**The 5 Main Components (Top Row):**

**1. Data Integration (Green cross icon)**

* **What it does:** Moves data from sources (databases, APIs, files) into Fabric
* **Replaces:** Azure Data Factory
* **AI Enhancement:** Copilot can create data pipelines by just describing what you need: "Get customer data from Salesforce every night"

**2. Data Engineering (Blue twisted ribbon)**

* **What it does:** Transforms and processes raw data at scale using Apache Spark
* **Replaces:** Azure Databricks / Synapse Spark
* **AI Enhancement:**
  + Auto-generates transformation code
  + Suggests optimizations: "Your join is slow, try partitioning by date"
  + Intelligent debugging

**3. Data Warehouse (Blue database icon)**

* **What it does:** Stores structured data for fast SQL queries
* **Replaces:** Azure Synapse dedicated SQL pools
* **AI Enhancement:**
  + Auto-indexing based on query patterns
  + Query optimization suggestions
  + Predictive caching (pre-loads data you'll likely query)

**4. Data Science (Dark blue diamond shape)**

* **What it does:** Build, train, and deploy machine learning models
* **Replaces:** Azure Machine Learning
* **AI Enhancement:**
  + **AutoML:** Automatically tries different algorithms to find the best model
  + **Feature engineering suggestions:** "Try adding day-of-week as a feature"
  + **Model monitoring:** Detects when model accuracy drops

**5. Power BI (Yellow bar chart)**

* **What it does:** Creates dashboards, reports, and visualizations
* **AI Enhancement:**
  + **Copilot:** "Show sales by region" → creates visuals automatically
  + **AI Insights:** Finds trends and anomalies automatically
  + **Natural Language Q&A:** Business users ask questions in plain English

**OneLake - Unified Data Storage (Blue lake icon)**

* **What it is:** Single storage location for ALL your data (like OneDrive but for company data)
* **Why revolutionary:**
  + **Before:** Data scattered across multiple storage accounts, hard to find, duplicate copies everywhere
  + **With OneLake:** One place, no data movement needed, everyone accesses the same data
* **AI Enhancement:**
  + Intelligent data organization
  + Auto-detects duplicate data
  + Smart access recommendations

**Purview - Data Governance (Shield icon)**

* **What it does:** Data catalog, security, compliance, and privacy controls
* **AI Enhancement:**
  + **Auto-discovery:** Scans all data and finds sensitive info (SSNs, credit cards, emails)
  + **Smart classification:** "This looks like PII, should we mask it?"
  + **Anomaly detection:** Alerts when someone accesses unusual data

**1. Speech/Audio (Speaker icon - left)**

* **Text-to-Speech:** Convert written text to spoken words
* **Speech-to-Text:** Transcribe audio recordings to text
* **Example use:** Transcribe customer service calls automatically

**2. Vision/Computer Vision (Eye/camera icon)**

* **Image recognition:** Identify objects in photos
* **Face detection:** Find faces in images
* **OCR:** Read text from images/documents
* **Example use:** Auto-tag product photos, extract text from invoices

**3. Form Recognition (Document icon)**

* **Extract data from forms:** Reads invoices, receipts, IDs automatically
* **Understands structure:** Knows what "Total Amount" means
* **Example use:** Process thousands of invoices without manual data entry

**4. Language/Translation (Japanese characters icon)**

* **Translation:** Convert text between 100+ languages
* **Language detection:** Identify what language text is written in
* **Example use:** Translate customer reviews from different countries

**5. Text Analytics/Search (Magnifying glass with nodes icon)**

* **Sentiment analysis:** Is this review positive or negative?
* **Key phrase extraction:** What are the main topics?
* **Entity recognition:** Find names, places, dates in text
* **Example use:** Analyze customer feedback sentiment automatically

**6. Anomaly Detection (Star/sparkle icon)**

* **Spot unusual patterns:** Find weird data points
* **Time series analysis:** Detect spikes or drops
* **Example use:** Alert when website traffic suddenly drops, detect fraudulent transactions

**LightGBM (Light Gradient Boosting Machine)**

* **What it is:** A super-fast machine learning algorithm for making predictions
* **What it's good for:**
  + Predicting customer churn (will customer leave?)
  + Fraud detection (is this transaction fake?)
  + Sales forecasting (how much will we sell next month?)
  + Credit risk scoring

**Why it's popular:**

* **Fast:** Can handle millions of rows quickly
* **Accurate:** Often wins data science competitions
* **Efficient:** Uses less memory than competitors

MLflow is an **open-source platform** for managing the complete machine learning lifecycle. Think of it as a "lab notebook + version control + deployment system" for ML models.

**Code version:** Which code created this model?

**Data version:** Which dataset was used?

**Parameters:** What settings were used?

**Metrics:** How accurate was the model?

**Artifacts:** The actual model files, charts, reports

**Environment:** Which libraries/versions were installed?

Managing a model from **birth to retirement**:

1. **Development:** Building and training the model
2. **Staging:** Testing in a safe environment
3. **Production:** Serving real predictions to users
4. **Monitoring:** Watching performance in the wild
5. **Retirement:** Replacing with a better model

**WITHOUT MLflow Model Management (Chaos):**

**Problems:**

1. **Version confusion:** "Which model is in production? v2.3 or v2.3\_final?"
2. **No rollback:** New model breaks → panic, can't quickly switch back to old model
3. **Lost models:** "Where did we save the model from Q2?"
4. **Deployment hell:** Different code to deploy locally vs. cloud vs. edge devices

**Feature:** Regression Pipeline  
**Library:** synapse.ml.featurize.text  
**Usage:** Loads the text features from parquet files

**What it means:**

* **Text featurization** = Converting text into numbers that machines can understand
* **Parquet files** = Efficient file format for big data (like compressed CSV)
* This prepares text data for analysis

**Row 2: Running the LightGBMRegressor**

**Feature:** Running the LightGBMRegressor  
**Library:** synapse.ml.lightgbm import LightGBMRegressor  
**Usage:** Runs the LightGBM regressor model

**What it means:**

* Uses the **LightGBM algorithm** (we discussed earlier - fast, accurate ML algorithm)
* **Regressor** = Predicts a continuous score (e.g., sentiment score from -1.0 to +1.0)
* Trains on your text data to learn patterns

**Row 3: Pre-built Solution**

**Feature:** Pre-built solution  
**Library:** synapse.ml.services.language import AnalyzeText(sentiment)  
**Usage:** Prebuilt solution to analyse sentiment

**What it means:**

* **No training needed!** Uses Azure's pre-trained AI service
* Just call the function, get instant results
* Powered by Azure Cognitive Services (the AI services we discussed earlier)

Fabric agents:

**1. Semantic Layer Integration**

**What it means:**

* Creates a "business-friendly" layer on top of technical databases
* Users ask questions in plain English, agent translates to SQL automatically

**2. Natural Language Query Support**

**What it means:**

* Business users can query databases without knowing SQL
* Agent converts questions → SQL → executes → returns results

**3. Multi-Source Data Execution**

**What it means:**

* Agent can query multiple databases at once (Lakehouse, Warehouse, KQL databases)
* Combines results automatically

**1. Enable Copilot in Tenant Settings**

**What it means:**

* Administrator must turn on Copilot in the Microsoft Fabric admin portal
* This is a one-time setup for the entire organization

**2. Paid Capacity Requirement**

**Important limitation:**

* Requires **F64 capacity or Power BI P1** license
* Free/trial versions don't support Copilot
* This is a premium feature

**3. Assign Capacity to Workspace**

* Create capacity in Azure portal
* Link it to your Fabric workspace
* Enables Copilot functionality

**4. Workspace and Lakehouse Link**

* Create a workspace
* Link it to a Lakehouse
* Enables Copilot's advanced features and integrations

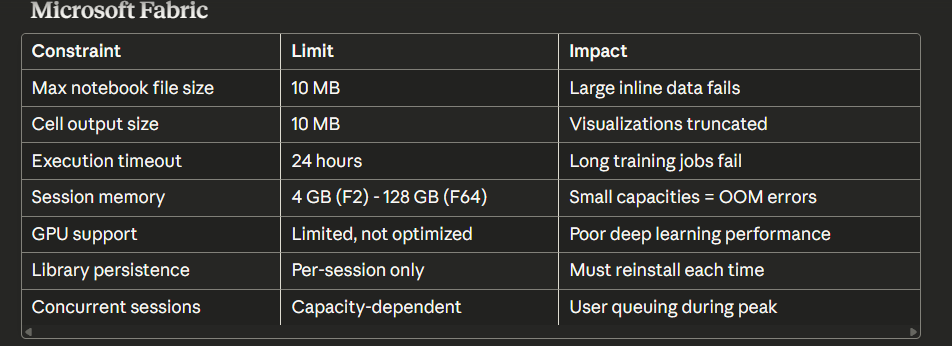
COMPARISON OF AZURE FOUNDRY V/S Fabric in mlops & model life cycle

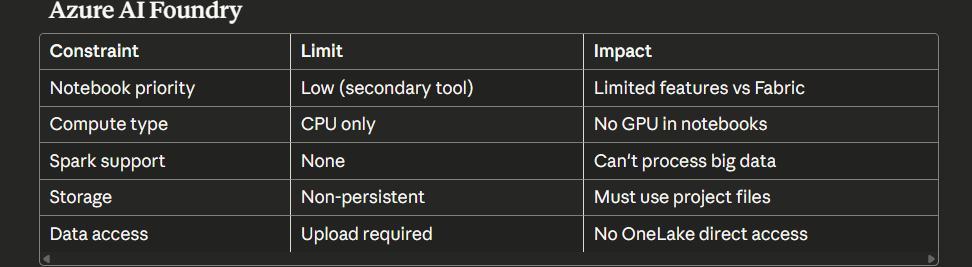
**The Bottom Line**

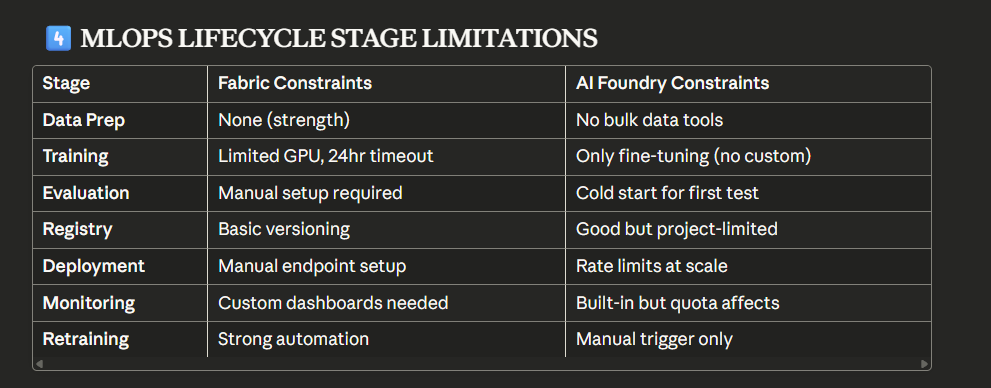
**Microsoft Fabric** is like a comprehensive data analytics workshop - great for businesses that need to analyze lots of data, create reports, and do some machine learning, but with less focus on cutting-edge AI.

**Azure AI Foundry** is like a specialized AI laboratory - built specifically for creating, deploying, and managing modern AI applications with stricter controls, better real-time capabilities, and more advanced AI features, but requires more technical knowledge.

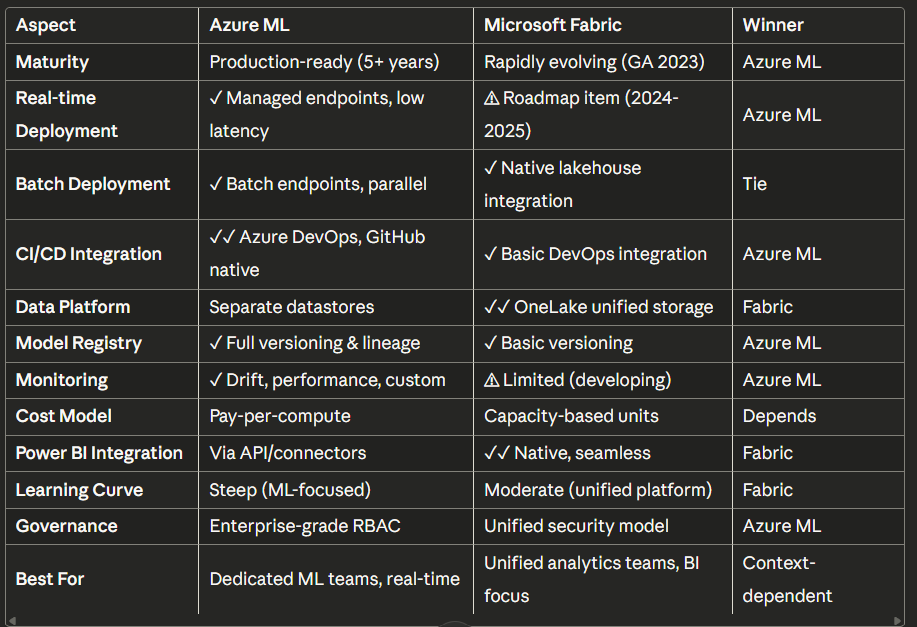
Choose Microsoft Fabric if you're focused on business intelligence and data analytics. Choose Azure AI Foundry if you're building AI applications and need modern AI capabilities.



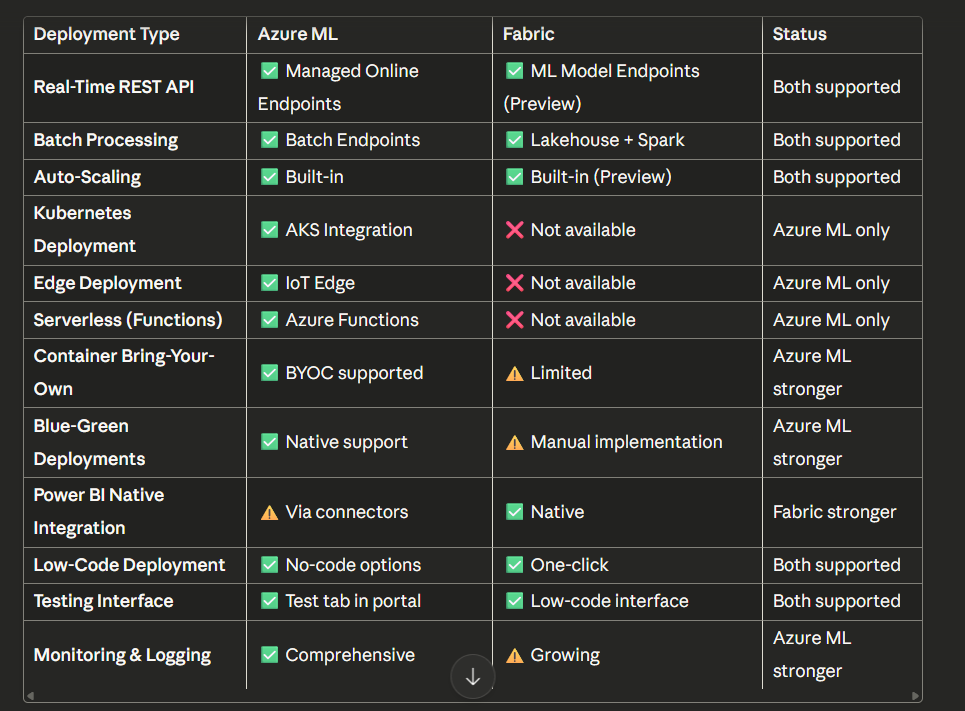


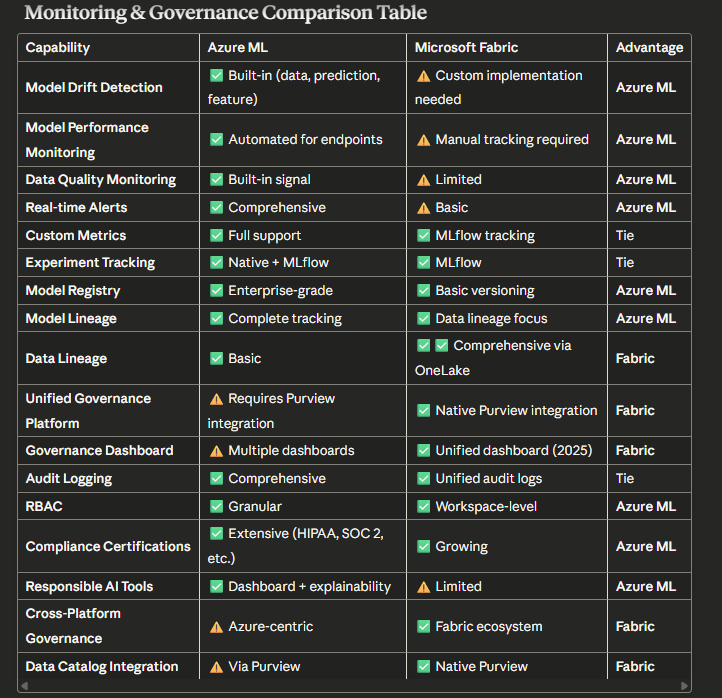


MLOPS LIFECYCLE COMPARISON:



DEPLOMENT OPTIONS:





**What is Semantic Link?**

Semantic link facilitates seamless collaboration between data scientists and business analysts by eliminating the need to reimplement business logic embedded in Power BI measures [Comparing Microsoft Fabric and Azure Machine Learning: Which is Right for Your Needs?](https://www.linkedin.com/pulse/comparing-microsoft-fabric-azure-machine-learning-which-kim-berg). Semantic link bridges the gap between the semantic models and the Synapse Data Science in Microsoft Fabric experience, providing a way for business analysts and data scientists to collaborate seamlessly and reduce data mismatch

**Key Capabilities:**

**1. Bidirectional Data Flow** The SemPy Python API can retrieve data and metadata from semantic models located in a Microsoft Fabric workspace, execute queries on them, and your notebook, Power BI dataset semantic model, and lakehouse can be located in the same workspace or in different workspaces [Top End to End MLOps Platforms and Tools in 2024 | JFrog ML](https://www.qwak.com/post/top-mlops-end-to-end)

**2. Business Logic Reuse**

* Access Power BI measures and calculations directly in Python
* No need to recreate DAX logic in Python
* Ensures consistency between BI reports and ML models

**3. FabricDataFrame Integration** Resolves column names in the FabricDataFrame to Power BI dimensions, ignoring any column names that can't be resolved within the given semantic model [Should I use Azure ML or Fabric Data Science?](https://robkerr.ai/should-i-use-azure-ml-or-fabric-data-science/)

**🎯 Use Case 1: Sales Forecasting with Business Logic Consistency**

**Scenario:** A retail company wants to build ML forecasting models that use the same business logic as their Power BI sales reports.

**Challenge:**

* Finance team has complex DAX measures for revenue calculations
* Data science team was recreating logic in Python (risk of inconsistency)
* Different definitions causing confusion in predictions vs. actuals

**What is SynapseML?**

SynapseML simplifies the creation of massively scalable machine learning pipelines [MLOps & LLMOps in MS Fabric - Microsoft Fabric Community](https://community.fabric.microsoft.com/t5/General-Discussion/MLOps-amp-LLMOps-in-MS-Fabric/td-p/3716622). It's built on Apache Spark and provides:

* **Distributed ML** - Process massive datasets across clusters
* **Pre-built Models** - Computer vision, NLP, anomaly detection
* **Azure AI Services Integration** - Cognitive Services at scale

**Key Features:**

**1. Scalable Training**

* Train models on terabytes of data
* Distributed hyperparameter tuning
* Support for deep learning frameworks

**2. Azure AI Services Integration** Enrich your data with artificial intelligence using pretrained models from Azure AI services [Microsoft Fabric in Action: An Architecture for Real-time and ML-Powered Data | by Vishal Khondre | Medium](https://medium.com/@vishal.khondre/microsoft-fabric-in-action-an-architecture-for-real-time-and-ml-powered-data-77362b5fd65c)

**3. MLOps-Ready**

* Pipeline orchestration
* Model versioning
* Automated retraining

**Use Case 3: Large-Scale Image Classification for E-commerce**

**Scenario:** Online marketplace has millions of product images that need automatic categorization.

**Challenge:**

* 50+ million product images
* Need to classify into 500+ categories
* Single-machine processing would take months
* Manual labeling is cost-prohibitive

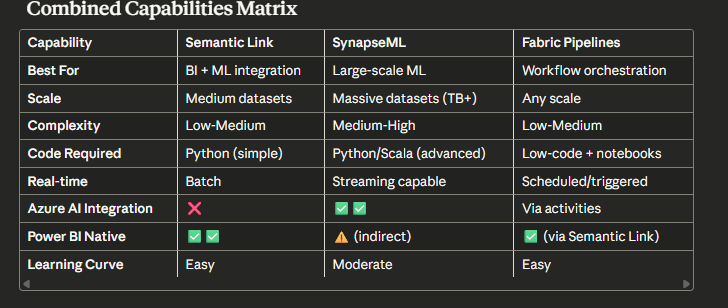
**Solution with SynapseML:**

**Benefits:**

* ✅ Process 50M images in hours instead of months
* ✅ Automatic scaling with Spark clusters
* ✅ Pay only for compute used
* ✅ Pre-trained models (no training required)
* ✅ Results immediately available in Lakehouse

**What are Fabric Pipelines?**

* Pipeline activities include Fabric notebook, HDInsight activity, Spark job definition, stored procedure, SQL scripts, and more, letting you run custom code or scripts to transform your data [Tutorial: Deploy a model - Azure Machine Learning | Microsoft Learn](https://learn.microsoft.com/en-us/azure/machine-learning/tutorial-deploy-model?view=azureml-api-2). Data pipelines enable data engineers to build complex workflows that can orchestrate many different types of data processing, data movement, data transformation, and other activity types [Deploy Machine Learning Models to Online Endpoints - Azure Machine Learning | Microsoft Learn](https://learn.microsoft.com/en-us/azure/machine-learning/how-to-deploy-online-endpoints?view=azureml-api-2).
* **Key Components:**
* **1. Copy Data Activity** - Move data between sources **2. Dataflow Gen2** - Low-code transformations (300+ connectors) **3. Notebook Activity** - Execute Python/Spark code **4. Control Flow** - If/else, loops, variables, parameters **5. Monitoring & Alerts** - Track pipeline health

  
  
**Best Practices for Combining These Capabilities**

**Pattern 1: The Complete ML Pipeline**

Fabric Pipeline → SynapseML Training → Semantic Link → Power BI

* Pipeline orchestrates the workflow
* SynapseML handles heavy ML processing
* Semantic Link publishes results to Power BI

**Pattern 2: Business-Driven ML**

Power BI (Business Logic) → Semantic Link → ML Model → Back to Power BI

* Ensures ML uses business-approved definitions
* Creates feedback loop between analysts and data scientists

**Pattern 3: Scalable AI Services**

Pipeline (Data Ingestion) → SynapseML (AI Services) → Lakehouse → Semantic Link → Power BI

* Use for vision, language, or speech at scale
* Results enriched and visualized

**Spark PREDICT for Batch Scoring**

**What is PREDICT?**

Microsoft Fabric allows users to operationalize machine learning models with the scalable PREDICT function. This function supports batch scoring in any compute engine. Users can generate batch predictions directly from a Microsoft Fabric notebook or from the item page of a given ML model [MLOps & LLMOps in MS Fabric - Microsoft Fabric Community](https://community.fabric.microsoft.com/t5/General-Discussion/MLOps-amp-LLMOps-in-MS-Fabric/td-p/3716622).

**Key Features of PREDICT**

**1. Supported Model Frameworks**

The PREDICT function is currently supported for this limited set of ML model flavors: CatBoost, Keras, LightGBM, ONNX, Prophet, PyTorch, Sklearn, Spark, Statsmodels, TensorFlow, and XGBoost [MLOps & LLMOps in MS Fabric - Microsoft Fabric Community](https://community.fabric.microsoft.com/t5/General-Discussion/MLOps-amp-LLMOps-in-MS-Fabric/td-p/3716622)

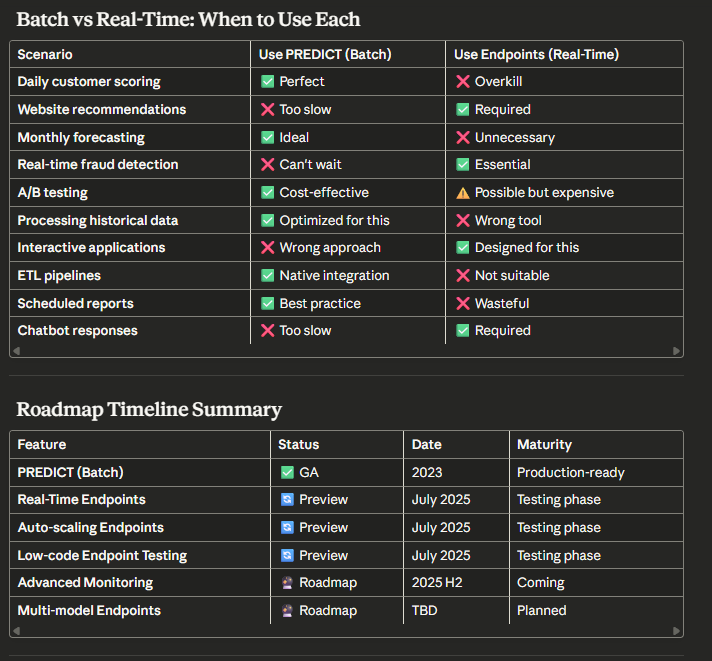
**2. Requirements**

PREDICT requires that you save ML models in the MLflow format, with their signatures populated [MLOps & LLMOps in MS Fabric - Microsoft Fabric Community](https://community.fabric.microsoft.com/t5/General-Discussion/MLOps-amp-LLMOps-in-MS-Fabric/td-p/3716622)

**3. Multiple Invocation Methods**

You can call PREDICT in three different ways:

* **Transformer API** - Native Spark ML pipeline integration
* **Spark SQL API** - SQL-based prediction generation
* **PySpark UDF** - User-defined function approach

  
  
**Spark PREDICT for Batch Scoring**

**What is PREDICT?**

Microsoft Fabric allows users to operationalize machine learning models with the scalable PREDICT function. This function supports batch scoring in any compute engine. Users can generate batch predictions directly from a Microsoft Fabric notebook or from the item page of a given ML model [MLOps & LLMOps in MS Fabric - Microsoft Fabric Community](https://community.fabric.microsoft.com/t5/General-Discussion/MLOps-amp-LLMOps-in-MS-Fabric/td-p/3716622).

**Key Features of PREDICT**

**1. Supported Model Frameworks**

The PREDICT function is currently supported for this limited set of ML model flavors: CatBoost, Keras, LightGBM, ONNX, Prophet, PyTorch, Sklearn, Spark, Statsmodels, TensorFlow, and XGBoost [MLOps & LLMOps in MS Fabric - Microsoft Fabric Community](https://community.fabric.microsoft.com/t5/General-Discussion/MLOps-amp-LLMOps-in-MS-Fabric/td-p/3716622)

**2. Requirements**

PREDICT requires that you save ML models in the MLflow format, with their signatures populated [MLOps & LLMOps in MS Fabric - Microsoft Fabric Community](https://community.fabric.microsoft.com/t5/General-Discussion/MLOps-amp-LLMOps-in-MS-Fabric/td-p/3716622)

**3. Multiple Invocation Methods**

You can call PREDICT in three different ways:

* **Transformer API** - Native Spark ML pipeline integration
* **Spark SQL API** - SQL-based prediction generation
* **PySpark UDF** - User-defined function approach

**Real-Time Scoring: Current Status & Roadmap**

**🎉 MAJOR UPDATE: Real-Time Endpoints Now Available (Preview)**

Fabric offers a wide variety of data-science capabilities, from automated machine learning with FLAML to batch inferencing with the SynapseML PREDICT function. We're pleased to announce that ML models can now serve real-time predictions from secure, scalable, and easy-to-use online endpoints [Machine learning model - Microsoft Fabric | Microsoft Learn](https://learn.microsoft.com/en-us/fabric/data-science/machine-learning-model).

**Announcement Date:** July 2025 (Preview)

**Real-Time Endpoint Capabilities (Preview)**

**Key Features:**

**1. One-Click Deployment** ML models in Fabric come prebuilt with online endpoints that can be used to serve real-time predictions. Each registered model version has a dedicated endpoint URL, which can be found under the "Endpoint details" heading in the Fabric interface [Operationalizing Machine Learning Pipelines in Fabric – Hitachi Solutions](https://global.hitachi-solutions.com/blog/operationalizing-ml-pipelines-in-fabric/)

**2. Auto-Scaling**

* Automatic scale-out based on request load
* No manual infrastructure management
* Pay-per-use pricing model

**3. REST API Access**

* Standard REST API for integration
* Supports JSON request/response format
* Authentication via Azure AD tokens

**4. Low-Code Testing Interface**

* Test endpoints directly in Fabric UI
* No code required for quick validation
* View response times and results

**PREDICT (Batch Scoring) - Simple Use Cases 🎯**

**1. Daily Email Campaign Targeting**

**What it does:** Every morning, score all 5 million customers to find who's likely to buy today.

**Why batch:** You don't need answers instantly - scoring overnight is fine.

**Example:**

* 11 PM: Pipeline starts
* 12 AM: PREDICT scores 5M customers
* 6 AM: Marketing team sees results in Power BI
* 8 AM: Emails sent to top 50K customers

**Like:** Grading all homework papers overnight vs. grading one while student waits.

**2. Loan Approval at Bank**

**What it does:** Tell customer "Approved!" or "Sorry, not now" immediately when they apply.

**Why real-time:** Customer is sitting in bank or on website waiting.

**Example:**

* Customer fills application → Submit!
* App calls Fabric endpoint → ⚡ 100ms
* "Congratulations! Approved for $10,000"
* Customer is happy and signs

**Like:** Getting your test score right after finishing vs. waiting 2 weeks.

**3. Chatbot Answers**

**What it does:** Help customer service bot answer questions with AI predictions.

**Why real-time:** Customer is typing and expecting instant responses.

**Example:**

* Customer: "Will my package arrive tomorrow?"
* Chatbot calls endpoint with order details → ⚡ 80ms
* Model predicts: "Yes, 95% likely by 3 PM"
* Chatbot: "Great news! Tomorrow by 3 PM!"

**Like:** Asking teacher a question and getting answer now vs. teacher emailing you tomorrow.