Google Cloud

The agentic enterprise



Scaling and governing autonomous agents with Gemini Enterprise.



Table of contents

1 The shift to autonomy

2 Gemini Enterprise

3 Model choice and connectivity

4 Building agents faster

5 Scaling agents effectively

6 Governing agents with confidence

7 Optimizing agent quality

8 Unlock the future

The shift to autonomy





For years, organizations focused on AI to accelerate one-off tasks like summarizing reports or drafting emails.

These early assistants required constant human instruction and manual oversight. Today, the industry is moving from prescription to delegation. Instead of juggling multiple disconnected AI apps, organizations need a single “front door to AI”—with agents capable of handling complex, multi-step business outcomes with full context.

However, this transition introduces significant operational and security challenges. Scaling and managing a digital workforce can lead to guesswork, where teams deploy AI agents without properly testing them. If you can’t track what your agents are doing, you introduce potentially catastrophic security, compliance and integration risks. Centralized oversight is required to avoid data leaks, prompt injection, and misaligned agents across fragmented systems.

Gemini Enterprise



Gemini Enterprise a secure-by-design architecture to the entire agent development lifecycle across your enterprise.

By combining advanced model reasoning with centralized identity, real-time security monitoring, and forensic observability, it transforms experimental agents into a governed, reliable, and auditable digital workforce.

Beyond security, Gemini Enterprise seamlessly connects AI research to your infrastructure—removing the trade-offs between model intelligence, infrastructure reliability, and agent orchestration. It gives technical teams a unified framework to build, scale, govern, and optimize autonomous agents, so you can confidently move from basic task automation to delegating complex business outcomes.

The four design layers of Gemini Enterprise

1

Native frontier intelligence

Gemini Enterprise combines Google DeepMind's multimodal reasoning and autonomous planning with enterprise-grade customization. This layer allows you to seamlessly tune and refine models based on your unique business needs.

2

Research-to-production lifecycle

Developers leverage systematic methodologies and quality frameworks to measure, test, and optimize agents. This structured process accelerates the transition from experimental prototypes to production-ready assets.

3

Verifiable trust

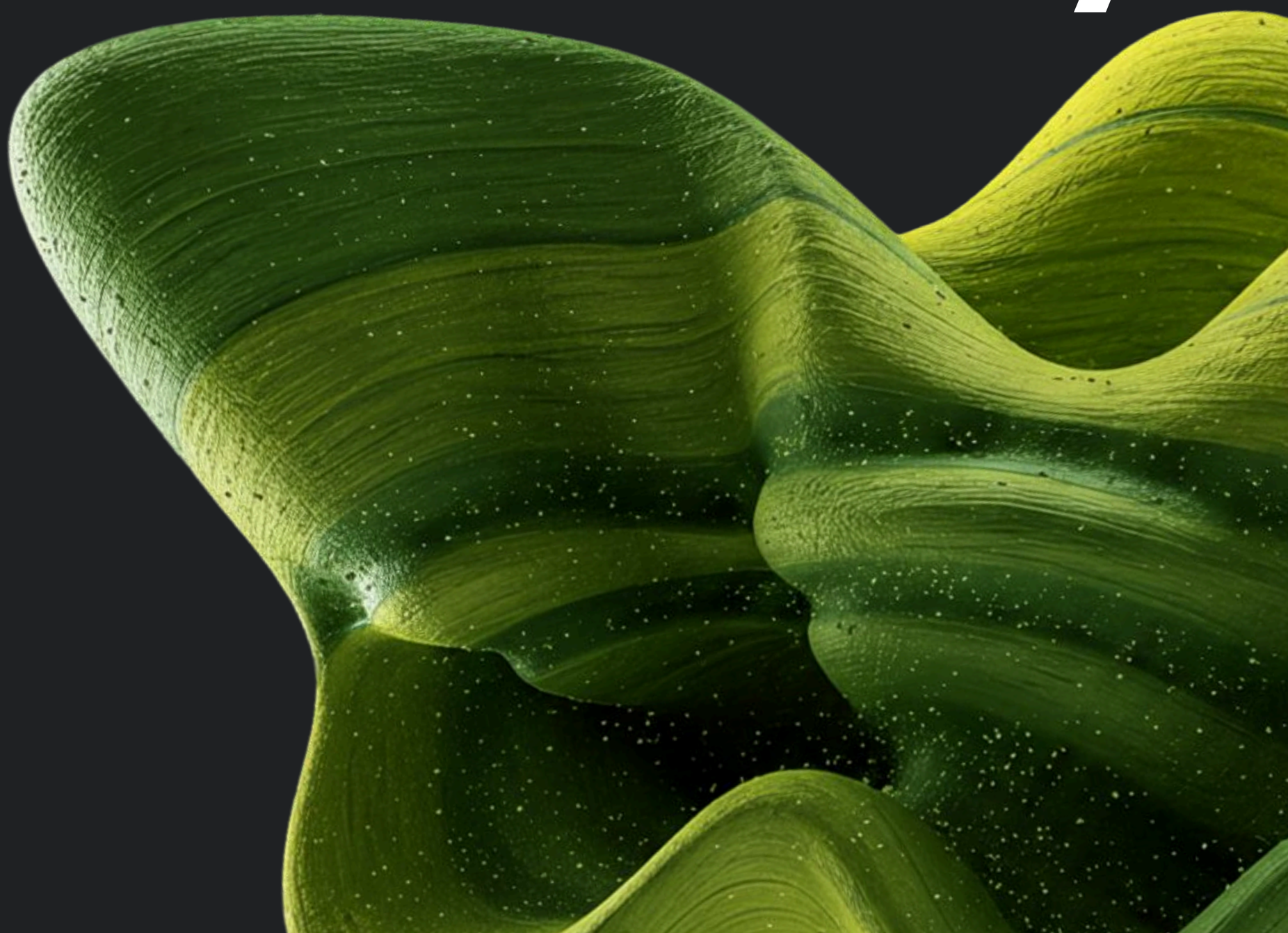
Secure-by-design foundation provides centralized identity, real-time security monitoring, and policy enforcement, allowing organizations to scale autonomous agents safely across global production environments while maintaining strict compliance.

4

Continuous enhancement

Advanced observability and evaluation tools turn agent development into a clear data-driven process. Teams get full visibility into agent execution paths, allowing them to analyze reasoning loops, trace tool lineage, and optimize performance.

Model choice and connectivity



Our shift towards autonomous fleets has been building for years—from operationalizing machine learning in 2021 through to gen AI in 2023 and building specialized agents in 2025.

Today, enterprises have the ability to orchestrate entire autonomous fleets of digital coworkers that make independent decisions with full business context.

To make this work, organizations require ultimate model flexibility. Model Garden provides access to over 200 leading models. This extensive selection allows developers to match the right model to the specific business task, running them as managed services that can be swapped or upgraded without rebuilding agent logic.

Yet model intelligence alone isn't enough. Moving from simple task automation to delegating complex business outcomes requires deep ecosystem connectivity.

Gemini Enterprise uses specialized protocols to connect models to external systems

1

Model Context Protocol (MCP)

This open standard accesses external resources and data sources to ground agent decisions in real-time enterprise context.

2

Agent-to-UI (A2UI)

This protocol generates rich interfaces directly for users, enabling intuitive human-agent collaboration and dynamic interaction.

3

Secure transaction execution

Gemini Enterprise secures data connections and enables transaction execution through the Agent Payments Protocol (AP2) and Universal Commerce Protocol (UCP), ensuring agents can safely execute payments and interact across diverse enterprise systems.

Building agents faster





Turning AI concepts into autonomous action requires tools that balance rapid prototyping with deep customization.

Gemini Enterprise delivers this balance by unifying visual design with code-first frameworks, allowing technical teams to build, test, and deploy agents at scale.

Visual to code-first development

Developers can start in **Agent Studio**, using a low-code visual interface to build and test agent prompts. When a project calls for deep customization, they can export this logic directly into the upgraded Agent Development Kit (ADK) to continue in a full-code environment. To further accelerate the agent development lifecycle (ADLC), technical teams use Google Antigravity within Gemini Enterprise.

Antigravity provides an agent-first development environment where developers can deploy autonomous subagents to write boilerplate, run browser-in-the-loop UX testing, and plan multi-step code refactoring asynchronously—all while running within your secure cloud boundary.

Secure workspaces and enterprise delivery

To ensure safety, Gemini Enterprise provides hardened, sandboxed workspaces. These isolated environments allow agents to run bash commands and manage files securely, completely separated from core systems. This secure-by-design infrastructure accelerates the delivery of production-grade agents, allowing organizations to deploy a digital workforce with total confidence.

COMMERZBANK

> At Commerzbank AG, we are building a secure foundation for responsible Agentic AI on Google Cloud. To bring this vision to scale, we are actively evaluating Google Cloud's new Agent Registry and Agent Gateway Services. These services are key to our governance strategy, offering vital controls for agent discoverability, policy enforcement, and access management. Furthermore, they provide the deep observability and auditability essential for us to scale our AI platforms in a compliant and trustworthy manner."



Scaling agents effectively





To scale intelligent decisions across an entire enterprise, organizations require an execution layer that matches the speed and persistence of human operations.

To scale intelligent decisions across an entire enterprise, organizations require an execution layer that matches the speed and persistence of human operations. Gemini Enterprise addresses this need through its re-engineered **Agent Runtime**. This delivers sub-second cold starts and allows organizations to provision new agents in seconds, deploying and scaling digital workforces instantly in response to real-time demand.

Unlike traditional stateless systems, Gemini Enterprise supports long-running autonomous agents capable of managing multi-day workflows. This persistence allows agents to execute complex, multi-step workflows and deep reasoning tasks—such as managing a sales prospecting sequence—without losing state or progress over days at a time.

To act as true teammates, agents must retain context over time. Gemini Enterprise achieves this through **Memory Bank** and **Memory Profiles**, which dynamically generate and curate long-term memories from past conversations. This persistent memory allows agents to recall high-accuracy details with low latency, eliminating repetitive prompts and establishing a continuous relationship with users.

Managing high-volume agent execution safely on Gemini Enterprise requires strict operational guardrails. Gemini Enterprise executes model-generated code within secure, hardened sandboxes, protecting core enterprise systems from risk while maintaining high-performance execution across the entire digital workforce.

AT&T

> At AT&T, as we are leveraging Agent Memory Bank for long-term memory, our autonomous and intelligent AI sales agents in the app channel can resume conversations after a gap by synthesizing key facts from prior customer interactions, effectively moving from guessing to remembering. As we extend this capability to IVR [Interactive Voice Response], we're building toward a seamless cross-channel sales journey where customers can continue conversations across app, voice, and web experiences without losing context or having to repeat themselves."



Governing agents with confidence

Deploying an autonomous digital workforce at scale requires centralized oversight to prevent organizational sprawl and mitigate operational risks.

Gemini Enterprise applies enterprise-grade governance and control to every agent in the fleet, whether built internally or sourced from the partner ecosystem.

Gemini Enterprise's secure-by-design architecture

1

Agent Identity

This capability assigns a unique cryptographic ID to every active agent. This creates a clear, verifiable, and fully auditable trail for every action an agent takes, mapping directly back to defined enterprise authorization policies and Google Cloud Identity IAM principals.

2

Agent Registry

This centralized catalog serves as the single source of truth for approved tools, agents, and skills. It indexes every internal asset, simplifying discovery while ensuring that users only access governed and approved resources across the entire organization.

3

Agent Gateway

Acting as the central control point for the ecosystem, Agent Gateway manages authentication and authorization across all environments. It enforces consistent security policies and Model Armor protections to safeguard against prompt injection, tool poisoning, and sensitive data leakage.

4

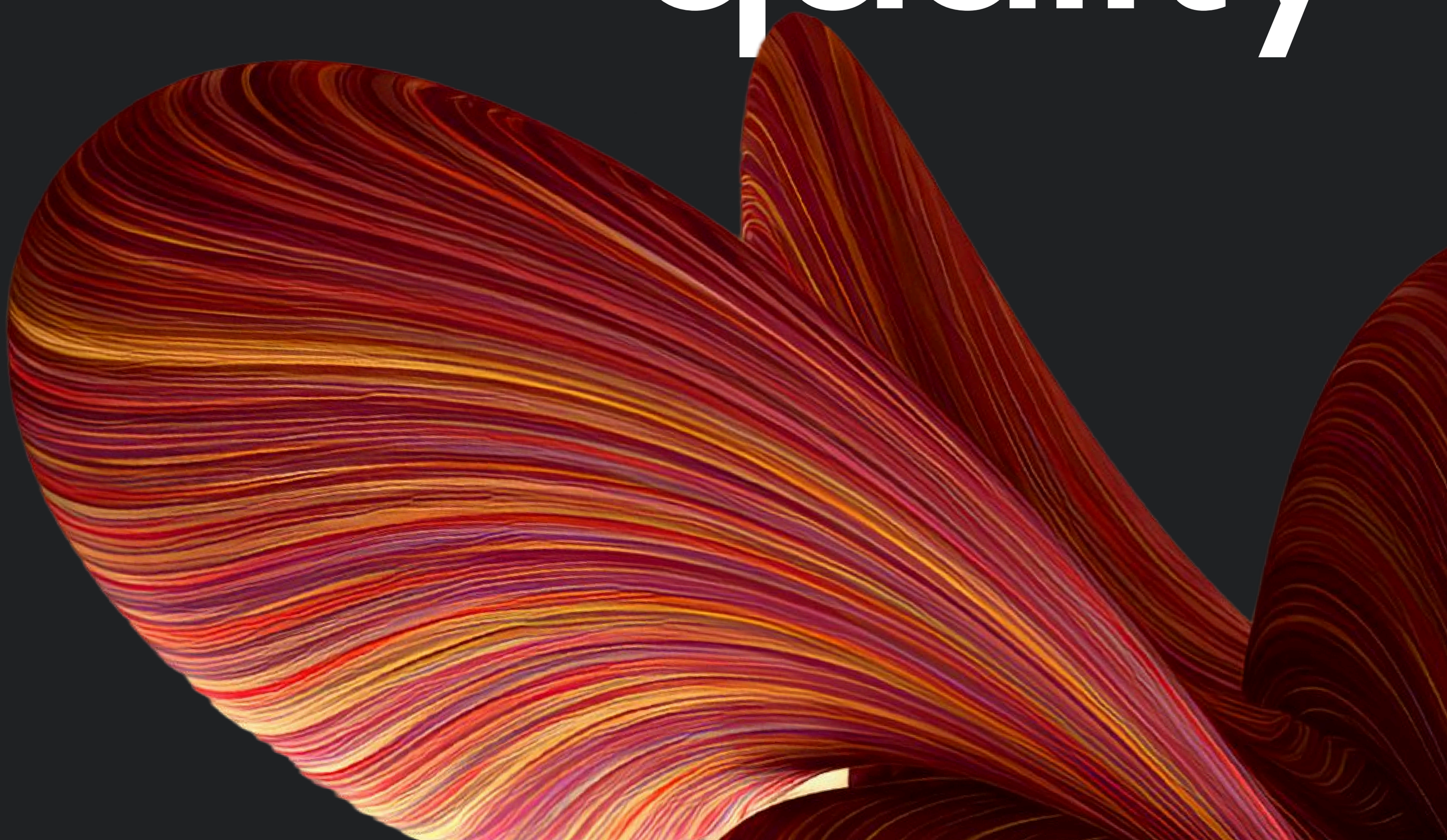
Agent security

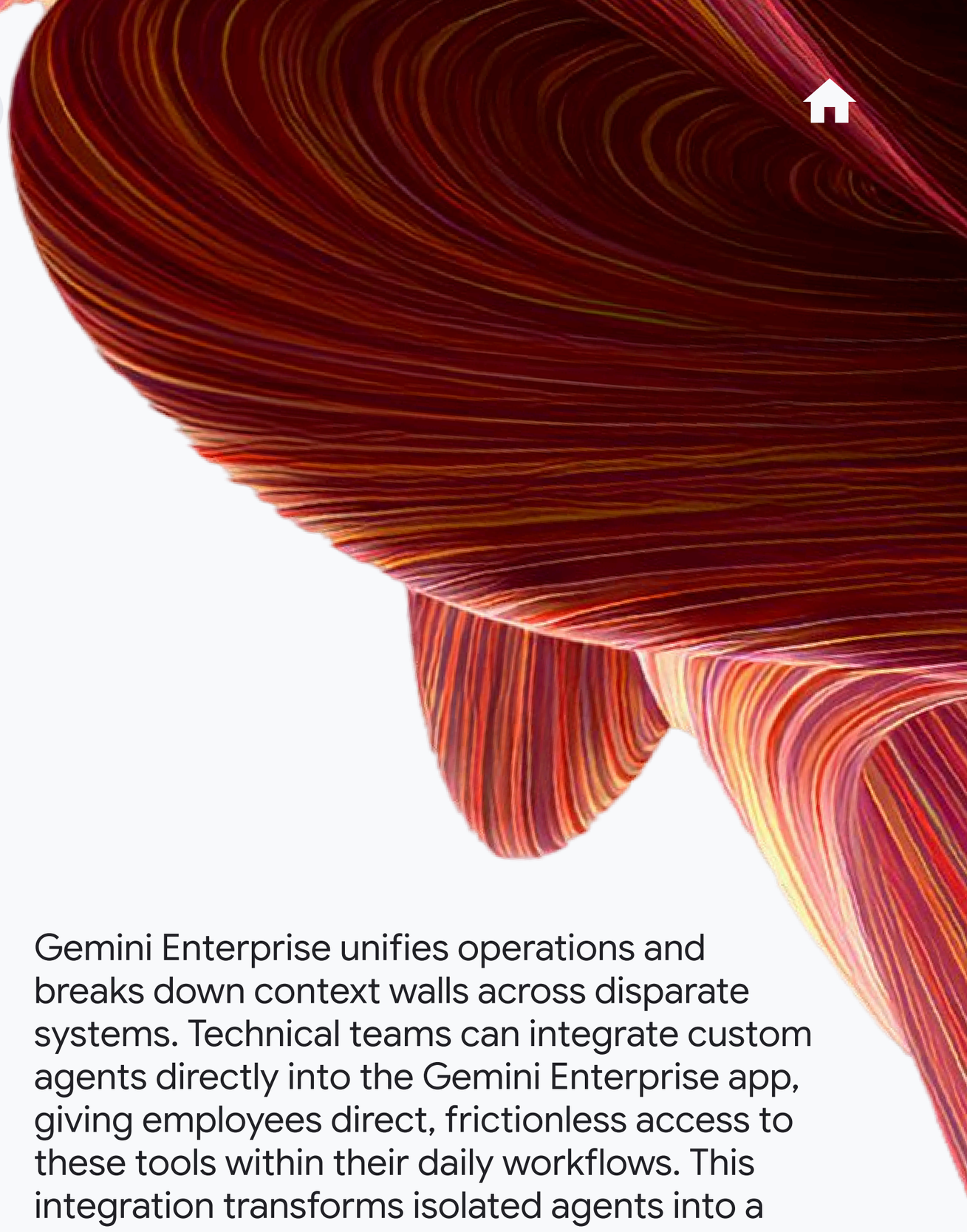
This layer monitors compliance and detects anomalies in real time. By using statistical models and an LLM-as-a-judge framework, Gemini Enterprise flags unusual reasoning, while Agent Threat Detection provides critical visibility into malicious activity to ensure compliant and predictable behavior.

BEST BUY

“At Best Buy, we see as more organizations adopt AI agents, agent identity is becoming just as important as human identity. In the past, we've struggled with orphaned service accounts, unclear ownership, and permissions that kept growing over time. Agent Identity helps bring accountability and governance to autonomous systems by making it clear who an agent is, what it can access, and who is responsible for it. From a security standpoint, applying least-privilege access to agents reduces risk while giving organizations the confidence to scale AI safely.”

Optimizing agent quality





Reliable agents require data-driven testing, not trial and error. Gemini Enterprise provides four core capabilities to help you move from guesswork to precision:

Gemini Enterprise unifies operations and breaks down context walls across disparate systems. Technical teams can integrate custom agents directly into the Gemini Enterprise app, giving employees direct, frictionless access to these tools within their daily workflows. This integration transforms isolated agents into a coordinated, enterprise-wide digital workforce that collaborates across Google Workspace and third-party systems.

1

Agent evaluation

Continuously monitor agent performance to proactively catch degradation and drift, using flexible metrics ranging from custom Python and LLM-as-a-judge to DeepMind adaptive rubrics.

2

Agent observability

Gain end-to-end visibility through comprehensive tracing and real-time dashboards that track agents' internal reasoning, tool usage, and execution speed.

3

Agent simulation

Test agents in controlled environments using virtual tools and synthetic users, while automatically grading multi-step conversations for task success and safety.

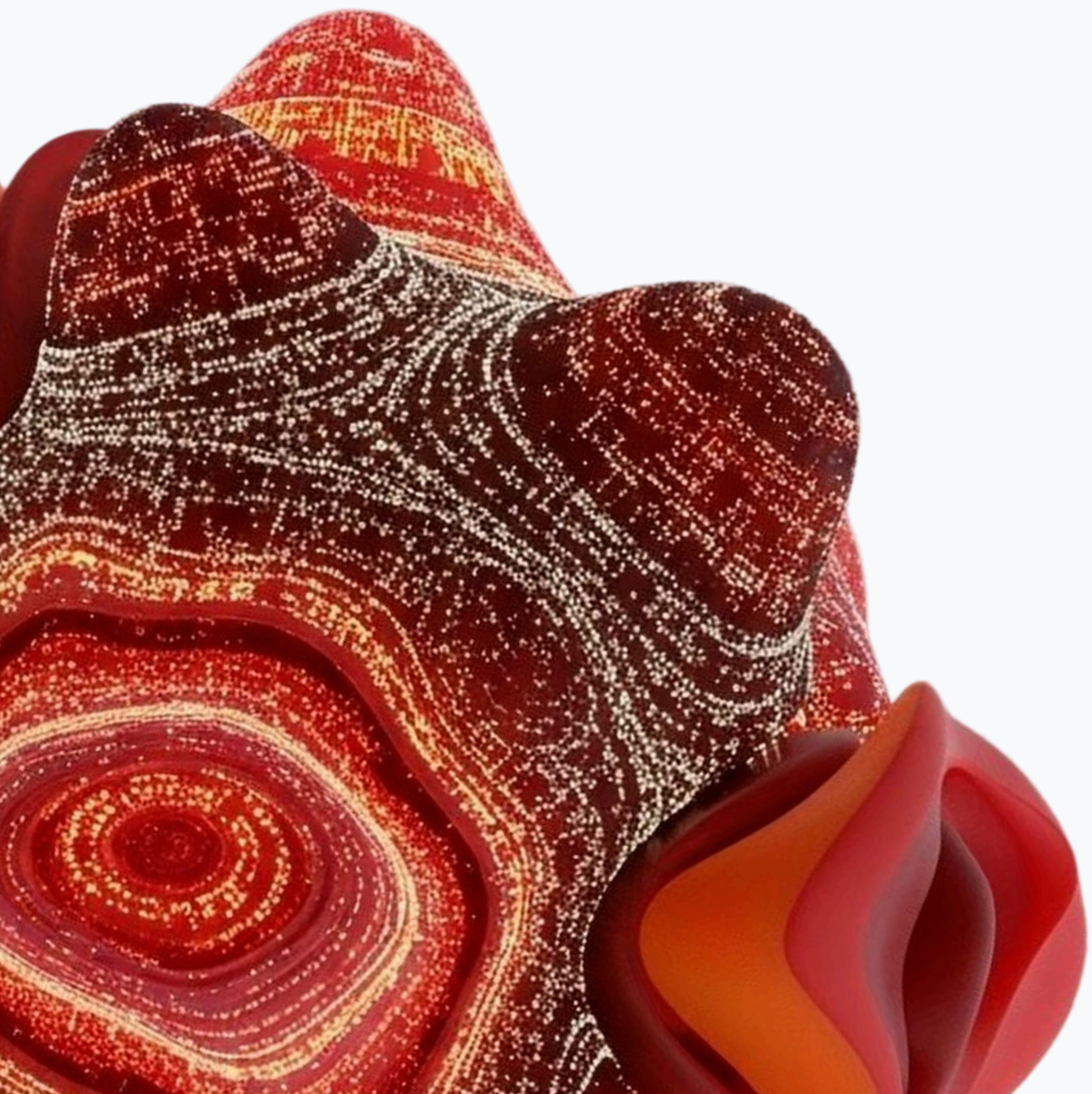
4

Agent optimizer

Automate agent refinement and eliminate manual log digging by clustering real-world production failures to generate targeted system instruction improvements.

LIBERTY GLOBAL

“At Liberty Global, Gemini Enterprise provides the high-speed engine our developers need to rapidly create and deploy specialized AI capabilities. When it comes to governance, which is paramount in a multi-entity environment, features like Agent Gateway and Agent Registry are absolute game-changers. They allow us to enforce strict security protocols and maintain centralized oversight, ensuring AI is deployed safely and compliantly across all our diverse companies.”



Unlock the future



The shift from using AI to carry out simple tasks to true agentic autonomy is now an operational reality.

Gemini Enterprise delivers a definitive competitive advantage through three core pillars: research-to-reality, multi-cloud flexibility, and open interoperability. By combining frontier intelligence with enterprise-grade infrastructure, Gemini Enterprise makes it easy to deploy trusted, autonomous digital workforces at scale across the organization.

Begin your journey today. Visit the Google Cloud console to start building. For detailed implementation guidance and the latest features, access the updated technical documentation and release notes.

Google Cloud

Activate your agentic enterprise

Contact us