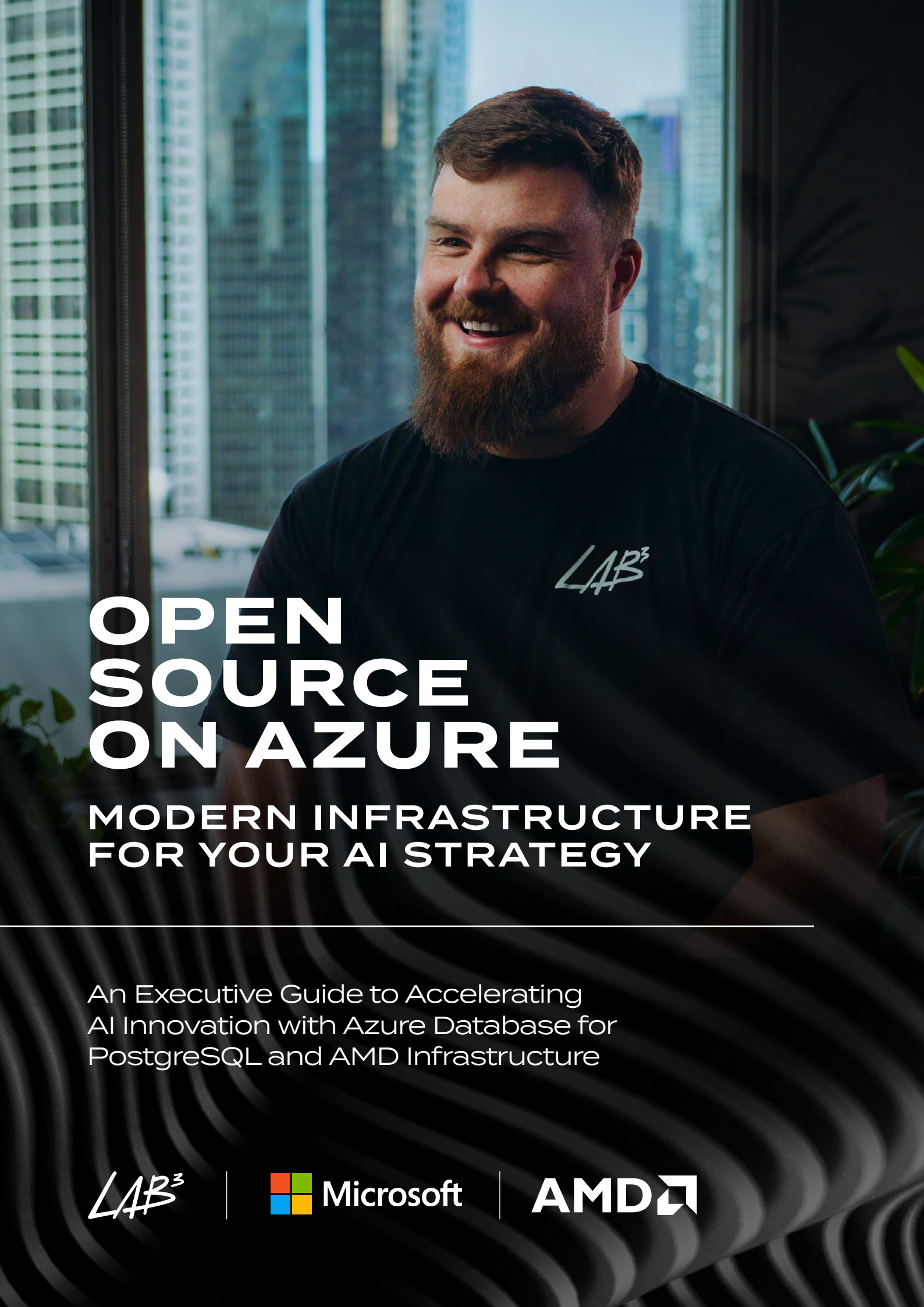


OPEN SOURCE ON AZURE

MODERN INFRASTRUCTURE FOR YOUR AI STRATEGY

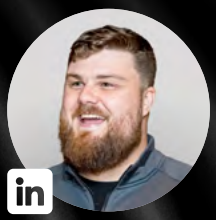
An Executive Guide to Accelerating AI Innovation with Azure Database for PostgreSQL and AMD Infrastructure

LAB3 | Microsoft | AMD

CONTENTS

FOREWORD

The CapEx Conversion: Leveraging Azure Database for PostgreSQL + AMD Infrastructure to fund innovation



By
LACHLAN WHITE | Chief Technology Officer, LAB[3]

IBM Champion 2026 & HashiCorp Ambassador 2021-2025

There's a shift happening in enterprise technology that a lot of organizations are overlooking in their AI strategy. Market leaders are changing their entire architectural philosophy to embrace open source. This shift is unlocking rapid modernization and innovation, and delivering early AI-wins for those that embrace it.

Why? Three letters: TCO. Total Cost of Ownership as an ongoing focus into future financial years. Organizations that have taken this approach consistently reclaim their CapEx to be used more effectively, converting migration and modernization savings into deployable investment into innovation for AI initiatives.

With mounting hardware and chip supply chain risks, and exponential growth to licensing and maintenance costs associated with on-premises solutions, there's never been a more compelling time for an enterprise to modernize their digital estate.

Specifically, we are advising many industry-leading clients to move to the open source stack on Azure, leveraging Azure Database for PostgreSQL on AMD infrastructure. This approach can cut infrastructure costs, provide a staged approach to modernization, and deliver quick wins by making existing data AI-ready.

For the last decade, Microsoft has invested heavily in adopting and supporting open source technologies natively on Azure.

Two-thirds of all compute cores running on Azure today are Linux. The most recommended managed database for new AI workloads is PostgreSQL. The container orchestration layer is Kubernetes, and the AI agent framework is also open source. In addition AMD invests heavily to ensure open source technologies run optimally on its processors and on Azure services. AMD is a top 10 contributor to Linux kernel development with a focus on day-zero enablement for new CPUs and GPUs, and has served as a longstanding member of the Linux Foundation

As a technology provider known for our unique accelerators which speed up infrastructure, AI and security deployments at scale, LAB[3] preferences the open source architecture and solutions set out in this Guide.

We've seen firsthand how the combination of open source flexibility with Azure gives enterprises the best of both worlds: innovation velocity without sacrificing governance, and cost-performance without sacrificing support. Together with Microsoft and AMD, we have developed this guide to help busy executives understand **how to leverage open source on Azure as the foundation for your AI ambitions.** Everything you read here is grounded in our real-world experience delivering for clients. Not theory, but practice.

Let's get into it.

FOREWORD

By
VARUN DHAWAN
Principal Product Manager - Azure Database for PostgreSQL, Microsoft

Artificial Intelligence is rapidly reshaping how organizations modernize applications, interact with data, and unlock value from their enterprise data. As organizations move from experimentation to production-scale AI adoption, open-source technologies are increasingly becoming the foundation for innovation across the AI stack.

PostgreSQL has emerged as one of the most important open-source data engines powering this transformation.

With innovations such as pgvector bringing native vector search capabilities directly into PostgreSQL, organizations can now combine relational data, embeddings, and semantic search in a unified platform.

This architectural simplicity is helping accelerate modern AI scenarios such as Retrieval-Augmented Generation (RAG), intelligent agents, and enterprise knowledge search without introducing unnecessary operational complexity.

At Microsoft, we believe Azure is the best place to run PostgreSQL workloads at enterprise scale. Azure Database for PostgreSQL combines the flexibility and innovation of the open-source PostgreSQL ecosystem with the operational capabilities enterprises expect from Azure (including high availability, security, governance, compliance, and deep integration with Azure AI services).

Together, this allows organizations to modernize existing workloads while building the next generation of AI-powered applications on a trusted cloud platform.

Partnerships across the open source ecosystem continue to play an important role in helping customers accelerate this journey. Microsoft, AMD, and partners like LAB[3] are working together to deliver scalable, secure, and cost-efficient infrastructure and platform solutions that help organizations move faster with confidence.

We hope this guide provides a practical perspective on how open-source technologies on Azure can help organizations build a strong foundation for enterprise AI innovation.

How Open Source Savvy Are You?

"In an uncertain world, there's tremendous safety in operating with known systems that are under your own direct control."

JEANNE BROOKS | Vice President – US Client Engagement, LAB[3]

Do you understand how Microsoft's open source strategy changes your platform decisions?

See page **6**

Have you evaluated open source alternatives to your existing proprietary data and compute services on Azure?

See page **11**

Are your AI workloads architected to take advantage of open source frameworks and AMD infrastructure?

See page **17**

Does your organization have the governance model to adopt open source at enterprise scale?

See page **21**

Microsoft + Open Source: A Platform Built In The Open

If you'd told an enterprise architect in 2005 that Microsoft would become one of the world's largest contributors to open source software, they'd have thought you were joking. And yet, here we are. The turning point is well documented. In 2009, Microsoft contributed over 20,000 lines of code to the Linux kernel. By 2015 Microsoft's chairman and CEO Satya Nadella put up a slide at a media briefing that simply read "Microsoft ♥ Linux.". Microsoft wasn't just saying it loved Linux. It was rebuilding Azure around it.

Today, 66% of customer cores in Azure run Linux. The platform supports every major distribution. From Red Hat Enterprise Linux to Ubuntu, SUSE, Debian, Oracle Linux, and more. Azure Kubernetes Service (AKS) runs on Linux. Azure Container Apps runs on Linux. Azure Database for PostgreSQL, Microsoft's recommended managed relational database for AI workloads, is open source.

"It's critical to understand what and how AI can be applied to drive change within your business. From a CFO perspective, it's not just about "cost-out" for IT, it's about identifying new revenue opportunities."

ANTHONY WALES
Director Strategic Growth, LAB[3]

DID YOU KNOW?

Microsoft's own COSMIC platform runs Microsoft 365 on Azure Kubernetes Service (AKS) across millions of cores.

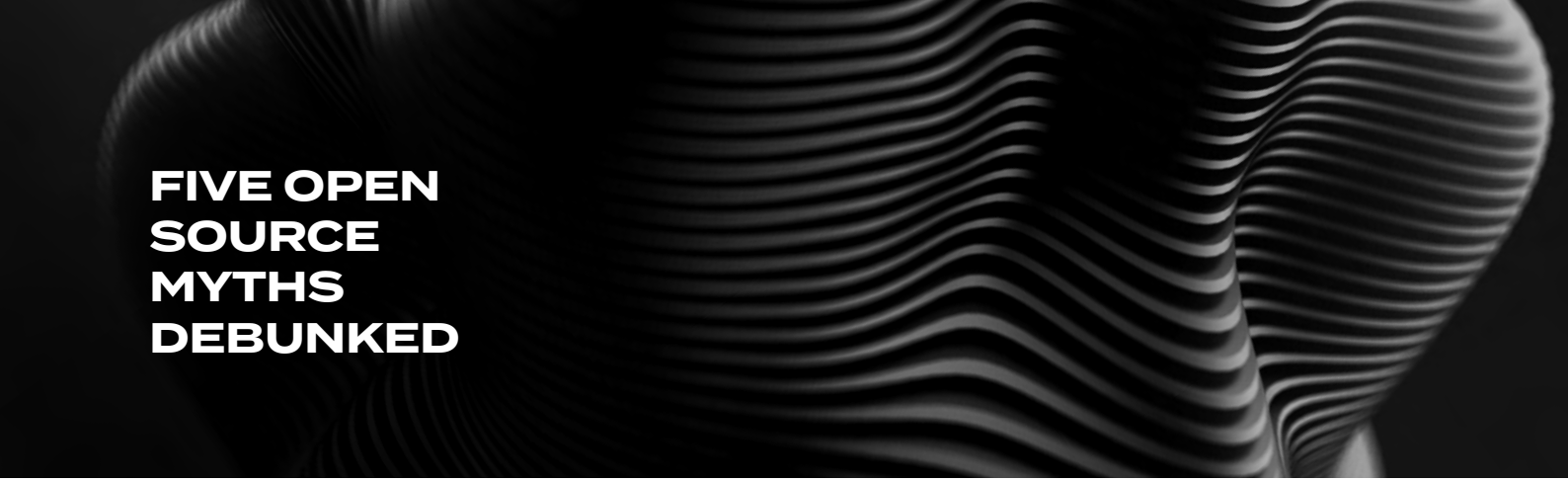
OpenAI's ChatGPT is built on Azure using AKS for container orchestration, Azure Blob Storage for content, and Azure Database for PostgreSQL and Azure Cosmos DB for globally distributed data.

The same open source technologies that power these platforms are available to all enterprises on Azure.

Microsoft's AI approach shares the same story.

The Microsoft Agent Framework, which reached its 1.0 production release in April 2026, is fully open source. It unifies the research innovations from AutoGen (a Microsoft Research project) with the enterprise foundations of Semantic Kernel into a single SDK. It's available on GitHub, integrates with Azure AI Foundry for managed deployment, and runs on open infrastructure.

The implication is straightforward: when you choose Azure for AI workloads, you're not choosing between Microsoft's ecosystem and open source. They're the same thing. Microsoft has recognized that open source is the most effective way to build cloud infrastructure at scale. Understanding this impacts how you should evaluate your technology investments, your skills strategy, and your AI roadmap.



FIVE OPEN SOURCE MYTHS DEBUNKED

MYTH #1

Open source is unreliable or lacks integrity

The transparency that open source brings is the new safety. You are not at the mercy of any one company's black box proprietary solution.

MYTH #2

Open source isn't secure or defensible

This myth has been well and truly debunked. You can control access to your technology environment and use open source architectures.

MYTH #3

Legacy systems provide stronger job security for IT teams

Legacy environments are slow, expensive, and block Data and AI innovation. In competitive markets, having a modern cloud native technology environment running at half the cost and twice the speed delivers true market agility.

MYTH #4

Buying from a big name vendor ensures predictable costs

As we've seen from several large providers, there's been an economic pivot from what had historically been acceptable licensing structures to new approaches that result in year-on-year exponential cost increases.

MYTH #5

Open source solutions lack professional, global support

These days companies are built on open source and they invest in ensuring support. In addition to their own teams, they leverage crowdsourced innovation via the global open source community. Nowadays, wherever you are in the world there is 24/7 support for Postgres.

PART 1

THE OPEN SOURCE AI STACK ON AZURE

Understanding how open source enables AI on Azure requires looking at the full stack. From the compute layer through to the application frameworks, each layer has a role to play, and each is anchored by open source technologies that Azure delivers as managed, enterprise-grade services.

Compute & Infrastructure

What compute foundation should your AI workloads run on?

Every AI workload needs data and compute power. Azure offers a cost-effective approach and native support to deliver a secure, scalable AI foundation.

- Azure Database for PostgreSQL and AMD infrastructure offer a quick, cost-effective approach to migration and modernization
- AMD EPYC™ processor options running Linux based workloads on Azure provide unmatched price performance for AI processing.

AMD EPYC™ on Azure

Azure's latest AMD-based virtual machines (Dasv7, Easv7, Fasv7 series) are powered by 5th Generation AMD EPYC (Turin) processors. These deliver up to 35% better CPU performance and price-performance compared to the previous generation, with workload-specific gains that are significant:

- up to 65% for in-memory cache applications
- up to 80% for cryptographic workloads
- up to 130% for web server applications[1].

For AI workloads specifically, AMD's architecture brings several advantages. The core density of EPYC processors makes them well-suited to AI inference tasks where you need to run many concurrent requests efficiently. Not every AI workload requires a dedicated GPU, and AMD's options prove this on all fronts.

AMD also powers GPU-accelerated VMs, including the ND MI300X v5, with Azure having been the first to launch this AI-focused offering in the cloud, and the industry's first confidential GPU VM, the NCC H100 v5.

Confidential Computing

AMD Infinity Guard features, including Secure Encrypted Virtualization (SEV), enable hardware-based trusted execution environments that protect AI models and data even from the cloud provider. Azure's confidential computing offerings (e.g. DCasv6, ECasv6 VMs) are built on this AMD technology. For organizations in regulated industries handling sensitive data in AI workloads, this capability is increasingly relevant.

DID YOU KNOW?

Azure was the first global cloud provider to deploy AMD EPYC-based virtual machines back in 2017.

Today, AMD and Microsoft have brought over 60 VM series options to market through years of deep engineering collaboration spanning general purpose, memory-optimized, compute-optimized, storage-optimized, high performance computing, GPU-accelerated, and confidential computing workloads.

[1] Announcing General Availability of Azure Da/Ea/Fasv7-series VMs based on AMD 'Turin' processors, Azure Compute Blog, January 2026

Container Orchestration

Above the VM layer sits containerization: Azure Kubernetes Service (AKS) and Azure Container Apps.

AKS is one of the largest managed Kubernetes deployments in the world. It provides managed Kubernetes with Azure Policy for governance, Defender for Containers for security, and Azure Monitor for observability. For teams that need full control over your container orchestration, AKS is the natural choice.

Container Apps provides a simpler, serverless container runtime for teams that don't need full Kubernetes control. It abstracts away cluster management while maintaining the same security and compliance posture. Both run on Linux, both run on AMD, and both are the natural deployment target for AI applications.

Critical Activities
Compute & Infrastructure

- **Evaluate AMD-based VM series for AI inference workloads.** Not every AI workload needs a GPU. AMD EPYC's core density and memory bandwidth make it well-suited to inference tasks where high concurrency matters more than raw single-thread performance.
- **Assess your container orchestration requirements.** Determine whether AKS (full Kubernetes control) or Container Apps (serverless simplicity) is the right fit for your AI workloads. Both run on Linux and AMD infrastructure.
- **Consider confidential computing for regulated AI workloads.** If your AI applications process sensitive data, AMD's SEV technology on Azure provides hardware-based protection that extends to the AI model itself.

Data

Cloud Data Platforms provide the data foundation to support AI Workloads

If compute is the engine, data is the fuel. Microsoft Fabric leverages Apache Spark (an open-source, distributed processing system designed for big data workloads) to make it easier for you to access large compute for enterprise reporting and more effective business decision making.

Azure Database for PostgreSQL has emerged as the open source database of choice for AI workloads and broader relational storage business needs on Azure.

Azure Database for PostgreSQL

Azure Database for PostgreSQL is a fully managed service that combines the flexibility of open source PostgreSQL with the operational guarantees of an Azure PaaS offering: automated backups, high availability with zone redundancy, managed patching, and integration with Azure's identity and security stack via Entra ID and managed identities.

A common key trend LAB[3] is seeing in 2026 across industries is that many organizations are migrating from their on-premises closed source databases to open-source cloud databases.

A known business rationale for doing so is that closed source databases often come with expensive licensing which includes patching and maintenance, often towards the databases end-of-life.

The move to cloud often coincides with a contract renewal for their databases and is considered as an option to mature the data estate and progress with a digital modernization program. The motivation for opting to proceed is often for access to the latest AI and agentic tools which can integrate with the data when hosted on Azure in a PostgreSQL database. Now organizations can have Agentic Models query their databases as part of providing decision support and recommendations to business users as part of the daily needs.

DID YOU KNOW?

PostgreSQL is not new, it's one of the most widely adopted relational databases in the world, with decades of enterprise use.

What has changed is its relevance to AI workloads.

The combination of relational data management with vector search capabilities makes PostgreSQL uniquely positioned as a unified data layer for AI applications.

pgvector: Vector Search in Your Database

What makes PostgreSQL particularly relevant for AI is pgvector, an open source extension that adds vector similarity search capabilities directly into the database. This means enterprises can store embeddings alongside their relational data, run semantic search queries using standard SQL, and build Retrieval-Augmented Generation (RAG) applications without introducing a separate vector database into their architecture.

Azure has invested heavily in making pgvector performant at scale:

- **DiskANN indexing** (a Microsoft Research innovation) enables efficient approximate nearest-neighbor search on billion-point datasets, scaling vector search far beyond what standard indexing approaches can handle
- **The Azure_AI extension** allows developers to generate embeddings directly within PostgreSQL using Azure OpenAI, eliminating the need for external embedding pipelines
- **The Azure_Local_AI extension** enables locally deployed LLMs to generate embeddings within the database itself.

Vectorizing a database is a common approach for large language models (LLMs) to access data quickly and efficiently. By converting text into relationships known as 'embeddings', what happens is the AI can make connections much faster based on understanding of the language. This is a key step in the process of building your AI models known as Retrieval-Augmented Generation (RAG). It grounds the models from their baseline that is received from OpenAI, Anthropic and similar, and grounds the LLM responses in your enterprise data.

Extensions such as pgvector allow vector storage and similarity search directly within Azure Database for PostgreSQL, providing scalable, performant retrieval while keeping data in a governed, relational environment.

"With Open Source, we see the waves of innovation occur more broadly and more rapidly at an ecosystem level compared to proprietary enterprise software.

pgVector is a great example of that, letting you add AI to your existing Postgres database without starting from scratch."

LACHLAN WHITE
CTO, LAB^3

Spark With Microsoft Fabric

Apache Spark is a mature and widely adopted open source project, with a strong ecosystem and long-standing enterprise usage. What has changed over time is its role within data and AI platforms. When combined with open data formats and shared storage layers, Spark provides a consistent and scalable processing engine for data engineering, analytics, and AI workloads within a unified data platform.

Spark and Microsoft Fabric provide easier access to large compute for enterprise reporting and more effective business decision making. Fabric leverages Spark as its core distributed compute engine for the Fabric platform. It combines the openness and scale of the Spark ecosystem with a SaaS-style operational model. What Spark enables is easy cluster management, while providing elastic scaling, integrated security, and native connectivity to your data lake. Governance is managed via identity and access controls aligned with Microsoft's broader security model. What this enables is Spark workloads to operate within existing enterprise controls.

A key industry trend we've observed in 2026 is the continued adoption of open-source analytics engines such as Spark to replace bespoke or proprietary data processing platforms organizations used to host on premises. Organizations are increasingly consolidating their Power BI, enterprise reporting, and data ingestions. This is often across batched, and streamed data sources onto Spark-based architectures as part of cloud modernization and data platform rationalization initiatives. These efforts often coincide with data estate consolidation, cost optimization, and the need to support advanced analytics and AI use cases that require scalable, distributed compute. Spark and Fabric make agentic interactive natural-language queries of your data possible.

The Architectural Simplicity of Open Source

The practical benefit for enterprises with open source is architectural simplicity. Instead of managing a relational database for structured data, a vector database for embeddings, and a document store for unstructured content, Azure Database for PostgreSQL with pgvector can serve as a unified data layer. Fewer moving parts, fewer integration points, fewer things to secure and govern.

Beyond PostgreSQL, Azure Cosmos DB offers open API compatibility (including PostgreSQL wire protocol) for globally distributed scenarios, and the broader data ecosystem, Apache Kafka, Apache Spark, Redis, is available through managed services that maintain open source compatibility.

DID YOU KNOW?

Azure Database for PostgreSQL with pgvector supports multiple indexing strategies for different performance profiles:

- IVFFlat for fast approximate search
- HNSW for high-recall scenarios
- DiskANN (Azure-exclusive) for billion-scale datasets.

This means you can tune your vector search performance based on your specific requirements. Move from initial prototype to production scale without changing your database technology.

Critical Activities
Data

- **Evaluate Azure Database for PostgreSQL as your AI data layer.** If you're currently running a commercial relational database, PostgreSQL with pgvector on Azure will likely provide equivalent or better capability for AI workloads at significantly lower licensing cost.
- **Assess whether a separate vector database is necessary.** For many enterprise AI use cases, PostgreSQL with pgvector eliminates the need for a standalone vector database, reducing architectural complexity and operational overhead.
- **Enable the azure_ai extension for embedding generation.** This allows your development teams to generate and store embeddings directly within Azure Database for PostgreSQL using Azure OpenAI to simplify the data pipeline for RAG applications.
- **Consider a cloud modernization program with Fabric.** Leverage Apache Spark for large scale compute and Power BI reports, and connect your data estate with the latest Microsoft AI and agentic tools for quick wins.

AI & Agent Frameworks

Unlocking AI readiness is integral to Migration & Modernization

Assuming your migration and modernization journey includes Linux on Azure and PostgreSQL as a data foundation, cultural AI adoption is then layered on top. AI projects need to build on a foundation. LAB[3] takes a layered approach based on Azure architectures. The bottom layer is Infra and Security with standard ways of establishing a virtual network with private endpoints (Landing Zone level). The next level up considers what AI capabilities are needed which can be standardized with our accelerators to speed up all subsequent AI projects. The top level is the specific project achieving the business outcome.

Make Infrastructure Selection a Part of Your AI strategy

Often overlooked but vital. AI initiatives require high level compute, which can come with an associated high cost. AMD compute options on Azure offer the best performance per dollar for PostgreSQL and can help control ongoing run costs for both database compute and process intensive AI workloads. Moving from non AMD compute to their latest v7 options on Azure can improve PostgreSQL...

- performance by 41%
- performance per dollar by 49%
- operational expenses up to 33% [2]

AMD's compute options also benefit from open source ecosystem enabling architecture, meaning the global support community are actively engaged in identifying even further optimizations, which could continue to reduce lifetime run costs as new research is incorporated.

"Each AI project can build its own foundation and reinvent the wheel for infrastructure, security and risk management. Or you can standardize those layers once and speed up every subsequent project by 30%."

JASON LEONARD
AI Practice Lead, LAB[3]

What frameworks should you build your AI applications on?

The AI framework layer is where Microsoft's open source commitment becomes most tangible for development teams building AI applications.

Microsoft Agent Framework as the Centerpiece

Released as a production-ready 1.0 in April 2026, Agent Framework is the convergence of two influential open source projects: AutoGen, which pioneered multi-agent orchestration patterns at Microsoft Research, and Semantic Kernel which provided enterprise-grade foundations for AI application development.

[2] When moving from E16sv5 to E16asv7, AMD claim 9xx5C-109

Agent Framework supports both single-agent and multi-agent patterns:

- Sequential orchestration: one agent hands off to the next
- Concurrent orchestration: multiple agents work in parallel
- Handoff orchestration: agents delegate based on capability
- Group chat orchestration: agents deliberate and reach consensus

Built-in support for streaming, checkpointing, human-on-the-loop approvals, and pause/resume for long-running workflows makes it suitable for production enterprise deployments. It's available in both Python and .NET, integrates natively with Azure AI Foundry for managed deployment, and supports open standards including Model Context Protocol (MCP) for dynamic tool discovery and Agent-to-Agent (A2A) protocol for cross-runtime collaboration.

Open Standards and Interoperability

The adoption of open standards is a critical differentiator. MCP allows agents to dynamically discover and invoke tools exposed by external services, meaning your AI agents can connect to your existing systems without custom integration code. A2A enables agents built on different frameworks to collaborate, preventing vendor lock-in at the orchestration layer.

DID YOU KNOW?

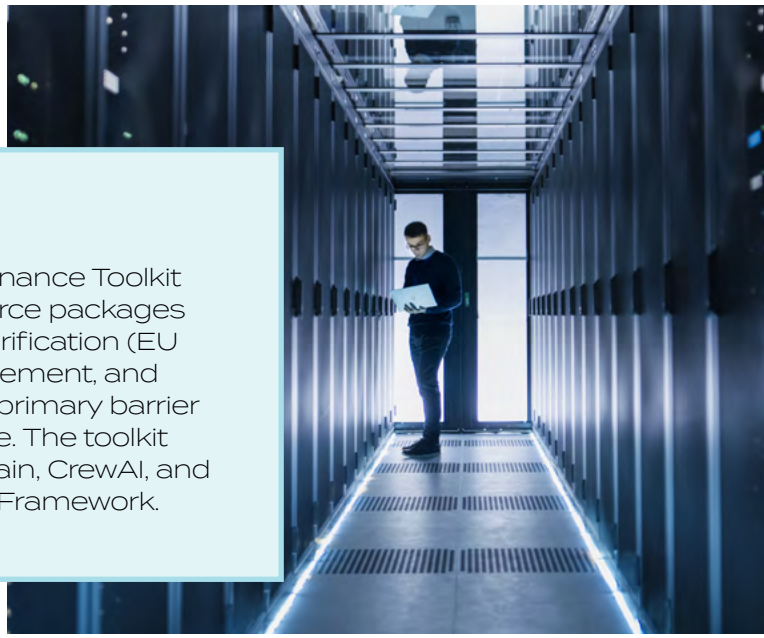
Microsoft recently released the Agent Governance Toolkit (April 2026) which includes seven open source packages covering policy enforcement, compliance verification (EU AI Act, HIPAA, SOC2), plugin lifecycle management, and RL training governance. This addresses the primary barrier to enterprise agent deployment: governance. The toolkit works across frameworks including LangChain, CrewAI, and Google ADK, not just Microsoft's own Agent Framework.

Microsoft Foundry

Foundry provides a unified platform for deploying and managing AI models including open models from Meta (Llama), Mistral, and others alongside popular proprietary models from OpenAI and Anthropic. This means enterprises aren't locked into a single model provider. You can evaluate and deploy the model that best fits each use case.

Whether that's a large frontier model for complex reasoning or a smaller, cost-effective model for specific and routine tasks, Foundry is designed to give executives a clear path from stable operations to safe AI experimentation at lower cost and risk.

Foundry provides a variety of ways to manage risk, including 'Human ON the Loop' risks. With traditional Human-IN-the-Loop structures, users are constantly prompted to check and confirm outputs. However, while most people are diligent in checking the first few outputs, over time as trust develops, efforts to verify naturally diminish. Foundry helps with guardrails including Identity and access controls so that the AI has only the same level of access as the person doing the role. This keeps the AI locked down to its specific, core function. Foundry also includes 'prompt shields' that protect generative outcomes and block bad faith actors through Content Safety components.



The broader ecosystem

Frameworks like LangChain and LlamaIndex integrate natively with Azure services. The Kubernetes AI Toolchain Operator (KAITO), a CNCF Sandbox project contributed by Microsoft, automates AI workload deployment on AKS, supporting large language models, fine-tuning, and RAG across cloud and edge.

The consistent theme across this layer is that Microsoft is building in the open. These aren't proprietary SDKs with open source wrappers. They're genuinely open source projects that happen to integrate deeply with Azure for enterprises that want managed infrastructure.

Critical Activities
AI & Agent Frameworks

- **Evaluate Microsoft Agent Framework for multi-agent workloads.** If you're building AI applications that go beyond single-prompt interactions, Agent Framework provides production-grade orchestration with enterprise governance built in.
- **Assess Azure AI Foundry for model management.** The ability to deploy and evaluate multiple models (including open models) from a single platform reduces vendor lock-in and enables model selection based on performance and cost.
- **Plan for agent governance early.** The Agent Governance Toolkit provides the policy enforcement and compliance tooling that enterprise deployments require. Implementing governance from the start is significantly easier than retrofitting it later.

DevOps & Governance

How do you operationalize the open source stack?

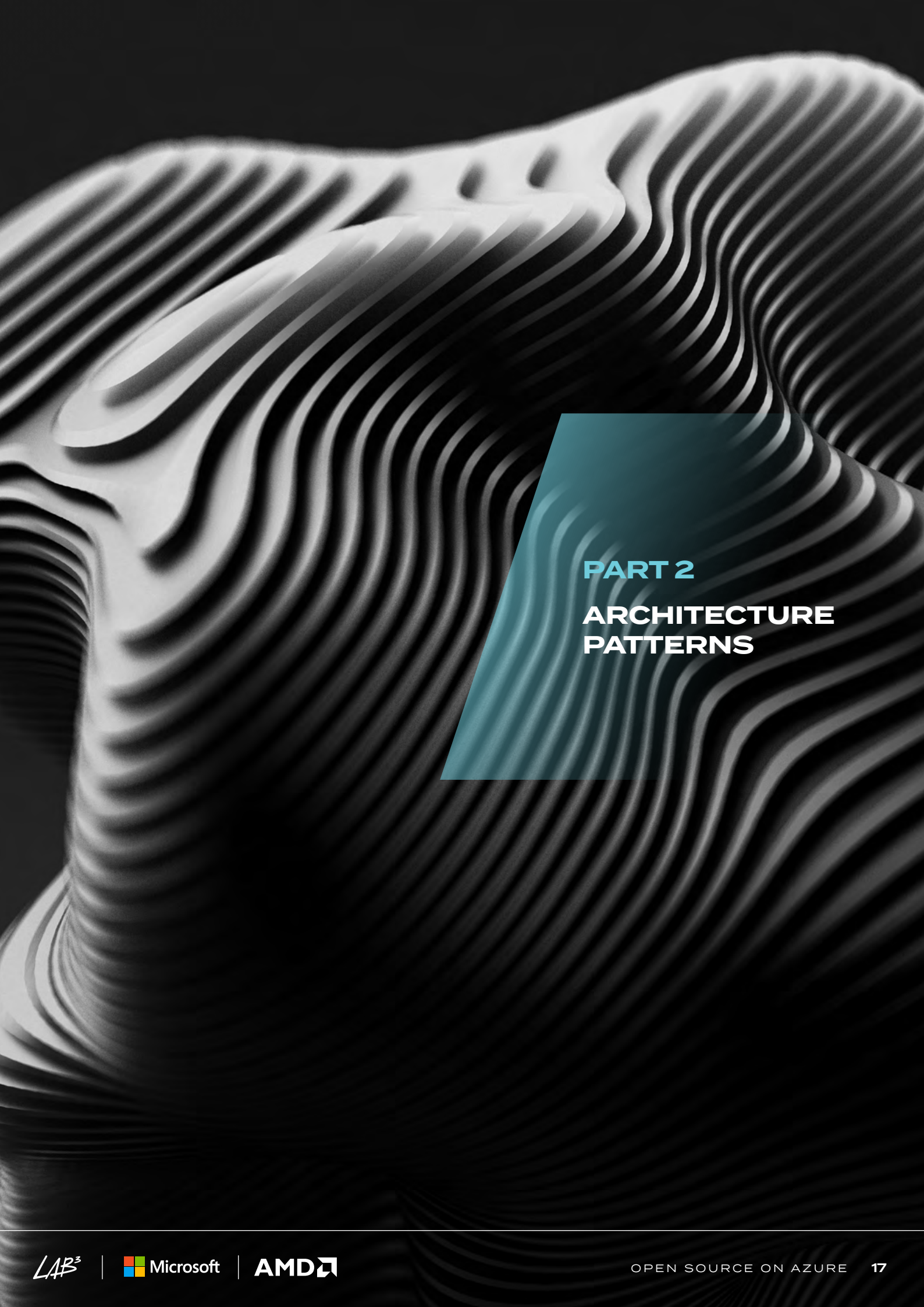
The operational layer that ties everything together is equally open source. Terraform and Bicep provide infrastructure as code. GitHub Actions provides CI/CD pipelines. Azure Monitor and the OpenTelemetry standard provide observability. Microsoft Defender for Cloud provides security posture management across the entire stack, including open source components.

This layer is where the "enterprise-grade" part of the equation lives. Open source gives you flexibility. Azure's managed services give you the governance, compliance, and operational guarantees that enterprise IT teams require. The combination is what makes this stack viable for regulated industries and large-scale deployments.

DID YOU KNOW?

LAB³ is a hyper-specialized HashiCorp partner and a verified GitHub partner. We leverage Terraform extensively for deploying and managing cloud foundations on Azure, including the open source AI stack covered in this guide. Our Landing Zone Accelerator deploys secure, scalable Azure foundations with built-in automation for Day 2 operations including patching, monitoring, and governance.

Get in touch to learn how LAB³'s Landing Zone Accelerator can deliver speed-to-value for your AI strategy.

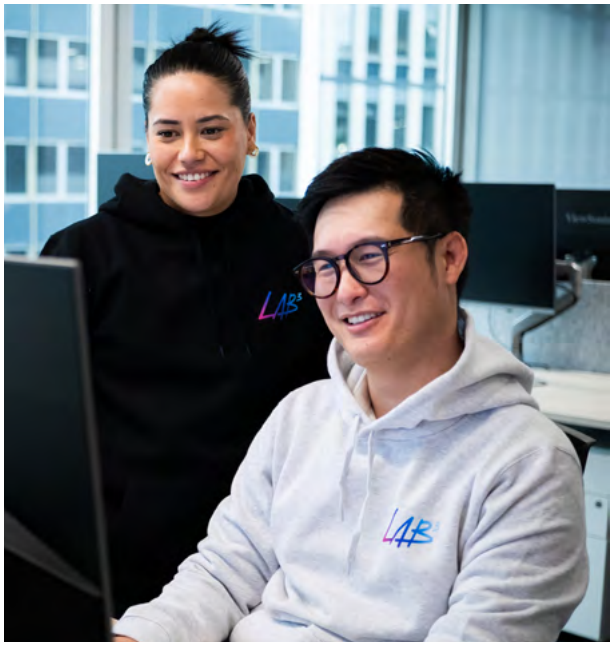


PART 2

ARCHITECTURE PATTERNS

Open Source Enabling AI Delivery

While understanding individual components is useful. Understanding how they work together is what matters. Following are three architecture patterns that represent how enterprises are deploying AI on the open source Azure stack today.



PATTERN #1

Retrieval-Augmented Generation (RAG)

The most common enterprise AI pattern and where the open source stack shines brightest.

RAG combines the knowledge retrieval capability of a search system with the natural language generation capability of a large language model. Documents and knowledge are ingested into an Azure Database for PostgreSQL option with pgvector applied, embeddings are generated via Azure OpenAI, and a retrieval layer queries for semantically relevant content to ground LLM responses in factual, enterprise-specific data.

Why this pattern works on the open source stack

What makes RAG compelling on Azure's open source stack is that PostgreSQL serves as both the transactional database and the vector store. There's no need for a separate Pinecone, Weaviate, or Qdrant deployment. Your structured data (customers, orders, configurations) and your unstructured embeddings live in the same database, query-able with standard SQL, secured by the same Entra ID policies, and backed up by the same managed service.

The compute layer (Linux VMs on AMD EPYC or Container Apps) handles the application logic. AMD's core density is particularly well-suited here for run efficiency with RAG applications that are typically I/O-bound and benefit from high concurrency rather than raw single-thread performance.

EXAMPLE USE CASES

- **Enterprise knowledge search:** Employees query internal documentation, policies, and procedures using natural language rather than keyword search.
- **Customer support automation:** AI agents retrieve relevant product documentation and case history to generate accurate, contextual responses.
- **Compliance and regulatory analysis:** Legal and compliance teams query regulatory documents to identify relevant obligations and precedents.

KEY COMPONENTS

Azure Database for PostgreSQL + pgvector → Azure OpenAI → Container Apps or AKS on AMD compute → Application layer

PATTERN #2

Multi-Agent Orchestration

The evolution from simple 'ask a question, get an answer' expectations to the more valuable 'define a goal, let agents collaborate to achieve it' approach.

Multi-agent AI represents the next evolution of enterprise AI with autonomous workflows that combine specialized agents that coordinate to complete complex tasks. This is the pattern that's enabling everything from sales pipeline analysis to compliance review to infrastructure management.

How it works on the open source stack

Microsoft Agent Framework provides the orchestration layer. Specialized agents (each with their own instructions, tools, and domain focus) coordinate through structured patterns. Each agent connects to external systems via MCP servers and to each other via the A2A protocol.

On Azure, these agents deploy as containerized workloads on AKS or Container Apps, with Azure AI Foundry providing the managed runtime for production deployments. Built-in observability through OpenTelemetry and governance through the Agent Governance Toolkit ensure enterprise-grade operations.

The open source nature of Agent Framework is critical here. Enterprises can inspect the orchestration logic, extend it with custom patterns, and avoid vendor lock-in on what is arguably the most strategically important layer of their AI architecture.

EXAMPLE USE CASES

- **Sales pipeline intelligence:** Agents analyze CRM data, market signals, and communication history to prioritize opportunities and recommend next actions.
- **Workforce demand forecasting:** Agents ingest project data, skills databases, and market benchmarks to forecast resource requirements and identify capability gaps.
- **Infrastructure operations:** Agents monitor, diagnose, and remediate infrastructure issues across cloud and hybrid environments.

KEY COMPONENTS

Microsoft Agent Framework → Azure AI Foundry → AKS on AMD compute → MCP/A2A for tool and agent interoperability

PATTERN #3

AI-Native Application Platform

The comprehensive pattern for building an entire application platform on the open source Azure stack.

Representing the most complex approach to tackle the most complex business goals, this approach simplifies both deployment as well as ongoing maintenance and costs. Deployed with Linux running on AMD for compute, AKS for orchestration, PostgreSQL for data (relational + vector), Microsoft Agent Framework for AI orchestration, GitHub for source control and CI/CD, Terraform for infrastructure, and Azure AI Foundry for model management.

Every layer is open source. Every layer runs on Azure as a managed service.

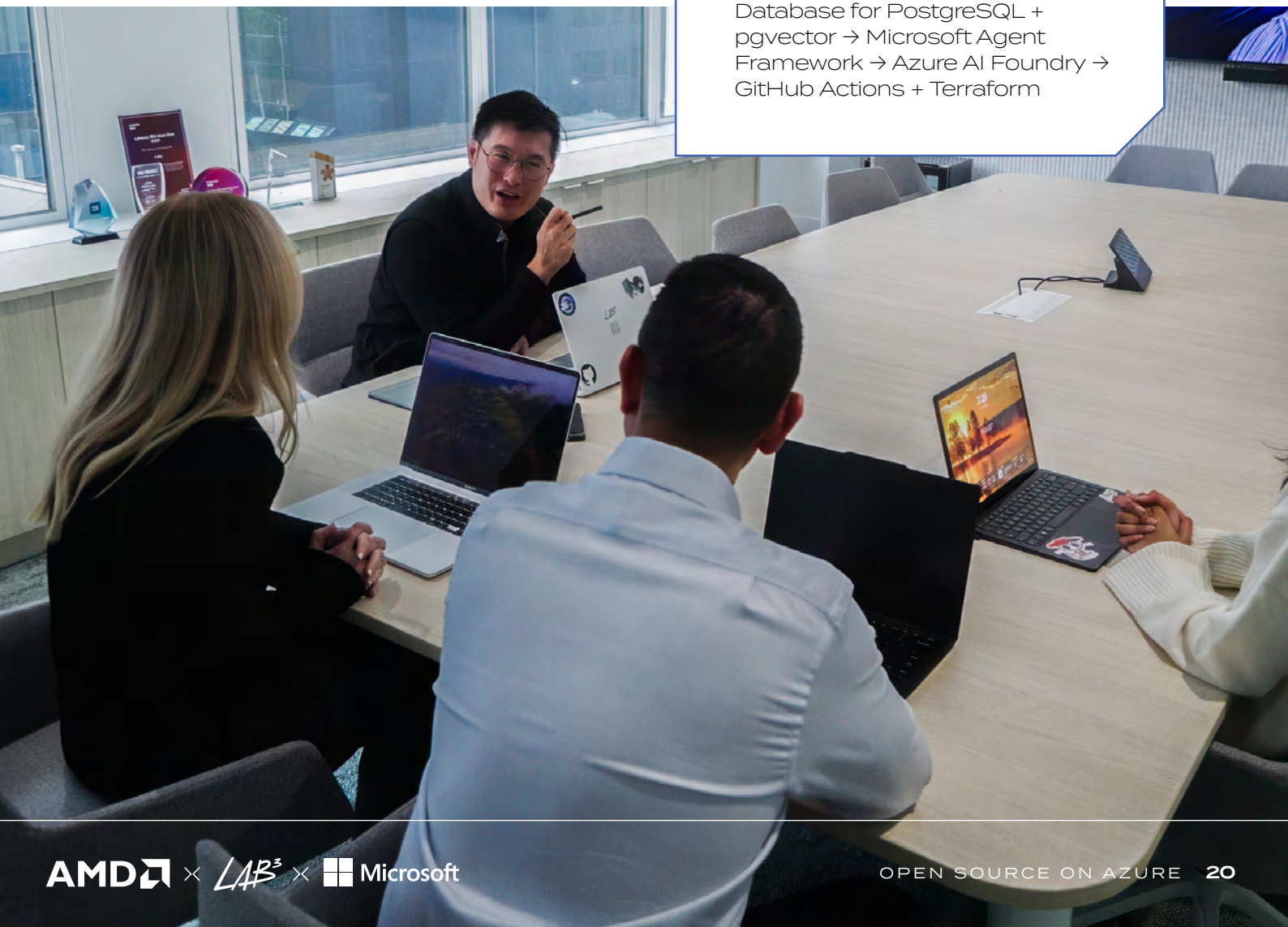
Why composability matters

The benefit of this pattern is composability. Each component can be swapped, extended, or migrated without rewriting the layers above and below it. An application built on this stack can run AI inference on AMD EPYC CPUs for cost-efficient workloads and burst to GPU-accelerated VMs for demanding tasks. It can use Azure OpenAI today and swap in open models via Foundry tomorrow. Natively deployable on Azure, with core open source technology that makes it also theoretically deployable anywhere (including hybrid and edge environments for specific use cases).

For enterprises building new AI-powered products or modernizing existing platforms, this pattern provides the foundation for long-term flexibility without sacrificing the operational benefits of a managed cloud.

KEY COMPONENTS

Full open source stack: Linux + AMD EPYC → AKS → Azure Database for PostgreSQL + pgvector → Microsoft Agent Framework → Azure AI Foundry → GitHub Actions + Terraform



PART 3

ENTERPRISE READINESS

Governance, Risk & Support

The executive objection to open source has always been the same:

“Is it safe? Is it supported? Can I bet my enterprise on it?”.

On Azure, the answer is unambiguously “yes”, and the reason is the integrated managed service model.

Introducing the managed service model delivered with Azure

When you deploy PostgreSQL on Azure, you’re not downloading an open source database and hoping for the best. You’re using Azure Database for PostgreSQL, which comes with:

- **99.99% SLA** with zone-redundant high availability
- **Automated backups** with point-in-time restore
- **Managed patching** for security updates
- **Integration with Microsoft Defender for Cloud** for threat detection
- **Entra ID authentication** with managed identity support

The database engine is open source. The operational wrapper is enterprise-grade.

The same model applies across the stack. AKS provides managed Kubernetes with Azure Policy for governance and Defender for Containers for security. Container Apps abstracts away even more operational complexity while maintaining the same security and compliance posture. Microsoft Agent Framework integrates with Azure AI Foundry for managed deployment with built-in governance and observability.

Compliance

Azure holds over 100 compliance certifications, including ISO 27001, SOC 2, HIPAA. These certifications apply to the managed services running open source components, meaning enterprises in regulated industries can adopt the open source stack without compromising their compliance posture.

Supply chain security

Azure gives clients a strategic advantage in mitigating global supply chain risks by securing continuous access to high-performance computing and the latest chipsets. As part of Microsoft's vast ecosystem, Azure benefits from deep, long-term partnerships with global industry leaders such as AMD to ensure consistent, early access to next-generation hardware and next-generation technologies. This integrated supply chain strength means clients can rely on Azure's global scale, data center diversity, and automation-driven provisioning to maintain resilience even in market volatility. Broadly, Microsoft Azure addresses supply chain security through multiple layers:

- Microsoft's contributions to the OpenSSF (Open Source Security Foundation)
- Integrated vulnerability scanning through Defender
- Managed patching model that keeps open source components current without requiring enterprise teams to track upstream releases.

Critical Activities
Enterprise Readiness

- **Map your compliance requirements against Azure's certification portfolio.** Identify which certifications are relevant to your industry and confirm that the managed open source services on Azure meet them.
- **Establish your AI governance framework before deploying agents.** The Agent Governance Toolkit provides the technical enforcement layer, but your organization needs the policy framework to drive it.
- **Leverage managed patching rather than self-managing open source components.** Azure's managed services handle security updates for PostgreSQL, Kubernetes, and other open source components, reducing your operational burden and security exposure.

PART 4

THE COST-PERFORMANCE ADVANTAGE

Enterprise AI strategies live and die on economics. The open source stack on Azure powered with AMD infrastructure delivers a cost-performance profile that proprietary alternatives struggle to match.

Licensing

The most obvious saving is licensing. Azure Database for PostgreSQL has no per-core licensing fees compared to common on-premises solutions.

For data-intensive AI workloads where you're storing and querying millions of embeddings, this difference is significant.

DID YOU KNOW?

In most enterprise cases, the licensing savings from Azure Database for PostgreSQL alone can fund a significant portion of AI development efforts.

Compute cost-performance

AMD EPYC processors extend the advantage at the compute layer. Azure's v7-series VMs powered by 5th Gen EPYC offer up to 35% better price-performance than the previous generation.

For AI inference workloads that don't require dedicated GPUs (most enterprise AI applications fall into this), AMD's high core density and memory bandwidth provide the throughput needed at a fraction of the cost of GPU-accelerated instances.

Scaling economics

AKS and Container Apps provide scaling economics that matter for AI workloads with variable demand. The Kubernetes autoscaler (including KEDA for event-driven scaling) ensures you're only paying for compute when workloads are active. Combine this with Azure Reserved Instances or Savings Plans on AMD-based VMs, and the unit economics of AI workloads improve significantly.

Avoiding lock-in

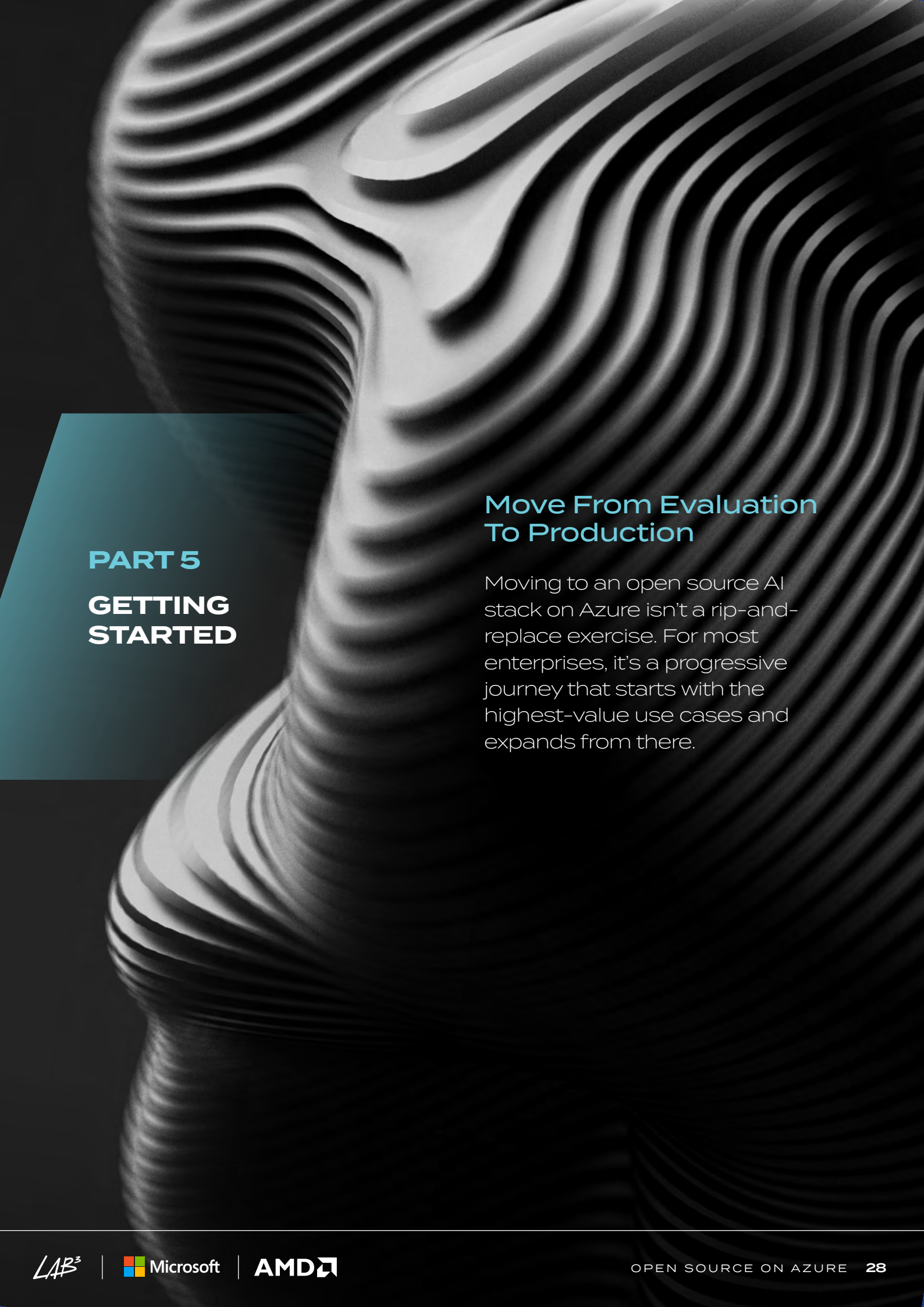
The broader economic argument is about avoiding lock-in. Every component in this stack has an open source foundation, which means enterprises retain optionality.

Your PostgreSQL skills transfer. Your Kubernetes configurations are portable. Your Agent Framework code isn't locked to a single vendor's runtime. This optionality has real economic value. It strengthens your negotiating position and reduces switching costs if your needs change.

The cost-performance story isn't about being cheap. It's about directing funds previously allocated to infrastructure and licensing towards the innovation that can differentiate your organization. Invest these savings back into the AI models, data, and use cases that will drive the most value for your business.

Critical Activities
Cost-Performance

- **Benchmark Azure Database for PostgreSQL against your current commercial database for AI workloads.** LAB[3] and Microsoft can provide a cost-free assessment of your environment to quantify the licensing savings and compare the vector search performance of pgvector against your current solution.
- **Evaluate AMD-based VM series for AI inference.** Run your inference workloads on AMD EPYC VMs and compare the cost-per-request against GPU-accelerated alternatives. Many inference workloads don't need GPUs.
- **Model your total cost of ownership.** Factor in ongoing maintenance and support costs. Consider how you quantify the economic value of portability and reduced vendor lock-in when comparing the open source stack against proprietary alternatives.



PART 5

GETTING STARTED

Move From Evaluation To Production

Moving to an open source AI stack on Azure isn't a rip-and-replace exercise. For most enterprises, it's a progressive journey that starts with the highest-value use cases and expands from there.

STEP 01

Assess your current digital estate

Start by understanding what you're already running. Many enterprises are surprised to find they're already using more open source on Azure than they think, including Linux VMs, AKS clusters, PostgreSQL instances. Map what's in place and identify where proprietary components are adding cost or complexity without adding value.

DID YOU KNOW?

LAB^3 leverages Microsoft's globally endorsed assessment tool, Dr Migrate, to provide a full mapping of your digital estate. Reporting identifies dependencies, suggests migration treatments, and provides detailed cost modelling and business case ready outputs.

Get in touch to organize your no-cost 90 day Dr Migrate license and get total visibility of your current state, as well as real-world evidence and actionable intelligence within days.

STEP 02

Identify your first AI use case

The RAG pattern is the natural starting point for most organizations. It's well-understood, it delivers immediate value (better search, better knowledge access, better customer support), and it maps cleanly to the PostgreSQL + pgvector + Azure OpenAI stack. Pick a use case with clear business value and a manageable data scope.

DID YOU KNOW?

LAB^3 offer no-cost AI Business Envisioning Workshops and reports proven to help identify high-value use cases, generate executive alignment and stakeholder buy-in, and drive momentum for enterprise scale AI adoption.

STEP 03

Stand up the foundation

Deploy Azure Database for PostgreSQL with pgvector enabled. Set up a Container Apps or AKS environment on AMD-based VMs. Configure Entra ID for authentication and Defender for security. Use Terraform or Bicep to codify the infrastructure so it's repeatable. This foundation serves every subsequent AI workload.

STEP 04

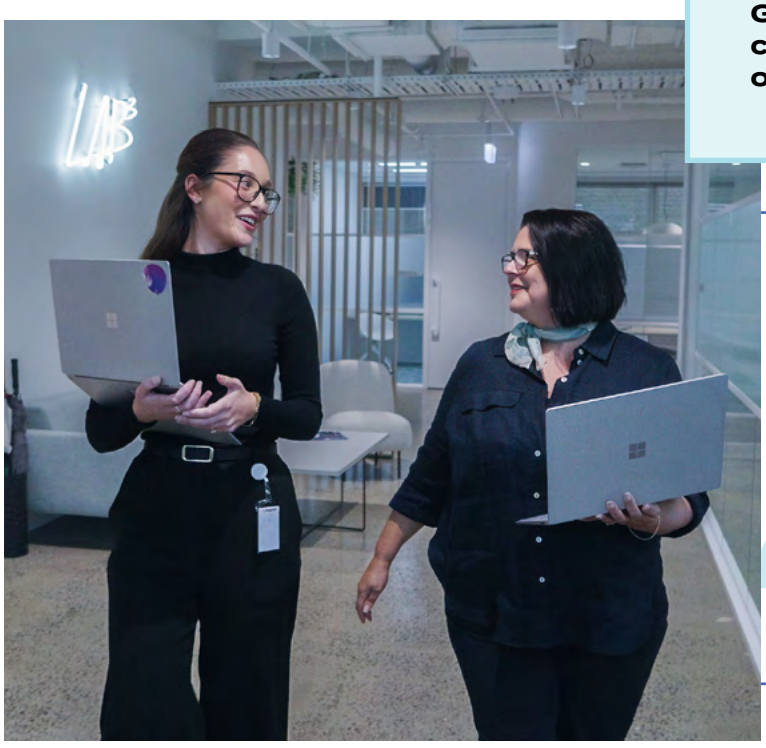
Build and validate

Implement your first RAG application on the open source stack. Validate the performance, security, and operational characteristics against your enterprise requirements. Use this as the proof point for broader adoption.

STEP 05

Scale with governance

As you move beyond the first use case into multi-agent patterns and production AI applications, layer in the governance tooling: Agent Governance Toolkit for policy enforcement, OpenTelemetry for observability, Azure Policy for infrastructure compliance. This is where the managed service model pays for itself as Azure handles the operational heavy-lifting while your teams focus on the AI applications and outcomes.



STEP 06

Engage a partner who has built on this stack

The open source AI stack on Azure is powerful, but it benefits from practitioners who've deployed it in production. Work with a specialized partner who understand the nuances of how to optimize PostgreSQL for vector workloads, how to structure multi-agent systems for enterprise governance, and right-sizing AMD compute for AI inference. The technology is mature. The implementation expertise is what accelerates time to value.

FIND OUT MORE

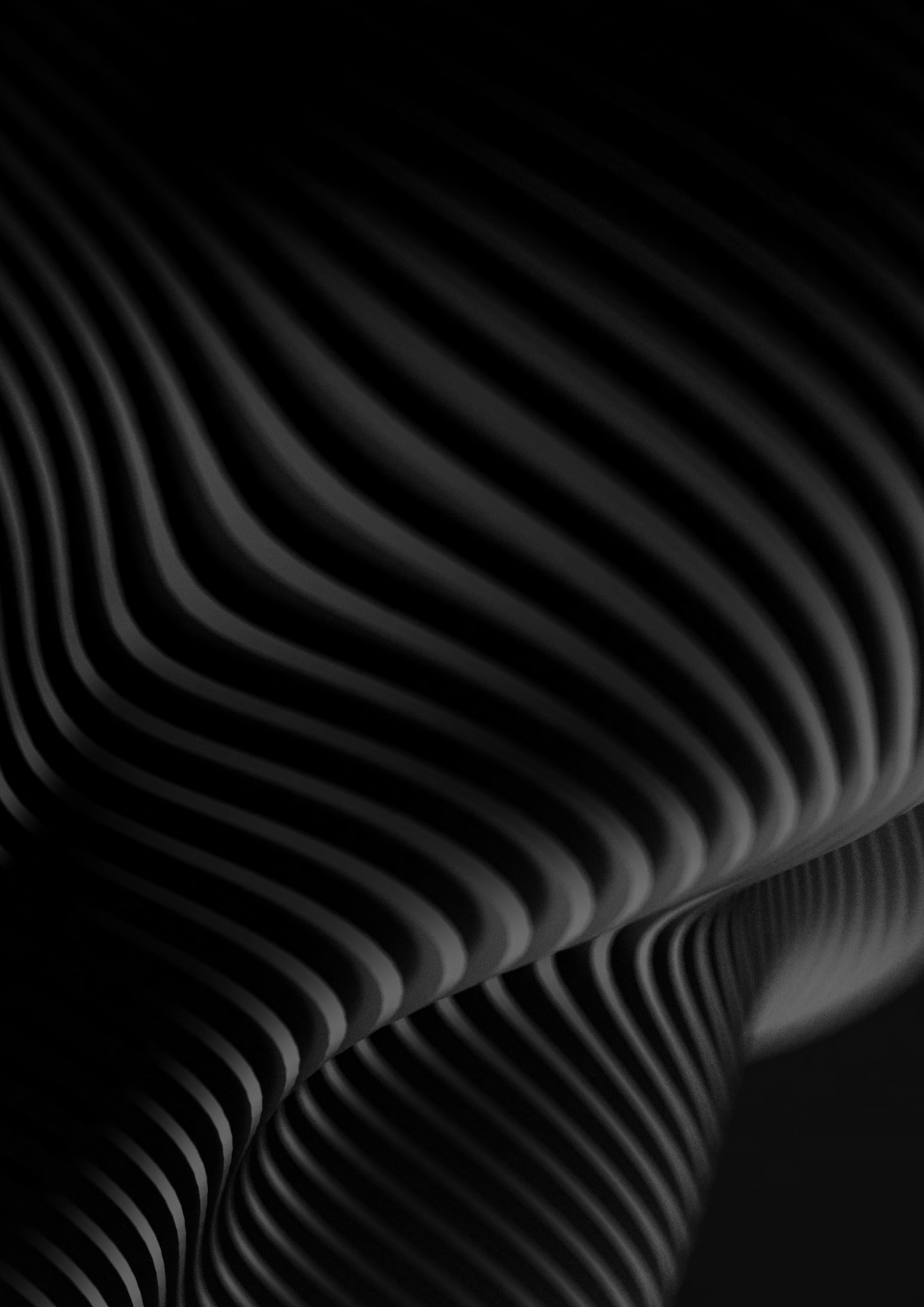
LAB³ has extensive experience deploying the open source AI stack on Azure for enterprises. Our Landing Zone Accelerator provides the secure, automated foundation for AI workloads on Azure, while our AI delivery practice helps organizations move from use case identification to production deployment.

Get in touch to discuss how we can accelerate your AI journey on the open source Azure stack.

CONTACT

JEANNE BROOKS
Vice President – US Client Engagement, LAB³
E jeanne.brooks@lab3.com

GET STARTED TODAY



LAB3.COM

ABOUT LAB³

LAB³ is an Azure-focused technology consultancy helping enterprises migrate, modernize, and accelerate AI adoption at scale. As one of only five Microsoft Global AI Discovery Partners, LAB³'s unique accelerators leverage automation and AI to deliver speed, certainty, and confidence in business outcomes for clients with highly regulated or complex technology environments.

The LAB³ catalogue includes an enterprise Landing Zone accelerator and an AI deployment platform, both built on the open source Azure stack described in this Guide. We don't just recommend this architecture. We build on it every day.

Operating across Australia, New Zealand, and the United States, LAB³ is a multiple Microsoft Partner of the Year award winner, with six Solutions Partner designations and nine Specializations. LAB³ is also a hyper-specialized HashiCorp partner and a verified GitHub partner.

If you're evaluating your AI platform strategy or looking to accelerate your first AI workloads on Azure, connect with our experts.

AMD.COM

ABOUT AMD

At AMD, we empower the future of AI with high-performance and adaptive computing solutions. As the leader in end-to-end computing, we accelerate AI breakthroughs through a broad ecosystem of technology partnerships and cutting-edge innovation.

Our advanced technologies power many of the workloads you run on Microsoft Azure, including general purpose, high performance, confidential compute, AI acceleration, visualization, and storage intensive. When you're building the next generation of AI models or deploying scalable solutions in Microsoft Azure, AMD is your trusted partner for driving transformation and delivering results.

This guide was prepared by LAB³ in partnership with AMD and Microsoft. The views and recommendations reflect LAB³'s experience delivering enterprise AI solutions on the Microsoft Azure platform.