

This Service Description describes the scope of services (including associated Customer requirements) for the Advisory Service identified above (the “**Service**”), and applies to the Service under Customer’s accepted Service Order.

### Service Overview

The GenAI Advisory Service offering is designed to empower organizations to harness the capabilities of generative AI, translating this cutting-edge technology into production-ready, impactful business applications. Over the course of the engagement, Databricks’ Data Scientists will assist in guiding Customer through the key architectural aspects and best practices of implementing generative AI solutions on Databricks, tailored to Customer’s business use case.

### Objective

The core objective of this Advisory Service is to help accelerate your organization’s ability to apply generative AI to solve a specific business use case using Databricks. Databricks will guide Customer in translating your business problem into a practical solution, while providing best practices. Databricks will advise on the following:

- Architecting a scalable, efficient, and secure generative AI solution on the Databricks platform
- Applying existing generative AI models (OSS or third-party APIs) to address your business use case
- Implementing an evaluation methodology to measure the performance of your generative AI solution
- Leveraging Databricks features to productionize your generative AI solution

The final deliverable will be a consolidated document of the advice and recommendations provided during the course of the engagement.

### Description of Services

Databricks will provide Services from the Technical Focus Areas and Representative Activities described below in advising Customer on implementing a generative AI solution on Databricks. Specific activities performed will vary, depending on Customer-specific objectives.

Technical Focus Areas	Representative Activities
<b>Use Case Requirements</b>	<ul style="list-style-type: none"> <li>• In-depth discussion to understand the business problem and identify how generative AI can provide a solution.</li> <li>• Review available data sets, assessing data quality, volume, and suitability for generative AI applications.</li> <li>• Assess the current architecture and outline the preparation requirements for integrating a generative AI solution within the current infrastructure.</li> <li>• Establish detailed use case requirements, such as data and compliance needs, latency/throughput specifications, scalability concerns, and end user interaction expectations</li> </ul>
<b>Architecture &amp; Design Decisions</b>	<ul style="list-style-type: none"> <li>• Discuss architectural patterns for generative AI workflows on Databricks, considering factors such as scalability, flexibility, and security</li> <li>• Outline how such workflows would fit within the current infrastructure, to facilitate developing, and deploying a generative AI solution to production</li> <li>• Review and recommend data storage and governance options on Databricks to optimize for model training and inference performance</li> <li>• Discuss components of the solution architecture where Databricks features such as Unity Catalog, MLflow, Model Serving, MosaicML, and Lakehouse Monitoring could be leveraged</li> <li>• Discuss plan for the integration of the solution with existing CI/CD pipelines and development workflow to streamline solution deployment</li> </ul>
<b>Development &amp; Evaluation</b>	<ul style="list-style-type: none"> <li>• Explore adapting generative AI models to address specific business challenges, e.g. when to use prompt engineering vs retrieval-augmented generation (RAG) vs fine-tuning</li> <li>• Examine Databricks functionality in the development workflow such as MLflow for experiment tracking and model management, Unity Catalog for data/model discovery and governance, and Model Serving for streamlined deployment</li> <li>• Deep dive on model training best practices for generative AI models on Databricks</li> <li>• Discuss selecting and tuning hyperparameters, in addition to balancing model performance with compute cost</li> <li>• Discuss techniques for reducing hallucinations</li> <li>• Discuss evaluation methodology, leveraging MLflow</li> </ul>
<b>Productionization and Next Steps</b>	<ul style="list-style-type: none"> <li>• Explore Databricks Model Serving &amp; Lakehouse Monitoring</li> <li>• Discuss the path to production and how to automate deployment</li> <li>• Discuss maintaining model performance over time, considering factors such as retraining schedules and data refresh cadence</li> <li>• Advise on next steps, including fine-tuning models or developing Customer-specific foundation model</li> </ul>

## Prerequisites

Throughout the engagement, Customer will assure that the following requirements are met, to enable the Services:

- Clear problem definition with a well defined business objective or expected goal/outcome
- Well understood data sources, stored in cloud storage and accessible from Databricks
- Engagement from Customer architecture, data science teams, and product/business owner(s)

## Out of Scope

- Defining the business use case, business logic, or end user experience
- Hands-on implementation of the solution, including but not limited to data preparation, data cleaning, solving data quality issues, model training/fine-tuning
- Setting up external services such as third-party API providers
- Follow-up consulting after the conclusion of this engagement. Please contact your Engagement Manager to schedule additional time.
- Work exceeding the allocation of Days and Services included in this engagement (see [Resources and Schedule](#))

## Resources and Schedule

Services consist of up to **5 consecutive Days of Data Scientist time**, and up to **½ Day of Project Management time**, applied against the Representative Activities in the Description of Services above. Databricks will work with you to mutually agree to a project schedule as part of the Project Management phase.

Resourcing assignments require a minimum 4-weeks advance request (while Databricks makes reasonable efforts to accommodate scheduling requests, personnel availability is subject to Databricks resourcing and discretion). Accordingly, Databricks recommends Customer coordinate with Databricks Services at least a month before placing its Service Order.

## Additional Definitions and Terms

- **“Agreement”** means your agreement with Databricks, providing general terms for our Services.
- **“Services Order”** may be any of these mutually-accepted formats placed under your Agreement: an Order, Success Credit redemption request, written statement of work, or similar document.
- **“we”, “us” or “our” or “Databricks”** means Databricks, Inc. or its Affiliates.
- **“you” or “your”** means the Customer organization that placed the Services Order.