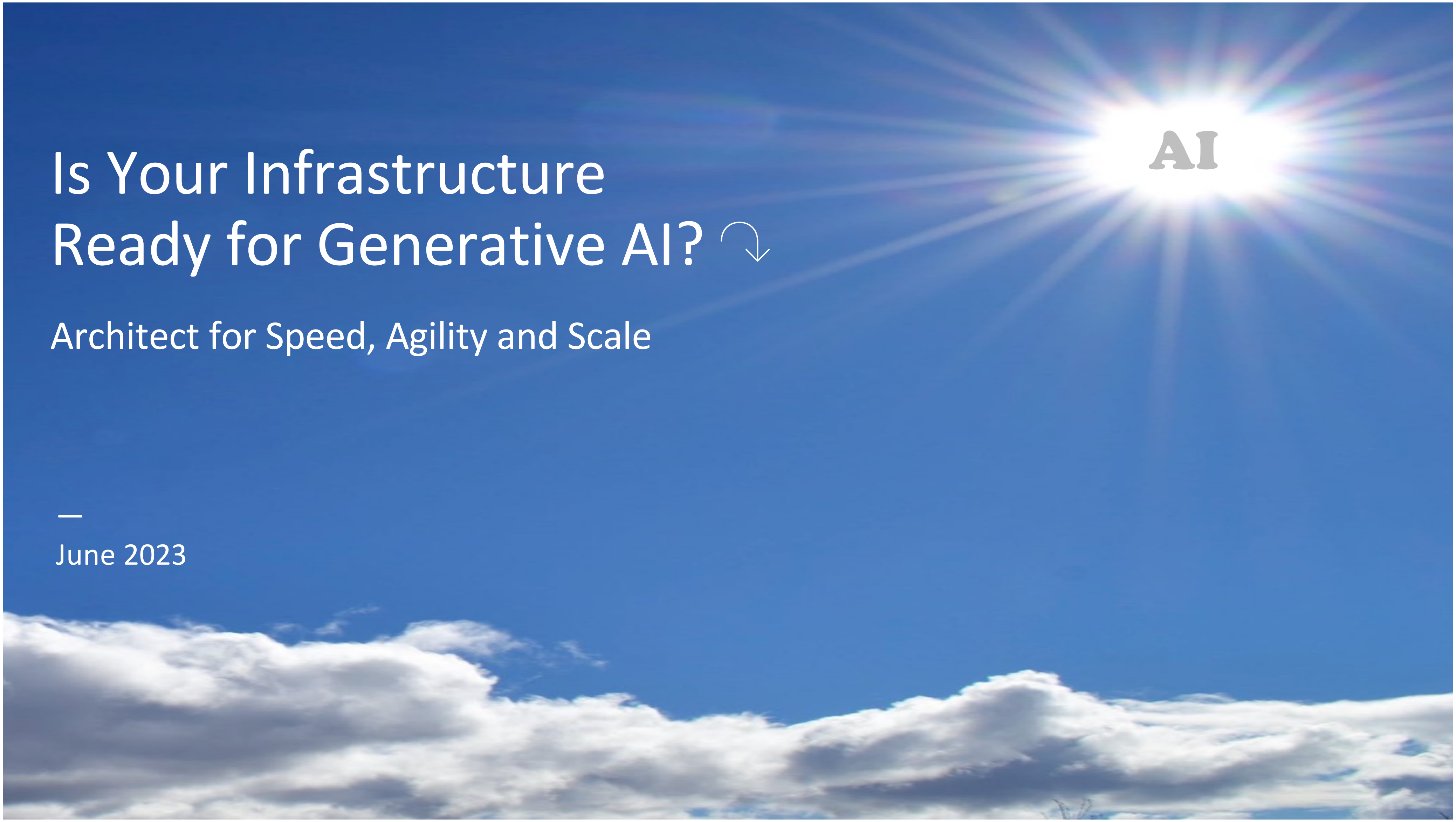
Is Your Infrastructure Ready for Generative AI? ⤵

Architect for Speed, Agility and Scale

—

June 2023



Key Attributes of an AI-Ready Infrastructure ↷

Top Use Cases

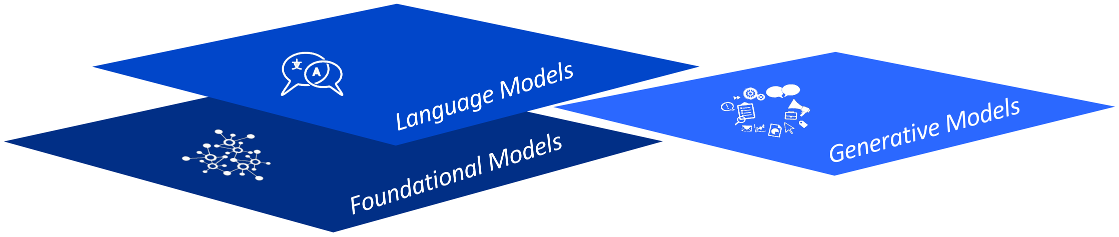
Create freely **Simply Hybrid Cloud**	*Create Securely* **De-risk data fluidity** .	*Create Intelligently* **Architect AI for Scale**	*Create Quickly* **Deliver IT on demand**
• Modernize applications with a containerized, micro-services architecture that can run anywhere • Distribute workloads across on-premises, multicloud locations within a centralized control plane • Integrate manage workloads across hybrid cloud digital ecosystems • Improve hybrid data workload agility with container-native storage	• Deploy risk & compliance controls across 3rd and 4th party digital value chains supporting hybrid cloud landscapes • Flexibility aggregate and store data for distributed workloads on/off premises • Safeguard data supply chains from cyberattacks, hardware failures, disasters and other threats • Enable real-time adaptable policies to personalize and manage data views	• Enable high-performance, fine-tuned inferencing for complex AI models • Securely scale AI data ingestion across distributed storage estates • Flexible distributed AI models to where the data lives across hybrid clouds • Safeguard AI models with AI-optimized encryption and cryptography • Agility train and run data-intensive models in the cloud as a service, reducing the cost and energy consumption	• Create a centralized control plane for delivering IT services across hybrid Cloud for any service • Adopt a consumption-based IT model consuming only what you need, when you need it • Compose fit for purpose IT optimized to the unique needs of any workload • Speed the introduction of new technologies with ease

Client Success Stories

AI Platform

Hybrid Cloud Platform

AI Infrastructure



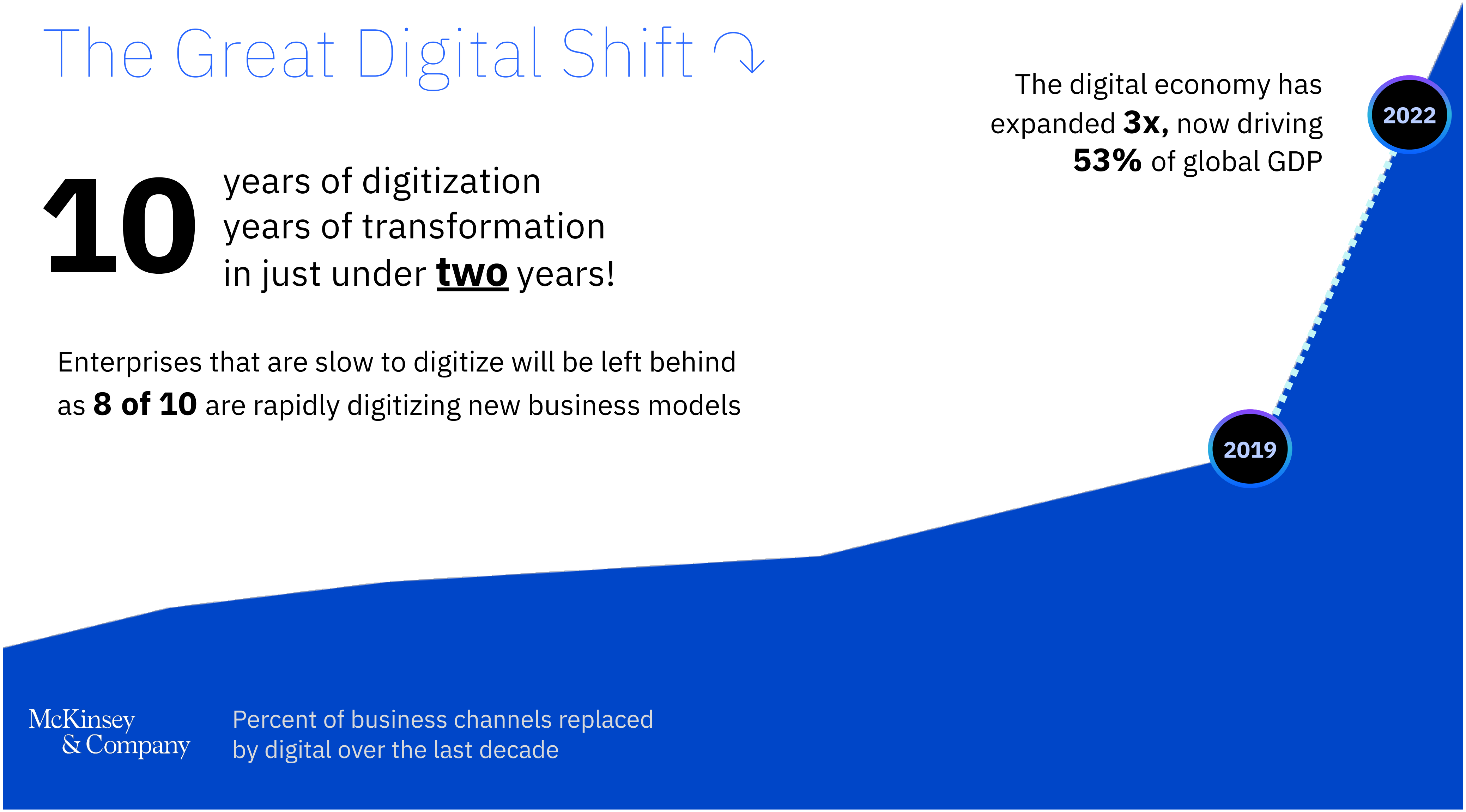
Narrative "At a Glance" ↷

- The Great Digital Shift
- Transformation, now reinvention
- Harnessing the digital technology flywheel
- The great AI awakening is underway
- AI for business is now essential
- Is your infrastructure ready for AI?
- Six critical considerations
- Architect AI for scale
- Deliver sustainable "green AI"
- Full-stack solutions for AI for business

The Great Digital Shift ↷

10 years of digitization
years of transformation
in just under **two** years!

Enterprises that are slow to digitize will be left behind as **8 of 10** are rapidly digitizing new business models



The digital economy has expanded **3x,** now driving **53%** of global GDP

Percent of business channels replaced by digital over the last decade

McKinsey & Company

Transformation, Now Reinvention ↷

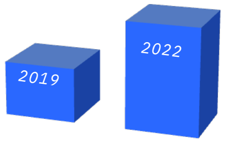
CEOs are now driving holistic, large-scale digital reinvention across their enterprises

To create "future-proof" digital businesses that are easier to scale, adapt and sustain

They envision highly personalized, frictionless experiences across all touchpoints that are:

- Data-driven, predictive
- Intelligently automated
- Secure, risk-aware
- Dynamically resilient

McKinsey & Company



6 in 10

CEOs rank new business building as a top priority

Half of revenue by 2026 will come from new digital products and services that do not exist today

Harnessing the Digital Technology Flywheel ↷

IBM is helping its clients make the right architectural decisions to speed outcomes and future-proof businesses

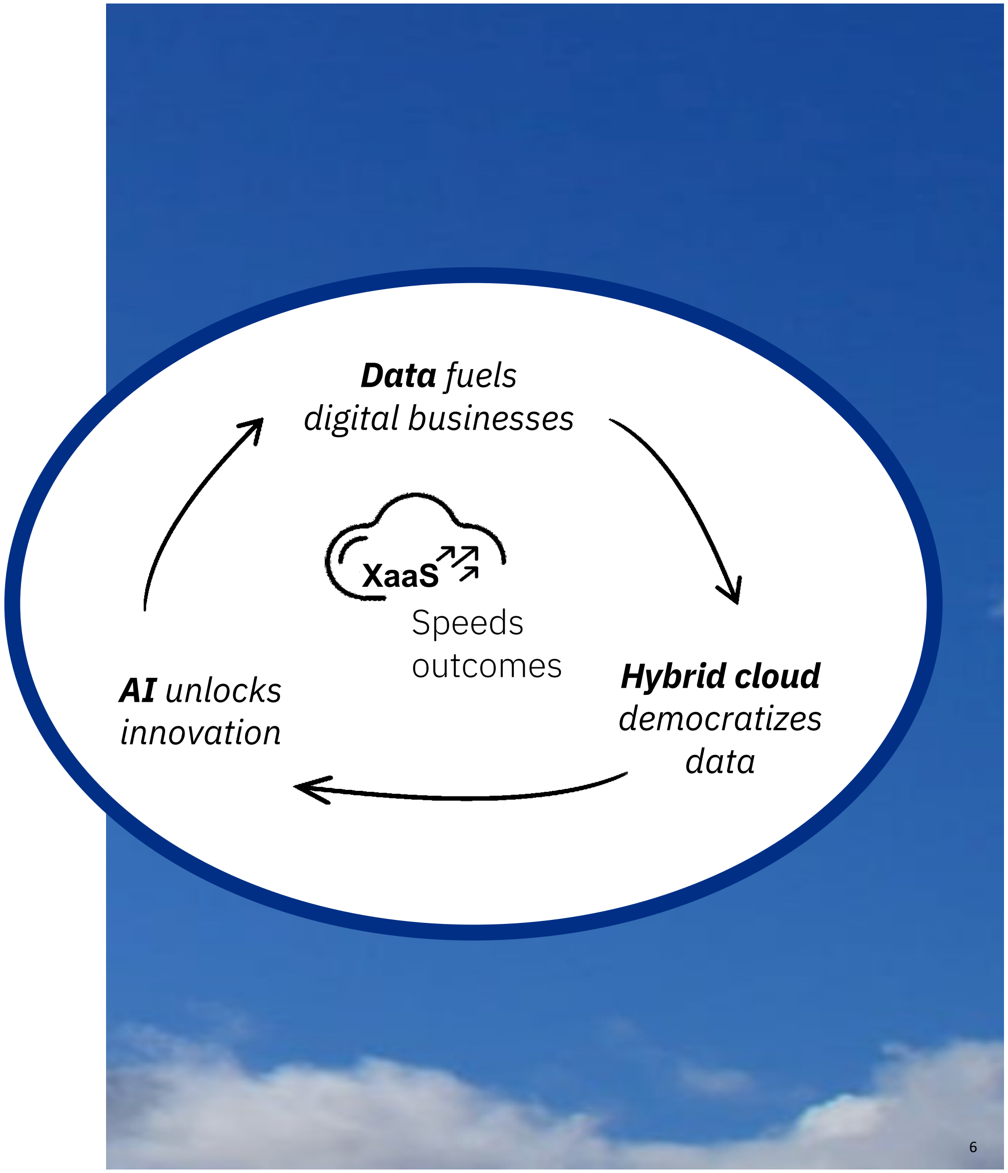
And to harness the "digital technology flywheel" to drive sustained innovation at lower cost and risk

As data is what fuels digital businesses, yet is of increasing velocity, variety and volume with shorter value lifespans

Hybrid cloud is what democratizes the use of data across ever expanding digital ecosystems and value chains

Enabling AI-powered workloads that unlock new innovations and infuse intelligence across organizations

Accelerated by XaaS consumption models that speed outcomes by lowering complexity, risk and cost


***Data** fuels digital businesses*
***Hybrid cloud** democratizes data*
***AI** unlocks innovation*
XaaS
Speeds outcomes

An AI Awakening is Underway ↷

8 in 10 of businesses by 2022 had prioritized AI and have rapidly expanded ROI by 7 points since 2020

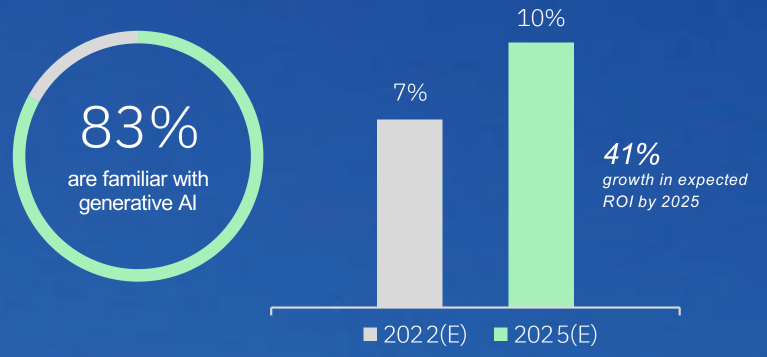
The advent of ChatGPT ”consumerized” AI for the masses and for business – just as how Netscape did for the Internet in the ‘90s

Business leaders now anticipate significant financial returns from Generative AI growing building upon the traditional AI ROI trajectory

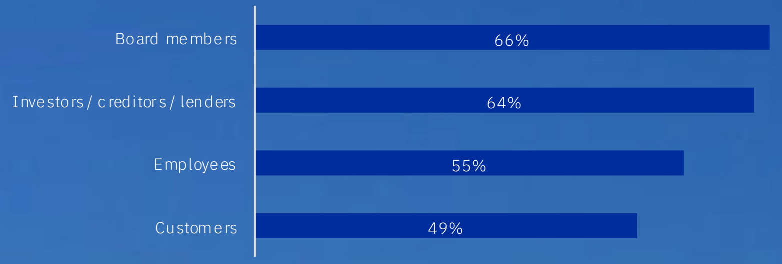
Generative AI use cases now driving hyper-spend for creating AI-ready infrastructures:

→ **91%** CAGR for generate AI compute to $11B
→ **76%** CAGR for generative AI storage to $2B
→ **47%** deployed as hybrid cloud services

Businesses view generative AI as a driver accelerated AI outcome



CEOs are facing considerable pressure to adapt Generative AI rapidly



AI for Business is Now Essential ↷

Businesses everywhere are focused on infusing predictive intelligence and value creation catalysts throughout their organizations

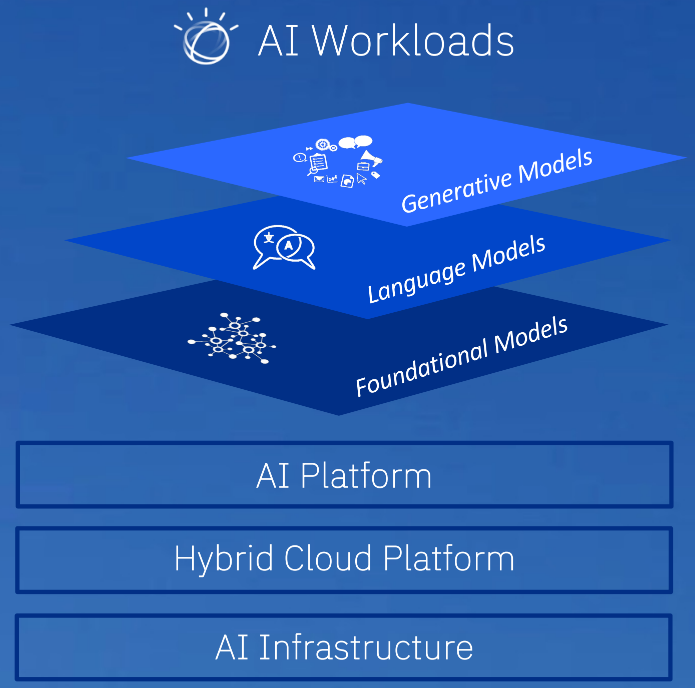
Anchored in a new era of AI models & technologies that will unleash unprecedented innovation

By understanding the unique language of business – data, docs, people, process, policy

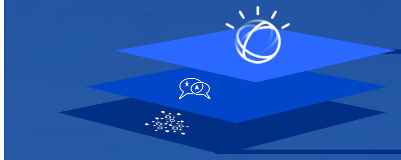
Intelligently automating customer care, IT operations, digital labor, cybersecurity, etc.

And building trust by helping people understand and explain complex decisions

63% are accelerating AI spend
30% on generative models



Is Your Infrastructure Ready for AI ⤵

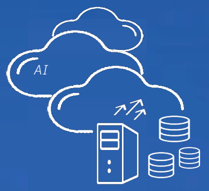


AI Platform

Design, tune, deploy AI

Integrate, manage hybrid data

Govern responsible AI



Enterprise-grade AI will require a highly sustainable, compute-and-data intensive distributed infrastructure

Why? →

100x

more model parameters

AI training and inferencing will require highly intensive processing of simultaneous computations

10x

growth in newly generated AI data

The iterative supply chain of synthetic, annotated and generative data that must be stored, secured, managed

7x

faster security threat lifecycles

Bad actors are using AI to identify new data and AI model vulnerabilities and to speed iterative attacks

7x

greater computational throughput

Elaborate data-intensive models will exponentially scale the rate of storage, memory and processors transactions

50x

performance degradation from distributed data

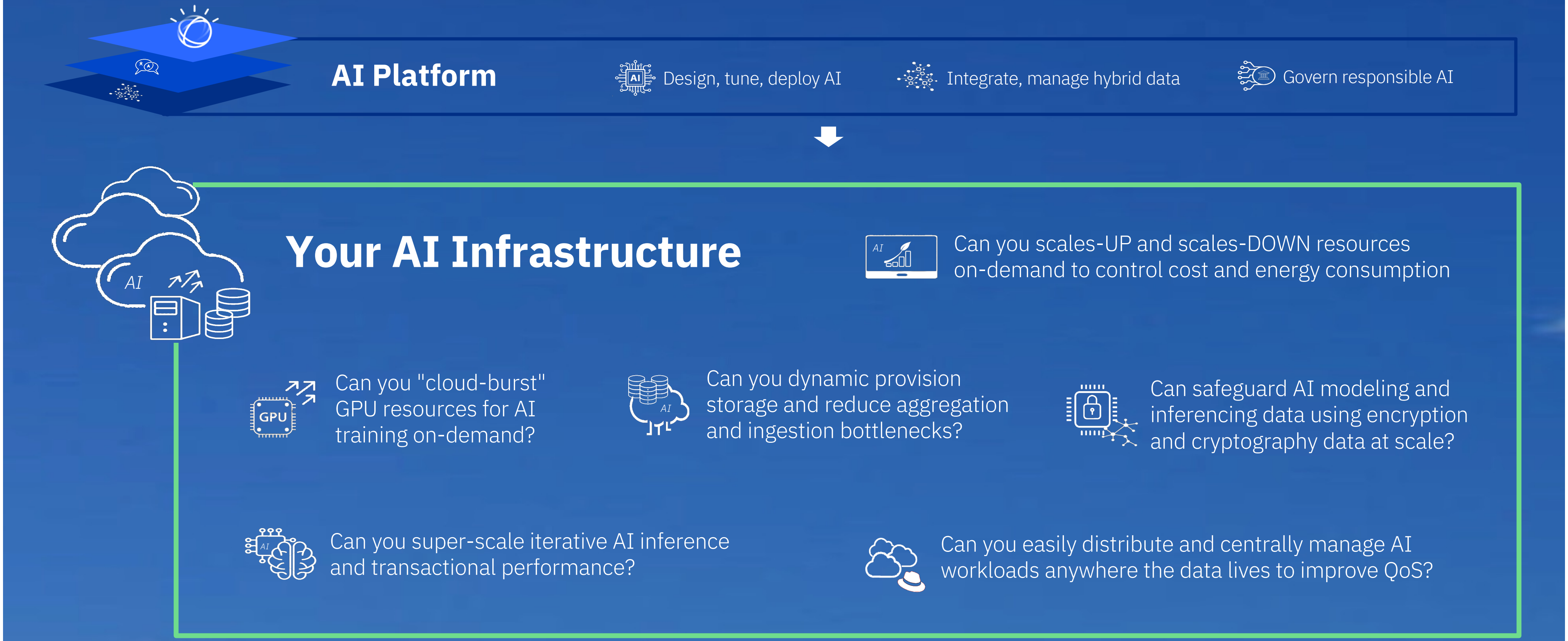
Driving the need to dynamically distribute AI models to where the data lives data lives to achieve near-zero latency and scalable responsiveness

2x

more energy consumption

As AI models rapidly consume more and more data, iteratively trained longer and longer and require faster and faster inferencing cycles

Six Critical Considerations ⤵



AI Platform

Design, tune, deploy AI

Integrate, manage hybrid data

Govern responsible AI

Your AI Infrastructure

Can you scales-UP and scales-DOWN resources on-demand to control cost and energy consumption

Can you "cloud-burst" GPU resources for AI training on-demand?

Can you dynamic provision storage and reduce aggregation and ingestion bottlenecks?

Can safeguard AI modeling and inferencing data using encryption and cryptography data at scale?

Can you super-scale iterative AI inference and transactional performance?

Can you easily distribute and centrally manage AI workloads anywhere the data lives to improve QoS?

Architect AI for Scale

Infrastructural modernization is inevitable to support the compute-and-data intensive nature of AI workloads

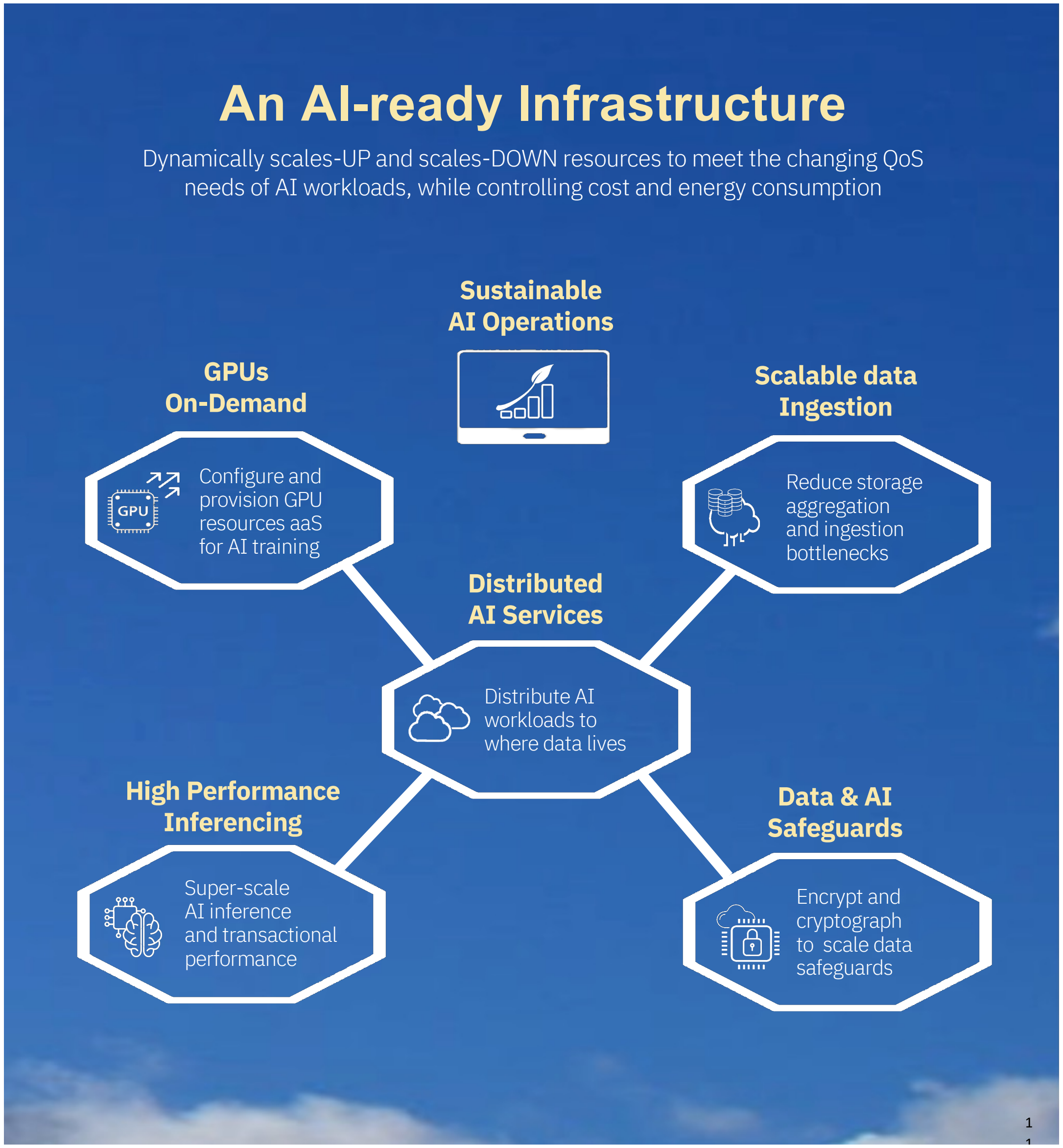
Making the right architectural decisions for model training, deployment and inferencing QoS is critical

Architected as a holistic interconnected hybrid cloud infrastructure that operates as an integrated system

Enabling you to scale resources and performance to support the highly dynamic nature of AI workloads

Ensuring you have the agility to distribute AI workloads anywhere, while securing sensitive data

And that allows you to achieve green AI operations that lowers costs, risk and energy consumption



An AI-ready Infrastructure

Dynamically scales-UP and scales-DOWN resources to meet the changing QoS needs of AI workloads, while controlling cost and energy consumption

Sustainable AI Operations

GPUs On-Demand

Configure and provision GPU resources aaS for AI training

Scalable data Ingestion

Reduce storage aggregation and ingestion bottlenecks

Distributed AI Services

Distribute AI workloads to where data lives

High Performance Inferencing

Super-scale AI inference and transactional performance

Data & AI Safeguards

Encrypt and cryptograph to scale data safeguards

Deliver Sustainable AI Operations

Given the highly compute-and-data intensive nature of foundational and generative AI models

It's essential to architect for "green AI" to lower costs, complexity and energy consumption

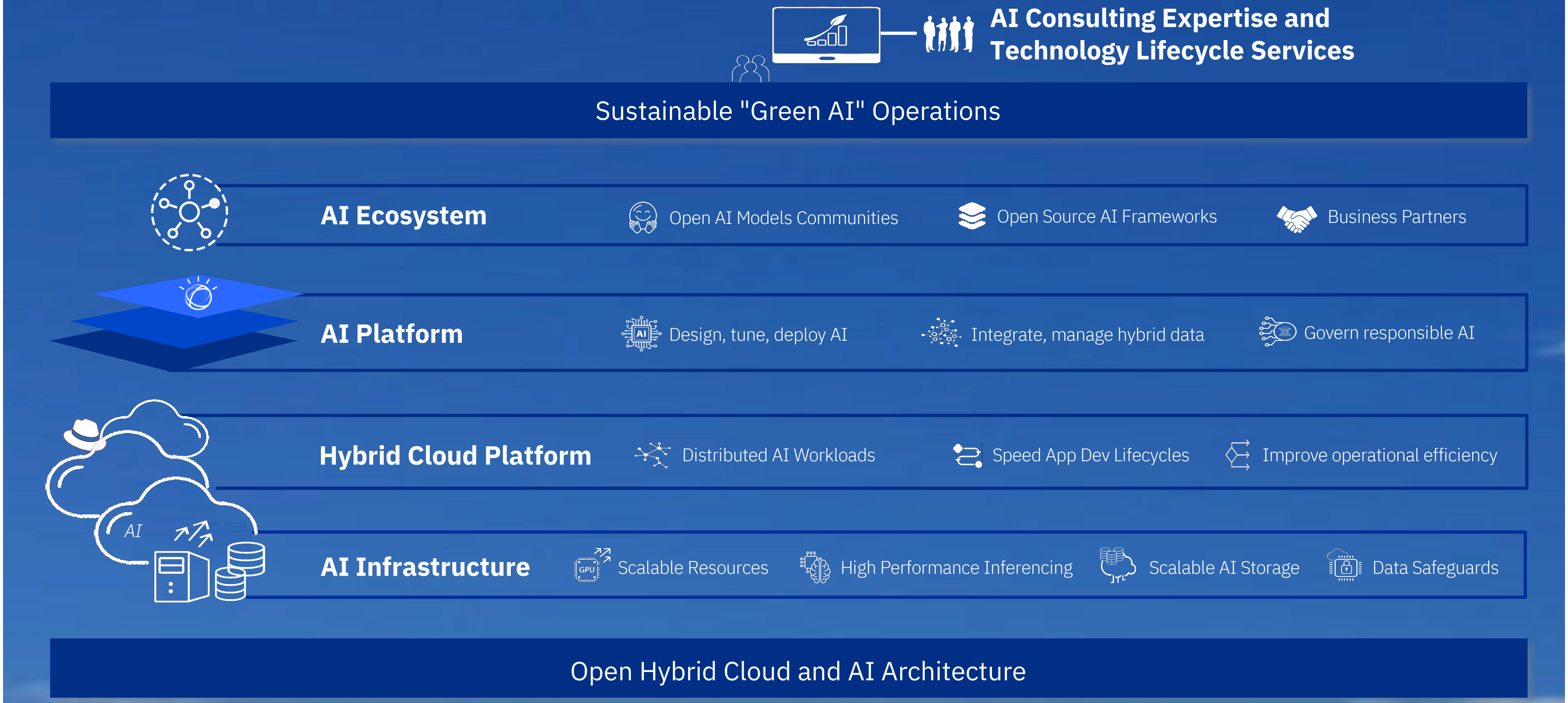
- **XaaS:** Create a unified XaaS operational console to consume only what you need, when you need it
- **Cloud-bursting**: Flexibly provision AI training resources without the need to build out your own infrastructure
- **AI Systems:** Invest in highly energy efficient systems built on leading edge inference chip designs
- **AI Storage**: Deploy high scalable, distributed storage optimized for low latency data aggregation and ingestion
- **Distribute AI**: Architect for centrally managed distributed AI to reduce data movement processing speeds

Architecting for Green AI is Not Really Optional

> "*Modern AI models consume massive amounts of computing resources and energy, and these requirements are growing at breakthrough speed*
>
> *The challenge in chasing progress in AI via ever-larger models is highlighted by the relationship of model size and model performance*
>
> *Computational requirements for best-in-class AI models have been doubling every **3.4** months*

Forbes

Full Stack Solutions
AI for Business ↷



AI Consulting Expertise and Technology Lifecycle Services

Sustainable "Green AI" Operations

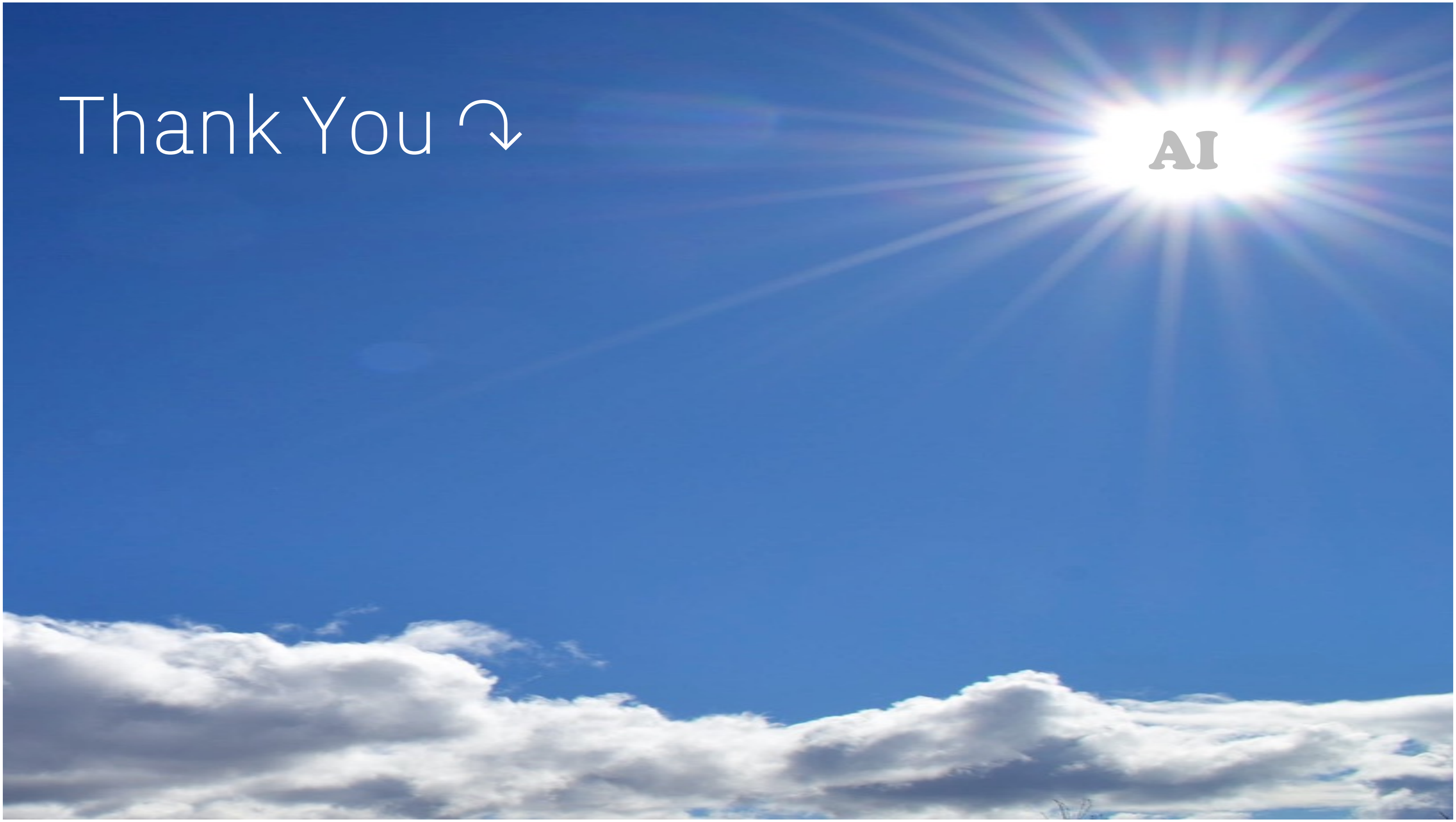
AI Ecosystem — Open AI Models Communities · Open Source AI Frameworks · Business Partners

AI Platform — Design, tune, deploy AI · Integrate, manage hybrid data · Govern responsible AI

Hybrid Cloud Platform — Distributed AI Workloads · Speed App Dev Lifecycles · Improve operational efficiency

AI Infrastructure — Scalable Resources · High Performance Inferencing · Scalable AI Storage · Data Safeguards

Open Hybrid Cloud and AI Architecture



Thank You ↷