White Paper

Azure AI Foundry vs Mosaic AI

Prepared by:

CloudFronts Technologies Private Limited

Prepared for:

CloudFronts Technologies Private Limited

CONTENTS

1. EXECUTIVE SUMMARY

This white paper provides a comparative analysis of **Azure AI Foundry** and **Mosaic AI**, two leading platforms for building and operationalizing AI-powered business solutions. Both platforms enable organizations to accelerate AI adoption, but they differ in focus, flexibility, and enterprise readiness.

2. INTRODUCTION

Azure AI Foundry is Microsoft's all-in-one platform for creating, running, and managing AI solutions. It is built with strong security and compliance, and it works smoothly with other Microsoft tools. This makes it a good choice for large businesses that need AI solutions that are safe, reliable, and ready to scale. The platform lets you do everything in one place—train models, fine-tune them, deploy, monitor, and manage them—with enterprise-level stability.

Mosaic AI, part of Databricks' Lakehouse platform, helps build AI using clean and well-managed data. It brings together data engineering, analytics, and AI model development in one place. It supports open-source tools, allows a lot of customization, and includes AI-assisted features. Mosaic AI is best for organizations that want to bring all their data together and experiment with AI to create business value.

For decision-makers, the choice depends on priorities:

- If a company wants strong control, compliance, and alignment with the Microsoft ecosystem, **Azure AI Foundry** is a better fit.
- If the focus is on data-driven AI, experimentation, and combining analytics with AI in one place, **Mosaic AI** works well.

In the end, both platforms can handle enterprise-scale AI, but they have different strengths. They are more **complementary than direct substitutes**.

3. PLATFORM OVERVIEW

Feature	Azure AI Foundry	Mosaic AI
Focus	Developer and Data Scientist	Data Engineers, Analysts, and Data Scientists
Core Use Case	Create and manage your own AI agent	Build, experiment, and deploy data-driven AI models with analytics + AI workflows
Interface	Code-first (SDKs, REST APIs, Notebooks)	No-code/low-code UI + Notebooks + APIs
Data Access	Azure Blob, Data Lake, vector DBs	Native integration with Databricks Lakehouse, Delta Lake, Unity Catalog, vector DBs
MCP Server	Only custom MCP servers are supported currently. The built-in option is not yet available, and the setup process is complex.	Native MCP support with Databricks ecosystem, simpler setup
Models	90 models available	Access to open-source + foundation models (MPT, Llama, Mixtral, etc.) + partner models
Model Customization	Full model fine-tuning, prompt engineering, RAG	Fine-tuning, instruction tuning, retrieval-augmented generation (RAG), model orchestration
Publish to Channels	Complex (Azure Bot SDK + Bot Framework + App Service to host API)	Direct integration with Databricks workflows, APIs, dashboards, and third-party apps
Agent update	Once an agent is installed in Microsoft Teams, any changes to the agent are reflected in real time.	Updates deployed seamlessly via Databricks workflows; versioning and rollback supported

4. KEY CAPABILITIES

Azure AI Foundry	Mosaic AI
• Prompt flow orchestration • RAG (Retrieval Augmented Generation) • Model choice • Vector search integration • CICD for deployment pipelines • Azure ML & responsible AI integration (deploy own model)	• **Data + AI unification** (native to Databricks Lakehouse with Delta Lake & Unity Catalog) • **Retrieval-Augmented Generation (RAG)** with Lakehouse data integration • **Fine-tuning & model customization** (instruction tuning, adapters, parameter-efficient tuning) • **Multi-model orchestration** (choose and route between multiple models: open-source, proprietary, or partner) • **End-to-end ML pipeline** (integrated with Databricks workflows for training, deployment, and monitoring) • **Secure governance** (data lineage, access control, audit with Unity Catalog • **Real-time deployment** (deploy models as APIs)

5. KEY COMPONENTS

Azure AI Foundry	Mosaic AI

• Workspace & agent orchestration
• 90+ models
• Models
  ○ Open AI – Pay-as-you-go
  ○ Other – Self hosting & Computation
• Security via Azure identity
	• Mosaic AI Agent Framework
	• Mosaic AI Model Serving
	• Mosaic AI Fine-Tuning
	• Mosaic AI Vector Search
	• Mosaic AI RAG Studio
	• Mosaic AI Evaluation & Monitoring
	• **Unity Catalog Integration** (Foundation Layer)

6. COST & LICENSE

Cost Area	Mosaic AI (Databricks)	Azure AI Foundry (Microsoft)
Vector Search	~$605–$760/month for 5M vectors (continuous use, ~720 hrs)	Not directly offered (done via external vector DBs)
Model Serving	Example: Llama 3.1 405B 142.857 DBU per million tokens ($90–$120 depending on DBU rate)	Token pricing: o3 ($2 input / $8 output per million tokens); o3-pro ($20 / $80 per million)
Fine-Tuning	**Llama 3.3 (70B model):** • 10 million words cost about **$146** • 500 million words cost about **$7,150** **Llama 3.1 (8B model):** • 10 million words cost about **$65** • 500 million words cost about **$2,860**	Not clearly published; typically enterprise pricing / case-by-case
Evaluation	Small batch: ~$0.88 (10 Q × 5 judges); Full pipeline: ~$1,000–$1,800/month	Not specified
Managed Compute	Built into Databricks DBU usage (autoscaling)	Cohere models: $2.94–$17+/hr depending on model & GPU
Total Agent Cost	End-to-end AI Agent (RAG + serving + eval + orchestration): **~$5,000–$7,000+/month**	Highly variable; token-based costs + GPU compute; reports of idle hosting fees (~$95/day/model)

7. USE CASES

For the past few months, we have been working on multiple AI projects. In this document, we will share our experience from our use cases where we first developed the agent in Microsoft Copilot Studio and then replicated it in Azure AI Foundry and then in Mosaic AI. This helped us understand the differences between the two platforms through real-world use cases.

Azure AI Foundry	Mosaic AI
Agents are intelligent and allow users to interact and modify responses.	Agents are intelligent and allow users to interact and modify responses.
Users can exit a flow at any time by simply providing a new intent.	Users can exit a flow at any time by simply providing a new intent.
Currently, only one AI search is allowed per agent.	AI search directly not available, but we can integrate it using Search APIs example: Google, Bing
Infrastructure setup is required (e.g., AI Agent SDK, App Service, Bot Framework).	Complex infrastructure setup and dedicate cost associate with same.
A custom MCP server must be registered as a tool.	In-built MCP server available, also we can integrate the third party / custom MCP server.
Response time is fast.	Response time is slow as it send all response together.

8. COMPARISON BETWEEN MOSAIC AI AND AI FOUNDRY

8.1. DEVELOPMENT

- In Databricks, you need to manually code the agent using the MLflow library. This allows more customization and API integration, but it takes more time and increases the chances of errors.
- In contrast, AI Foundry focuses more on the concept and purpose of agent development. It's a low-code platform, making it faster and easier to build agents.

8.2. MODELS

- Databricks offers only **10 base models**, compared to **around 90 in AI Foundry**. It also doesn't provide access to **Azure OpenAI models**. In addition, many of its models **don't support tool usage**, which limits extra functionality.

Attribute	Databricks Models	Azure OpenAI Foundry Models
Access Type	Model-specific; pay-per-token via Databricks	Deployed Azure models (GPT-3.5, GPT-4, etc.)
Origin	OSS (GPT-OSS), third-party (Claude, Llama, Gemma)	OpenAI-developed (GPT-3.5, GPT-4), Azure offerings
Context Window	Moderate, not specified in list	8K–128K tokens (most), up to 1M for GPT-4.1 Microsoft LearnThe Verge
Modalities	Likely text-only	Text, code, vision, audio in some models
Special Enhancements	Varies: Llama, Claude, Gemma architectures	Turbo efficiency, reasoning, tool calling, multimodal
Availability	Through Databricks platform	Deployed via Azure AI Foundry

8.3. KNOWLEDGE SOURCE

- In **Azure AI Foundry**, you can add only **one knowledge source of each type** (for example, a single Azure AI Search resource, file, or SharePoint). If you add a new one, the previous source is automatically removed.
- In **Mosaic AI**, there is **no such limitation**. It supports data cleaning and preparation through the **Medallion Architecture**. However, this data can only be accessed using **SQL queries**—attempts to use Spark or PySQL within agents will fail.

8.4. MEMORY ISSUE

In **Mosaic AI**, the model server may crash or reach its request limit, usually because it runs out of memory or cannot handle the current request load.

In contrast, **AI Foundry** offers built-in monitoring, auto scaling, and self-healing capabilities. This ensures that workloads adjust automatically to demand without manual effort, reducing downtime and freeing up the team to focus on innovation instead of maintenance.

8.5. AGENT UPDATES

AI Foundry

- Unlimited modifications with **real-time autosave**, so no version control issues.
- Provides **version history and rollback**, making it easy to restore older changes.
- Safe for **iterative development**, experimentation, and frequent updates by multiple team members.
- Supports **collaboration** across teams without risk of overwriting or losing functionality.

Mosaic AI

- **Hard limit** of ~15 modifications per notebook, which restricts iterative development.
- If an older version is deleted, the agent can become **unstable or unusable**.
- No strong built-in **rollback/version history**, so restoring lost work is difficult.
- Risk of **losing productivity** if frequent edits are needed (e.g., tuning prompts or logic).
- Requires extra caution in managing versions, which adds to the **operational overhead**.

9. RECOMMENDATIONS

1. Start with Your Team's Skillset and Use Case

- If your organization has business users or functional teams looking to quickly build productivity tools, Copilot Studio is ideal.
- If you have a technical team (developers, data scientists) aiming to build custom, scalable AI apps, choose Azure AI Foundry.
- If your team works heavily with data engineering and analytics on Databricks, and wants to leverage data pipelines, medallion architecture, and SQL-based access, Mosaic AI is the better fit.

2. Assess AI Maturity and Long-Term Goals

- For organizations new to AI, Copilot Studio is an easy way to start using AI with simple, low-code tools.
- For companies that want to build large-scale AI solutions, Azure AI Foundry gives more control, flexibility, and room for innovation.
- For data-driven organizations already invested in Databricks Lakehouse, Mosaic AI allows seamless integration with existing data, governance, and ML workloads.

3. Use All Platforms Strategically

- You can use Copilot Studio for simple internal tasks and quick AI tools.
- Use Azure AI Foundry for complex, customer-facing, or industry-specific AI solutions.
- Use Mosaic AI when you need direct access to enterprise data, SQL-based queries, and AI agents tied closely to your data lakehouse.

4. Focus on Governance and Scalability Early

- As adoption grows, ensure your AI solutions follow responsible AI guidelines.
- Standardize model operations (MLOps) using Azure AI Foundry capabilities.
- With Mosaic AI, focus on data governance and lineage via Unity Catalog to maintain security and trust in data-driven AI.

11. CONCLUSION

For companies that want to build powerful, flexible, and large-scale AI solutions, Azure AI Foundry is the best choice. Unlike low-code tools that are made for small or basic tasks, AI Foundry gives your technical team the tools to build, test, and launch advanced AI applications with full control over the models, data, and workflows.

It's made for businesses that want to do more than just try out AI—they want to make it a key part of how they work, from customer service to industry-specific tools to smart automation.

By choosing Azure AI Foundry, you're not just testing AI—you're building a strong, future-ready AI system for your entire organization.

On the other hand, Mosaic AI is ideal for organizations that are data-centric and already invested in Databricks. It allows teams to directly connect AI with enterprise data using the Lakehouse and Medallion Architecture, ensuring data is cleaned, governed, and ready for AI-driven insights. Mosaic AI is best suited for scenarios where data engineering, analytics, and AI need to work hand-in-hand—for example, building agents that can query enterprise data, generate insights, and integrate with existing data pipelines.

By choosing Mosaic AI, you're embedding AI directly into your data ecosystem, making it easier to scale insights, maintain governance through Unity Catalog, and create AI systems that are tightly connected to your organization's data foundation.

Let's Connect

We hope you found this blog useful. If you would like to discuss anything further, please reach out to us at transform@cloudfronts.com