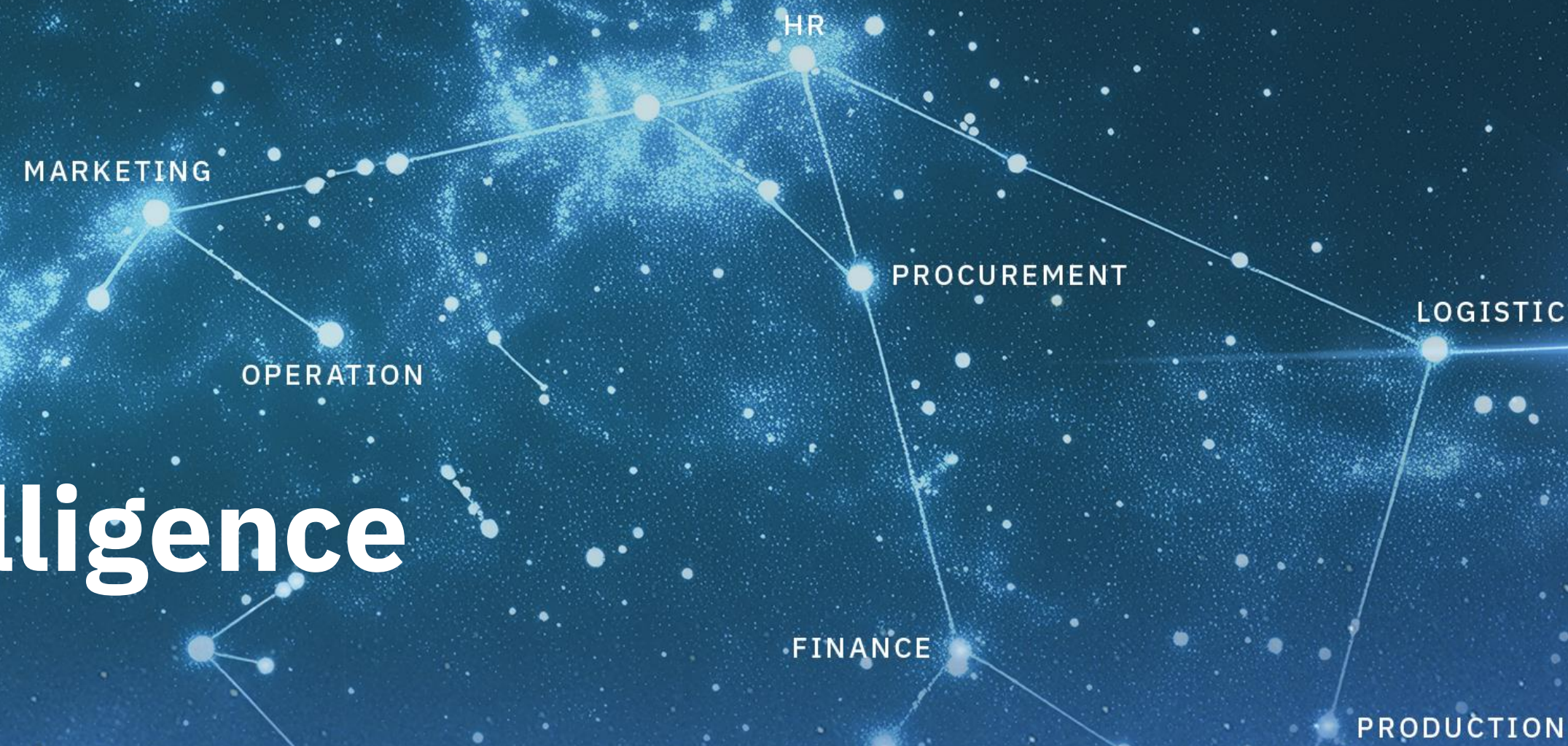


IBM's Artificial Intelligence Strategy and Vision

Jaysen Kratz

Data & AI Specialist,
IBM Technology



Today's talk

Accelerate AI
productivity across
the enterprise

Propel the next wave of AI Productivity

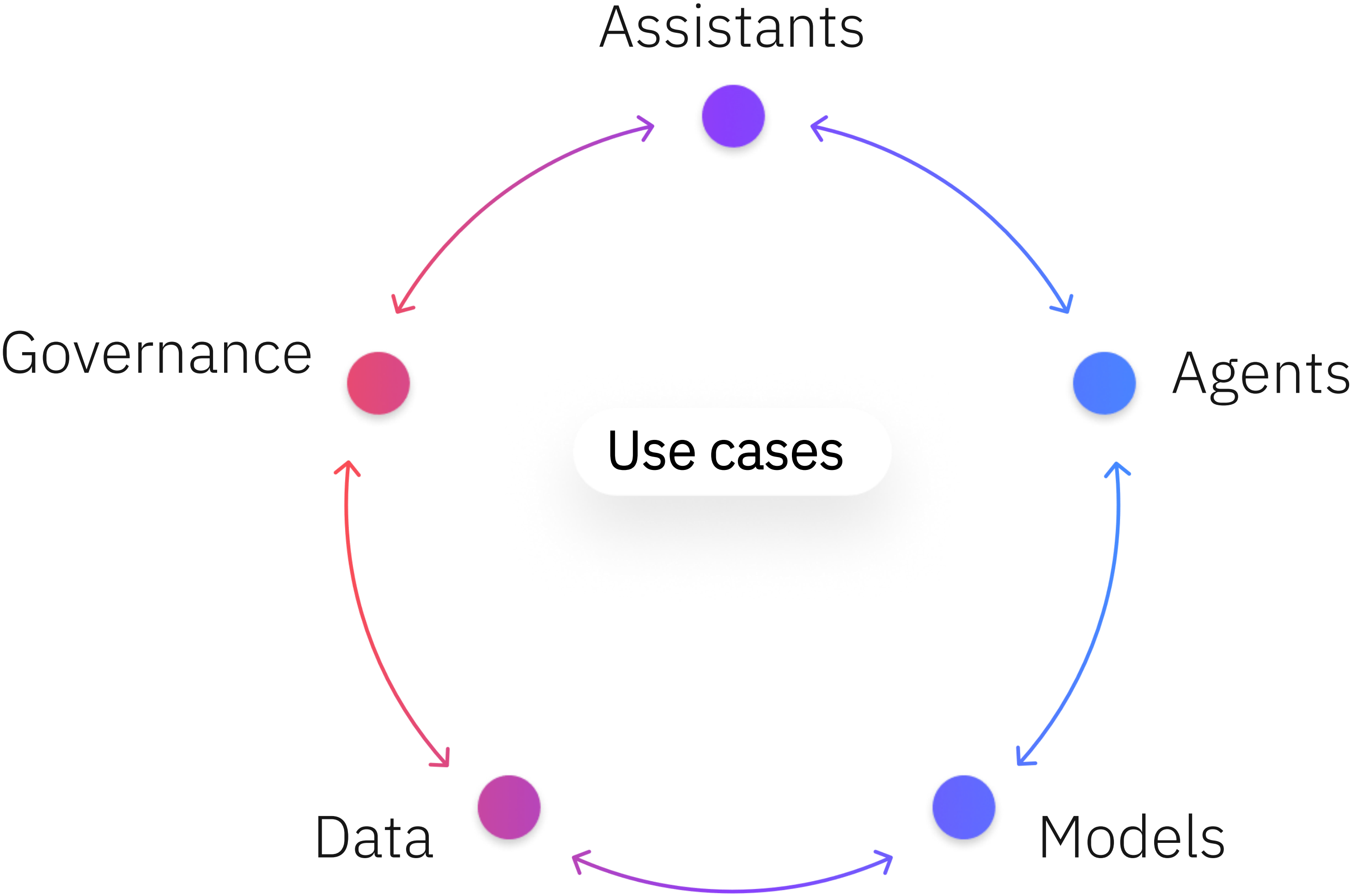
Key areas IBM is focusing on with clients and partners

Why IBM for a Data Platform

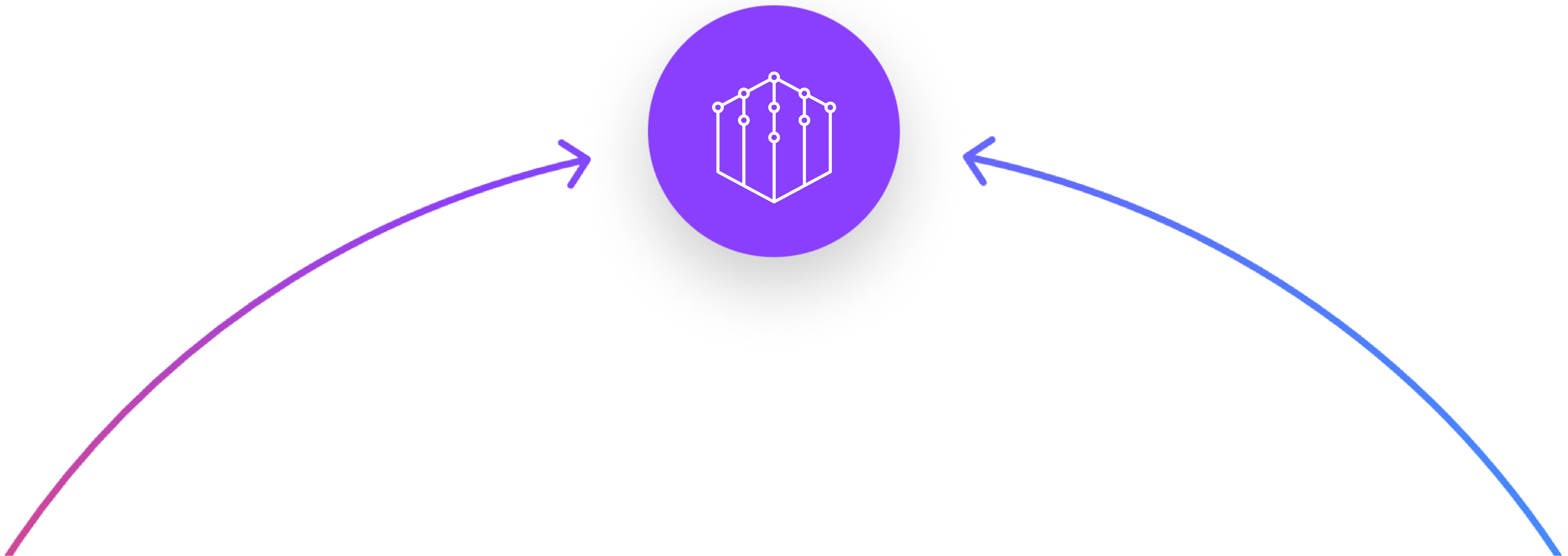
The path to \$3B
(and counting)
through unlocking
productivity gains
across IBM

AskEPM 100% of questions answered watsonx.ai	Support case summarization USD 1.1 million annual operational savings watsonx	AI sales assistant 50-70% of repetitive tasks automated watsonx Orchestrate
IT Operations 70% of Ansible Playbook content generated 6x faster watsonx code assistant	Supply Chain USD 150 million reduction in supply chain costs watsonx	Marketing 67% content creation efficiency gain for supported asset types watsonx.ai
HR 94% inquiries resolved by AskHR watsonx assistant	Finance 95% benchmark accuracy of touchless forecasting watsonx	IT Support 90%+ reduction in time to solve blocked invoices watsonx.ai Watsonx Orchestrate

AI building blocks
to the future



Assistants



Leading domains
and use cases for
AI assistants

Human Resources

Transform HR operations
and employee self-serve.

- Employee support
- Talent acquisition
- Onboarding
- Performance management

Customer Care

Transform CX and enable
customers to self-serve.

- Customer self-serve
- Modernize contact center
- Contact center insights
- Agent assist

Sales

Boost sales performance and
automate lead qualification.

- Sales enablement
- Prospecting support
- Opportunity management

Procurement

Simplify everyday workflows.

- Supplier assessment
- Source to pay
- Procurement insights

Finance

Expedite order processing,
fulfillment, invoicing,
and dues.

- Order to cash
- Expense management
- Payroll reconciliation

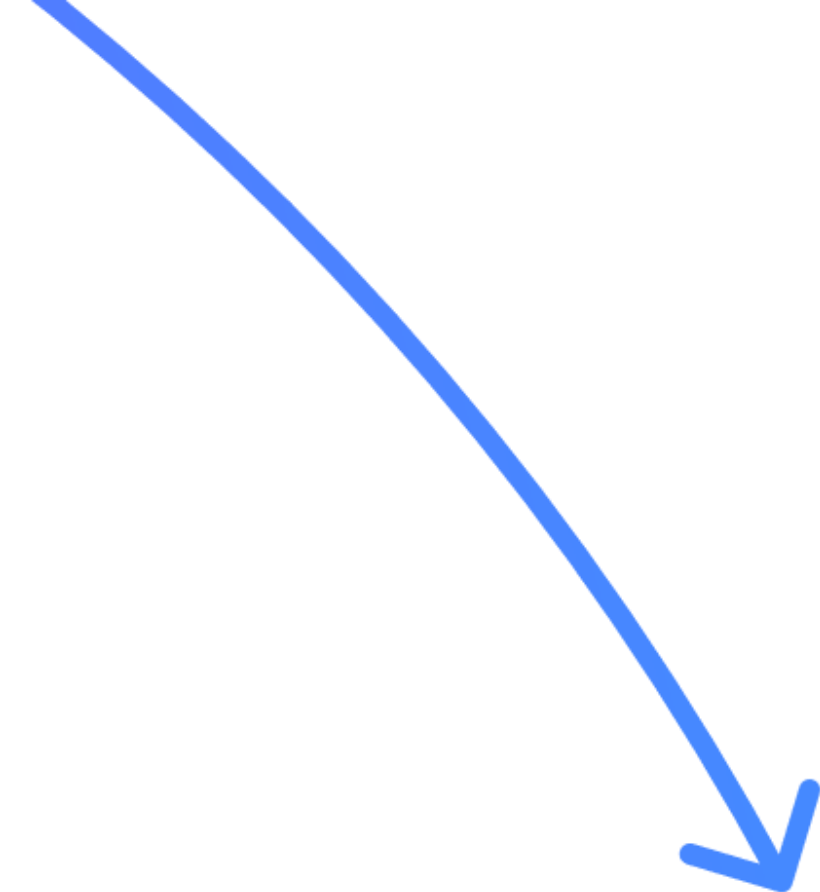
Supply Chain

Boost employee productivity
by orchestrating daily tools.

- Sourcing support
- Inventory management
- Sustainability



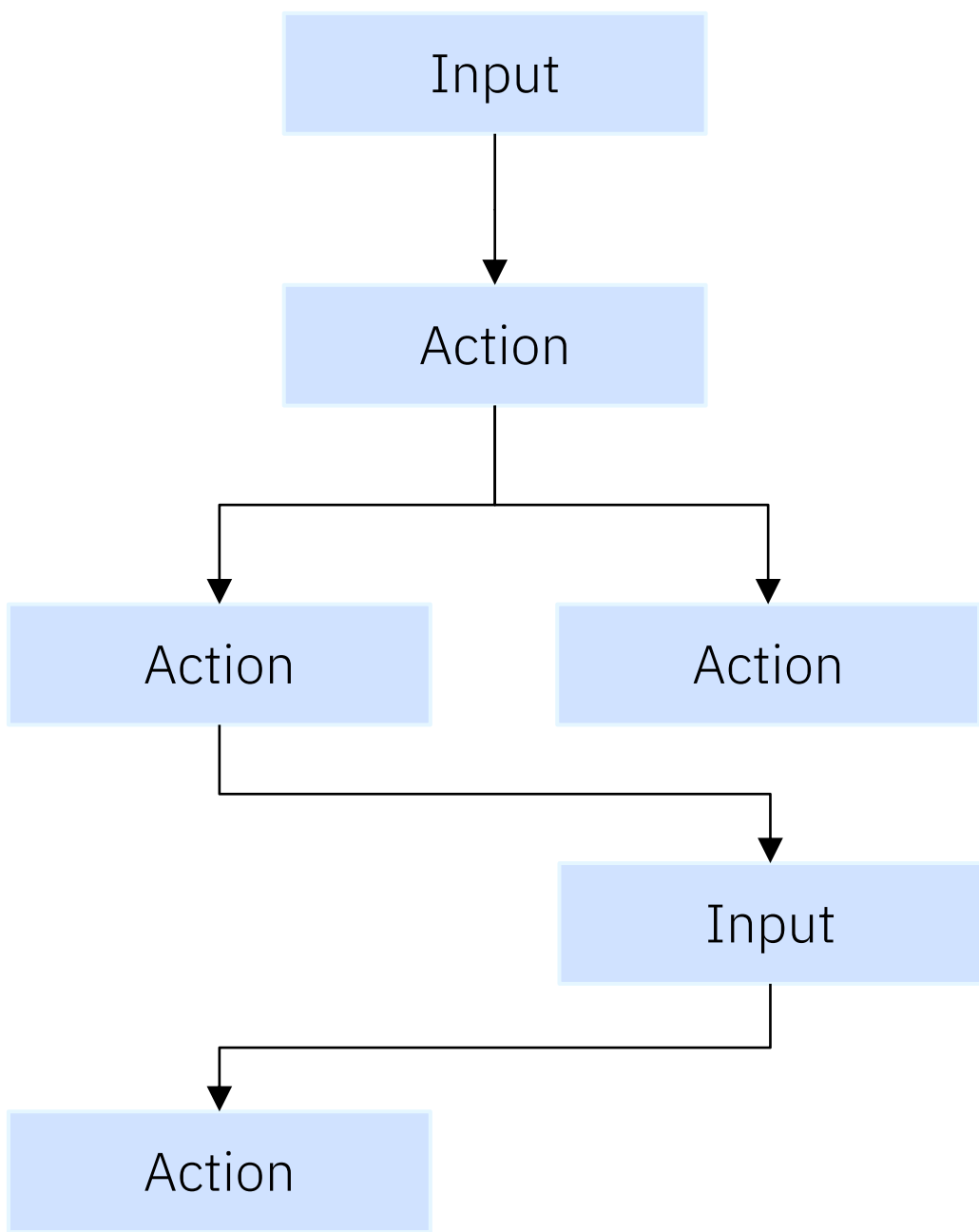
Agents



Evolution of AI assistants

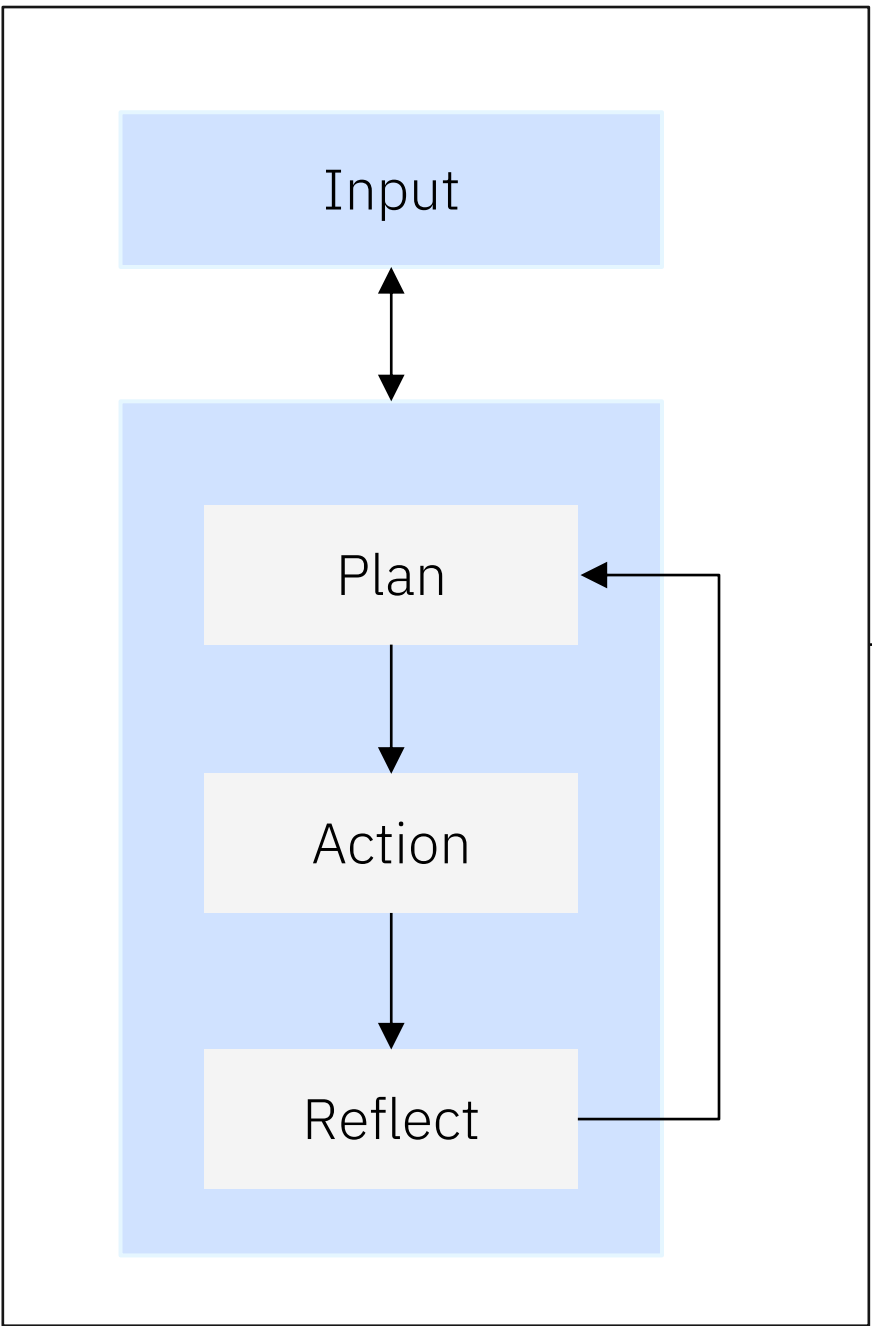
Traditional assistants

- Rule based (if x, do y)
- Predefined action paths



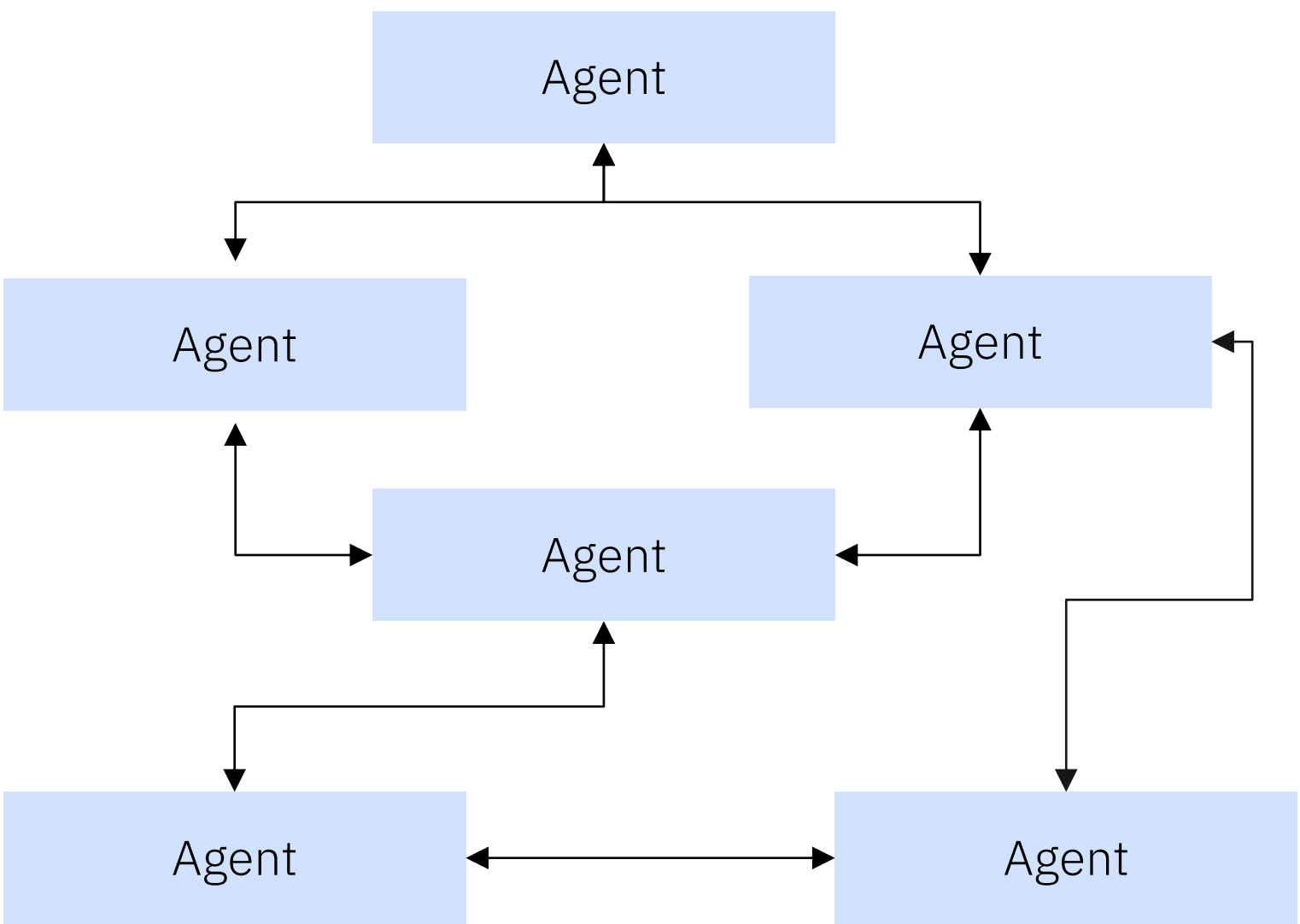
Single-agent assistants

- Task based (e.g., flight booker)
- Performance constraints as scope of task increases
- Limited control



Multi-agent assistants

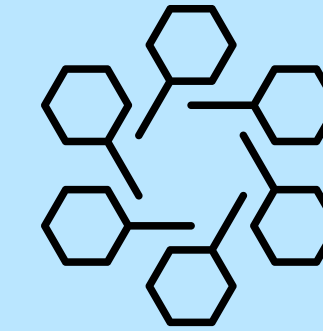
- Domain based (e.g., travel agent)
- Specialized agents (planner, flight booker, hotel booker, etc.) work together, improving system performance
- Control over how agents communicate
- Multiple architecture options (network, hierarchical, supervisor, custom)



Scale AI agents across the business

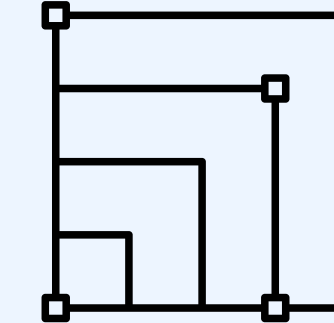
Multi-agent orchestration

Manage all agents in one place



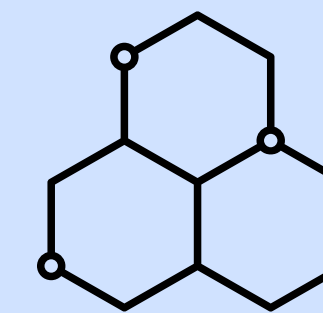
Prebuilt agents

Accelerate AI agent deployment



Custom-built agents

Build custom designed agents

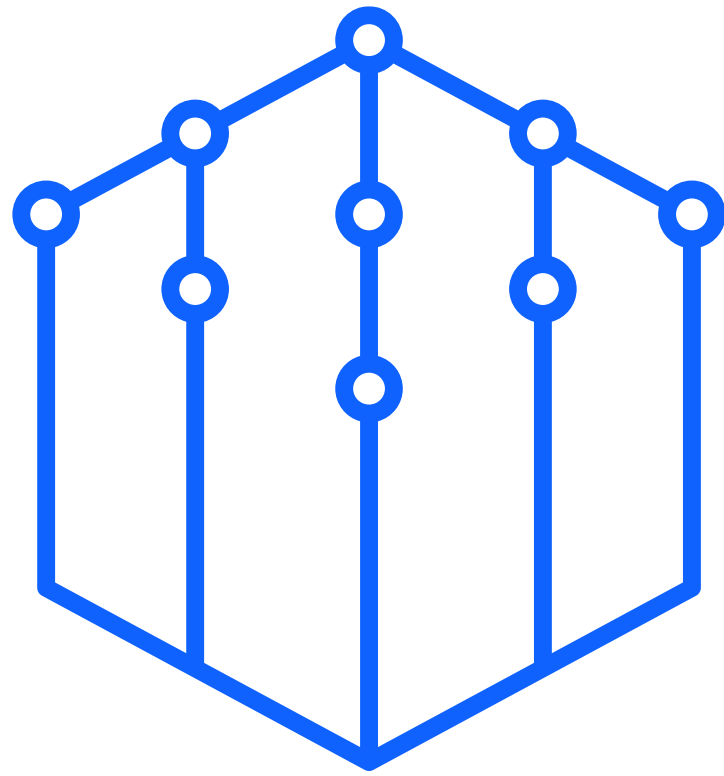




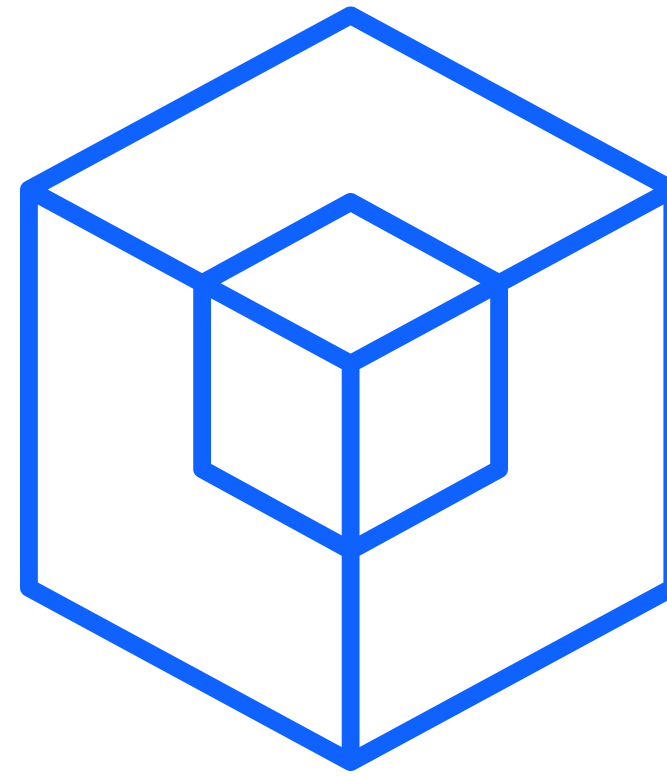
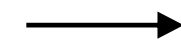
Models

The IBM approach: fit-for-purpose models

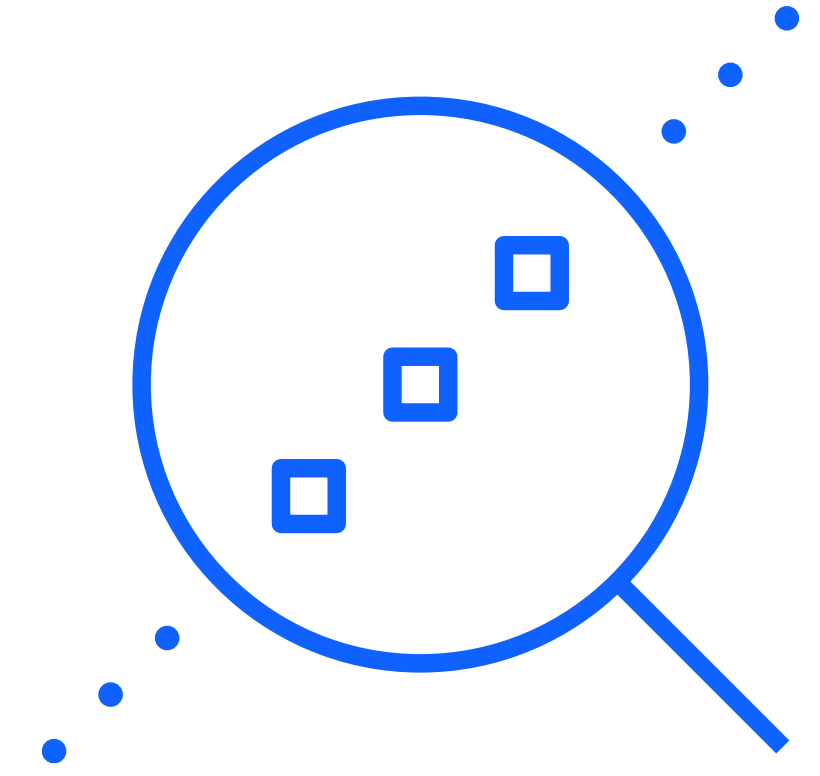
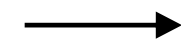
Up to 42x lower
inferencing costs



The right data



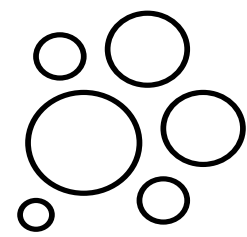
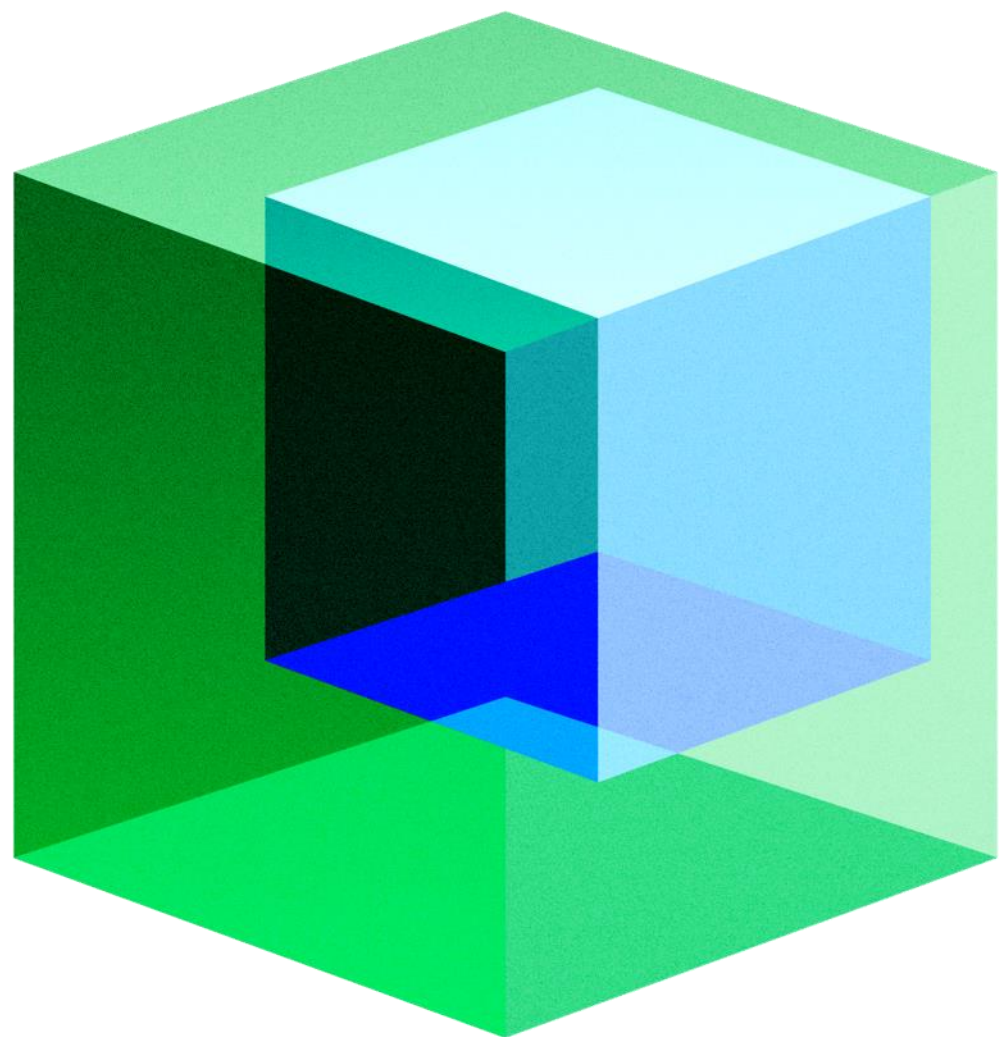
The right model



Targeted use case
fine tuning

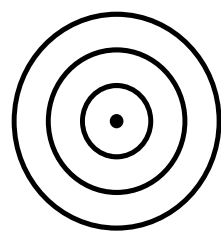
IBM Granite

A family of **open, performant and trusted** AI models to accelerate enterprise AI adoption



Open

- Open sourced under Apache 2.0
- Transparency of data, training methods
- Customize with business data



Performant

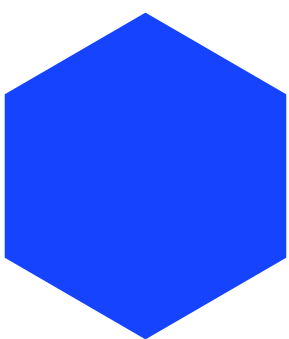
- Diverse range of fit-for-purpose models
- Designed for scalability



Trusted

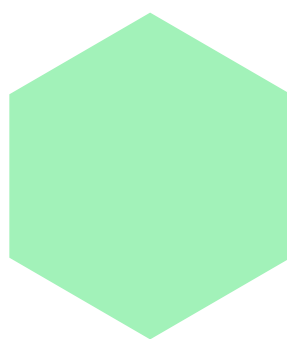
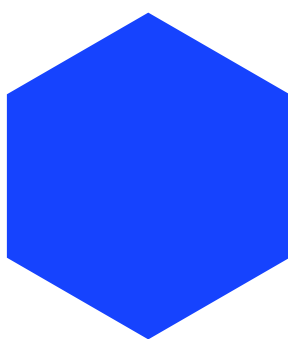
- IP indemnification
- Responsible and safe AI
- Guardrails to secure data and mitigate risks

Granite family of models



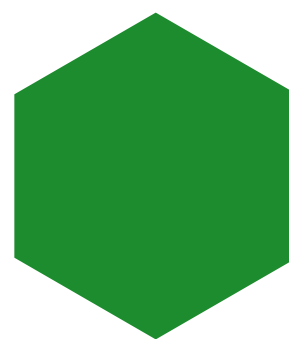
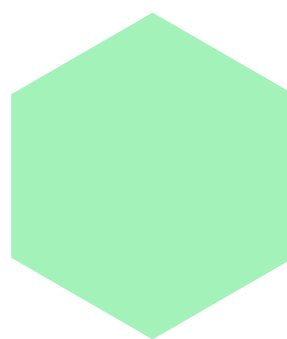
Large Language Models (LLMs) for enterprise

Granite-3.0-8B-Instruct
Granite-3.0-2B-Instruct



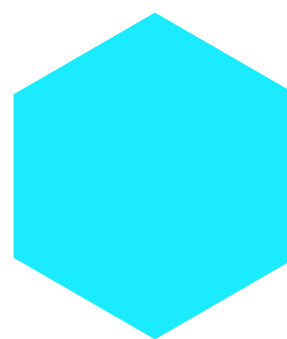
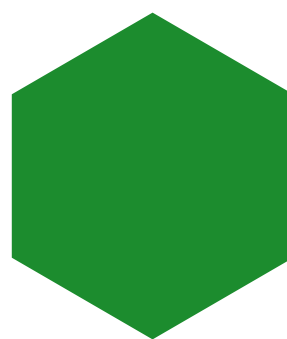
Inference-efficient Mixture of Experts (MoE)

Granite-3.0-3B-A800M
Granite-3.0-1B-A400M



Guardrail models

Granite-Guardian-3.0-8B
Granite-Guardian-3.0-2B

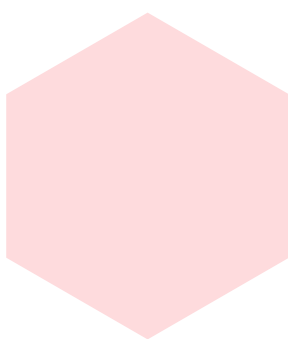


Speculative decoding models

Granite-3.0-8B-Instruct-Accelerator



Time-series models



Geospatial models

Customizing an LLM with InstructLab to achieve
model performance at a fraction of the cost

Up to 98%
cost savings

Add skills and knowledge to the
model, generate synthetic data

Use Lab techniques to
merge contributions

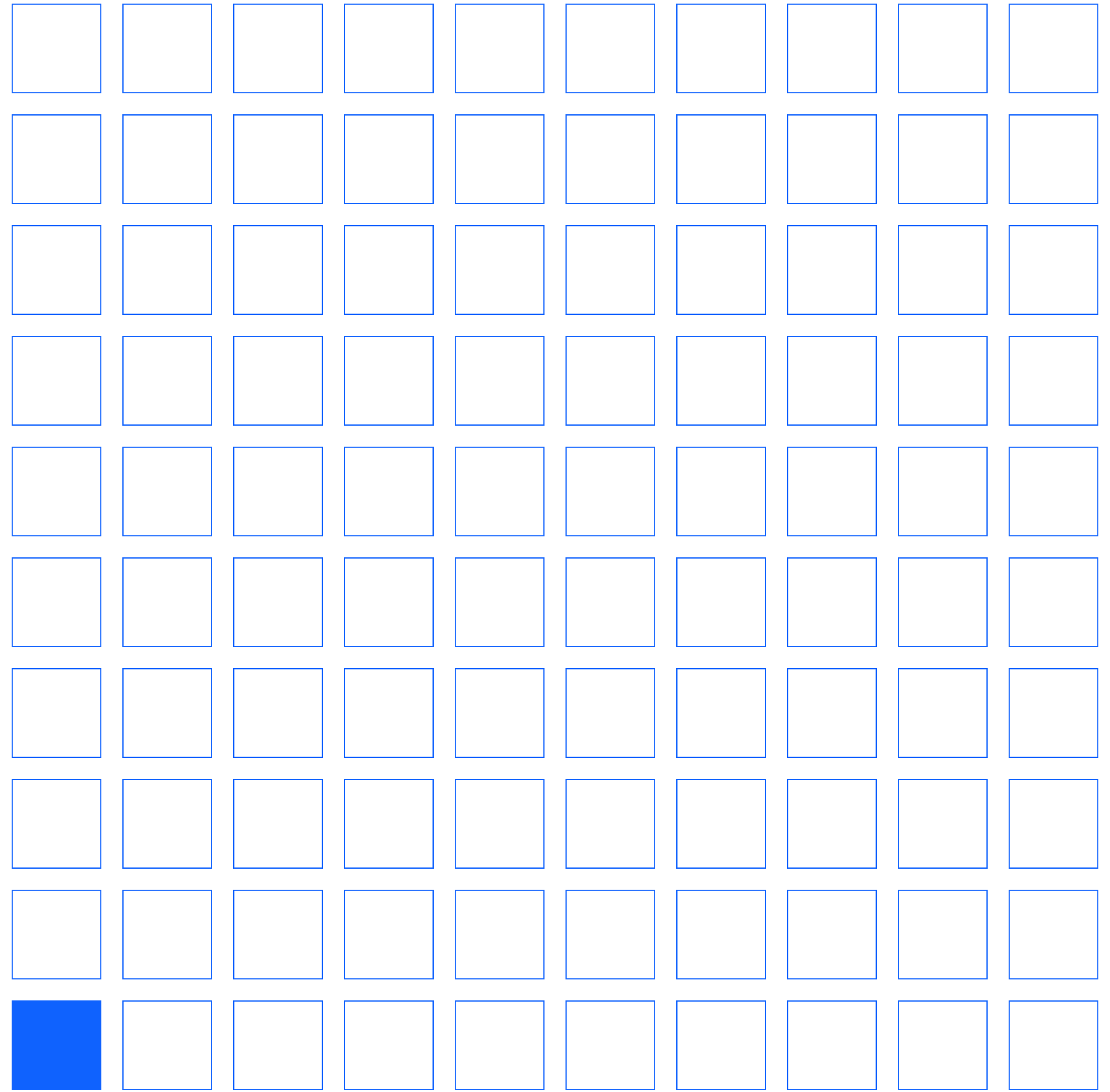
Deploy
customized model



Data



Today, less than 1%
of enterprise data
is currently being
leveraged by
generative AI



However, accessing trusted data and data complexity remain top barriers to AI implementation

Accessing trusted data

< 1/3

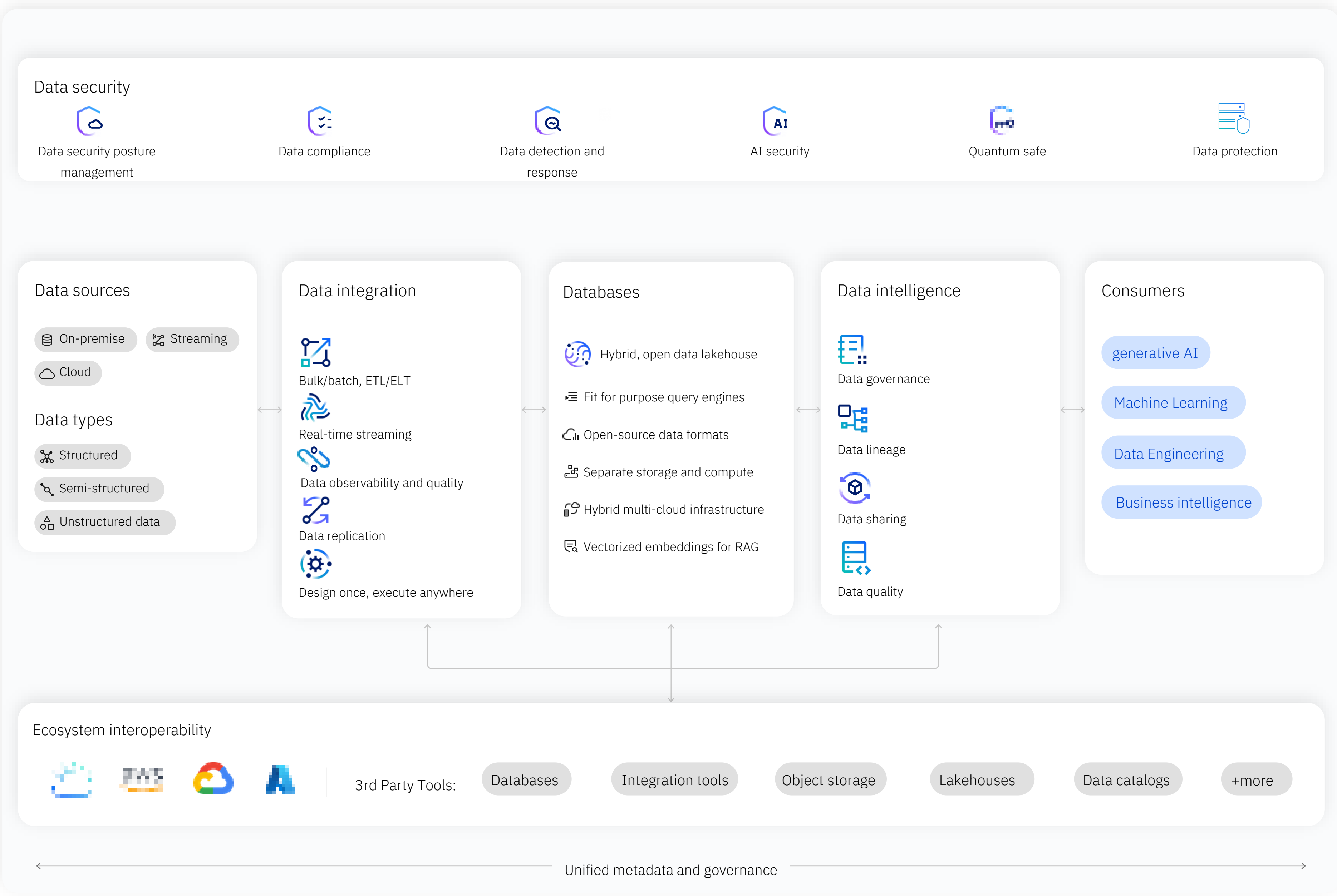
of tech leaders believe their enterprise data meets the [accessibility](#), [quality](#) and [security](#) standards that support the efficient scaling of generative AI

Growing data complexity

90%

of all enterprise-generated data is [unstructured](#) growing at [3x](#) the rate of structured data

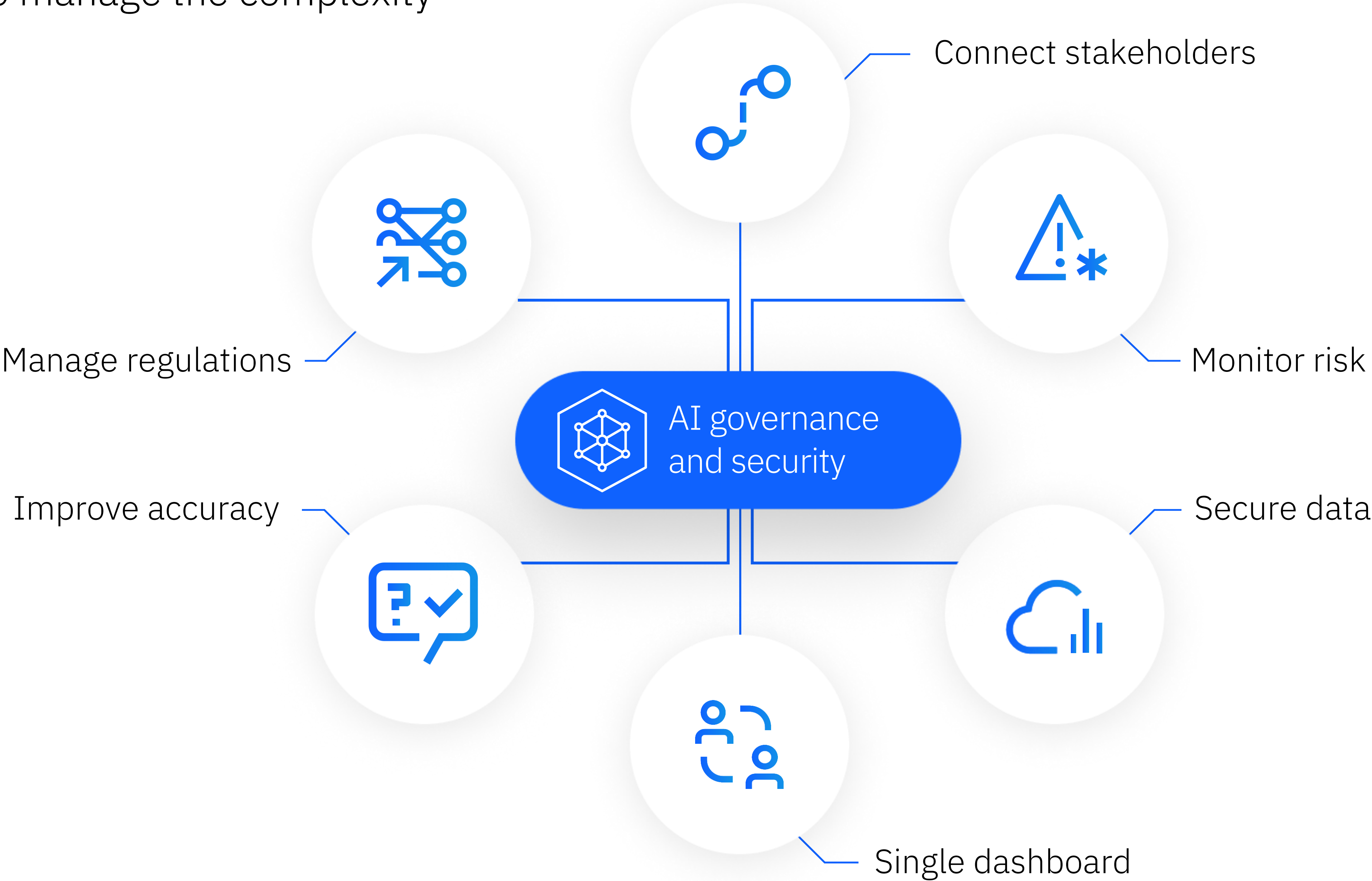
Integrate, access,
govern, and secure
all data types
with an open
and hybrid
data architecture



Governance

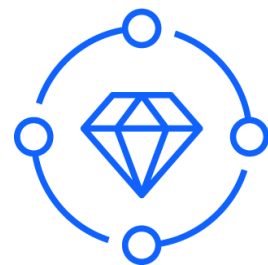


How to manage the complexity



watsonx.governance

Accelerate responsible,
transparent and explainable
AI workflows



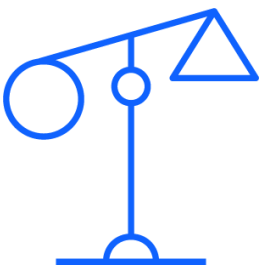
Centralized AI lifecycle governance

Manage, monitor and govern
any AI: model, app or agent;
across IBM and 3rd party like
OpenAI, AWS, and Meta



Proactive AI risk and security management

Proactively detect and
mitigate AI risks, evaluate
AI assets, and secure AI
deployments with Guardium
AI security



Trustworthy and dynamic compliance

Manage AI for safety
and transparency with
our regulatory library,
automation and
industry standards

Platform agnostic: Govern any AI, deployed anywhere



Why IBM watsonx for scaling enterprise AI to drive productivity ?

Open

Offers choice to train the right foundation models, including open-source models, and the choice of data, tools, and frameworks to achieve desired business outcomes.

Run AI wherever the business needs to, across any cloud, at scale.

Trusted

Built with open and transparent technology to give enterprises confidence in their AI and meet regulatory compliance demands.

Responsible AI and protected data backed by enterprise governance and security controls.

Integrated

Integrates technology seamlessly into existing infrastructures, systems, and processes with choice of cloud to transform the enterprise and drive productivity from within.

Embedded AI for targeted use cases that drives enterprise scale productivity.

IBM Software

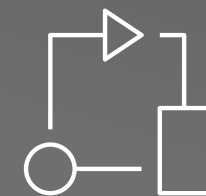
Hybrid Cloud

Unify on-prem, public, private clouds and edge to scale virtualization and AI across environments



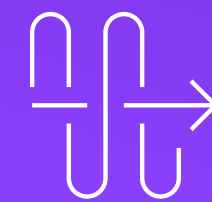
Transaction Processing

Deliver unmatched transactional performance, security and reliability



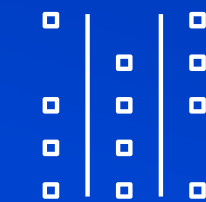
Automation

Automate technology lifecycle management with AI for productivity, resiliency and spend optimization



Data

Access trusted and secure data to drive AI productivity



AI

Open and Trusted

Today's talk

Accelerate AI
productivity across
the enterprise

Propel the next wave of AI Productivity

Autonomous agent capabilities are here to bring business to the next level. IBM has the tools to overcome the costs and challenges

Key areas IBM is focusing on with clients and partners

- Re-invent how work is done with AI agents and assistants
- Work with AI tools, models, and governance to run the AI lifecycle
- Optimize how to access, store, prepare, and protect data

Why IBM for a Data Platform

- *Open*: Run AI wherever the business needs to, across any cloud, at scale
- *Trusted*: Responsible AI and protected data backed by enterprise governance and security controls
- *Integrated*: Embedded AI for targeted use cases that drives enterprise scale productivity

Analysts validate IBM's leadership

Gartner Magic Quadrant for Data & Analytics Governance Platforms
First time published in 2024



- 

Leader in Gartner’s MQ for Cloud AI Developer Services 2024:
"Its technology solutions reach across different regions, ...meeting customers’ unique needs and regulatory requirements."
- 

Leader in Gartner’s MQ for Cloud Database Management Solutions, 2024: *"Trusted database software which handles billions of transactions each day"*
- 

Leader in Gartner’s MQ for Augmented Data Quality Solutions, 2024:
"...help users discover, augment, visualize, & cleanse data for AI through a self-service...conversational, natural language interface"
- 

Strong Performer Forrester Wave™: AI/ML Platforms, Q3 2024:
"one-stop AI platform that can run in any cloud"
- 

Strong Performer Forrester Wave™: AI Foundation Models For Language, Q2 2024: *"some of the most robust and transparent insights into the underlying training data"*
- 

Leader in Forrester Wave for Data Management for Analytics, Q1 2023
"IBM’s superior vision focuses on simplifying data management across hybrid cloud, multi-cloud, & the edge,"
- 

Leader in IDC MarketScape Worldwide AI Governance Platforms, 2023
"[IBM] streamlines the AI governance process by delivering a unified experience and eliminates the need for various tools or platforms"
- 

Recognized as Gartner Peer Insights Customer’s Choice, 2024
"Top notch technical support that I haven’t seen with other vendors"

