building systems that learn from every interaction, improve with every cycle and expand into new domains without proportional increases in cost or complexity. These are the “AI-first” enterprises, and unlike the ceiling they are replacing, this frontier does not hold still. It advances with every breakthrough, every deployment and every new capability that enters the field. What these enterprises demonstrate is not a new fixed limit, but a different order of what an organization can do.

The World Economic Forum’s AI-First Enterprises workstream, part of the AI Global Alliance, brought together more than 50 of the world’s most advanced enterprises and thinkers with a shared mission: unlock what enterprises can do for industry, global ecosystems and society by exploring the business and operating model breakthroughs that AI is making possible and shaping the next era of enterprise. Over the past year, the workstream went deep into the frontier of architectures, decisions and the patterns that set genuine AI-first organizations apart. What follows is drawn directly from that community.

What we found is both clarifying and demanding. It's "clarifying", because we have discovered a coherent set of fundamentals across companies that seem different from one another – from drug discovery to financial infrastructure, and from autonomous logistics to legal AI. It's "demanding", because the gap between organizations that have restructured around intelligence and those that have not is widening fast.

The five building blocks in this paper – the intelligence engine, the adaptive technology stack, redesigned operations, human-AI teaming and new value creation – are not a roadmap for transformation. They describe how the frontier already operates. The opportunity for leaders willing to engage seriously is to use them as a lens: to see what is now possible and to design accordingly.

The structural limits are gone. What comes next, how far this frontier moves, who it reaches and what it makes possible for industries and societies are the questions we will keep pursuing.

Executive summary

A new enterprise operating system is emerging: organizations designed around intelligence as a core capability for value creation.

A new class of organizations is emerging, designing their operations around intelligence, using data as a key resource and embedding artificial intelligence (AI) into processes, decision-making and execution. These "AI-first" enterprises are demonstrating enhanced performance: faster innovation cycles, the ability to scale with smaller teams and the creation of new products built on continuously learning systems.

This white paper examines how these organizations are built, how they operate and what separates them. Drawing on deep engagement with more than 50 of the world's most advanced AI-first enterprises and thinkers, it maps the patterns and principles that define how the frontier actually operates.

Five building blocks capture the fundamentals of how AI-first enterprises are designed and run:

- **Intelligence engine at the core**: Enterprises place intelligence at the centre of their operations and business models, turning every interaction into a compounding source of advantage. This accelerates speed, extends scale and expands scope in ways that grow stronger with use.
- **Adaptive AI technology stack**: Organizations build modular, adaptable technology architectures that integrate data, models and workflows while being designed to evolve as the frontier moves.
- **Operations redesign**: Enterprises systematically digitize their workflows with intelligence – not optimizing them at the margins but connecting all of their operations to the intelligence engine.
- **Human-AI teaming**: Roles, teams and organizational structures are designed to enable continuous collaboration between people and AI systems, defining what human contribution looks like at the frontier.
- **New value creation**: Every AI-first organization, regardless of sector, must decide how to position intelligence integration in the market. From AI as a product feature to AI as invisible infrastructure, positioning choice determines what customers pay for, where value accrues and how the business competes.

Together, these building blocks form a blueprint for operating in a world where intelligence is abundant and continuously advancing. To make this actionable, this paper extends the traditional business model canvas with AI-first fundamentals, providing leaders with a practical framework for designing and operating as AI-first enterprises.

The rise of AI-first enterprises is in its early stages. It remains unclear which approaches will prove most effective across industries, how broadly they will scale and what their long-term implications will be for productivity, employment and competition.

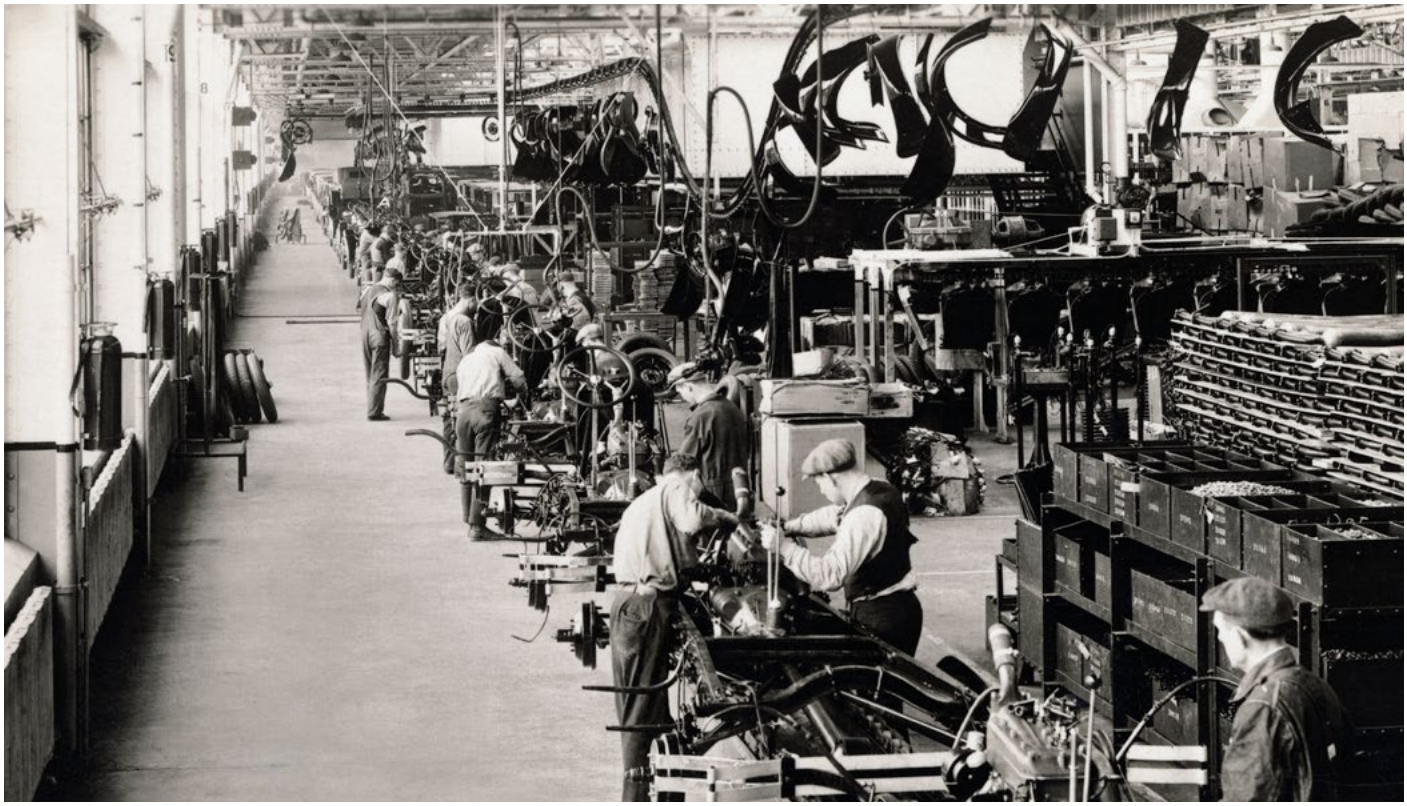
For leaders, the imperative is to build the capabilities required to test and scale AI-first operating systems across the enterprise. For entrepreneurs, the opportunity lies in turning intelligence into differentiated products, services and systems at scale. For policy-makers, the challenge will be staying ahead of these rapidly evolving models and ensuring development translates into broad-based economic and societal value.

Introduction

Artificial intelligence (AI) is the latest class of general-purpose technology, like electricity, computers and the internet. Historically, these technologies had a limited impact when incumbent organizations applied them within existing systems. Real breakthroughs came when pioneers rebuilt their businesses around the technology itself.

In electricity, early manufacturing adopters replaced steam engines with electric motors while retaining the same layouts and production processes, resulting in no productivity gains but reducing energy costs by 20–60%.[1]

The breakthrough came when pioneering factories redesigned their entire blueprints around electricity; one example is Henry Ford reworking his assembly line system between 1919 and 1926.[2] These leaders reconfigured workflows, distributed power to individual workstations and rebuilt production systems,[3] unlocking step-changes in operating models and systems and laying the foundation for the modern factory.



What this means for AI

The speed and breadth of adoption are what make AI different. In October 2025, Wharton estimated that 82% of decision-makers use AI weekly, up from 37% in 2023.[4] The factors driving this industry adoption are:

1. **Democratized access**: Innovations in interfaces and large language models (LLMs) have enabled non-technical users to adopt, build and operate AI, meeting the user at their level of proficiency.

2. **Reasoning and agency**: Systems can interpret, decide and act across workflows with increasing autonomy.

3. **Multimodal intelligence**: AI can process or generate text, images, audio and more, expanding the scope of application.

4. **Parallel execution**: Tasks that were sequential can now be executed simultaneously, compressing time to scale.

For today's market, intelligence is changing what kinds of organizations can be built. AI-first organizations are applying this lesson from the outset, designing operations, decision-making and business models around intelligence.

TABLE 1 | **Enterprise archetypes by AI adoption**

Archetype	Definition	Attributes	Litmus test
AI-enabled	Enterprises that apply AI to automate or augment discrete tasks within existing workflows, improving performance without materially redesigning the operating model	– AI applied at the task level – Performance improves through localized efficiency and productivity gains – Talent model focused on domain operators experimenting and using AI tools – Tech stack integrates third-party AI tools into existing systems – Unit economics improve through cost reduction or productivity gains, not model-driven differentiation	If the AI tools were removed, would your workflows and organization structure collapse? **If the answer is "no," your enterprise is AI-enabled**
AI-first	Enterprises that systematically (re)design workflows, roles and decision rights so AI becomes the strategic lever for creating and delivering value at industrial scale	– Step-change (re)design of core operating workflows around AI – Tech stack (re)structured to embed AI into execution, decision-making and governance at scale – Systematic allocation of tasks between humans and AI – Built for industry standards, compliance and production-grade scale – Augmentation and automation enhance human capabilities, improving operational leverage beyond what humans or AI can do alone – Unlocks new capabilities that could not be designed, delivered or scaled without AI	If AI systems were removed, could your business still operate (workflows, decisions, and delivery)? **If the answer is "no," your enterprise is AI-first**
AI-native	New enterprises created with AI as a core production capability, with products, services and competitive advantage fundamentally dependent on AI systems	– AI embedded in the value proposition, not just operations – Value scales through intelligence compounding (data > model > better outcomes) – Talent model centred on AI researchers, machine learning (ML) engineers and model evaluators – Tech stack intentionally built for continuous model training, deployment and rapid iteration at scale	If AI systems were removed, would your value proposition still exist? **If the answer is "no," your enterprise is AI-native**

AI-enabled approaches limit the full value creation potential of the technology

Despite AI investment estimated at over $250 billion globally in 2025,[5] many organizations are seeing only incremental gains. A global survey found that only 25% say AI is having a transformative effect on their company. Meanwhile, 84% of companies have not redesigned jobs around AI capabilities.[6]

Like electricity, the primary reason is structural. In many cases, AI is deployed as a layer on top of existing processes and embedded narrowly in tools. This approach improves efficiency at the margins but does not unlock the technology's full potential.

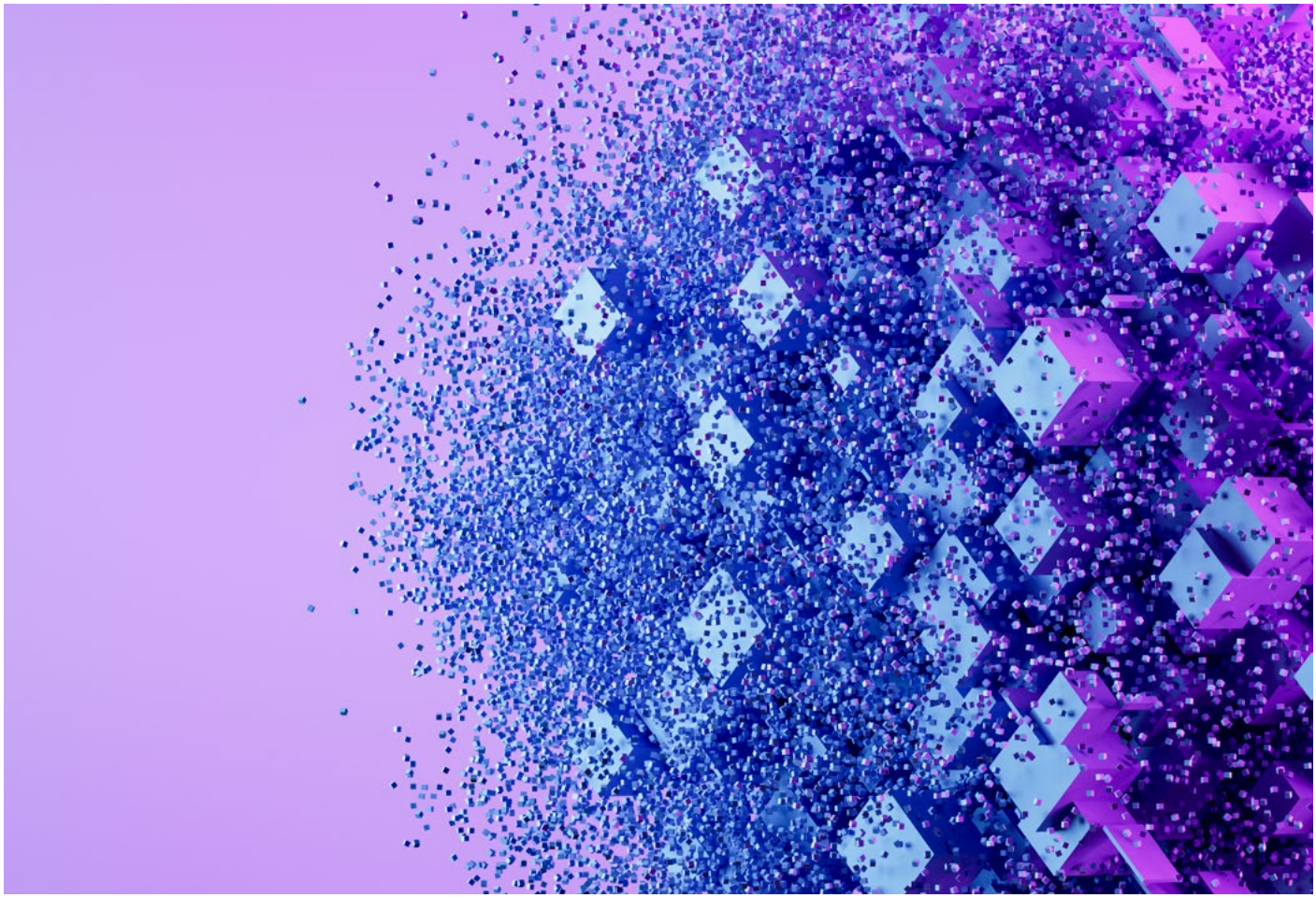
In contrast, AI-first performance gains are early, but striking:

- **Compressed cycle times**: Commercial insurance workflows went from 28 days to 2.8 hours.[7]
- **Shorter roads to scale**: $100 million annual recurring revenue (ARR) in months, compared to four to eight years previously.[8,9]
- **Moonshots within reach**: A 2.5-day continuous autonomous inference loop identified the first viable drug candidate for a previously intractable HIV target.[10]
- **Small teams achieving the output of entire functions**: Some teams are achieving 15 times more productivity with AI tools than without them.[11]
- **New market dynamics**: AI is reshaping how companies position and deliver value to customers.[12]

Innovations in business and operating model logic enable this impact. To inform this step change, the World Economic Forum launched the AI-First Enterprises Workstream as part of the AI Global Alliance. By bringing together over 50 executives leading AI-first companies and departments and drawing on insights from more than 150 executives and experts from interviews, workshops and expert consultations, this group explored:

- How are AI-first enterprises changing the nature of business and the way it creates and delivers value?
- What are the emerging patterns and innovations behind their success?

Together, these patterns and innovations are enabling a new class of AI-first enterprises built around continuous learning, real-time decision-making and scalable intelligence. The following blueprint outlines the core building blocks of these organizations.

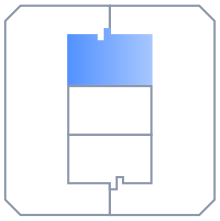


Fundamental five building blocks

Five building blocks define the fundamentals of how AI-first enterprises deliver value. They should be understood not as a linear roadmap but as an integrated system. Progress in one area reinforces the others. Taken together, they form an emerging blueprint for the AI-first enterprise operating system.

FIGURE 1 **Fundamental building blocks for an AI-first operating system**

BLOCK 1



Intelligence engine

When intelligence sits at the core of an enterprise, it becomes a durable source of speed, scale and scope.

AI-first enterprises build intelligence engines as an operating system around which the firm thinks, runs and creates value. By strategically embedding intelligence into the foundational business logic, these organizations ensure that operational data, proprietary intellectual property (IP) and sources of advantage compound.

At the centre is an **intelligence engine** (Figure 2), a self-reinforcing and data-driven flywheel that embeds all of the organizational knowledge, learns from every interaction and transaction and grows smarter with use. It connects customer data, operational context, business objectives and real-world performance into continuous feedback loops that learn, improve and scale with each cycle:

- **The speed loop**: Accelerating discovery, experimentation and time-to-value through autonomous inference
- **The scale loop**: Operationalizing intelligence as a multi-use platform across business outcomes
- **The scope loop**: Expanding capabilities into new domains, products and markets

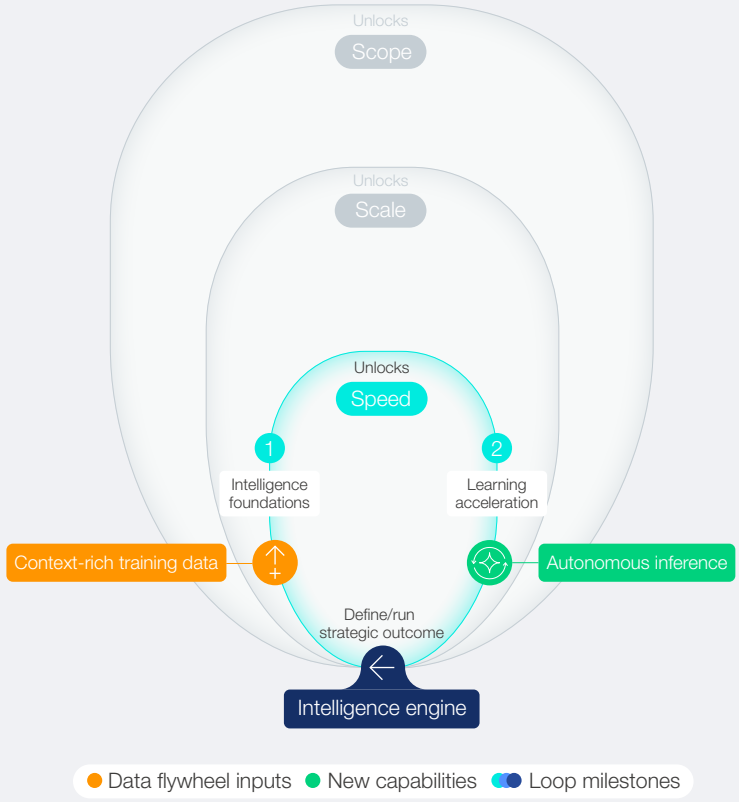
FIGURE 2 **The intelligence engine**

1.1 The speed loop

Objective: AI-first enterprises use the speed loop to run more experiments, generate and test hypotheses, and validate ideas before committing resources, lowering the cost of experimentation while increasing the pace of learning. Failures narrow the search space; successes become building blocks for the next cycle. Over time, this creates a system where proprietary learning compounds.

In practice, one example is a drug discovery company moving from candidate hypothesis to validated simulation in days, not months. Others include a software company shipping functional prototypes from plain-language briefs in seconds, and a robotics firm testing and refining physical behaviours in simulation before a single real-world deployment. What these organizations share is not just speed but a **structural** advantage.

FIGURE 3 **The speed loop**



Data inputs: Builds a learning base from context-rich data tied to outcomes and key performance indicators (KPIs), improving learning performance with every cycle

1 **Intelligence foundations**: Anchors the system in defined business outcomes and domain-specific context, calibrating what it learns from and optimizes for

2 **Learning acceleration**: Compresses the path from proof of concept to proof of value by using AI to rapidly generate, simulate, test and validate ideas before scaling

New capabilities – autonomous inference: Enables systems to run the full discovery-to-deployment process independently, generating, testing, evaluating, adapting and refining outputs against defined objectives while using persistent memory and human feedback to improve with every cycle

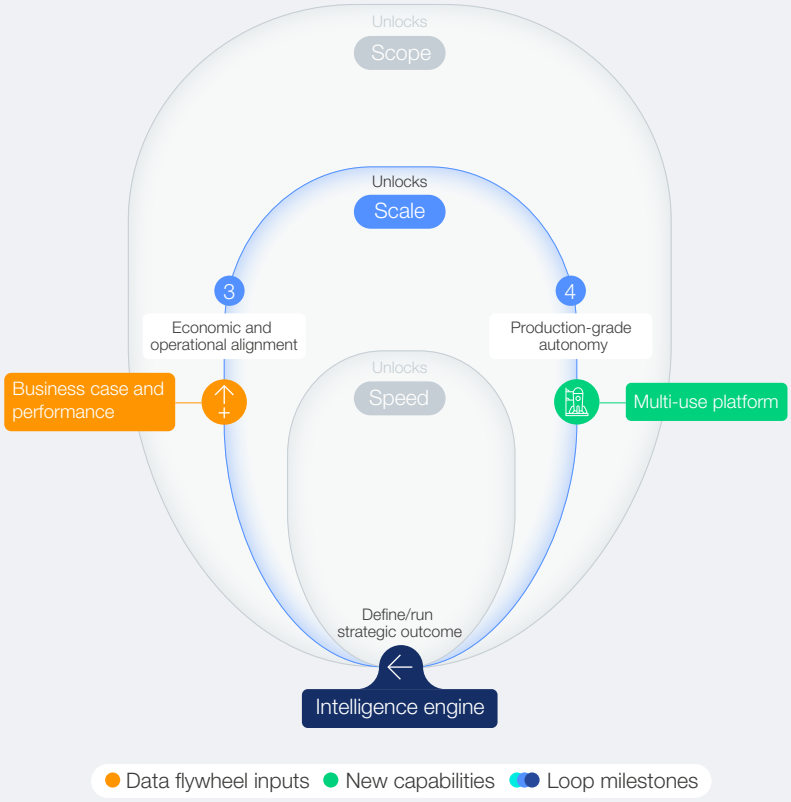
TABLE 2 | **How the speed loop runs**

1	
Intelligence foundations	**What it enables** Aligns the system to clear objectives and structured, domain-specific learning bases
Mechanisms	
Objective-led context building	**How it works** Translate business goals into clear, measurable AI objectives, linking outcomes (e.g. revenue, accuracy) to system-level metrics (e.g. binding affinity, accuracy, cost per output), bringing together relevant structured and unstructured data to build context, reveal patterns and give meaning to system-level signals. **Example: Workera** Workera defines good context as personalization. By connecting to systems like Workday and ingesting role data, resumes and strategic priorities, it builds two layers: individual context around a user's skills and experience, and company context around the capabilities the organization needs to develop. Workera uses this context to assess skills, identify gaps and recommend targeted learning paths. Context is considered "good" when it improves the precision of these assessments, tailors recommendations to the individual's role and ensures skill development aligns with business objectives.
2	
Learning acceleration	**What it enables** Speeds up experimentation by generating, testing and refining solutions before real-world deployment
Mechanisms	
Discover solutions and rapidly prototype	**How it works** Identify high-value opportunities and test hypotheses in simulations before committing resources. **Example: Formation Bio** Formation Bio built a tool called Cassini that can be used to map how the drug market may evolve over the next few years. It analyses clinical pipelines, competitor timelines and scientific evidence to estimate where a new drug is most likely to succeed. The system can even predict which Phase II clinical trials are likely to work before the trials begin. By seeing these outcomes earlier, researchers can focus time and capital on the most promising drug programmes.
Evaluations, safety and governance	**How it works** Apply automated evaluation systems with defined thresholds, enabling safe and controlled autonomous execution. **Example: ServiceNow** ServiceNow built an AI agent evaluation framework with quantitative thresholds that agents must clear before going live. The framework scores agents against defined task completion rates (typically 80–95%) with allowable variance depending on objective. When outputs fall outside these bounds, a human-in-the-loop reviewer is engaged at predefined checkpoints. Once live, ServiceNow runs statistical evaluation across agentic systems, where the volume of decisions is 1,000 times greater than in preproduction and exhaustive human review is impossible.

1.2 The scale loop

Objective: The scale loop operationalizes intelligence as a multi-use platform, applying the same models and workflows to increase output across adjacent functions. This requires connecting AI systems directly to business metrics, integrating them into live workflows and ensuring they operate reliably in real-world conditions. Once deployed, each use strengthens the system, improving its efficiency and quality of output.

FIGURE 4 **The scale loop**



Data inputs: Shifts from training data to live business performance within a domain, using live operational data tied to business outcomes and system economics

3 **Economic and operational alignment**: Links system performance directly to business metrics and embeds AI into how work is executed, with live feedback

4 **Production grade autonomy**: Ensures systems operate reliably at scale within defined boundaries, with continuous improvement in production

New capabilities – multi-use platform: Builds outcome-oriented platforms that compound capability, reduce marginal cost of growth and expand into adjacent use cases

In practice, fewer custom model builds and retraining cycles are required to scale intelligence as a company grows. A warehouse platform built for operations monitoring might expand into bottleneck detection, intervention and throughput optimization. A fraud platform could detect more suspicious activity while learning from downstream investigation results and resolution outcomes.

TABLE 3 | **How the scale loop runs**

3	
Economic and operational alignment	**What it enables** Aligns intelligence with how the business operates and creates value, focusing it on the decisions that most affect business outcomes and making performance fully observable.
Mechanisms	
Performance linked to outcomes	**How it works** Translate business goals into optimizable and granular key performance indicators (KPIs) directly tied to revenue, cost, risk, etc. **AI-first example: Claryo** Claryo reflects a shift from monitoring warehouse activity to linking operational performance directly to daily profit and loss (P&L) and square-foot economics. Its spatial generative AI platform processes roughly 50 terabytes (TB) of visual data per warehouse each month to identify work stoppages, backlogs and bottlenecks in real time. For example, the system alerts teams when pallets sit idle in express lanes for an extended period and quantifies the resulting financial loss. This visibility enables more proactive management during high-demand periods. During a Black Friday surge, one customer used Claryo's visual monitoring capabilities to scale its workforce from 150 to 300 employees, increase floor management from two leaders to six, and four supervisors to 12, and maintain stable unit economics under peak pressure. Claryo reports an annualized P&L gain of $2.56 per square foot.[13]
Making leverage points observable	**How it works** Focuses AI on high-impact decisions and implement them with full visibility, enabling faster iteration and control. **AI-first example: Taktile** In regulated financial services, high-stakes decisions often sit inside workflows such as onboarding, credit underwriting, anti-money laundering (AML) and fraud prevention. Taktile's customers systematically deploy agents to target time-consuming tasks that were previously impossible to automate (e.g. financial spreading and "know your customer" checks) and enhance overall decision system intelligence by automatically filtering out fraud false positives. Taktile enables institutions to layer rules, agents and human judgement to make faster, more accurate decisions while maintaining governance and oversight in a common decisioning layer.
4	
Production-grade autonomy	**What it enables** Runs AI systems reliably at scale under real-world conditions, with clear boundaries and continuous improvement
Mechanisms	
Industrial-grade reliability	**How it works** Define acceptance criteria, test against edge cases, regulations and compliance and validate systems under real-world conditions before deployment. **Example: Waymo** Before operating without a safety driver in a new city, Waymo vehicles must satisfy 12 acceptance criteria that assess readiness for deployment. These are evaluated through a combination of quantitative thresholds and simulation testing – including over 20 billion miles in virtual environments, where real-world scenarios are replayed and varied to stress-test the system before any public launch.[14]
Continuously improve without breaking production	**How it works** Run a live system ("production") and a test system ("challenger") in parallel. The challenger is evaluated against key metrics such as accuracy, latency and edge cases and promoted only when it outperforms the live system. **Example: BforeAI** BforeAI applies this approach to cybersecurity by running new detection models in parallel with its live system. Challenger models analyse the same network data and are evaluated on their ability. Only models that outperform production and meet strict accuracy thresholds (e.g. over 90%) are deployed. This system enables the company to predict threats up to 80% of the time before an attack begins while allowing it to proactively mitigate threats without disrupting live operations.

Scaling through reuse

How it works

Treat each deployment as a variation of the same system, embedding learnings into a multi-use platform and reducing marginal cost of scale.

Example: Osmo

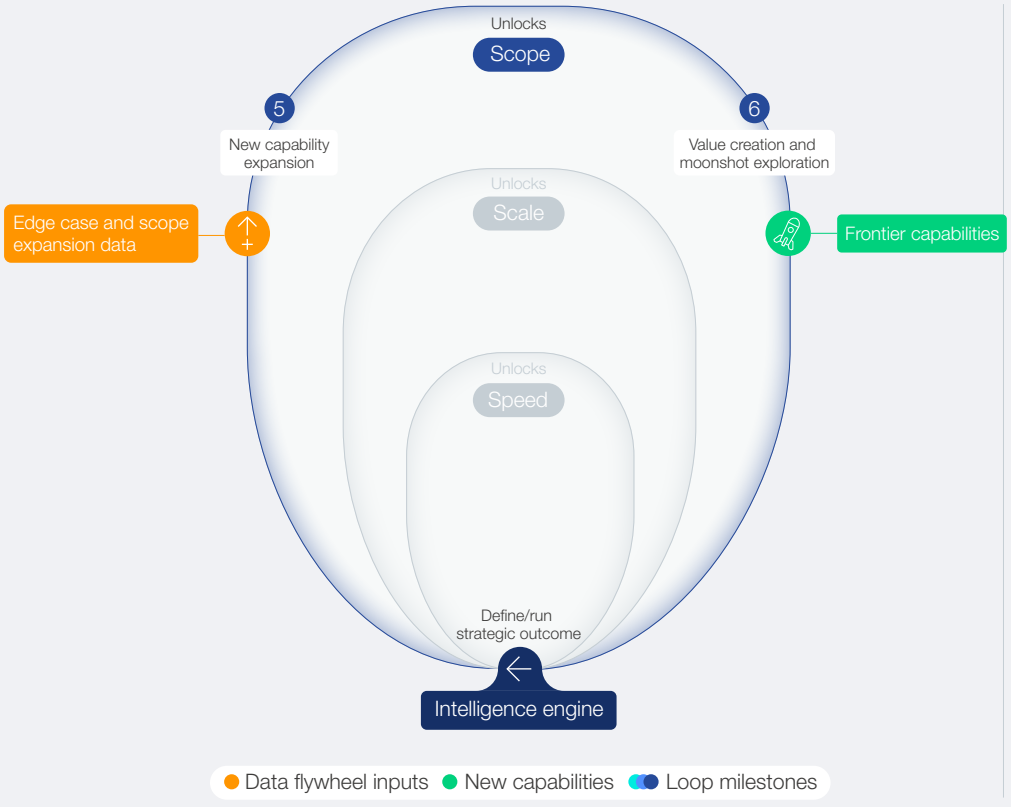
Osmo built an olfactory intelligence platform trained on over 3 billion molecules and more than 5 million fragrance classifications, predicting how molecular combinations will smell and perform in a product. Rather than building a new model for each fragrance product or customer, Osmo uses the same platform to develop new formulations, predicts market performance and personalizes individual preferences. Osmo then feeds the resulting formulation back into the system, making each subsequent fragrance development cycle faster and more personalized to the customer. What once took six months now takes 60 seconds.

1.3 The scope loop

The scope loop expands what the organization can do with intelligence. Organizations can reuse capabilities across workflows, share them among teams and recompose them into net-new products, services and markets.

In practice, this means organizations can begin using intelligence built for one purpose to create value in other domains. A legal AI system can expand from contract analysis into tax and deal work. A shared intelligence layer can become the foundation for fraud detection, personalization and risk decisions across the business. Over time, organizations can build frontier capabilities such as world models that push ambition towards scientific, technical and commercial moonshots that no prior combination of data, talent or coordination could unlock.

FIGURE 5 **The scope loop**



Data inputs: Shifts from performance optimization to expansion data, including new domains, edge cases and market contexts

5 **New capability expansion**: Reuses and extends proven capabilities across workflows, teams and domains, with systems learning and improving with use

6 **Value creation and moonshot exploration**: Recombines capabilities to identify new high-upside opportunities, create new products and enter new markets

Capability frontier: Builds a system where each deployment expands capabilities into new domains; the constraint shifts from building intelligence to deciding where to apply it.

TABLE 4 | **How the scope loop runs**

5	
New capability expansion	**What it enables** Extends proven capabilities across domains, allowing intelligence to generalize beyond its original use case
Mechanisms	
Transfer across domains	**How it works** Apply the same underlying system to new domains by adapting context while keeping the core architecture constant. **AI-first example: Harvey AI** Harvey built a domain-specific LLM for legal workflows: contract analysis, case law research and litigation support. The same architecture was extended to tax compliance in 2024, when Harvey co-built a tax AI assistant with PwC and OpenAI – trained on legislation and revenue guidance across more than 10 jurisdictions. In 2025, they expanded the platform into mergers and acquisitions, covering financial document review, due diligence "red flag" detection and deal analysis. One workflow alone has been executed over 10,000 times.[15,16]
Shared intelligence layer	**How it works** Build a common platform where models, features and learnings are reused across the organization. **AI-first example: Stripe** In 2025, Stripe built a payments foundation model (PFM) trained on tens of billions of transactions. The PFM captures the structure of every transaction as a shared embedding layer the entire organization builds on top of, rather than each product team training bespoke, task-specific models. Every new product is developed against the same shared representations – spanning fraud detection (Stripe Radar), authorization, card-testing prevention and dispute management. On its first deployed application, the PFM identified over 95% of card-testing attacks in real time – a 22% improvement over prior methods.[17]
6	
Value creation and moonshot exploration	**What it enables** Turns accumulated intelligence into new products, markets and high-impact opportunities
Mechanisms	
Recomposing capabilities	**How it works** Combine existing capabilities into new product offerings without rebuilding underlying systems or building net-new infrastructure underneath. **AI-first example: Anthropic** Anthropic launched Claude in 2023, strategically deploying the same underlying intelligence across different product categories. Claude Code recomposed coding and tool-use capabilities into an agentic command-line coding tool that reached $1 billion in annualized revenue within six months of launch. In January 2026, Claude Cowork recomposed Claude Code's agentic patterns into a desktop application for non-technical knowledge workers, with a plugin system for sales, legal, finance, human resources (HR) and design. In April 2026, Claude Design recomposed Claude's vision and reasoning capabilities into a visual prototyping tool. In May 2026, Claude for Small Business recomposed Claude Cowork's agentic capabilities into 15 ready-to-run workflows for finance, operations, sales and HR.[18]

Unlock moonshot exploration

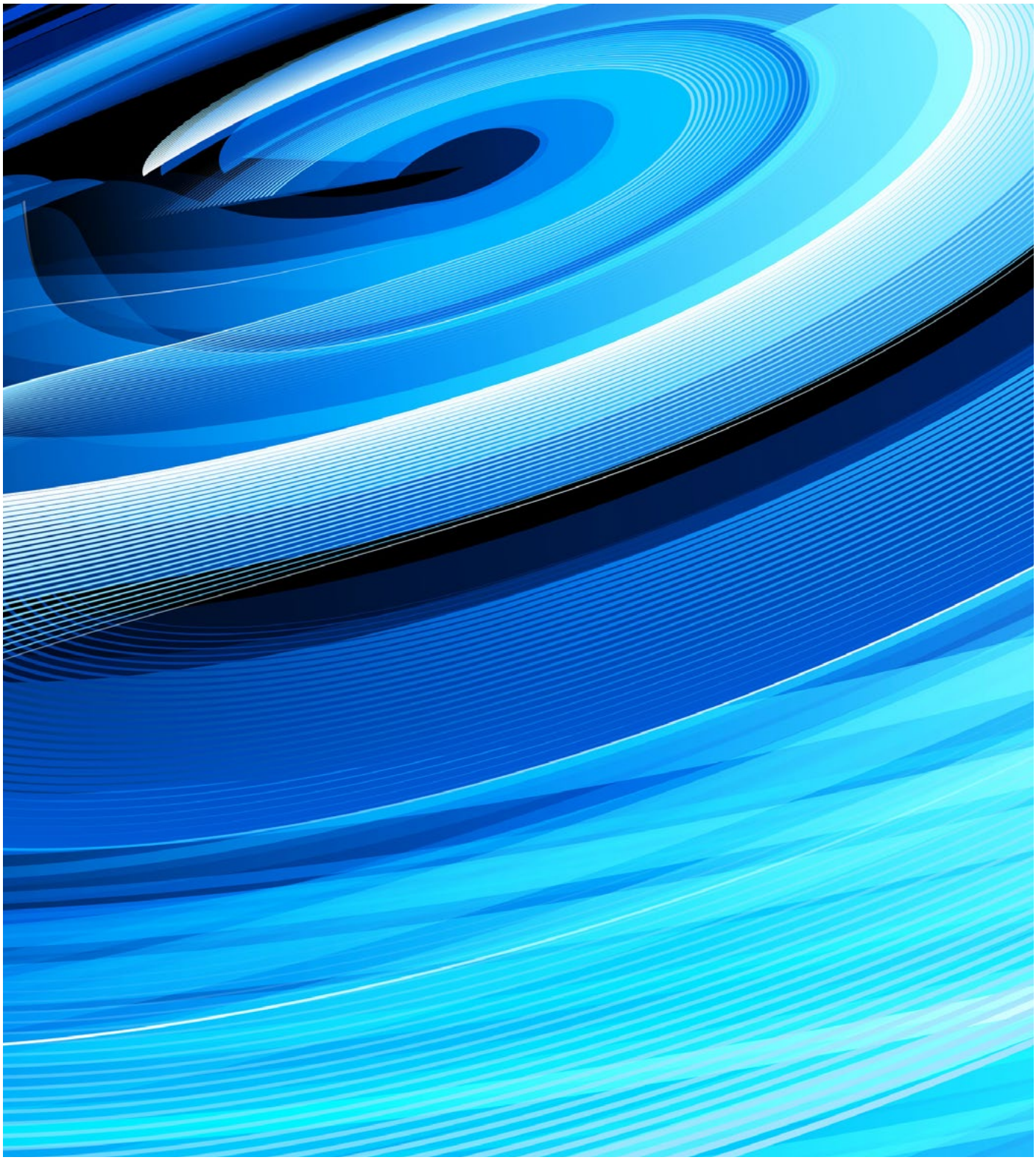
How it works

Apply intelligence to problems that were previously unsolvable due to complexity, cost or coordination constraints.

AI-first example: Isomorphic Labs/DeepMind

In 2026, Isomorphic Labs released the Drug Design Engine (IsoDDE), a unified system that moves beyond structure prediction into end-to-end drug discovery. Building on AlphaFold's breakthroughs, IsoDDE predicts protein structures, actionable binding sites and drug candidates, bridging the gap between prediction and real-world therapeutics.

Historically, scientists struggled to develop drugs for around 85% of human proteins because they could not identify sites where drugs could attach effectively. IsoDDE addressed this limitation by identifying both a known and a previously hidden binding site on cereblon – a protein targeted in cancer therapy – using only its amino acid sequence as input, with no prior knowledge of where drugs might attach.[19]

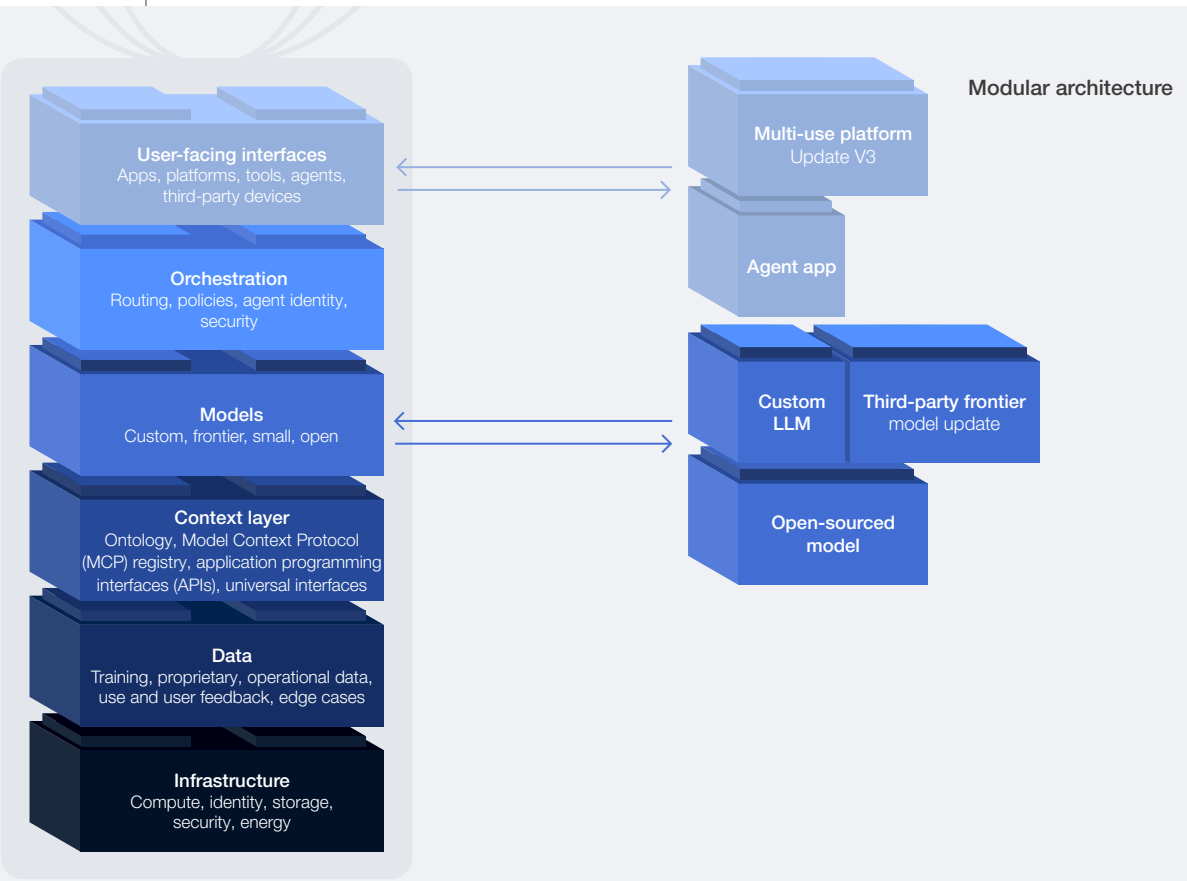


BLOCK 2

Adaptive AI technology stack

When enterprises design modular technology stacks, the intelligence engine can evolve as fast as the technology beneath it.

FIGURE 6 **AI-first modular technology stack**



An AI-first technology stack is the architecture that allows the intelligence engine to run stably at scale. **This architecture must be modular**, connecting data, models, context and orchestration, allowing users to access intelligence across workflows, improve it through feedback and adapt it as models and tools change.

An AI-first technology stack does not necessarily start with replacing existing systems of record – such as customer relationship management (CRM) or enterprise resource planning (ERP) – but with redesigning how intelligence is layered and accessed across them. Core systems remain the source of truth, but AI sits on top, orchestrating workflows, sharpening decisions and improving interactions. Over time, as the intelligence engine matures, users may reshape or replace parts of the stack where they no longer provide an advantage.

A modular architecture design allows for:

- **Interoperability**: Swapping components without rebuilding the whole system
- **Identity and authority**: Governing who reaches what (models, data, connectors) so AI acts within guardrails
- **Open architecture**: Maintaining stability in production and keeping control over the parts that matter most
- **Modernization**: Adapting reliably as new models, tools and infrastructure emerge

FIGURE 7 | **AI-first modular technology stack plays**

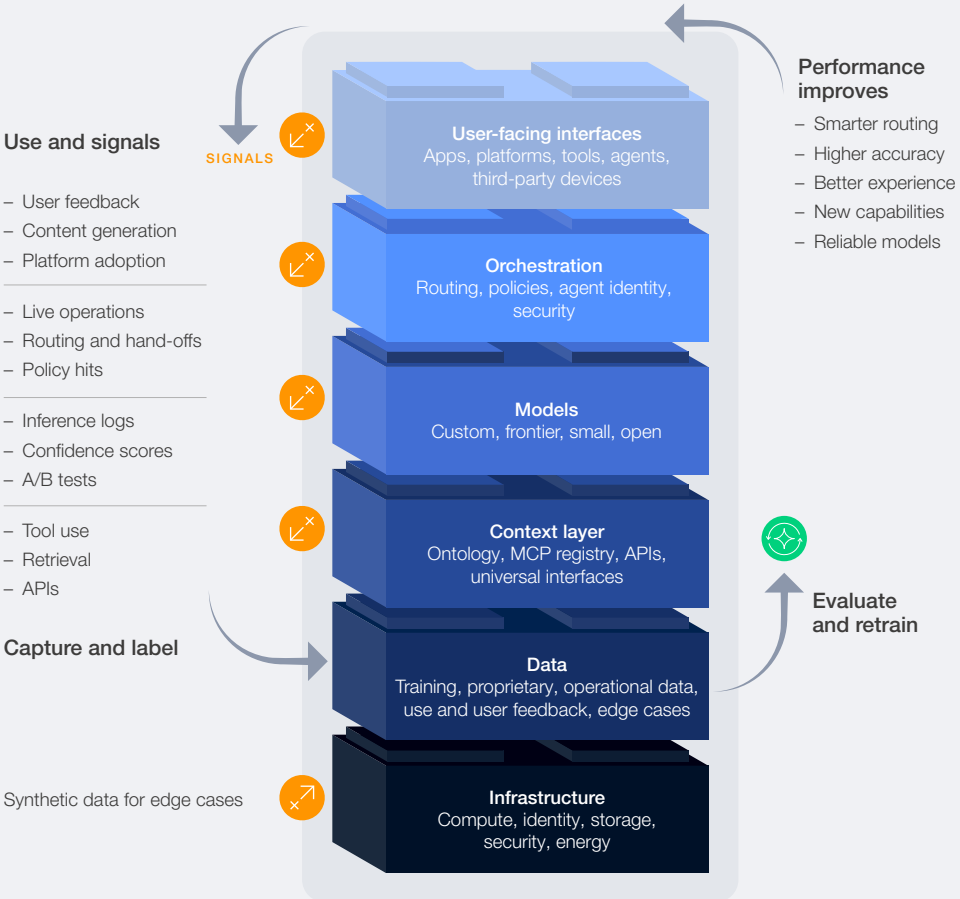
Play 1	Play 2	Play 3	Play 4
Turn use into fuel for the intelligence engine	**Own the control layers**	**Compose a model-agnostic portfolio**	**Make context dynamic, not static**
How the data flywheel compounds intelligence	How the stack stays stable	How the model layer is managed	How context becomes an advantage

2.1 Turn data into fuel for the intelligence engine

Objective: The intelligence engine only compounds if the data feeding it gets better with every cycle. AI-first enterprises integrate feedback, training and evaluation into a single system, enabling performance to improve over time rather than relying on one-off fine-tuning efforts.

In practice, data quality and structure determine a system's intelligence capabilities. In fraud operations, false positives and investigator overrides are labelled training data that accelerate discovery; live operational data deepens reliability at **scale**; and synthetically generated attack patterns expand the **scope** of detection into scenarios real data alone could never cover.

FIGURE 8 | **How the data flywheel works**



Creating a data flywheel across the full technology stack is critical: Every instance of use, whether a model output is accepted, overridden or flagged, becomes a signal. Run repeatedly across the stack, those signals compound: models improve, orchestration tightens and the system grows more capable with every cycle of use.

Labelling data for learning and evaluations: Data only compounds when it is structured for learning. This means separating day-to-day operational data (e.g. use activity, transactions, system events) from the data used to train AI models, and storing that training data in consistent, labelled and trackable formats. While AI can work with unstructured data, business context still needs to be coded in.

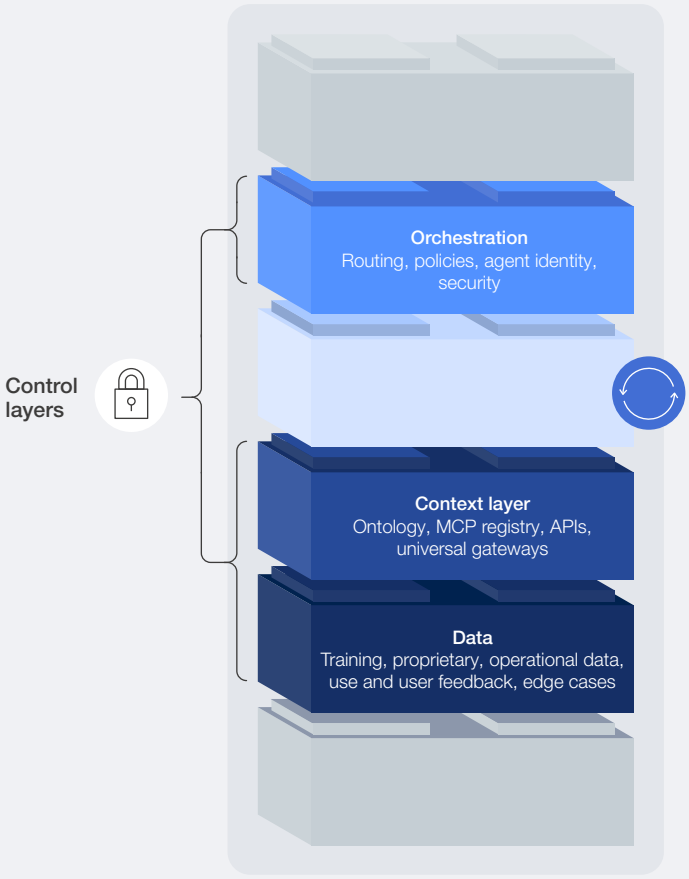
Synthetic data for edge cases: Synthetic data can be useful for highly specific needs: rare edge cases, low-frequency scenarios or workflows where labelled data is hard to obtain. If used well, it helps teams train and evaluate systems on the cases that matter most when production data is not enough.

2.2 Own the control layers

Objective: By owning the control layers around the model, a modular architecture can absorb technology changes without constantly rebuilding workflows or giving up critical architectural control.

In practice, a customer support team that runs thousands of AI-assisted interactions across chats, email and team handoffs should be able to continue as normal. If the company upgrades a model or replaces a vendor, service quality shouldn't change or require a redesign of the workflow. By owning orchestration and how context connects with internal application programming interfaces (APIs), the company can make upgrades while keeping routing, policies and customer data flows stable.

FIGURE 9 **How control layers govern the technology stack**



Own orchestration, not just models: The orchestration layer routes tasks to the right model based on cost, performance evaluations and policy. This gives the enterprise control over how intelligence is deployed as model economics and capabilities change.

Manage the gateways between your data and your models: To perform a task well, models need context from systems such as CRM, HR and finance. Connectors such as the MCP let them access that information through a common layer rather than one-off integrations. This reduces integration fragility and limits workflow rebuilds when vendors change.

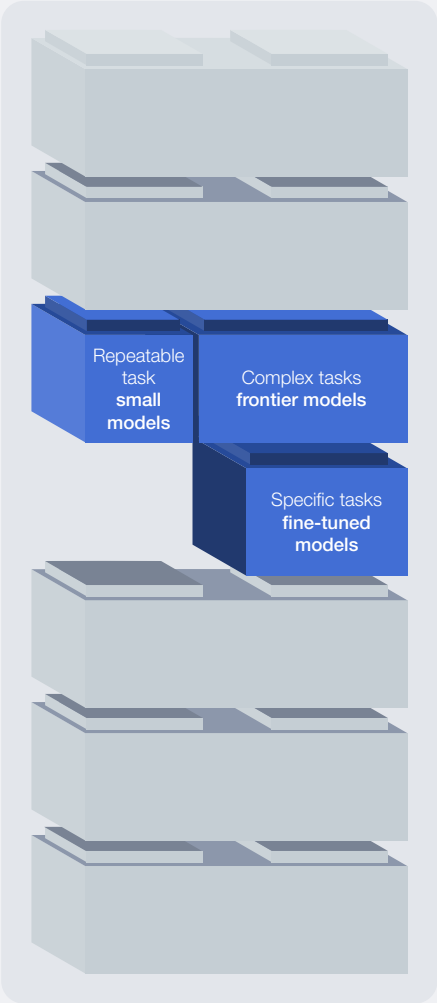
Build APIs around proprietary data: Keep proprietary data behind internal APIs instead of wiring it directly into third-party tools. This maintains a stable interface even as vendors change. It also preserves built-in controls such as caching, permissions and audit logs.

2.3 Compose a model-agnostic portfolio

Objective: Treat the model layer as a portfolio, managing model-task trade-offs across performance, cost and reliability, and routing work to the right model rather than relying on a single model across the business.

In practice, the model choice changes across creative workflows. A lower-cost model may handle repetitive production work, such as tagging, formatting or variant generation, while a frontier model can be used for more difficult tasks, such as concept development, narrative framing or synthesis. Brand-critical work may rely on a tuned system grounded in internal guidelines and assets.

FIGURE 10 **How to manage and apply multiple AI models**



Manage models as a portfolio: A portfolio approach reduces overreliance on any one vendor or model class, improves flexibility as model economics shift and allows the enterprise to invest in building or buying models with gaps in key workflows.

Match the model, validate the fit: Frontier models are best suited to complex, high-judgement work; smaller models to repeatable, lower-cost tasks; and fine-tuned models to use cases where proprietary data materially improves performance. The ability to swap confidently across all three depends on evaluation: testing whether the quality of output remains consistent when the model changes. This learning needs to be updated regularly throughout the organization.

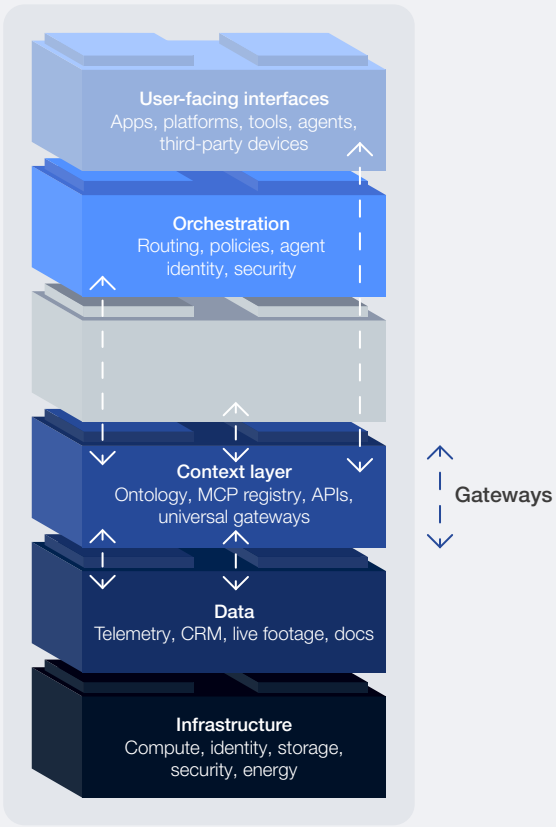
Build custom or domain-specific models where the advantage is real: These models matter when proprietary context is important, training data is strong or general-purpose systems do not perform well. They are most useful when the task depends on institutional knowledge, cultural context or sensitive IP.

2.4 Make context dynamic, not static

Objective: **Assemble context in real time.** Instead of relying on static copies of data, the system should pull live information from the source, preserve its authority and tailor the retrieved information to the task at hand.

In practice, a field technician does not rely on a static prompt loaded with old manuals and policies. The system pulls the current equipment history, latest service bulletin and site-specific operating context in real time. This allows the recommendation to reflect what is true now, rather than what was copied into the model weeks earlier.

FIGURE 11 **How context flows through the technology stack**



Keep context modular and centrally accessible: Instead of copying everything into one storage location, organize context into discrete purpose-built clusters, each encoding its own schemas, rules and examples, and make them discoverable through a centralized registry. Agents pull only what is relevant to the task rather than relying on static context loaded upfront.

Encode source-authority as weights in the context layer: Agents need more than tagged data; they need to know which source takes precedence when sources conflict. Those rules travel with the data, so downstream retrieval and reasoning reflect the right hierarchy by default.

Assemble context just in time: Rather than loading every known fact into a static prompt, a context router assembles only what is needed in the moment. Corrections from users flow back into the codification layer.

CASE STUDY 1

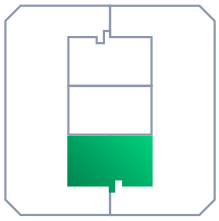
Indeed building a modular AI operating layer

Indeed transformed its stack by building a centralized agent platform based on the Model Context Protocol (MCP), which modularized how models, data and applications connected across the enterprise. The platform virtualized underlying services across internal and external products, enabling secure access through three-legged Open Authorization (3LO) authentication at scale.

This architecture powers Indeed's conversational app in ChatGPT, where job seekers can use natural language and their Indeed profile to discover relevant opportunities and navigate directly to jobs on Indeed across more than 50 countries and 27 languages.

In addition to extensible, secure MCP access, the Indeed platform also includes automated risk guardrails, evaluation systems that continuously track adoption, accuracy and productivity, and the ability to swap models to optimize across cost, performance and latency. The goal is for every interaction to feed back into the system – improving orchestration, accelerating learning and allowing Indeed to continuously expand capabilities for job seekers and employers without disrupting operations.

BLOCK 3



Operations redesign

When intelligence allocation shapes how operations are designed, workflows stop being processes to execute and become systems that learn.

The ambition for an AI-first enterprise is not AI applied to selected workflows. It is an organization where every workflow has been digitized, codified and made legible to the intelligence engine – so that the engine can learn from it, improve it and ultimately run it. Operations and intelligence become inseparable. Every decision, process and action feeds a system that continuously gets smarter, faster and more capable. The enterprise does not allocate AI to the business; the business runs on AI. At that point, competitive advantage is no longer about which tools a company uses – it is about how deeply intelligence is embedded into how the organization thinks and operates.

Most enterprises today are not yet at the starting line of that journey. They are running pilots – dozens of disconnected experiments that demonstrate local value but never compound, because they are not connected to a shared intelligence layer and not designed to scale. The leap from pilots to AI-first operations requires a deliberate intermediate step: treating intelligence like capital and deciding strategically where to invest it first. Rather than letting AI proliferate across the organization without coordination, AI-first leaders identify the most influential workflows, redesign them end-to-end around intelligence capabilities and ensure that every execution cycle feeds data and signals back into the engine. Over time, the operating model is rebuilt – not in one transformation programme, but workflow by workflow, until the business runs on AI.

This block maps that discipline across three levels – from identifying where to allocate intelligence first, to redesigning work around it, to building the systems that make intelligence-native operations run at scale. They are not a sequence to complete, but the practices that separate enterprises compounding at the frontier from those still running isolated experiments.

Gamma illustrates what this discipline looks like in practice.

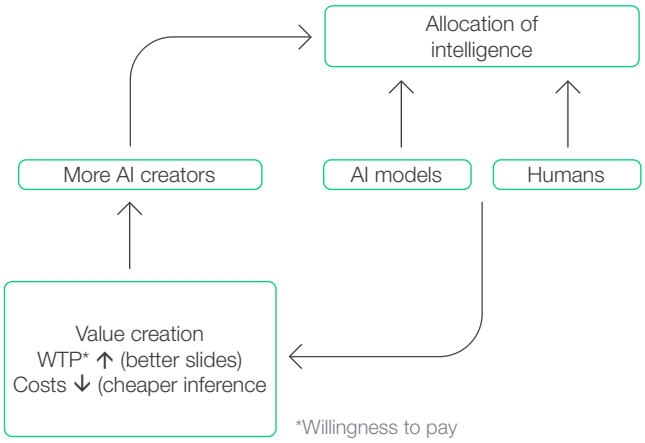
CASE STUDY 2

Gamma's allocation of intelligence

By late 2025, Gamma, an AI-native presentation company, had crossed **$100 million in ARR**, with a team of roughly **50 people**, while users had created over **400 million assets** on the platform and were generating **more than 1 million pieces of content daily**. The company treated every step in content creation as an allocation problem, asking questions such as:

- Which tasks should frontier models handle and which should cheaper specialized models handle?
- Which prompts required orchestration rather than fine-tuning?
- Which outputs still required human judgement about taste, workflow and product direction?

The allocation of intelligence



Source: Koning, R., Keller, N. T. and Kim, H. (2025). *Gamma: Slides in the Blink of AI*. Harvard Business School. https://store.hbr.org/product/gamma-slides-in-the-blink-of-ai/826001?sku=826001-PDF-ENG.

Gamma's edge came from continuously reassigning work to the lowest-cost, highest-quality combination of model, system design and human input. Six months after launch, inference-related gross margin moved from approximately 31% to approximately 77%. The intelligence engine improved because the allocation decisions became more deliberate. Each improvement made the next cycle faster and cheaper.

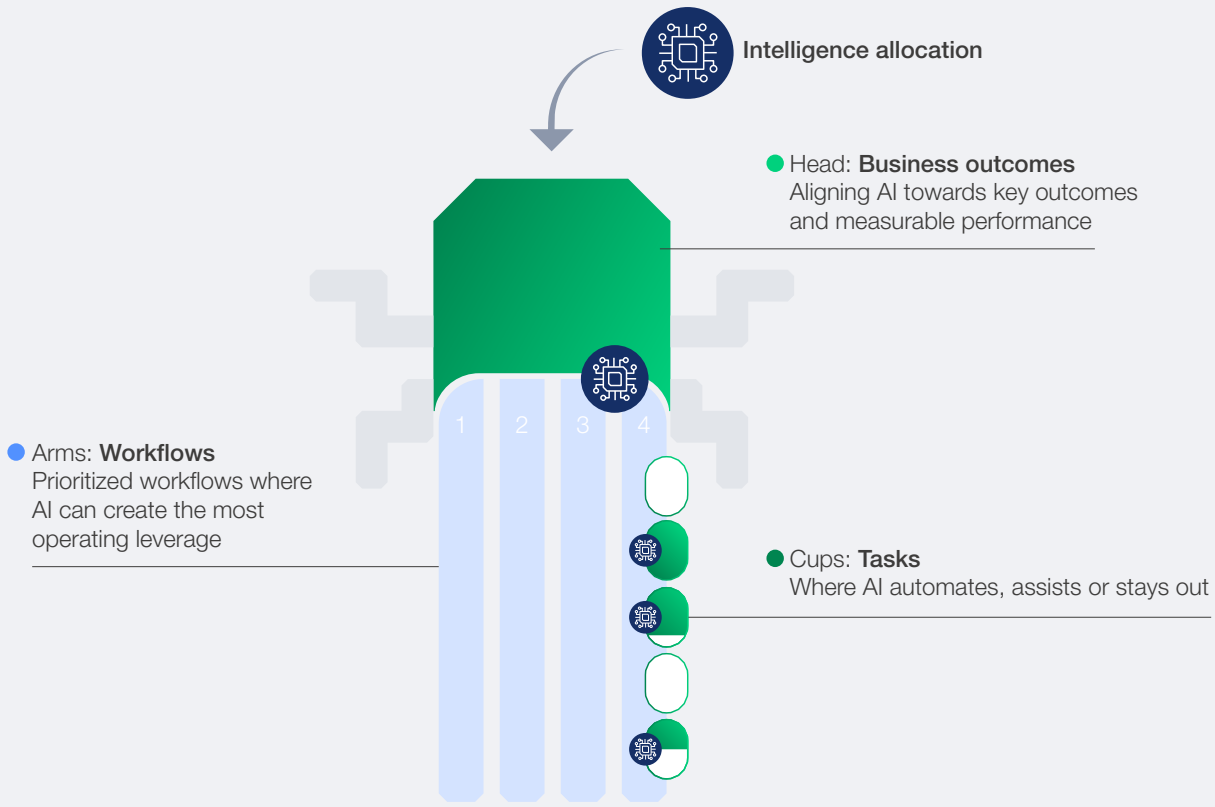
This logic scales. As organizations move from pilots to AI-first operations, the challenge shifts from running isolated experiments to systematically digitizing and redesigning workflows so that every cycle of execution feeds the intelligence engine.

The octopus is an icon and apt metaphor for distributed intelligence. This is because its intelligence is not concentrated in a single brain; of its roughly 500 million neurons, about two-thirds are allocated to its arms. The brain sets the intent; the arms sense and coordinate and the cups execute.[20,21] AI-first enterprises operate similarly: the intelligence engine sets direction and provides context, while intelligence is distributed across workflows, agents and teams to act.

Three levels connect this intent to execution:

- **Level 1**: Where to allocate intelligence
- **Level 2**: How to redesign work
- **Level 3**: Building intelligence-native operations

FIGURE 12 **From allocation to operational leverage**

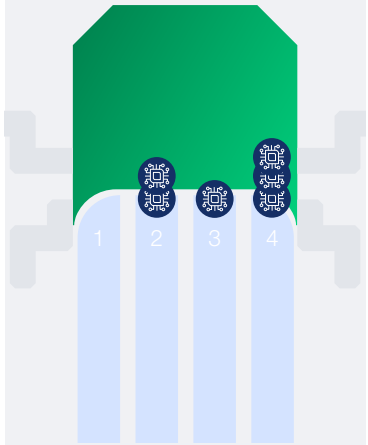


Three levels of intelligence allocation

1
Where to allocate intelligence

Aligning AI to business outcomes and prioritizing high-impact workflows

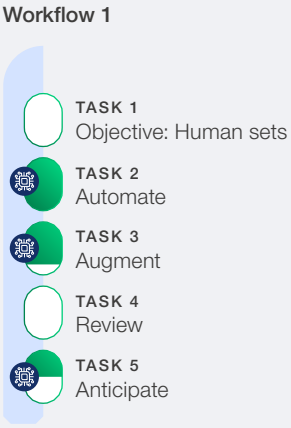
Outcomes to workflows



2
How to redesign work

Defining AI's role across processes and tasks and matching them to the right models

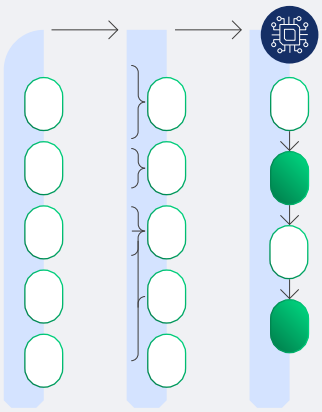
Workflows to tasks



3
Building intelligence-native operations

Building the systems of ontology, redesigning workflows and ensuring visibility to coordinate execution

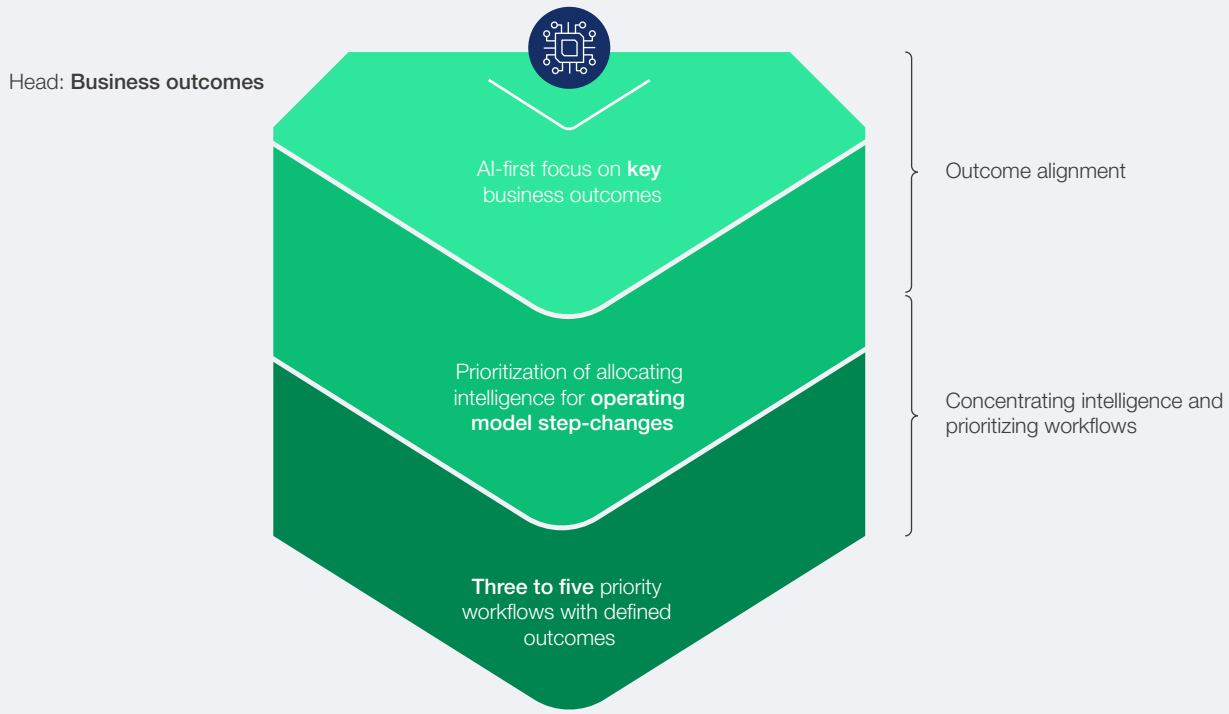
Orchestration



As organizations begin to treat intelligence like capital, the challenge shifts to allocating intelligence responsibly across all levels and thousands of tasks in million-dollar workflows.

3.1 Where to allocate intelligence

FIGURE 13 **Head – outcome-driven allocation**



Who leads this: The C-suite, chief executive officers (CEOs) and business unit leaders

Unless they are AI native companies, AI-first enterprises do not try to digitize everything at once. They identify the workflows where redesigning around intelligence will create the greatest step-change in performance and start there. The strongest starting points share three characteristics: **scale and repetition** (high-volume, continuous processes where improvements compound over time), **workflow friction** (processes slowed by handoffs, multi-step coordination or manual review) and **cognitive complexity** (decision-heavy or information-intensive tasks where AI can materially improve judgement and speed).

These characteristics create conditions where embedding intelligence delivers meaningful gains – and where the data generated by each cycle is rich enough to feed the engine and improve the next. This is why workflows such as customer support, procurement, underwriting, claims and troubleshooting often emerge as early priorities. Enterprises do not choose them because they are easy, but because redesigning them generates compounding returns.

CASE STUDY 3

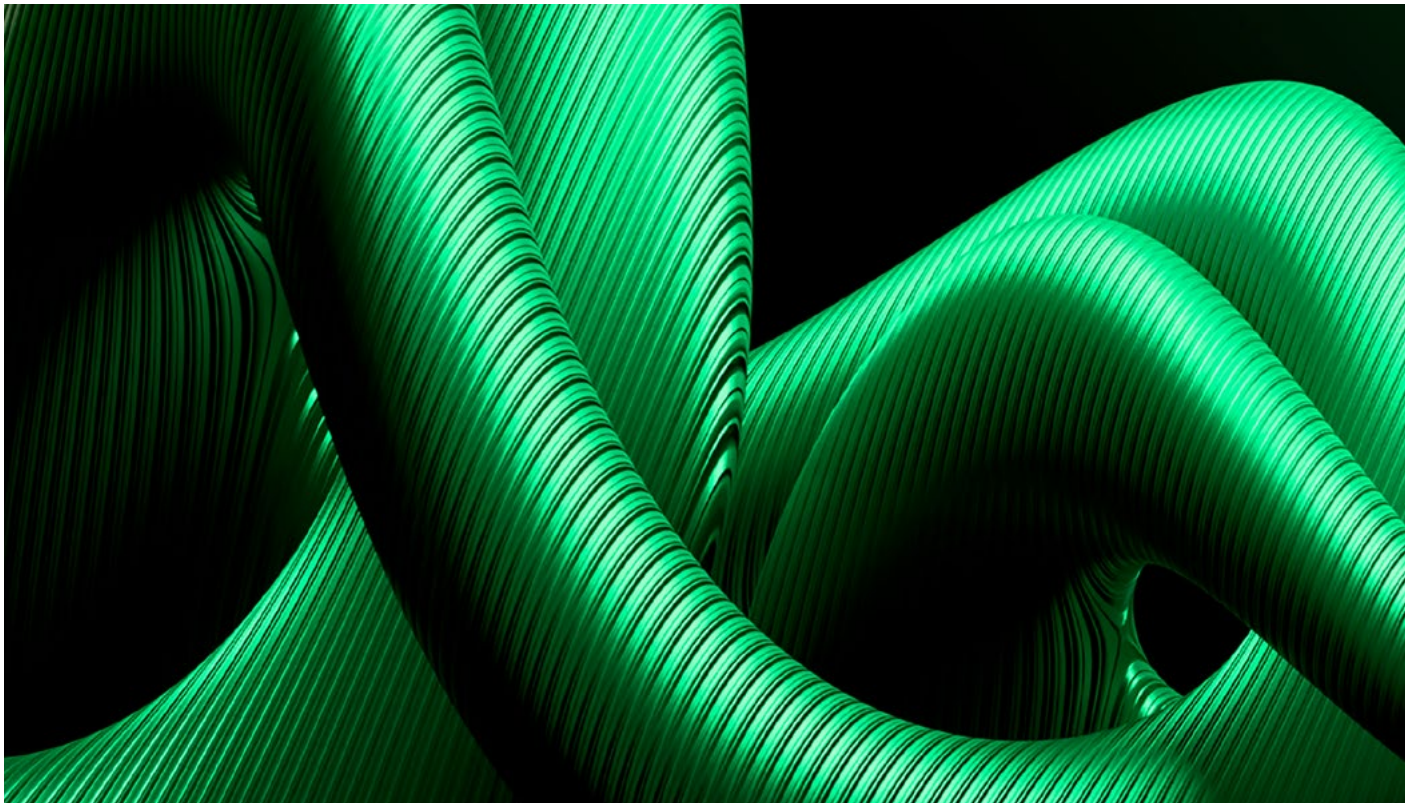
How Genesys operationalized AI through distributed execution

Genesys began with an AI task force of operational leaders, embedded across functions, responsible for ensuring initiatives were grounded in real workflows. An initial effort inventoried over 200 use cases and developed ideas about which tasks in each role AI could automate, creating a shared view of where AI could drive impact.

Genesys' focus has now shifted to execution through a standardized playbook that guides how teams:

- Identify and prioritize workflows
- Redesign how work is done
- Evaluate solutions (build, buy or activate)
- Measure value

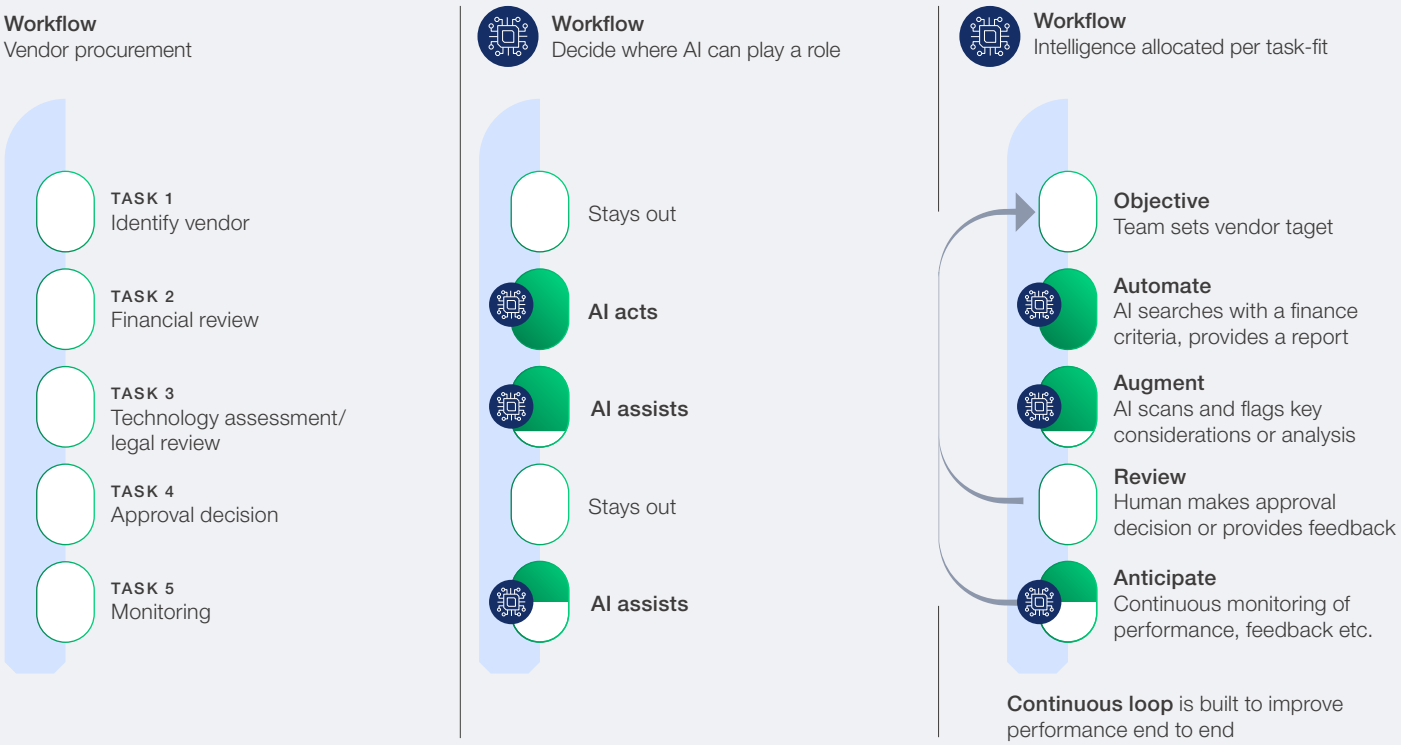
Execution is distributed. Teams move quickly within their domains. Demand generation has deployed agents across business development workflows at scale, while IT is embedding AI into service management. Progress happens in parallel, anchored on high-impact workflows.



3.2 How to redesign work

Who owns this allocation: Chief information officers (CIOs), business unit leaders, workflow owners

FIGURE 14 **Arms – workflow redesign illustrative example of vendor procurement**



Once the right workflows are identified, the work is to redesign them – not to layer AI on top of existing processes, but to rethink how work is structured, who owns what and where intelligence acts. **AI-first leaders start with two questions:** what can be reliably handled by AI and what must remain with people?

In practice, AI may handle high-volume, repeatable tasks, assist with complex decision-making where speed and consistency matter and support long-term monitoring and exception detection. Humans remain responsible for judgement, accountability and the decisions that carry the highest stakes or require contextual knowledge that AI cannot yet replicate. The boundary between the two is not fixed; it shifts as the intelligence engine improves.

Taktile illustrates what this looks like in regulated financial services, as mentioned in Block 1; AI agents are deployed at defined decision nodes across workflows such as underwriting, AML and fraud prevention. By layering rules, agents and human judgement, institutions make complex decisions more observable and governable while reducing the manual burden. Each output feeds back into the system.

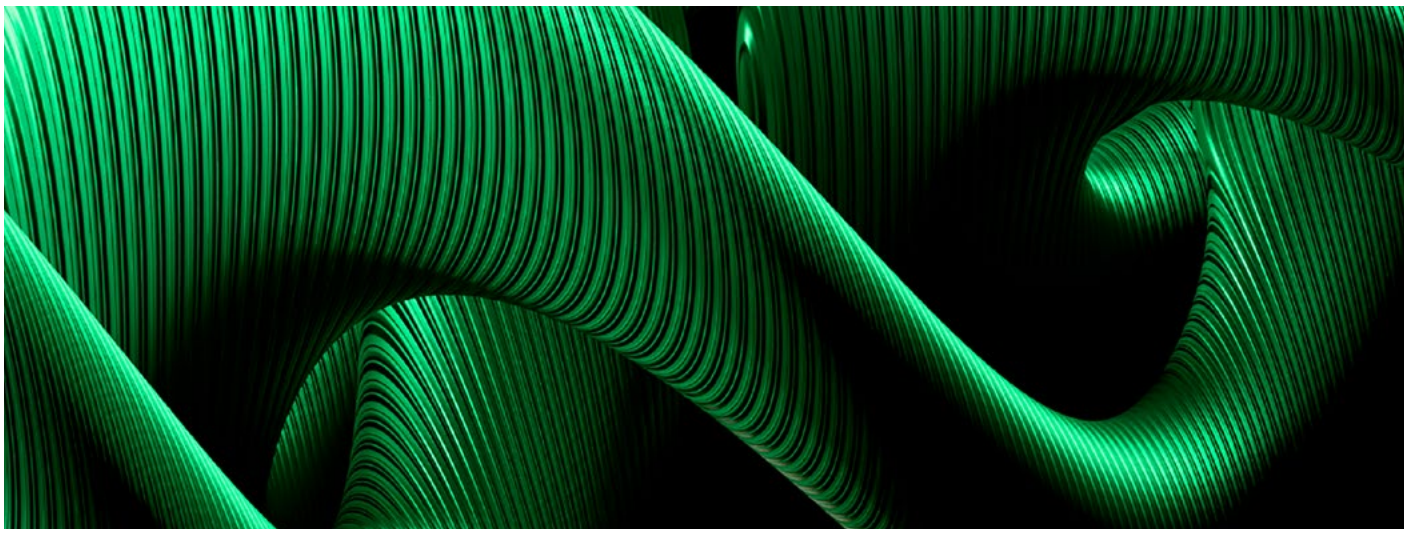
Levels of agency

Defining clear boundaries between automation and human augmentation is key when identifying how intelligence is allocated at the task level and workflows are redesigned.[22,23]

FIGURE 15 **Levels of human agency scale (adapted)**

Levels of human agency scale (HAS)
Defines levels of automation across team dynamics and degree of human involvement required

CUPS: **TASKS**	AI role	HAS level	Required human involvement	Collaboration dynamics
AI Acts	**Automation** – AI completes tasks	H1	AI agent handles the task entirely on its own without human involvement	**AI agent drives task completion** – the AI agent takes primary responsibility for task execution with no or minimal human oversight
		H2	AI agent requires input at key points to achieve better task performance	
AI Assists	**Augments** – AI enhances human capabilities	H3	Human and AI agent work together to outperform either alone	**Equal partnership** – human and AI agent collaborate closely throughout the task
		H4	AI agent requires human input to successfully complete the task	**Human drives task completion** – the human takes primary responsibility for task execution with varying levels of AI assistance
AI Stays out		H5	Task completion fully relies on human involvement	



Model task fit

Getting this right also requires matching each task to the right model. Not every task requires the most powerful or expensive model, and it is critical to balance compute constraints. AI-first enterprises build multi-model architectures that route work based on what each task actually requires – weighing accuracy, cost and latency. High-stakes decisions in regulated contexts demand near-zero error rates and human review. High-volume, repeatable tasks are best handled by smaller, cheaper models running at scale. The goal is not to use the most capable model everywhere. It is to use the right model for each task, and to keep improving that judgement as the system learns.

Accuracy:

Task type	Accuracy target	Why
Finance or health care example	No error	Major safety or regulatory risk
Customer-facing outputs	More than 90%	User trust; brand risk
Internal developer copilot	70–80%	Human review is a natural checkpoint

Cost:

Task type	Right model	Logic
Routing, classification, scheduling	Small language models (SLMs) (cheapest)	High volume, narrow scope
Complex reasoning, synthesis	Frontier LLM	Requires generalist capability; low frequency
Domain-specific	Own/tuned model (can be most expensive)	Privacy control IP-dependent

Latency:

Task type	Latency tolerance	What changes if slow
Real-time threat/fraud detection	Continuous	Threat already deployed
Customer-facing support agent	Less than two seconds	Experience degrades; abandonment
Internal research synthesis	Minutes	Workflow delayed; human waits
Drug discovery inference loop	Hours to days	Acceptable; depth matters more

3.3 Building intelligence-native operations

Who owns this allocation: Chief operating officers (COOs), CIOs, chief technology officers (CTOs), chief digital officers, business unit leaders, managers

The hardest part of the journey is not identifying the right workflows or redesigning individual tasks. It is building the operational infrastructure that allows intelligence to coordinate across workflows, improve continuously and run reliably at scale. This is what separates organizations that have successfully redesigned a handful of workflows from those that have rebuilt their operations around intelligence.

Three ways to orchestrate and design intelligence-native operations:

1. Making the business legible to the intelligence engine

AI-first enterprises digitize and codify how the business works – its workflows, decisions, rules, handoffs and actions – so that intelligence can reason about them, act on them and improve them over time. Building an ontology is central to this: it defines how the business operates, how data is used, how decisions are made and what actions can be taken. Without it, handoffs become inconsistent, decisions are harder to trace and the intelligence engine cannot operate reliably across workflows. With it, the engine can reason, improve and act consistently, and every new workflow added to the system strengthens it.

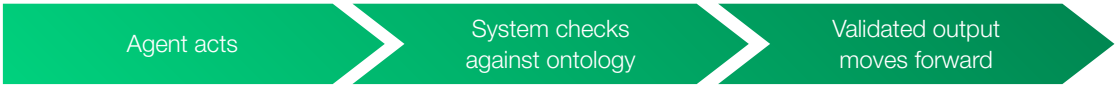
This enables:

- **Interpretability**: Outputs are expressed in business terms and can be traced to underlying data and rules.
- **Governability**: Decisions are checked against explicit rules, constraints and policies.
- **Connectivity at scale**: Teams, systems and agents can execute workflows consistently.
- **Economies of scale**: One ontology supports multiple workflows, reducing duplicates and making workflows easier to monitor and improve.
- **Operational digital twin/company world model**: Ontology turns enterprise activity into a live representation of the organization, showing how work flows and helping leaders monitor changes.

TABLE 5 **What ontology codifies**

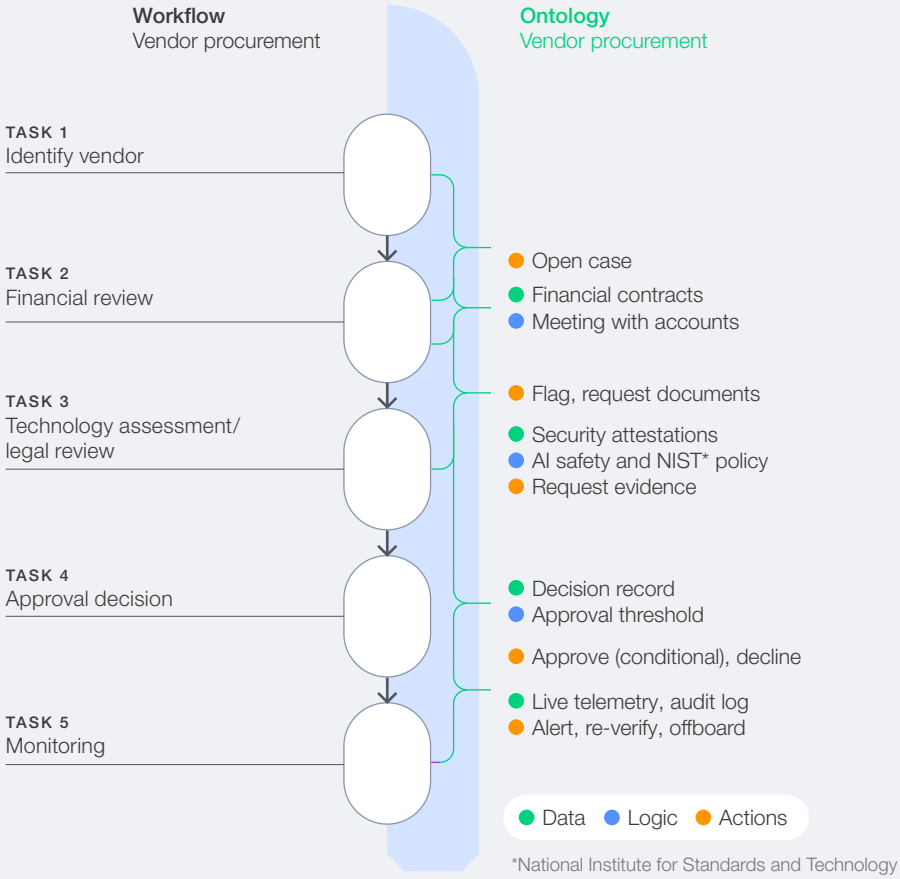
Layer	Question it answers	What is codified	What it enables
Data	What is the current state of the business?	Entities such as customers, documents, assets, policies and systems	Shared context for humans and AI
Logic	How should the business reason about it?	Relationships, rules, constraints and decision criteria	Consistent reasoning and verification
Actions	What can be done next?	Allowed actions, task routing, escalation paths, workflow triggers	Coordinated execution

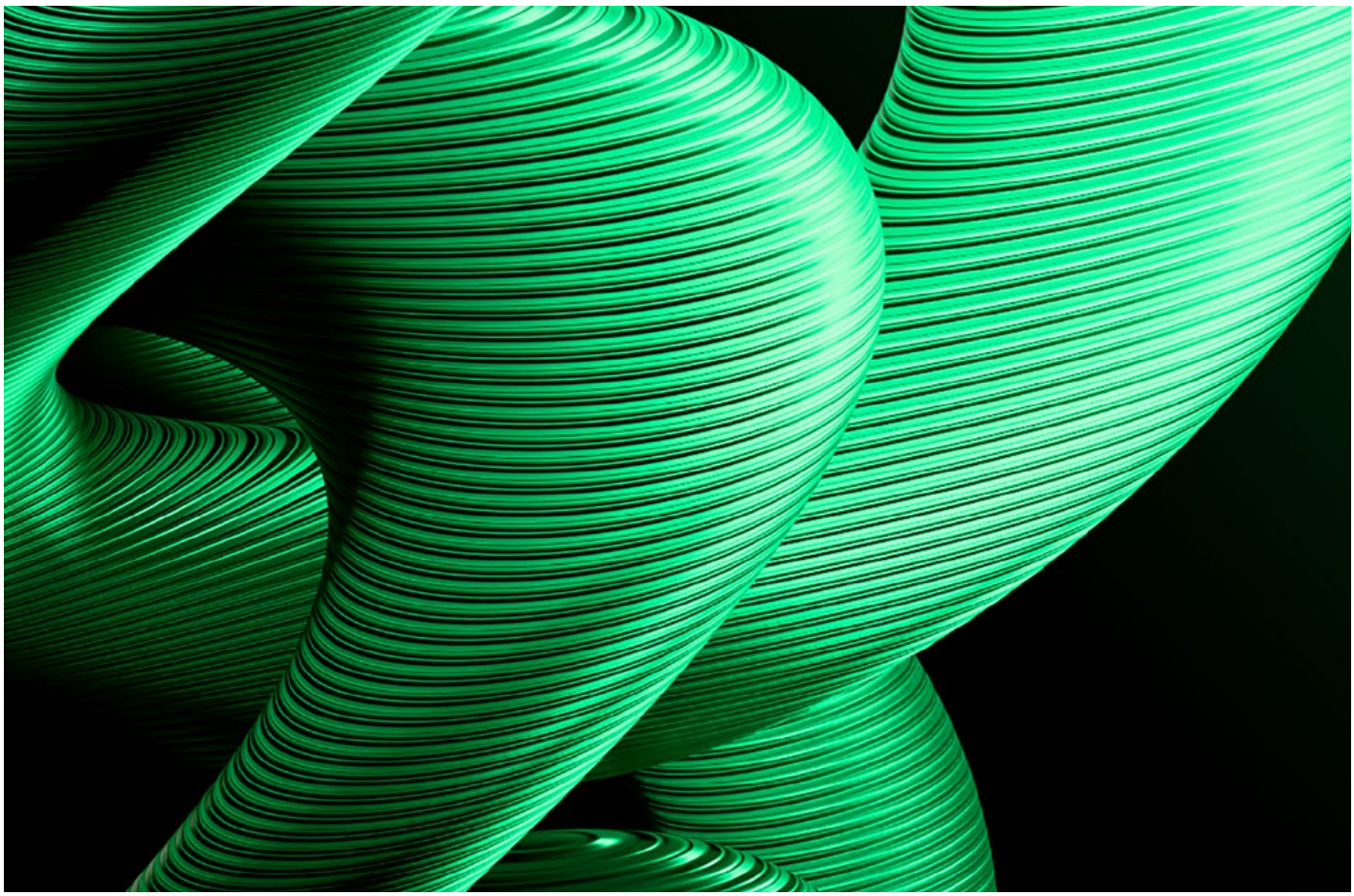
Source: Palantir. (2026). *The Ontology system.* https://www.palantir.com/docs/foundry/architecture-center/ontology-system.



By embedding business context and rules directly into the ontology, organizations move from prompt-driven automation to structured, verifiable workflows that allow the intelligence engine to run.

FIGURE 16 | **Workflow ontology – vendor procurement illustration**

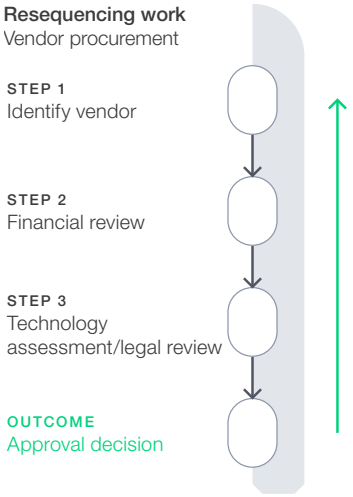
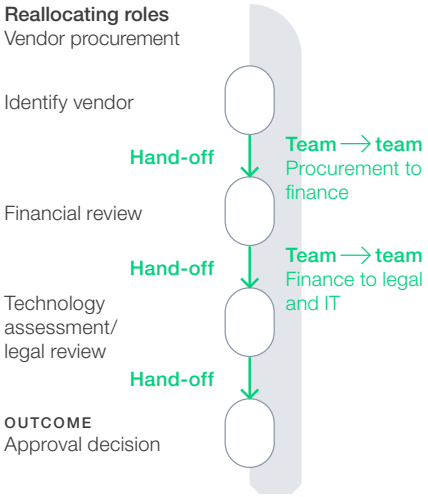
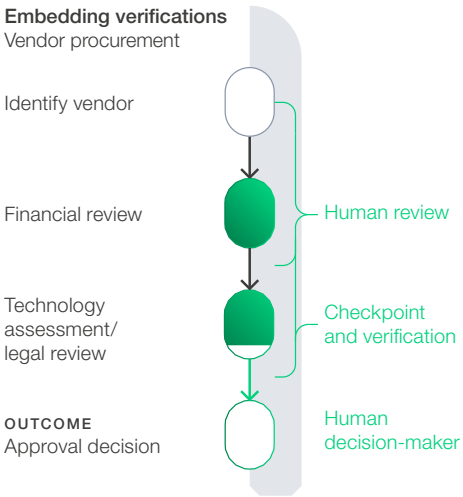




2. Redesigning workflows end-to-end

AI-first enterprises redesign end-to-end workflows around desired outcomes. This means resequencing tasks, redefining handoffs, embedding verification and assigning clear roles to humans and AI.

TABLE 6 **Tactics for redesigning workflows**

	Strategy	Tactics
	1 Resequencing work	Work backwards from the outcome to map the workflow. At each step, assign the right AI system or human and redesign the workflow around that structure.
	2 Reallocating roles	Identify handoff points across the workflow. These transitions often introduce friction, such as delays, duplication and inconsistent decisions. AI can improve flow through routing, low-stakes reviews and batch coordination.
	3 Embedding verifications	Verification is embedded into the workflow by using ontology to define who or what can act at each stage and preserving critical quality and compliance checkpoints with clear escalation when work returns to a human.

3. Designing for trust and operational visibility

As workflows become AI-driven and continuously evolve, visibility is essential to maintain control. Unlike traditional systems, agentic workflows do not follow fixed sequences. They adapt in real time. Without visibility, organizations lose the ability to understand performance, intervene when needed and improve the system.

What good visibility looks like:

- **Real-time execution**: Dashboard and monitoring systems that show how work is progressing across agents, humans and systems in real time
- **Traceability**: Clear visibility into decisions, including inputs, reasoning and outcomes
- **Intervention capability**: The ability to pause, override or redirect active workflows
- **System evolution tracking**: Visibility into how workflows change over time as agents learn and adapt

All of these factors become possible when organizations build agentic systems on clear evaluation, risk assessment and governance foundations, as outlined in the World Economic Forum's 2025 white paper, *AI Agents in Action: Foundations for Evaluation and Governance.*[24]

CASE STUDY 4

Cognizant: creating operational visibility at scale

Cognizant's approach to enterprise AI shows how organizations can build operational visibility and trust across both experimentation and day-to-day workflows. The organization launched a global, record-breaking "vibe coding" initiative (where users describe their intent to AI in natural language) to rapidly upskill its workforce and understand how AI could be deployed across the enterprise with greater visibility and control.

More than 53,000 employees across 40 countries experimented with AI-assisted development tools in sandboxed environments, providing visibility into use patterns, compute consumption, security risks and where AI was creating value across workflows. At first, intelligence was allocated by providing broad, lightweight token access before directing additional compute towards the strongest ideas for deeper development.

This emphasis on visibility and controllable scale use is also reflected in 1C, a multi-agent platform that connects hundreds of enterprise applications and AI agents into one searchable interface for 350,000 employees. The platform allows HR, IT and operational tasks to run through a single system, helping reduce fragmentation and create a more trusted operating layer for enterprise AI, reducing support tickets by 50% and increasing operational efficiency by 50%.

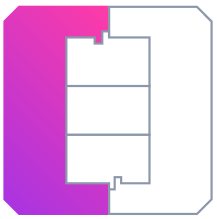
3.4 Operational leverage: the new business economics of intelligence

From operations to operational leverage

AI-first enterprises treat intelligence like capital and allocate it effectively across the organization. When intelligence is embedded across workflows and the engine runs reliably at scale, organizations can get more from the same operating base – increasing throughput, accelerating cycle times, expanding scope and improving quality simultaneously. But this does not mean costs automatically go down. As intelligence scales, companies must manage a new economic equation: the performance gains from intelligence-native operations versus the variable costs of delivering them at scale, including compute, model access, licenses and specialized AI roles.

The goal, however, is not to sustain this economic equation indefinitely – it is to reach a point where the distinction between "AI in workflows" and "the intelligence engine running operations" collapses entirely. When every workflow has been digitized and codified, when every cycle of execution feeds signals back into the engine and when the engine continuously improves what the organization can do, allocation becomes less a decision and more a property of the system itself. Intelligence is no longer deployed selectively. It is how the business operates.

BLOCK 4



Human-AI teaming

When structure follows collaboration, human-AI teams achieve what neither could design, decide or deliver alone.

Today's pioneers are crossing revenue thresholds with one-fifth to one-twentieth the headcount of prior software leaders:

- **Anthropic: $10 billion** in annualized run-rate revenue with **2,300 employees**[25]
- **Cursor: $1 billion** annualized run-rate revenue with **150–300 employees**[26,27]
- **Google: $30 billion** in annualized run-rate revenue with **32,000 employees**[28,29]
- **Salesforce: $30 billion** in annualized run-rate revenue with **79,000 employees**[30,31]

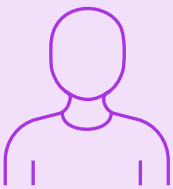
These early performance signals are compelling, but they also point to significant workforce disruption and the possibility of new enterprise formation and employment growth. Existing workforce metrics may not fully capture the impact of technologies that are still continuously developing. As intelligence becomes embedded across workflows, the differentiator shifts to talent: who is hired, how they're developed and how the organization is designed to direct intelligence towards outcomes.

This means enabling people to orchestrate, supervise and scale AI capabilities into measurable outcomes. If executed well, human-AI teaming can remove mundane work and shift human contribution towards judgement, creativity and orchestration. It unlocks otherwise hidden capacity, and the limiting factor shifts to how well human-AI teams are structured to collaborate, govern and compound performance.

An AI-first strategic talent advantage is built through three dimensions:

1
Growing human-AI talent

Developing the skills required to direct, evaluate and improve AI systems

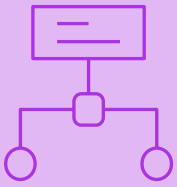


2
Organizing for human-AI teams

Structuring teams to integrate technical, domain and operational capabilities; organizing with new norms in leaner, cross-functional pods

3
Evolving the enterprise for AI-first execution

Designing the organization to scale intelligence across business units and workflows – federated AI ownership, flatter structures and the ability to operate continuously



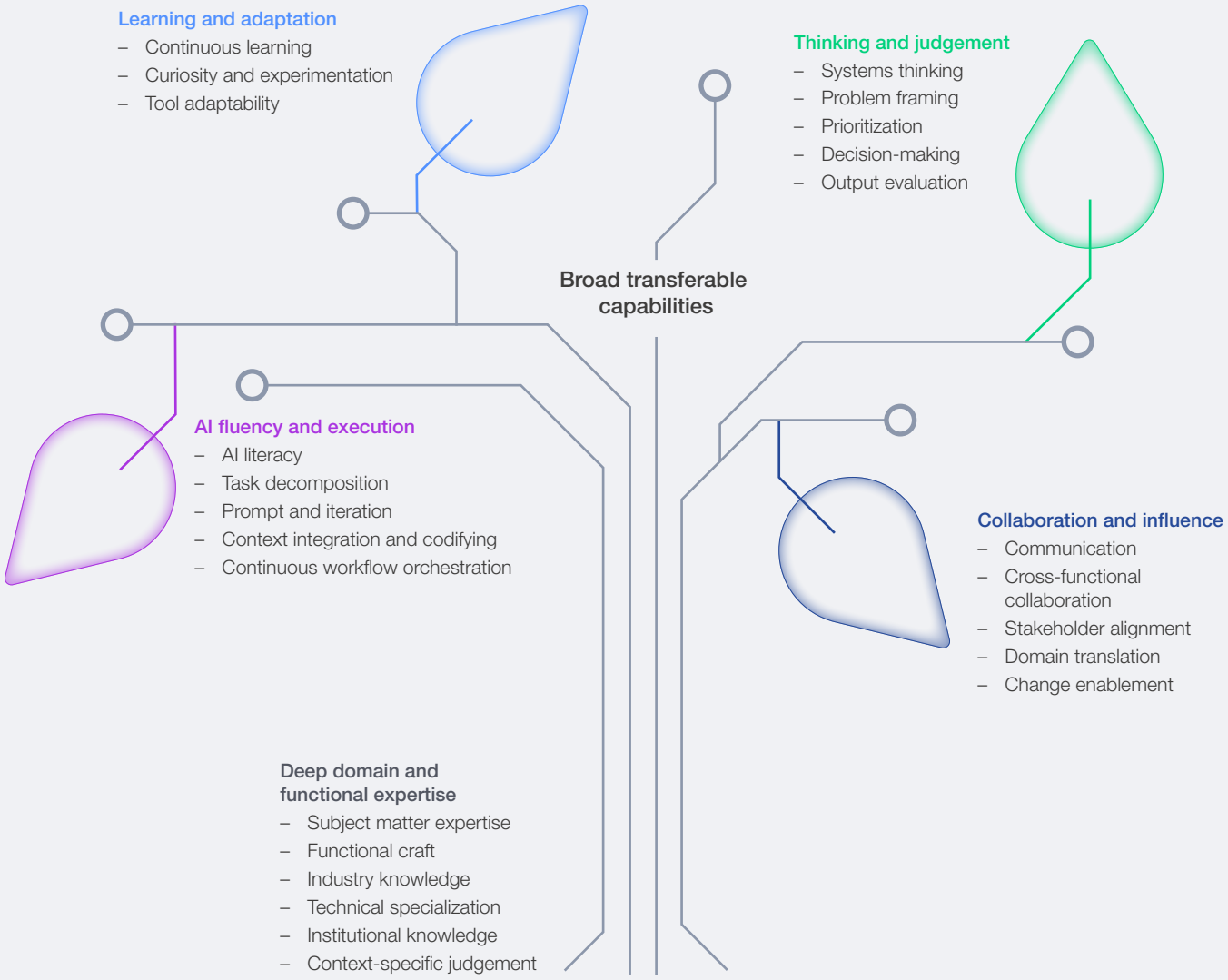
4.1 Growing human-AI talent

While most companies are not creating new roles in their organizations, AI-first organizations are already hiring for new profiles, such as design engineers, forward deployment engineers, evaluation specialists, AI researchers, safety engineers – all of which didn't exist a few years ago. Not every new title is relevant for every business. But this signals that the capability requirements for teams are shifting, reflecting how work is changing.

In these new roles and profiles, a consistent combination of capabilities is emerging, combining **breadth and depth**:

- Breadth enables individuals to frame problems and adapt to new tools, gaining system thinking to direct AI across domains.
- Depth provides domain expertise and judgement that AI cannot substitute.
- The executional layer between them, the technical "middle layer", is increasingly absorbed by AI, which is why both breadth and depth now carry more weight in who gets hired.

FIGURE 17 **The AI-first T-shaped talent tree**



Note: The AI-first T-shaped talent model was developed with AI innovation leaders on key skills represented in their organization during the AI-first workshop in October 2025.

How the skill life cycle changes

In a traditional skill life cycle, a person learns a skill, applies it and eventually retires it when the work changes. AI transforms that linear progression in two ways:

1. **Tools evolve faster than skills can stabilize**: Prompting habits, orchestration patterns and tool-specific techniques embedded into daily workflows can become outdated in months, sometimes overnight. Approaches that felt durable may need to be rebuilt as models, interfaces and capabilities shift.

2. **Deep domain expertise gets locked to old context**: Expert capabilities are deepened based on their context: roles held, tools used, work done. When AI changes the context, expertise requires refreshing.

Practices that sustain advantage

- **Codify individual expertise into reusable AI skills**: The best AI-first talent embeds their judgement, context, workflow logic and quality bar into the intelligence engine. This turns individual expertise into reusable capabilities that can be applied repeatedly, shared and improved over time.

- **Embed organizational practices into shared AI systems**: Team learnings, functional processes and ways of working are codified into the intelligence engine, reflecting how the organization operates. These shared "skills" can be collectively contributed to and continuously improved, allowing institutional knowledge to be applied and scaled across teams.

- **Manage skill atrophy**: As AI takes on more execution, human capabilities begin to weaken through disuse. Organizations often only discover how critical that tacit knowledge was when AI breaks. AI-first enterprises guard against this by keeping durable capabilities such as problem framing, judgement, systems thinking, exception handling and evaluation active, even when AI can perform much of the visible work. Skills can be retrained; judgement is harder to restore.

4.2 Organizing for human-AI teams

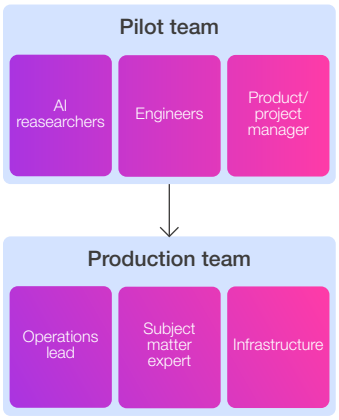
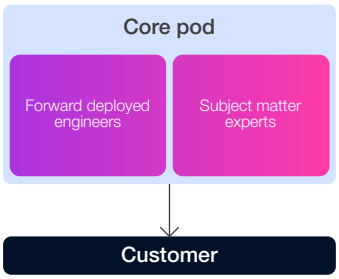
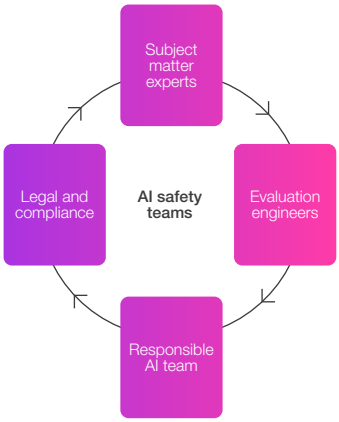
AI-first leaders realize that achieving production-grade AI requires new team structures and norms. The most effective AI-first teams are small, fast-moving units that bring together technical, domain and operational capabilities to build, maintain and harness the intelligence engine in real workflows.

For example, **a prototype may require around 1,000 skills to build, but getting it into production reliably can demand 10 times that.**[32] Getting from a promising prototype to a reliable product depends less on the technology than on how teams are organized around it. The most effective structure is a cross-functional team, typically fewer than 10 people, built around one product, workflow or customer problem, and operating with flat hierarchies.

AI-first enterprises organize teams such that the intelligence engine sits at the core while people operate at the edge. The engine provides every team with the context and route information needed for execution in real time, enabling them to navigate uncertainty, exercise judgement and bring subject-matter expertise, turning intelligence into business outcomes.

In practice, three new team types are emerging as must-haves for the AI-first enterprise.

TABLE 7 **AI-first team archetypes**

AI-first team archetypes	Pilot and production	Customer-embedded	AI safety
What is this team solving?	Scaling AI reliably across products, services and internal solutions	Accelerating speed-to-market and personalizing customer solutions	Operating within defined boundaries, internal policies and industry standards
Team composition			
Teamwork	– Built for production from day one – Continuous exchange, not a handoff	– Requirements shaped live – Live feedback collapses silos	– Safety engineered from the start – Predefined evaluations and guardrails

AI-first team norms that enable performance:

Every team has its own context and approach to AI. However, AI-first teams have a few common dynamics:

FIGURE 18 **AI-first team norms**

1

Group norms

- AI to be attempted as the first tool for problem solving
- Teams to always ask: can AI do this better, quicker or more accurately?
- AI outputs to always be tested against known data/figures, good business sense and common sense
- AI drives, but humans set the direction, answering the question “where are we going?” while AI determines how to get there

2

Metrics of success

- Percentage adoption for relevant tasks (goal of 50%)
- Improved productivity or efficiency (goal of more than 20%)
- De-siloing – does work across teams take less time?
- Quality control and evaluations become essential

3

Core values

- AI-at-scale is the objective and most important problem to solve
- AI is a tool to supercharge key goals
- Learn from others deploying AI solutions and break silos
- Be vigilant on quality (report hallucinations)
- Continuous learning and agile practices unlock speed

4

Role

- AI is a tool, not a friend; it will either produce content or delegate work as predefined by its function
- AI supports and redesigns processes, but does not define what gets done at this point (human interaction needed)
- Multiplier effect disappears without true domain experts being able to guide AI output

5

Quality standards

- AI outputs or process must be better than human performance
- Human craft and taste is required to ensure differentiation and quality
- Review everything AI makes with human eyes/perspectives
- Prioritize safety and privacy during product build and refinements
- Adhere to guidelines on disclosing what is AI-generated and what is human-generated

4.3 Evolving the organization for AI-first execution

Designing how intelligence moves through the enterprise is becoming a new source of organizational advantage. As organizations use AI to make context, work and decisions more visible and machine-readable, they can evolve beyond structures built to route information and execute work at fixed times. Shared intelligence layers, distributed ownership and autonomous operations allow organizations to coordinate faster across silos, move decisions closer to value creation and scale AI without losing control. For this reason, the CEO must own the AI strategy. Redesigning work across the enterprise cannot be delegated to only a centralized technical team or treated as a side initiative.

Decentralized and coordinated execution: the federated organization design model

The most effective operating model for an AI-first enterprise is a **federated one**. A CEO-led central intelligence team provides shared tooling, infrastructure and governance, while business units retain ownership of budgets, execution and AI decisions so implementation stays close to the problem.

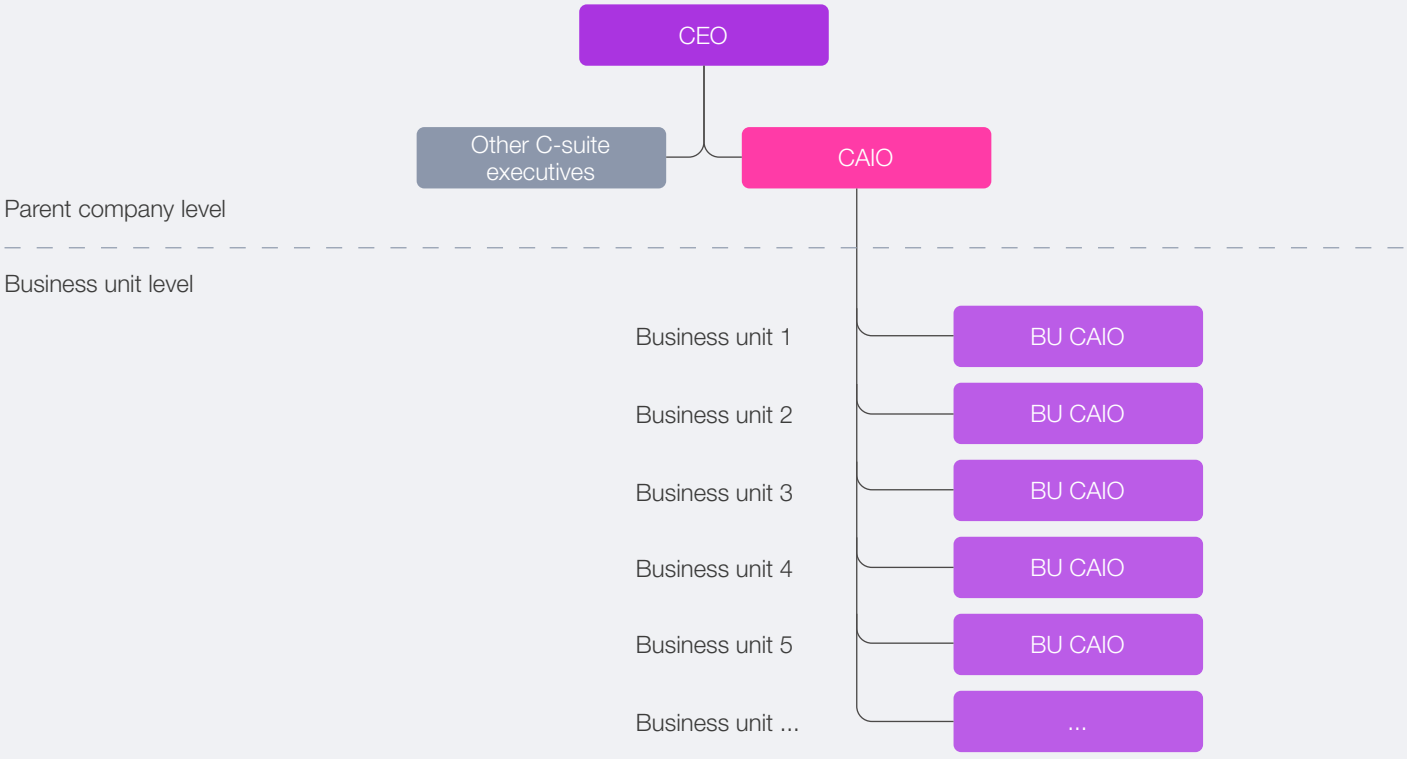
The model works because it separates two decisions:

- **The centre decides** what infrastructure the enterprise runs, building shared capabilities such as approved models, data platforms, evaluation tooling and safety guardrails.
- **Business units decide** where to apply AI, fund their own use cases and own the resulting return on investment in their P&L.

This separation increases speed and accountability. When a central team owns the roadmap, requests pile up, delivery slows and teams closest to the customer lose control over prioritization. In practice, chief AI officer (CAIO) leads are embedded within each business unit leadership team. They bring central standards, tooling and deployment playbooks into the business while feeding lessons back to the centre. This allows them to scale successful solutions across units, fund common model needs and resolve adoption bottlenecks with expert support.

This approach supports enterprise-wide coordination without sacrificing speed, ownership or creativity.

FIGURE 19 **AI-first federated deployment model**



CASE STUDY 5

Rakuten: enabling over 70 business services through a federated approach

Rakuten's AI & Data Division (AIDD), which reports to the CEO, operates using a federated model designed to accelerate AI deployment across more than 70 business units. This model is supported by three reinforcing mechanisms:

- **AIDD functions as an internal AI platform provider**, developing shared AI capabilities including customized LLMs and SLMs, AI-powered search platforms, recommendations, personalization, customer service and multilingual agents that business units can adopt through self-service or with support from the AIDD.
- **Embedded CAIOs in each business unit** maintain dual reporting lines to both the business unit head and the group chief AI and data officer, helping translate enterprise AI strategy into business-specific execution while coordinating across the broader Rakuten AI ecosystem.
- **Business-led investment prioritization enables business units**, beyond centrally funded shared infrastructure, to prioritize and fund their own AI initiatives through their respective P&Ls, keeping AI investments closely tied to business impact and operational needs.

Leveraging the intelligence engine as a coordination backbone

The intelligence engine becomes the company's coordination backbone when tribal knowledge, documents and operational capabilities are connected through a shared context layer, enabling live signals from work to be captured, interpreted and acted on. It functions as a real-time system of record for company activity, aggregating and routing information that was traditionally coordinated across teams and managerial layers. This shifts organizations away from layered structures towards outcome-oriented roles, where individual contributors and directly responsible individuals can access the context and operational assets needed for execution. As the system takes on more of the coordination burden, teams can operate with greater autonomy, pooled intelligence and faster decision cycles.[33]

BLOCK 5

New value creation

When intelligent products and experiences define how value is delivered, it creates market dynamics.

The fundamentals still apply – having a strong value proposition

The intelligence engine does not stop at the boundary of internal operations. How a company positions its capabilities in the market – what it offers, how customers experience it and how they engage with it – determines what flows back into the engine. AI-first enterprises design their market-facing layer not just to acquire and retain customers, but to generate the data that makes the engine smarter with every cycle.

This changes what it means to have a strong value proposition. AI's novelty may drive trials among first-time users, but only clear and repeatable value drives retention. AI greatly reduces switching costs, so keeping customers is the greatest challenge. AI apps have high customer churn, with fewer than 1 in 16 monthly subscribers paying, which is 33% worse than the rate for non-AI apps.[34] Regardless of whether a company is AI-native, AI-first or AI-enabled, customers still ask the same questions: whether the AI tool solves their problem, whether they can trust it, whether it is worth what they are paying or why they should choose it over the alternative.

AI-first enterprises position the new capabilities of the intelligence engine for their market and customers through AI-native product design, trust engineered into the system and new forms of entering the market.

5.1 Designing product

Intentionally design how intelligence capabilities are packaged so people can use them intuitively, effectively and confidently in flow. This means looking beyond the chat interfaces to the user's and customer's actual objectives, journey and environment, and designing experiences where intelligence is introduced only where it adds value.

The objective is not to maximize AI use, but to apply intelligence where it improves outcomes. As AI capabilities increase, they do not automatically create better results. They can also create noise, friction and "AI slop" – low-quality, generic output that erodes rather than adds value – when intelligence is poorly applied.

TABLE 8 **AI-first design principles for packaging new capabilities**

Goal of design	What it means in practice
Task to modality interaction	The right AI modality to use depends on context, whether that means photo and annotation in the field, mobile guidance for frontline managers or hands-free voice in clinical settings.
Create starting points	Most users do not begin with a perfect instruction; contextual actions, templates and guided entry points often work better than an empty prompt. Prebuilt structures, templates, mapped components and examples help new AI users start.
Progressive disclosure	Good AI interfaces reveal complexity and workflows step by step instead of overwhelming users with long outputs all at once. Queuing, scaffolding and staged guidance are ways of achieving this.
Personalized feedback and retrieval	Learn from prior interactions and offer context-aware guidance that nudges deeper thinking rather than blind trust. AI systems should help users organize, revisit and build on prior work instead of burying useful outputs in chat history.
Collaborative environments	AI should support how teams diverge, converge, review and align – improving communication and collaboration around shared goals without adding noise or fragmentation.
Intentional personality	Users may anthropomorphize AI and build distinct relationships with different systems, so teams should be deliberate about how much personality the interface projects.
Cultural interaction preferences	Interaction patterns are not universal. What feels intuitive in one market may feel unfamiliar in another, so the intelligence engine should be tested and adapted across cultural contexts.

CASE STUDY 6

Contextere bringing AI-first troubleshooting to the factory floor

Contextere, an industrial AI company, improved factory-floor troubleshooting with a voice-first interface built around how frontline workers already operate. By replacing a 47-minute information-gathering process with near-instant context assembly, the system reduced troubleshooting time by up to 80%, showing that AI performance gains come from meeting users where they are rather than forcing them into new workflows.

5.2 Prioritizing trust

Build the controls, boundaries and transparency that make customers willing to adopt and rely on it. For AI-first companies, trust is not a layer added after launch; it is part of how the product is packaged from the beginning.

In practice, teams make trust visible in the product itself. In a shopping assistant, that means showing why an item was recommended, what source the sizing or review information came from, where uncertainty remains with a confidence score and requiring the user to confirm before checkout.

TABLE 9 **Four principles of trust in AI**

Trust principle	Why it matters	What it looks like in practice
Protect identity, credentials and control	When AI acts on a customer's behalf, trust depends on limiting exposure of sensitive data and preserving customer control.	Scope credentials and permissions to each interaction, minimize data exposure and make it clear what the system can access and do. For sensitive deployments, keep models inside customer-controlled infrastructure.
Match accuracy to what is at stake	The higher the consequence of failure, the higher the need for accuracy, oversight and human review.	Apply stricter thresholds and more human intervention in regulated or customer-facing use cases, while allowing more autonomy in low-risk tasks like brainstorming or scheduling.
Make confidence and boundaries visible	Trust grows when users understand what the system knows, how reliable it is and where its limits are.	Show confidence levels, communicate degradation and define clear boundaries for what the system is designed to do.[35]
Make it verifiable and auditable	Users are more likely to trust AI when they can verify how outputs were produced and what evidence supports them.	Provide tracking, visible reasoning or progress signals and auditable evidence for high-stakes decisions.

CASE STUDY 7

A unified platform spanning the patient journey

Ant Group developed a multimodal AI platform designed to support clinical decision-making and patient consultation across different stages of care.

At its core is a medical-grade LLM capable of processing text, medical images, audio and video. This allows the system to reason across symptoms, laboratory results and imaging data within a single framework. A privacy-preserving data layer encrypts and harmonizes clinical information from hospitals, primary care clinics and consumer health devices, enabling secure use in regulated environments.

The platform assists both clinicians and patients. Before a visit, patients can use the AI assistant to help record symptoms and schedule appointments. During the visit, it provides convenient medical treatment for users throughout the entire process and offers customized intelligent triage services to enhance the user's healthcare experience. Following diagnosis and treatment, the system also delivers follow-up care, medication guidance and rehabilitation planning services, thereby establishing a continuous feedback loop rather than a one-time interaction.

5.3 Positioning in the market

Every AI-first company, regardless of sector or business model, faces the same question: how do you position intelligence in the market so that customers adopt it, trust it and return to it? The answer depends on where intelligence sits in the offer: whether it enhances an existing product, becomes the product itself, reshapes a platform or workflow, or operates as invisible infrastructure underneath the customer experience. These choices determine what customers pay for, where value accrues and what kind of business is being built.

How to position intelligence in the market today

This applies to any AI-first company, whether it is a manufacturer embedding AI into after-sales service, a financial institution redesigning underwriting or a software company shipping AI-native products. The question is not whether to position AI, but how deliberately to do so. The five archetypes listed in Table 10 capture the range of choices, from AI as an enhancement layer to AI as the invisible medium through which an entirely new product category becomes possible.

TABLE 10 **Five archetypes for positioning AI in the market**

Archetype	What it means in practice	Source of value
AI as a feature	Makes an existing product smarter	Better user experience and performance
AI as the product	The model or application is what is purchased	Direct utility, output or expertise
AI as process and platform	Reshapes workflows and coordination	Productivity, standardization and increasing returns of scale
AI as infrastructure	Powers rails, protocols and execution layers	Enablement and ecosystem dependency
AI is invisible	What the user experiences is not “AI” – it is a new interface, behaviour or medium made possible by it	Creation of new user experiences and product categories made possible

Illustration through agentic capabilities

Beyond their current positioning, AI-first enterprises must also prepare for a market in which agents increasingly participate in discovery, evaluation and transactions on behalf of users. This does not require being in the business of selling agents – it requires understanding how agent-mediated interactions will reshape where and how customers engage, decide and buy. Whether a company sells industrial equipment, financial services or software, the channel through which customers arrive and commit is changing.

To illustrate how this evolution unfolds, consider how agent participation in markets matures over time. At the earliest stages, agents complete routine transaction steps after a user has already made a decision. As capability increases, agents translate user needs into search and options, then begin remembering preferences and prior interactions. At more advanced levels, agents evaluate and select within user-defined guardrails. At the most advanced level – which is still emerging – the system anticipates needs and acts before the user has made an explicit request. Most of the market today sits at the second level. The trajectory towards higher levels of agent autonomy is progressing, however, and how companies position themselves in that progression will increasingly shape where and how they capture value.[36]

FIGURE 20 | **Five levels of agentic commerce maturity (adapted from Stripe)**

Level		Description
1	**Eliminating web forms**	The agent completes routine transaction steps after the user has decided.
2	**Descriptive search**	The user describes needs; the agent translates them into search and options.
3	**Persistence**	The agent remembers preferences, constraints and prior interactions.
4	**Delegation**	The agent evaluates and selects within user-defined guardrails.
5	**Anticipation**	The system predicts needs and acts before explicit demand.

The business-to-business (B2B) context is the natural starting point for companies navigating this shift. Enterprise transactions operate within clearer rules, procurement workflows and decision criteria, giving agents well-defined roles to play. Agents can be trusted to source, compare and act within established parameters – reducing friction without requiring the level of personal trust that consumer contexts demand. Consumer markets are more complex: preferences are less stable, personal data is more sensitive and trust must be built across a wider range of users, platforms and vendors.

Looking ahead, three models of agent ownership are emerging, each with different implications for how companies position themselves:

- In the first, users own and control their own agent, which acts on their behalf across providers and platforms – open-source personal agent ecosystems such as OpenClaw already point in this direction.
- In the second, a vendor or marketplace embeds an agent within its own ecosystem to guide discovery and purchase within that environment – Amazon's Rufus is an early example of this model in practice.
- In the third, an independent intermediary operates across merchants and providers, comparing options and orchestrating decisions on the user's behalf – OpenAI's shopping research in ChatGPT and Google's AI shopping experiences point towards this direction. These models are in the early stages and will likely coexist, but understanding which model a company is building for, or being built against, is the relevant strategic question.

The central question for new value creation is how intelligence is designed, trusted and positioned in the market. This remains a still-emerging frontier, as new interfaces, experiences, agentic capabilities and business models come to market. For AI-first enterprises, the strategic challenge is to connect these products and services back to the intelligence engine, so that customer interactions create immediate value while also generating feedback that strengthens performance, improves business outcomes and compounds over time.

Design your own AI-first blueprint

This work was conducted to understand how becoming AI-first changes how a business creates and delivers value.

Strategyzer's **Business Model Canvas**[37] was a vital resource for exploring this. It provides a universal language for founders and leaders to describe how organizations create, deliver and capture value, and a structured lens to examine what is truly changing in an AI-first enterprise.

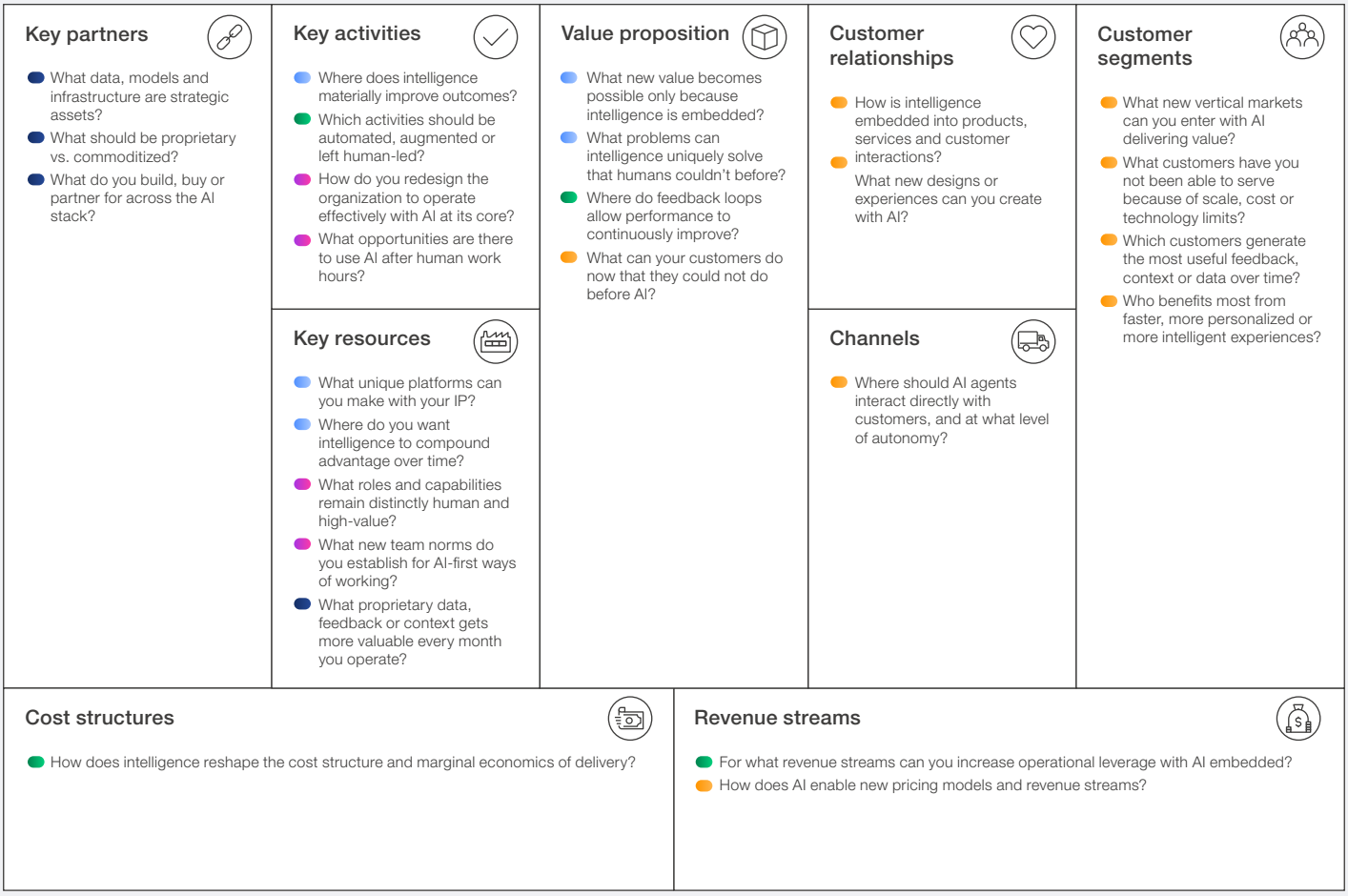
What emerged is that **the building blocks of the Business Model Canvas remain intact**: value propositions, customers, operations and revenue models still define every business.

What changes for an AI-first enterprise is **how each block is designed, what it contains, how they interact** in a system shaped by intelligence and **the questions asked to define and redesign them.**

Extending the canvas with AI-first fundamentals

The five AI-first fundamentals extend the Business Model Canvas by updating it for a world where intelligence is abundant and programmable. Each fundamental maps directly to a core part of the canvas. Taken together, the Business Model Canvas and the AI-first fundamentals form a practical blueprint for designing AI-First businesses.

FIGURE 21 **AI-first business model canvas**



Business model canvas

AI-first business model canvas instructions: Use the questions on the previous page to reflect on how AI will reshape every aspect of how you think about your business.

Take 30 minutes to reflect and answer these questions, then review to identify key themes you have surfaced. These observations will power a robust AI strategy for your organization.

AI-first fundamentals

- Intelligence engine
- Adaptive AI technology stack
- Operations redesign
- Human-AI teaming
- New value creation

Conclusion

The emergence of AI-first enterprises is one of the most consequential organizational shifts of today, and it is still in the early stages. While AI-first and AI-native organizations are demonstrating new forms of speed, scale and operational leverage, it is not yet clear which models will prove most resilient or dominant over time.

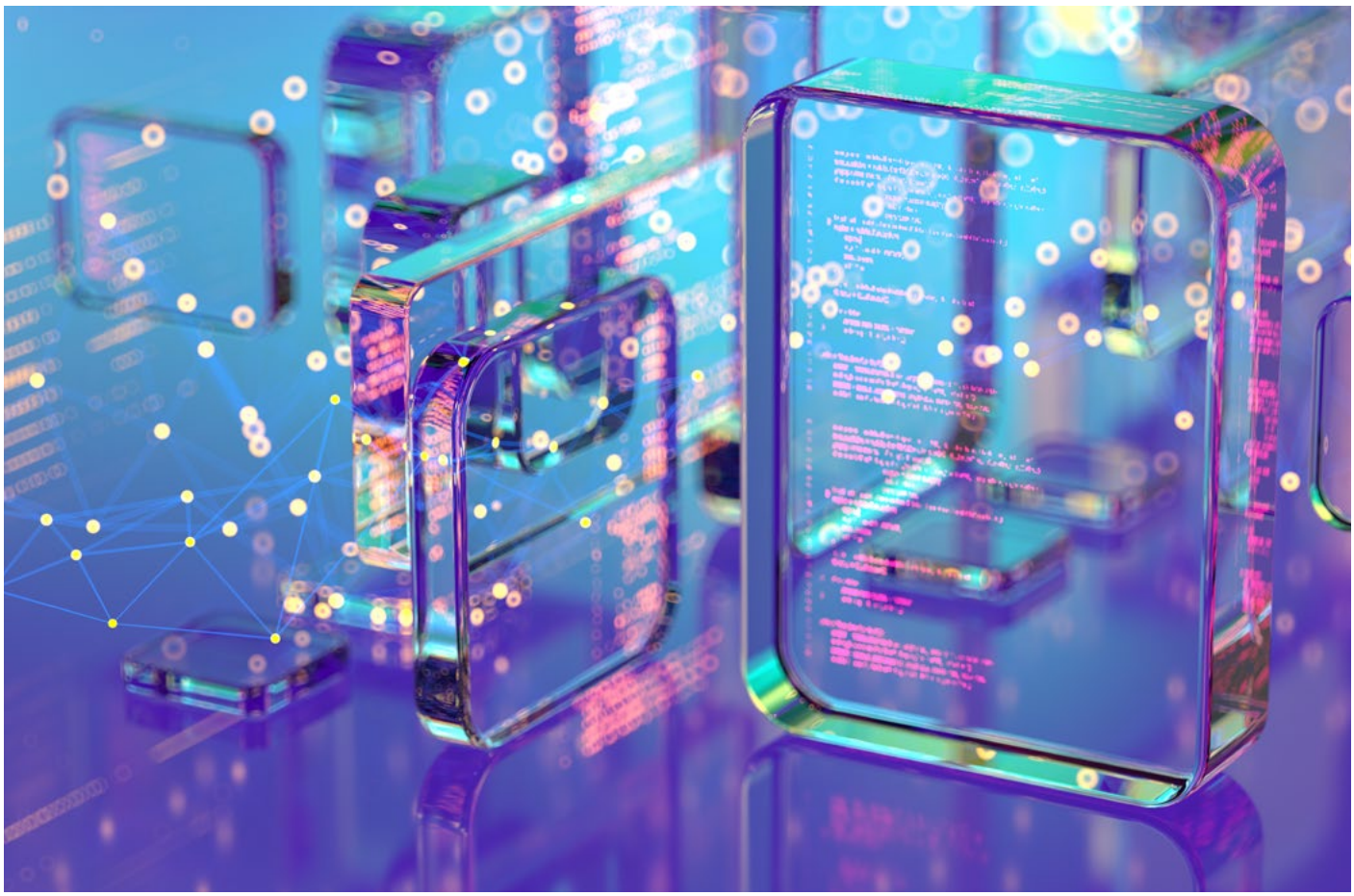
The question for leaders is not whether to become AI-first, but how fast and where to start. The window to build the foundations – the intelligence engine, the operating model, the talent, the market positioning – is open, but it will not stay open indefinitely. Organizations that treat this moment as an experimentation phase risk finding themselves structurally behind competitors that are already compounding.

For business leaders, the challenge is not to assume a single winning model, but to build the capacity to learn and adapt. As a bridge to the end state, one practical approach is to run parallel operating models – developing AI-first workflows, teams and systems alongside existing ones – and systematically measuring the differences in performance. This allows organizations to identify where intelligence creates real advantage and where traditional models remain effective.

For entrepreneurs, the opportunity is expanding as intelligence becomes more accessible. As access to AI capabilities broadens, differentiation will increasingly come from how organizations design their products, workflows and business models and their unique IP and expertise. Advantage will be defined by distinctive execution, domain insight and the ability to translate intelligence into real-world outcomes.

For policy-makers, the priority is to closely monitor how these operating approaches evolve and their broader impacts. It is critical to understand where they deliver value, where risks emerge, how competition evolves when intelligence compounds differently across organizations and geographies, and how they affect different sectors and populations over time.

The World Economic Forum's AI-First Enterprises workstream will continue to convene leaders across business, government, civil society and academia to track these developments, share emerging practices and deepen understanding of how intelligence is transforming organizations and markets.



Contributors

World Economic Forum

Abhishek Balakrishnan
Initiatives Lead, AI Innovation, Centre for AI Excellence

Maria Basso
Head, AI Applications and Impact, Centre for AI Excellence

Cathy Li
Head, Centre for AI Excellence; Member, Executive Committee

Kearney

Michael Römer
Senior Partner & Global Lead, Activate: Digital and Analytics, Kearney

Roland Scharrer
Partner

Matthew Totlis
Manager

Silicon Foundry

Prutha Atre
Director

Mark Menell
Managing Director

Acknowledgements

Mohamed Amer
Co-Founder and Chief Executive Officer, Claryo

Henrique Arutin
Head, Data and Artificial Intelligence, Banco Bradesco

Amir Banifatemi
Chief Responsible AI Officer, Cognizant

Brandon Barbello
Co-Founder and Chief Operating Officer, Archetype AI

Gabriel Batstone
Co-Founder and Chief Executive Officer, Contextere

Arya Bolurfrushan
Founder and Chief Executive Officer, Applied AI Corporation

Hannah Calhoon

Vice-President of AI Innovation, Indeed

Alexia Cambon
Senior Director of Research, Microsoft

Trotter Cashion
Head of AI Product, Riot Games

Sebastian Cesario
Co-Founder and Chief Technology Officer, BforeAI PreCrime

Navin Chaddha
Managing Director, Mayfield Fund

Paul Cheek
Co-Founder and Chief Executive Officer, AI-Driven Enterprise Institute

Kevin Chung
Chief Strategy Officer, Writer

Michael Costigan
Vice-President, Salesforce Futures, Salesforce

Scott Crawford
Vice-President of Product and Analytics, GoDaddy

Lauren Dunford
Co-Founder and Chief Executive Officer, Guidewheel

Carlos Eduardo de Almeida Mazzei
Chief Technology Officer, Itaú Unibanco

Yan-David (Yanda) Erlich
General Partner and Head of Tech Investing, B Capital

Matthew Fitzpatrick
Chief Executive Officer, Invisible Technologies

Sarah Franklin
Chief Executive Officer, Lattice

Marta Ghiglioni
Head, Endeavor San Francisco Hub

Mark Gorenberg
Chair, MIT Corporation, Massachusetts Institute of Technology (MIT)

Matt Hawes
Founding Engineer, AIP, Palantir

Andrew Hogan
Head of Insights, Figma

Xuedong Huang
Chief Technology Officer, Zoom

Deepti Kalra
Co-Founder and Chief Executive Officer, Echos

Kian Katanforoosh
Founder and Chief Executive Officer, Workera

Rem Koning
Associate Professor and Co-Director for D3 Institute, Harvard Business School (HBS)

Guillaume Lardeux
Senior Vice-President and Head of Transformation, Genesys

Benjamine Liu
Chief Executive Officer and Co-Founder, Formation Bio

Natalie Mao
Vice-President, Artificial Intelligence and Data, Rakuten

Bryan McCann
Co-Founder and Chief Technology Officer, SuSea

Ryan Meadows
Chief Revenue Officer, Lovable

James Milin
Co-Founder and Chief Executive Officer, Workhelix

Arjun Prakash
Co-Founder and Chief Executive Officer, Distyl AI

Priyanka Rajkumar
Senior Vice-President, Strategy & AI for Payment Services, ExlService

Keshav Reddy

Founder and Chief Executive Officer, Equal Identity

Umesh Sachdev
Chief Executive Officer and Co-Founder, Uniphore

Dan Saffer
Assistant Professor, Carnegie Mellon University

Emily Sands
Head of Data and AI, Stripe

Erin Schaefer
Chief Operating Officer, General Catalyst

Tim Sears
Chief AI Advisor, HTEC

Rudina Seseri
Founder and Managing Partner, Glasswing Ventures

Uljan Sharka
Founder and Chief Executive Officer, Domyn

Mihir Shukla
Chief Executive Officer, Chairman and Co-Founder, Automation Anywhere

Michael Spranger
President, Sony AI

Murali Subbarao
Global Vice-President, AI Platform, ServiceNow

Melissa Valentine
Tenured Associate Professor of Management Science and Engineering, Stanford University, and Senior Fellow, Stanford HAI

Jason Warner
Chief Executive Officer and Co-Founder, Poolside

Maik Taro Wehmeyer
Chief Executive Officer, Taktile

Andrew Wells
Chief Data and AI Officer, NTT DATA North America

Alex Wiltschko
Founder and Chief Executive Officer, Osmo Labs

Timothy Young
Chief Executive Officer, Jasper AI

World Economic Forum

Audrey Duet
Head, Data and AI Innovation, Centre for AI Excellence

Tarik Fayad
Lead, Strategic Integration, MENA, Centre for AI Excellence

Connie Kuang
Lead, Technology Convergence, Centre for AI Excellence

Stephanie Smittkamp
Coordinator, Centre for AI Excellence

Production

Laurence Denmark
Creative Director, Studio Miko

Will Liley
Editor, Studio Miko

Jean-Philippe Stanway
Designer

Endnotes

1 Solano, J. P. (2018). *What electricity revolution tells us about our future with A.I*. Medium. https://medium.com/data-science/what-electricity-revolution-tells-us-about-our-future-with-a-i-69eff44295d4.

2 The Henry Ford. (n.d.). *Highland Park Plant Engine- Generator, 1915-1916*. https://www.thehenryford.org/collections/explore/artifact/167313?AssetId=THF18671.

3 Damron, W. (2025). Gains from factory electrification: Evidence from North Carolina, 1905–1926. *Explorations in Economic History*, vol. 96. https://www.sciencedirect.com/science/article/abs/pii/S0014498325000014?via%3Dihub.

4 Wharton University of Pennsylvania. (2025). *Accountable Acceleration: Gen AI Fast-Tracks into the Enterprise*. https://ai.wharton.upenn.edu/wp-content/uploads/2025/10/2025-Wharton-GBK-AI-Adoption-Report_Full-Report.pdf.

5 Maslej, N., Fattorini, L., Perrault, R., Gil, Y., et al. (2025). *Artificial Intelligence Index Report 2025*. https://arxiv.org/abs/2504.07139.

6 Deloitte. (2026). *State of AI in the Enterprise: The untapped edge*. https://www.deloitte.com/content/dam/assets-shared/docs/about/2025/state-of-ai-2026-global.pdf.

7 Derived from 35 AI leader interviews conducted by the World Economic Forum's Centre for AI-Excellence.

8 Heath, A. (2025). *AI startups are turning their revenue into recruiting bait*. The Verge. https://www.theverge.com/column/826172/ai-startup-arr-numbers-sierra-bret-taylor?utm_source=chatgpt.com.

9 Bessemer Venture Partners. (2024). *From Startup to Centaur: The founder's roadmap to $100 million ARR*. https://www.bvp.com/assets/uploads/2024/04/From-Start-to-Centaur-The-founders-roadmap-to-100-million-ARR-Bessemer-Books-Edition-040924.pdf.

10 Derived from 35 AI leader interviews conducted by the World Economic Forum's Centre for AI-Excellence.

11 Ibid.

12 Brohan, M. (2025). *Gartner: AI agents will command $15 trillion in B2B purchases by 2028*. Digital Commerce 360. https://www.digitalcommerce360.com/2025/11/28/gartner-ai-agents-15-trillion-in-b2b-purchases-by-2028/.

13 This information was derived from expert interviews and validated with the chief executive officer of Claryo.

14 Waymo. (2025). *Safe to Deploy: How We Know The Waymo Driver Is Ready For The Road.* https://waymo.com/blog/2025/06/safe-to-deploy/.

15 Harvey. (2024). *Harvey Co-Builds Custom Model for Tax with PwC*. https://www.harvey.ai/blog/harvey-co-builds-custom-model-for-tax-with-pwc.

16 PwC. (2025). AI-powered deals: PwC's latest M&A technology platform reimagines the way clients do deals. https://www.pwc.com/gx/en/news-room/press-releases/2025/aipowered-deals-pwcs-latest-mna-technology-platform-reimagines-the-way-clients-do-deals.html.

17 Stripe. (2025). *Applied AI: Delivering on growth*. https://stripe.com/gb/sessions/2025/applied-ai-delivering-on-growth; Affaki, J. (2025). *Using AI to optimize payments performance with the Payments Intelligence Suite*. https://stripe.com/blog/using-ai-optimize-payments-performance-payments-intelligence-suite.

18 Anthropic. (2025). *Anthropic acquires Bun as Claude Code reached $1B milestone*. https://www.anthropic.com/news/anthropic-acquires-bun-as-claude-code-reaches-usd1b-milestone; Anthropic. (2026). *Introducing Claude Design by Anthropic Labs*. https://www.anthropic.com/news/claude-design; Anthropic. (2026). *Introducing Claude for Small Business*. https://www.anthropic.com/news/claude-for-small-business.

19 Isomorphic Labs. (2026). *The Isomorphic Labs Drug Design Engine unlocks a new frontier beyond AlphaFold*. https://www.isomorphiclabs.com/articles/the-isomorphic-labs-drug-design-engine-unlocks-a-new-frontier.

20 Sivitilli, D. M., Strong, T., Weertman, W., Ullmann, J., et al. (2023). Mechanisms of octopus arm search behavior without visual feedback. *Bioinspiration & Biomimetics*, vol. 18, issue 6. https://pubmed.ncbi.nlm.nih.gov/37793413/.

21 Shaw, A. and Berndt, L. C. S. (2026). *The Embodied Octopus: Distributed Intelligence and Active Inference in a Flexible Body*. Computational Psychiatry and Neuropharmacological Systems Lab Department of Psychology, University of Exeter.

22 Shao, Y., Zope, H., Jiang, Y., Pei, J., et al. (2025). *Future of Work with AI Agents: Auditing Automation and Augmentation Potential across the U.S. Workforce*. Stanford University. https://arxiv.org/abs/2506.06576.

23 Ibid.

24 World Economic Forum. (2025). *AI Agents in Action: Foundations for Evaluations and Governance*. https://www.weforum.org/publications/ai-agents-in-action-foundations-for-evaluation-and-governance/.

25 Garfinkle, A. (2025). *How Anthropic grew—and what the $183 billion giant faces next*. https://fortune.com/2025/12/04/how-anthropic-grew-what-the-183-billion-giant-faces-next/.

26 AI Funding Tracker. (2026). *Cursor Revenue: How the $29B AI Coding Tool Makes Money*. https://aifundingtracker.com/cursor-revenue-valuation/.

27 Capoot, A. (2025). *AI startup Cursor raises $2.3 billion funding round at $29.3 billion valuation*. CNBC. https://www.cnbc.com/2025/11/13/cursor-ai-startup-funding-round-valuation.html.

28 Statista Research Department. (2026). *Annual revenue of Google from 2002 to 2025*. https://www.statista.com/statistics/266206/googles-annual-global-revenue/.

29 Stock Analysis. (2026). *Alphabet Inc. (GOOGL)*. https://stockanalysis.com/stocks/googl/employees/.

30 WallStreetZen. (2026). *Salesforce Revenue*. https://www.wallstreetzen.com/stocks/us/nyse/crm/revenue.

31 Macrotrends. (2026). *Salesforce: Number of Employees 2012-2026*. https://www.macrotrends.net/stocks/charts/CRM/salesforce/number-of-employees.

32 Derived from an AI-first workshop conducted by the World Economic Forum, with 40 participants of AI innovation executives and experts.

33 Claude Public Artifacts. (n.d.). *The Intelligence Layer as Coordination Substrate: A Pressure Test*. https://claude.ai/public/artifacts/11378d88-0c0b-4a55-bc10-89962969728a.

34 RevenueCat. (2026). *State of Subscription Apps 2026.* https://www.revenuecat.com/pdf/state-of-subscription-apps-2026-sosa.pdf.

35 Li, J., Yang, T., Zhang, R., Liao, Q. V., et al. (2025). Understanding the Effects of Miscalibrated AI Confidence on User Trust, Reliance, and Decision Efficacy. https://arxiv.org/pdf/2402.07632.

36 Collison, P. and Collison, J. (2026). *Stripe's 2025 annual letter.* Stripe. https://stripe.com/annual-updates/2025.

37 Strategyzer. (2026). *The Business Model Canvas.* https://www.strategyzer.com/library/the-business-model-canvas.

WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744
contact@weforum.org
www.weforum.org