

Last updated Q3FY25

The material in this GenAI pitch deck is intended to help you:

1. Learn the Generative AI for Comms value proposition and key capabilities
2. Position Generative AI for Comms using Databricks
3. Ensure your joint go-to-market solutions reflect the same value proposition

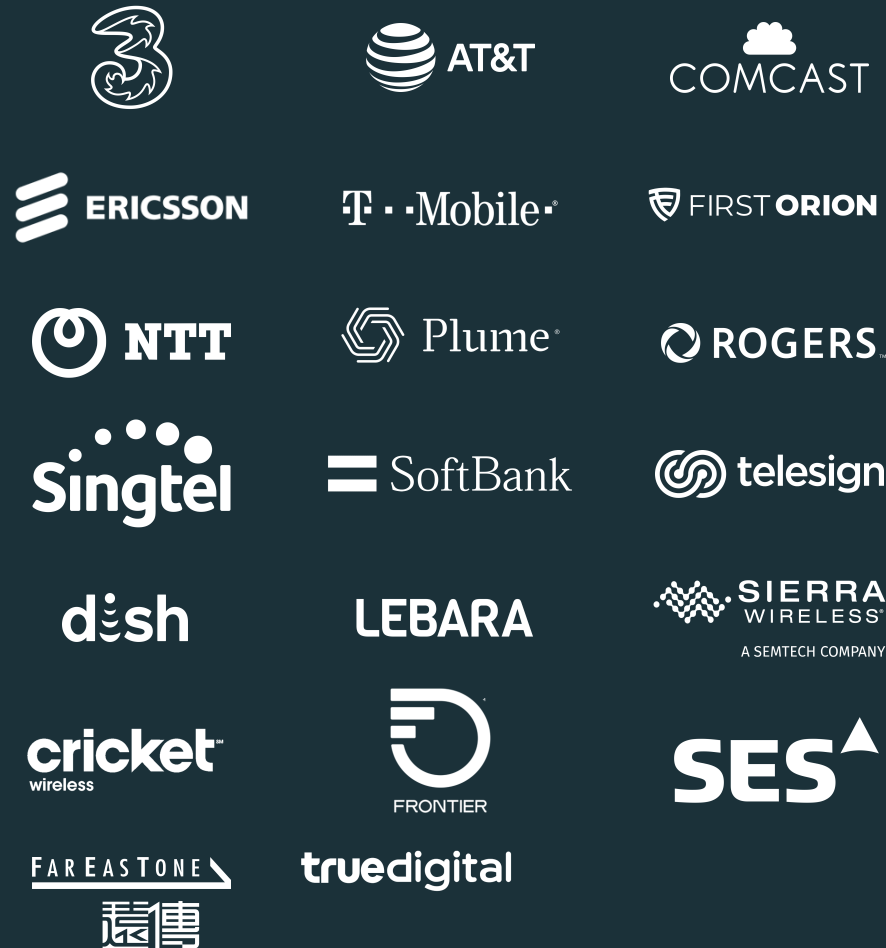


Generative AI in Communications

databricks.com/communications



Building data-driven businesses in Communications



Databricks will unlock over **\$5.6B of value in Communications, Media, & Ent** over the next three years with **AI**

Databricks unlocks value from high impact AI use cases across Communications, Media & Entertainment

Use Case	Audience segmentation & targeting	Dynamic Pricing & Revenue Mgmt	Customer Sentiment, Satisfaction Analysis	Supply Chain Optimization	Call Center Augmentation
Databricks Customer	\$2B+ Media Conglomerate	\$10B+ Entertainment & Vacation provider	\$20B+ Entertainment Operator	Top 3 US Telco Provider	\$8B+ Call Center operator
Value Driver	<ul style="list-style-type: none"> ▲ Higher engagement ▲ Increase CPM 	<ul style="list-style-type: none"> ▲ Increase ARPU ▲ Increase Customer volume ▲ Increase Upsell 	<ul style="list-style-type: none"> ▲ Quality & velocity of insight ▲ Improve advertising sales ▼ Reduce cost to analyze ▼ Reduce time to analyze 	<ul style="list-style-type: none"> ▼ Reduce supply chain logistics costs with end to end visibility 	<ul style="list-style-type: none"> ▲ Increase upsell ▲ Improve CSAT ▲ Reduce Churn
Total Validated Impact	\$140-160M	\$100-150M	\$15-35M	\$200-250M	\$50-100M

▲ Generate Revenue ▼ Save Cost

Source: Validated case studies and analysis of top Databricks Telco customers



Transforming Communications with Generative AI

Market Research

94%

CSP leaders believe GenAI will have a significant impact in next 5 years¹

86%

CSPs have GenAI in production or have GenAI experiments in the last year²

60%

CSPs plan to use open source models for GenAI/LLMs to enable scalability³

Domains expected to bring the most value from GenAI

based on MIT survey of 600+ CIOs, percentage ranked in top 2

Data Quality & Network Optimization	43%
Real-time Insights	40%
Automation & Efficiency (Contact Center)	33%
Security, Compliance & Risk Management	23%
Internet of Things (IoT) & Machine-to-Machine (M2M)	17%
Customer Experience	13%

1) Bringing Breakthrough Data Intelligence to Industries, MIT Technology Review ([link](#)). 2) CEO decision-making in the age of AI, IBM Institute for Business Value ([link](#)). 3) How generative AI can boost highly skilled workers' productivity, MIT Management Sloan School ([link](#)). 4) McKinsey: The economic potential of generative AI: The next productivity frontier ([link](#))



GenAI use cases across Communications



Data & Network Optimization

- Network operations monitoring roaming analysis
- QoS speed and reset analysis
- Capacity and predictive usage
- Trouble ticket resolution
- CDR/PCMD/LSR pipelines
- Traffic/demand forecasting
- Network anomaly detection



IoT & M2M

- Smart home analytics
- Predictive maintenance
- Equipment fault detection
- Real-time resource management
- Engineering deployment prediction
- Data monetization—location based campaigns, road traffic management, tourism, etc.



Security, Compliance, and Risk

- Fraud detection
- Autonomous monitoring for DoS and robo-calling
- Brute force attacks
- Bot detection
- GDPR/CCPA compliance

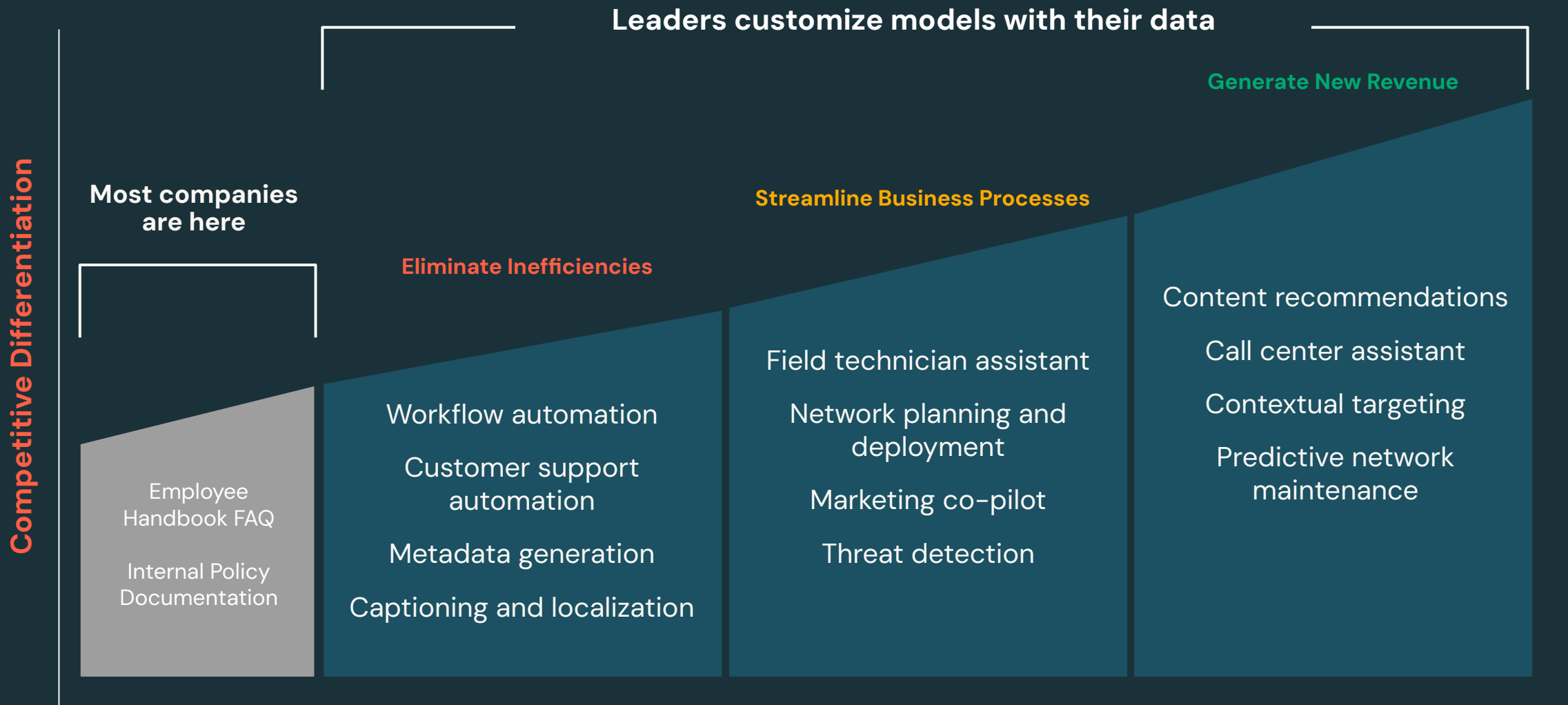


Customer Experience & Service

- Customer 360
- Next Best Action/Offer
- Personalization/consumer journey
- Churn prediction
- Customer lifetime value
- Call center automation



How communications service providers are achieving results with Generative AI



Challenge:

Building and deploying
production-quality
Gen AI solutions is
difficult

90%

of enterprises not
confident going
to production

Why is it hard to get to production

Control, Quality, Cost



No **control** over the data or the models



Production-level **quality** is challenging



High **cost** at scale



Concern over data leakage
Lack of control and ownership



Unpredictable performance
Need automation and scale



Foundation models are expensive at scale
Expensive to build LLMs

Build better GenAI solutions on Databricks

You need more than just good models

Complete control



Complete ownership over models and data



Reduce potential privacy & reputational risk

Production quality



Accurate, safe and governed GenAI apps



Only platform with native evaluation, monitoring and governance

Lower cost



Cost-effective to build LLMs at scale



Up to 90% less expensive for RAG and to train your own LLMs

Databricks Mosaic AI is the only end-to-end platform for all your GenAI use cases

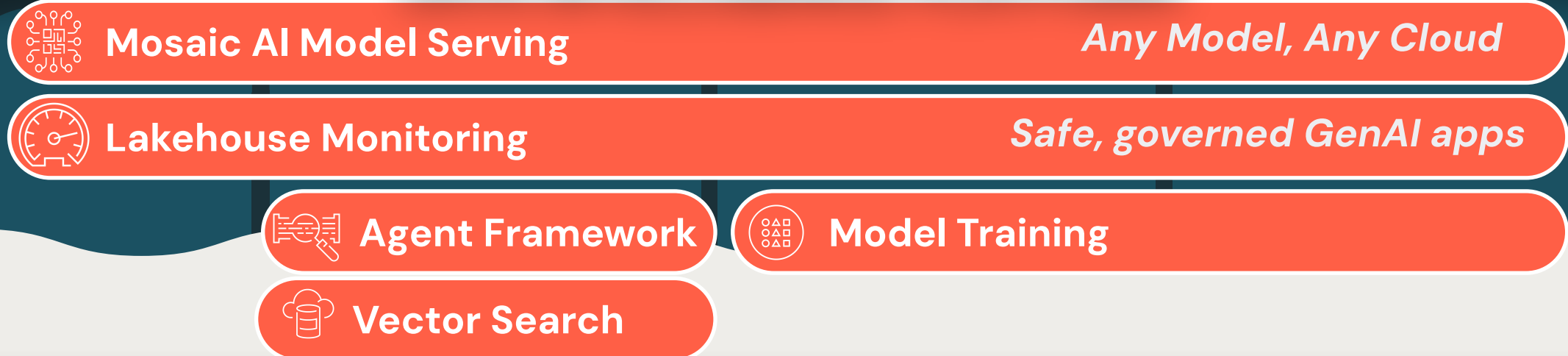
Out-of-the-box models

RAG

Fine-tuning




Pretraining

DATA INTELLIGENCE PLATFORM



Communications: Where to get started

Key use cases where we are seeing the most traction

<u>Use Case</u>	<u>Description</u>	<u>GenAI Type</u>	<u>Effort / Cost</u>	<u>Value¹</u>
 Customer Support Automation	Automatically categorize, summarize, and draft a response for every inbound customer message (email/call) using your brand tone to quickly and efficiently remediate customer concerns.	Fine-tuning + RAG	Complexity: 100% Cost: 50% Support: 100%	\$480-960m Estimated annual impact
 Call Center Assistant	Instant and interactive access to a customer's recent pre-call experience and recommended next best actions/offers at the fingertips of every call center agent	Fine-tuning + RAG	Complexity: 50% Cost: 50% Support: 100%	\$480-960m Estimated annual impact
 Field Technician Assistant	Diagnose, troubleshoot, and resolve cell tower issues more efficiently by equipping field technicians with a chat interface that replaces the need to manually sift through hundreds of tower manuals (PDFs) and internal knowledge bases (wikis)	RAG	Complexity: 50% Cost: 50% Support: 100%	\$360-720m Estimated annual impact



Field Technician Assistant

Global telecommunications provider with \$78B in revenue leverages Databricks to diagnose, troubleshoot, and resolve call tower issues

Challenges

- Field technicians are a limited resource
- Technicians ability to resolve issues timely has direct impact on customer satisfaction and revenue

Use case

- Field Technician Assistant

Solution

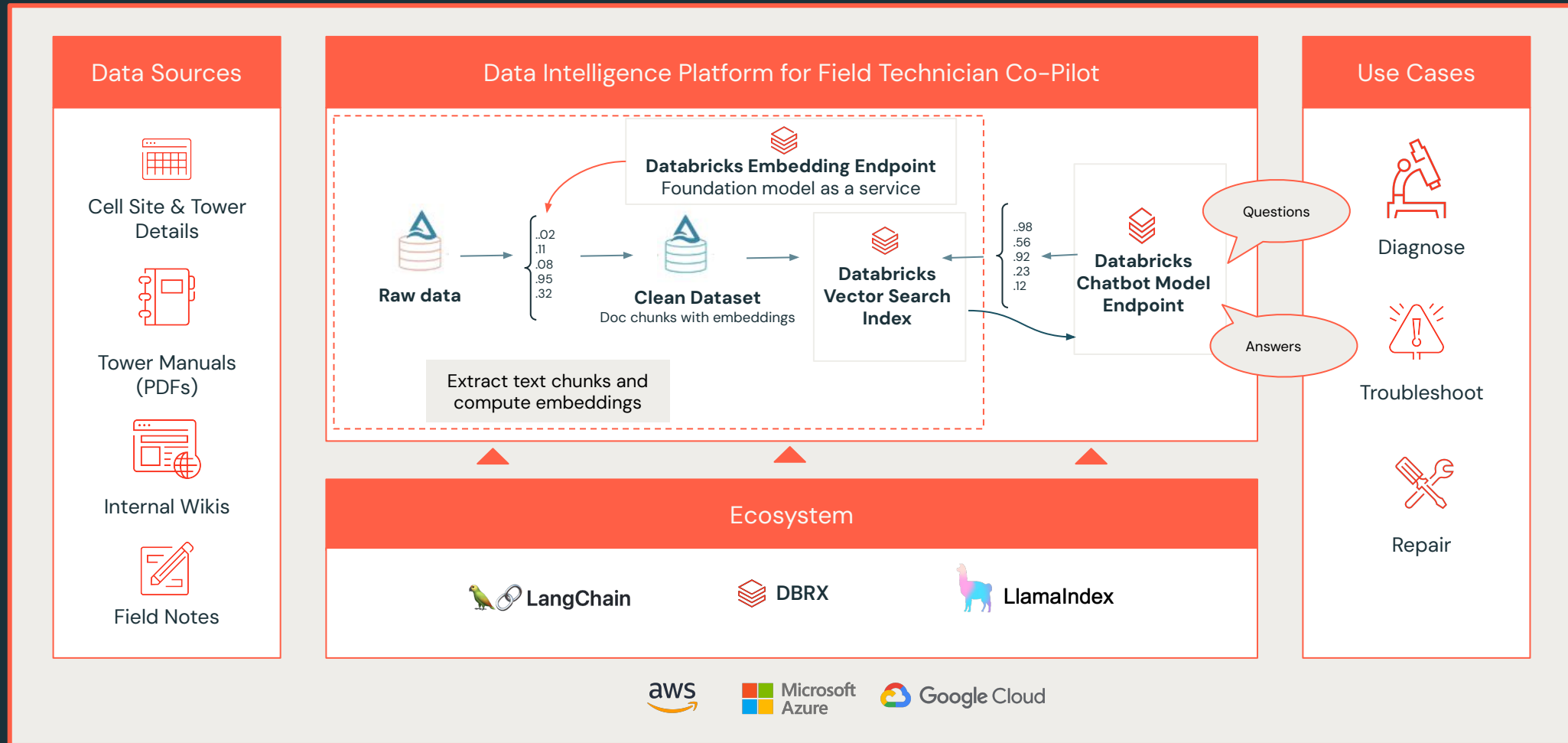
- LLM-powered chat interface empowers field technicians to better diagnose, troubleshoot, and resolve cell tower issues more efficiently
- This replaces the need to manually sift through hundreds of tower manuals (PDFs) and internal knowledge bases (wikis)

Impact

- Enables faster resolution of cell tower issues, reducing downtime and service disruptions.
- Streamlines the troubleshooting process
- Decreases the need for on-site technician visits by enabling remote diagnostics and resolution
- Minimizes service disruptions and improves network reliability by quickly resolving tower issues

Field Technician Assistant Architecture

RAG Model





Customer Support Automation

Global communications provider with \$120B in revenue uses Databricks to automate customer support responses, increasing accuracy while reducing time to resolution

Challenges

- As the volume of customer support queries increases, so does the strain on human support resources
- You want human agents focused on the most critical tasks

Use case

- Customer Support Automation

Solution

- Automatically categorize, summarize, and draft a response for every inbound customer message (email/call) using your brand tone to quickly and efficiently remediate customer concerns.
- Augment human support teams by providing guidance that improves the speed, consistency and accuracy

Impact

Leverage the complete knowledge base of support responses to draft more accurate and timely answers. This helps to improve customer satisfaction through improved time to resolution for support queries.



Call Center Assistant

Global telecommunications provider with \$122B in revenue uses Databricks to integrate intelligent chatbots alongside human support agents to improve win-back and upsell

Challenges

- Call centers are critical opportunities for win-back and up-sell with customers
- Need solutions that improve the accuracy and speed to resolution

Use case

- Call Center Co-Pilot

Solution

- Instant and interactive access to a customer’s recent pre-call experience and recommended next best actions/offers at the fingertips of every call center agent
- Augment human support teams by providing guidance that improves the speed, consistency and accuracy
- Accelerate ramp-up of new customer support agents

Impact

Tool provides customer support agents with direct access to real-time information, and personalized recommendations based on customer history. This helps to:

- Improve the time to resolution
- Increase the accuracy of recommendations
- Improve customer win-back and upsell rate
- Increase the value of each human agent



Achieve quick results in Generative AI with Databricks

1

Conduct Workshops & Build Rapid POCs

Get to POC in weeks

- Gen AI Use Case Identification
- Architecture Workshop (optimal approach to Gen AI, eg. RAG)
- LLM PoC (RAG)

2

Leverage Trusted Partners to Accelerate

Accelerate your path to Production

PS Packages

- LLM Reference Implementation (RAG)
- Fine-tune model & strategy
- Pre-training Enablement

SI + Databricks GenAI Offerings

- Accenture, Slalom
- Tredence, Koantek, Aimpoint

3

Training & Solution Accelerators

Enable your teams

Training Services

- [Generative AI foundation course](#)
- In-person and virtual training

Solution Accelerators

- [Building recommendations](#)
- [Product review summarization](#)
- [Creating Brand-Aligned Images](#)
- [LLMs for RAG-based Q&A](#)
- ...and more

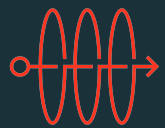




Appendix

Influence of GenAI on industry trends

Transforming operations, customer interactions, and service offerings



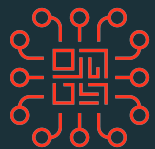
Personalization At Scale

Increased engagement and retention at scale through messaging and recommendations



Enhanced Customer Service

Improve customer experience and reduce burden on human agents



Efficient Network Optimization

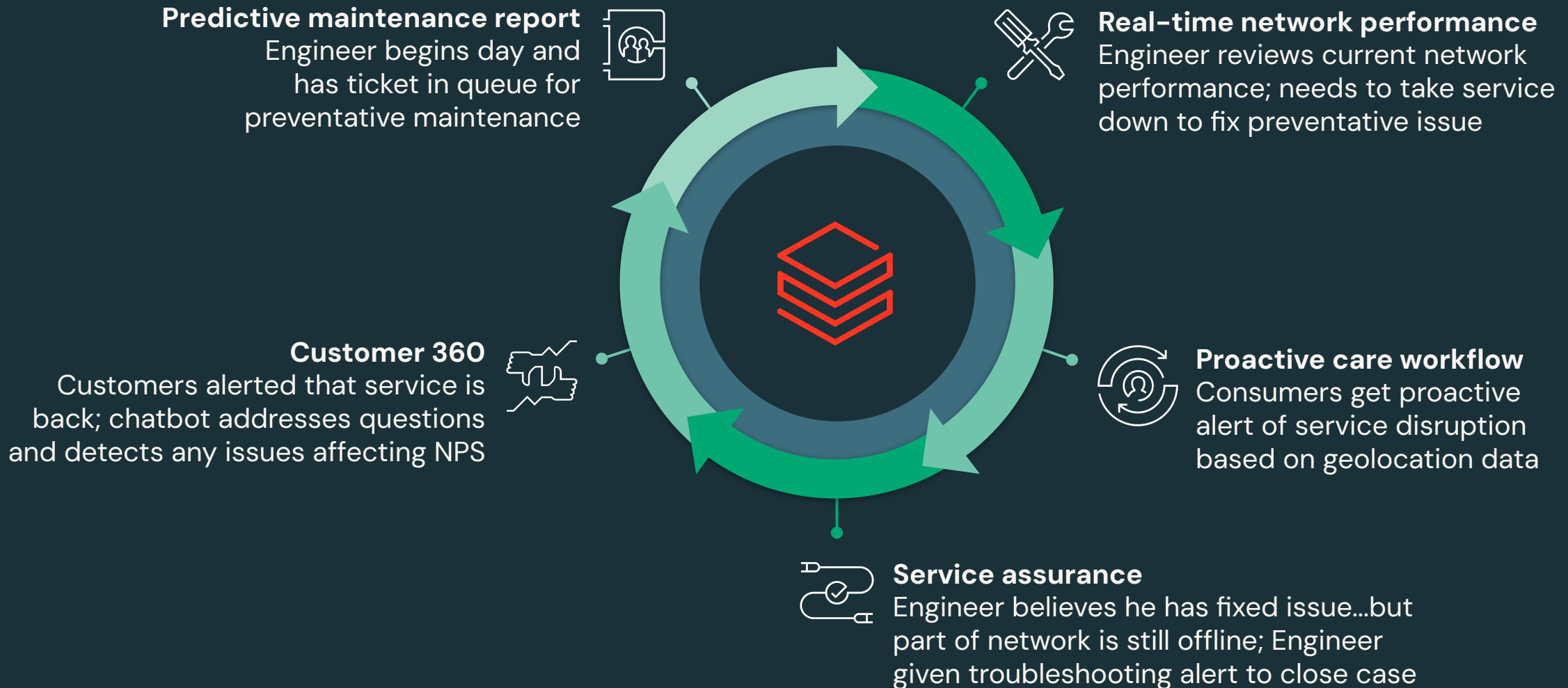
Reduce the cost to serve and operate through network optimizations



Predictive Maintenance

Increase reliability and efficiency of networks, improving service quality while reducing downtime

Bringing GenAI to Life



Driving value with LLMs

Using LLMs and generative AI across the entire CSP lifecycle



Transforming
Personalization



Network Optimization
and Efficiency



Enhancing Customer
Service and Support



Unlocking Revenue Streams
through Value-Added Services



Network Security and
Threat Detection



Predictive Maintenance
and Crew Scheduling



Solution Accelerators

Enhanced Product Search

Foster transformative customer interactions

Large language models (LLMs) can be used to harness the rapidly growing range of content and goods to ensure customer searches yield the desired results. With the Databricks Data Intelligence Platform for Retail, organizations can:

- Unify product, query and label data within a retailer’s product catalog
- Enable rapid search with analytics against numerical arrays
- Train and deploy an LLM model with Databricks Model Serving

Business Value:
Better customer experience
Higher conversion rate
Increased cart size

DEPLOY MODEL AS MICROSERVICE TO INTEGRATE WITH APPLICATIONS

Model	Version	Name	State	Compute
	Version 1	wands_tuned_search-1	Ready	Medium 8-16 concurrent requests (8-16 CPU)

UNIFY STRUCTURED AND UNSTRUCTURED DATA FOR LABELING

```
labels_schema = StructType([\n  StructField('id', IntegerType()),\n  StructField('query_id', IntegerType()),\n  StructField('product_id', IntegerType()),\n  StructField('label', StringType())\n])\n\nspark\n  .read\n  .csv(\n    path='dbfs://wands/downloads/Label.c\n    sep='\\t',\n    header=True,\n    schema=labels_schema\n  )\n  .write\n  .format('delta')\n  .mode('overwrite')\n  .option('overwriteSchema', 'true')\n  .saveAsTable('Labels')\n\ndisplay(spark.table('Labels'))
```

LEVERAGE DEEP LEARNING TO TRAIN GENERATIVE AI AND LLM MODELS

```
graph TD\n    DS[Data storage] --> MD[Model development]\n    MD --> MR[Model registry]\n    MR --> MS[Model serving]\n    MS --> U[Users]\n    U --> LSC[Live customer searches]\n    U --> RP[Recommended products]
```



Product Recommendations with LLMs

Deliver intuitive product recommendations that drive customer journeys

Product recommendations play a central role in guiding customers through their shopping journey with tailored suggestions based on their buying behaviors and preferences. With LLMs, Retailers can automate the delivery of personalized suggestions that adapt to evolving customer preferences — enhancing user engagement, increasing sales and fostering long-term customer loyalty.

Use this [Solution Accelerator](#) to develop product recommendations based on common sense linkages for new-to-market products and optimized recommendation engines:

- Convert all of your specific product descriptions and metadata into embeddings and store them in a searchable index
- Task an LLM to recommend products based on their connection to other relevant products



Automating Product Review Summarization

Keep track of customer feedback at scale

The breadth of digital communication channels has made it increasingly effortless for customers to write product reviews that can significantly impact the perception of a business. With large language models (LLMs), retailers can easily extract and summarize insights from huge volumes of customer feedback to enhance decision-making and product development strategies.

Use this [Solution Accelerator](#) to streamline the summarization of customer feedback, allowing your organization to:

- Process a high volume of reviews at a lower cost
- Collect feedback from a wider range of products and summarize it regularly
- Task an LLM to extract different sets of information from each high-level category of reviews

UNIFY CUSTOMER REVIEWS TO TRAIN MODELS

EXTRACT ACTIONABLE INSIGHTS FROM CUSTOMER REVIEWS

DEPLOY MODEL AS MICROSERVICE TO INTEGRATE INTO APPLICATIONS

Build Summary Reviews

This dataframe will have positive and negative reviews placed in the same row rather than having separate rows for each. We will use a pivot function for this.

```
# Imports
from pyspark.sql import functions as SF

# Build meta reviews df
summary_reviews_df = (
    reviews_df.groupBy("asin", "title", "author", "week_start")
    .pivot("star_rating_class")
    .agg(SF.first("final_review_summary"))
    .withColumnRenamed("high", "positive_reviews_summary")
    .withColumnRenamed("low", "negative_reviews_summary")
    .orderBy("asin", "title", "author", "week_start")
)
```



Creating Brand-Aligned Images using GenAI

Analyze unstructured data with LLMs to improve the customer experience

Image-generating technologies offer significant benefits for retail and consumer goods companies. By using generative models that produce both stylized and photo-realistic images from user prompts, marketing professionals, designers, and product development teams can quickly and effectively explore new ideas and designs.

While using models pre-trained on large volumes of generic images is great for producing cohesive imagery, most organizations seek to mimic patterns, designs and aesthetics specific to a particular brand or domain.

In this [Solution Accelerator](#), users can expedite the end-to-end development of personalized image generation applications:

- Pre-process training images for fine-tuning
- Fine-tune a model to understand specific brand/domain elements to produce outputs better aligned with the needs of the organization
- Manage & deploy the fine-tuned model behind a Databricks model endpoint to make it available for downstream applications.

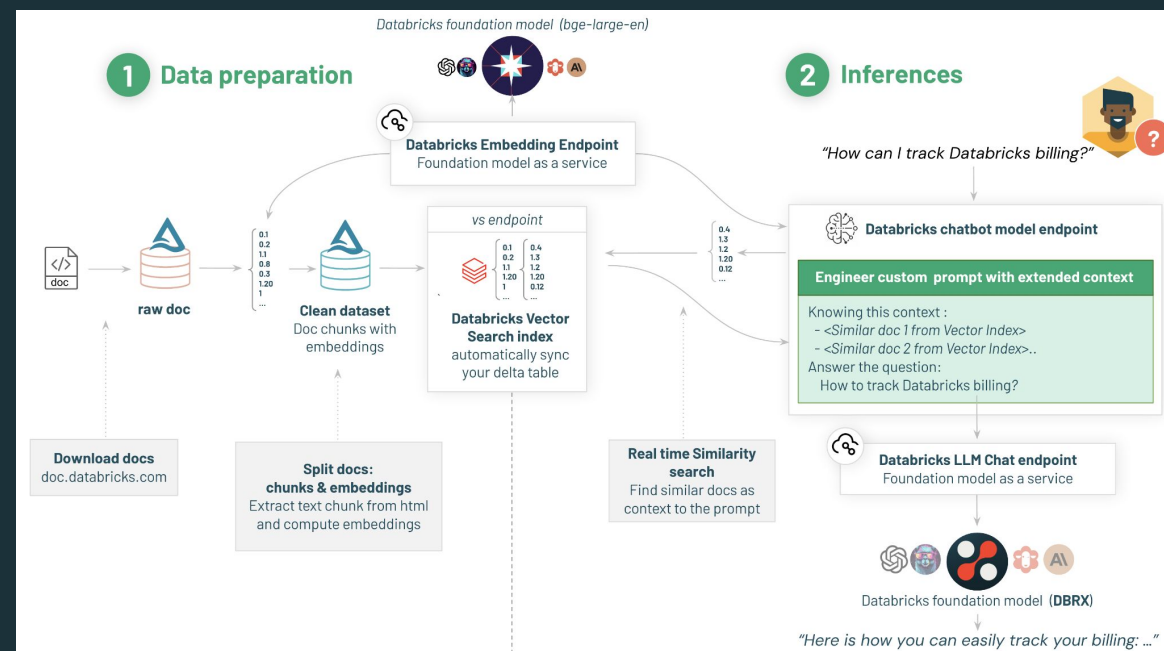


Deploy LLM Chatbots with Databricks

Incorporates RAG, DBRX Instruct Foundation Models, Vector Search

In this [tutorial](#), we will cover how Databricks is uniquely positioned to help you build a chatbot using Retrieval Augmented Generation (RAG) and Databricks DBRX Instruct Foundation Model:

- Prepare clean documents to build your internal knowledge base and specialize your chatbot
- Leverage Databricks Vector Search with our Foundation Model endpoint to create and store document embeddings
- Search similar documents from our knowledge database with Databricks Vector Search
- Deploy a real-time model using RAG and providing augmented context in the prompt
- Leverage the DBRX instruct model through with Databricks Foundation Model endpoint (fully managed)



LLMs with Better Data using Cleanlab Studio

Improve training data to boost LLM performance

Data powers AI in the enterprise, but real-world data sets have been found to contain 7%–50% annotation errors. Unsurprisingly, erroneous data — from imperfectly labeled data to outliers — hampers the training (and evaluation) of ML models across tasks like intent recognition, entity recognition and sequence generation.

Our joint [Solution Accelerator](#) with Cleanlab helps to improve training data to boost LLM performance by 37% without spending any time or resources to change the model architecture, hyperparameters or the training process.

- Fine-tune state-of-the-art LLMs using OpenAI's APIs
- Evaluate trained LLMs to achieve high test accuracy (65%)
- Improve data quality using confident learning and DCAI without writing any code or having any ML expertise using Cleanlab Studio

FINE-TUNE OPENAI'S DAVINCI MODEL ON A BASELINE DATA SET

Improving OpenAI Davinci LLM by Improving Data with Cleanlab Studio

original data

improved data (data and label issues fixed)

Test Accu

IMPROVE DATA QUALITY USING CLEANLAB STUDIO

```
pip install --upgrade cleanlab-studio
import cleanlab_studio

cleanlab_studio
Collecting cleanlab-studio
  Downloading cleanlab_studio-1.0.5-py3-none-any.whl (56 kB)
    Requirement already satisfied: aiohttp>=3.8.1 in /local_disk0/ephemeral...
Collecting Pillow>=9.2.0
  Downloading Pillow-9.5.0-cp39-cp39-manylinux_2_28_x86_64.whl (3.4 MB)
Collecting json>=3.1.4
  Downloading json-3.2.0.post0-cp39-cp39-manylinux_2_17_x86_64.manylinux...
Collecting click>=8.1.0
  Downloading click-8.1.3-py3-none-any.whl (96 kB)
  Downloading click-8.1.3-py3-none-any.whl (96 kB) 96 kB 9.6 MB/s eta 0:00:01
```

OPTIMIZE LLM PERFORMANCE AND EVALUATE RESULTS

Project: Stanford Politeness

Issues resolved: 100 | Label Issues found: 514 | Outliers found: 0 | Total Examples: 1916

Given	Suggested	Corrected	Quality	Action	Tags
polite	neutral	neutral	0.142	auto-fix	-
polite	neutral	neutral	0.158	auto-fix	-
polite	neutral	neutral	0.177	auto-fix	-
impolite	neutral	neutral	0.180	auto-fix	-
polite	neutral	neutral	0.181	auto-fix	-
polite	neutral	polite	0.181	keep	-
polite	neutral	impolite	0.182	re-label	-
polite	neutral	neutral	0.182	auto-fix	-
impolite	polite	polite	0.190	auto-fix	-
polite	neutral	neutral	0.193	auto-fix	-
polite	neutral	polite	0.194	keep	-
polite	neutral	neutral	0.197	auto-fix	-
polite	neutral	neutral	0.200	auto-fix	-
polite	neutral	neutral	0.201	auto-fix	-
impolite	neutral	neutral	0.201	auto-fix	-
polite	neutral	-	0.203	exclude	-



LLMs for Customer Service & Support

Elevating customer satisfaction with LLM-powered chatbots

With the advancements in large language models (LLMs), the integration of intelligent chatbots has emerged as a transformative force, offering agent-led customer service and support organizations unrivaled efficiency, scalability and personalized interactions. [Using pretrained LLMs on the Databricks Lakehouse](#), organizations can:

- Ingest enterprise data from various knowledge bases to build a context-enabled LLM-based chatbot
- Augment human support teams by providing guidance that improves the speed, consistency and accuracy of their work
- Accelerate ramp-up of new customer support agents



LLM for Customer Service Analytics

Analyze unstructured data with LLMs to improve the customer experience

Improve call center performance by automating tasks such as customer service, claims processing, and underwriting. With this [Solution Accelerator](#):

- Detect customer intents based on textual data to improve customer experience
- Leverage large language models (LLMs) to optimize chatbot performance
- Unify internal data with external unstructured data to minimize underwriting risk

ANALYZE UNSTRUCTURED DATA AT SCALE WITH LLM

Generating a data profile

```
# By using the display function, we can easily generate a data profile for our dataset
display(dataset["train"].to_pandas())
```

index	topic_en	question_en
1	life-insurance	Can Creditors Take Life Insurance After Death?
2	renters-insurance	Does Travelers Insurance Have Renters Insurance?
3	auto-insurance	Can I Drive A New Car Home Without Insurance?
4	life-insurance	Is The Cash Surrender Value Of Life Insurance Taxable?

TRAIN MODELS TO CLASSIFY CUSTOMER INTENT

Calculating Performance Metrics

```
# We calculate the amount of correct predictions and divide by the total

import seaborn as sns
from matplotlib import pyplot as plt
%matplotlib inline
%config InlineBackend.figure_format='retina'

pred_df = predictions.toPandas()
pred_df["hit"] = pd.to_numeric(pred_df["accuracy_per_intent"])
accuracy_per_intent = pred_df.groupby("topic_en").agg({"hit": "sum", "accuracy_per_intent": "mean"})

fig, ax = plt.subplots(nrows = 1, ncols = 1)
plt.xticks(rotation = 45)
sns.barplot(x = accuracy_per_intent["hit"], y = accuracy_per_intent["accuracy_per_intent"], palette = "mako", ax = ax)
plt.title("Prediction Accuracy per Intent")
```

DEPLOY REAL-TIME MODELS USING DATABRICKS MODEL SERVING

Querying the endpoint through REST API

```
def test_prediction_endpoint(questions):
    endpoint_url = f"serving-endpoints/{endpoint_name}/invocations"
    payload = {"instances": questions}

    data_json = json.dumps(payload)
    print(data_json)
    headers["Content-Type"] = 'application/json'
    response = requests.request(
        method="POST",
        headers = headers,
        url = f"{databricks_url}/{endpoint_url}",
        data = data_json
    )

    if response.status_code != 200:
        raise Exception(f"Request failed with status {response.status_code}")
    return response.json()

test_questions = [
    "my car broke, what should I do?",
    "what is my life insurance coverage?",
    "can you send me my health insurance cover?"
]
```



Communications accelerators



Streaming Quality of Service/Experience

Measure platform stability and performance to prevent churn

[Blog](#) - [Notebook](#)



Customer Lifetime Value

Understand customer lifetime and estimate future spend

[Blog](#) - [Notebook](#)



LLMs for Customer Support

Elevate customer satisfaction with LLM-powered chatbots

[Blog](#) - [Notebooks](#)



Telco Network Analytics

Increase network reliability and reduce customer churn

[Blog](#) - [Notebooks](#)



Sales Forecasting and Ad Attribution

Leverage offline and alternative data to understand advertising effectiveness

[Blog](#) - [Notebook](#)



Recommendation Engine

Create a personalized experience for your customers to drive engagement and monetization

[Blog](#) - [Notebooks](#)



Behavioral Segmentation

Advanced segmentation to target the right people with the right messages and advertising

[Blog](#) - [Notebooks](#)



Telco Churn Prediction

Predict which telco customers are likely to churn and empower teams to take action

[Blog](#) - [Notebooks](#)

Value Potential

Leveraging Databricks expertise **saves cost** and can **accelerate time to market** by up to **6 months**

Training large language models is difficult; most teams typically take multiple training runs and iterations to successfully converge a large language model

Training Model Runs

Impact


databricks
*Fewer, lower
cost runs*



- **Model training expertise** → no unnecessary runs
 - Efficient ML software stack / configuration,
 - Hyperparameters / data filtering / mixture techniques
 - instruction fine tuning strategies / recommendations
- **Lower cost** to get started
- **Average time to market <6mo**

Others
*Several
expensive
runs*



- Lack of expertise → **unnecessary runs**
- Lengthy time to market → **Longer debug cycles and less experimentation**
- **Average time to market >12mo**

Training an LLM with Databricks requires **83% fewer GPU hours**, resulting in **52% cost savings**

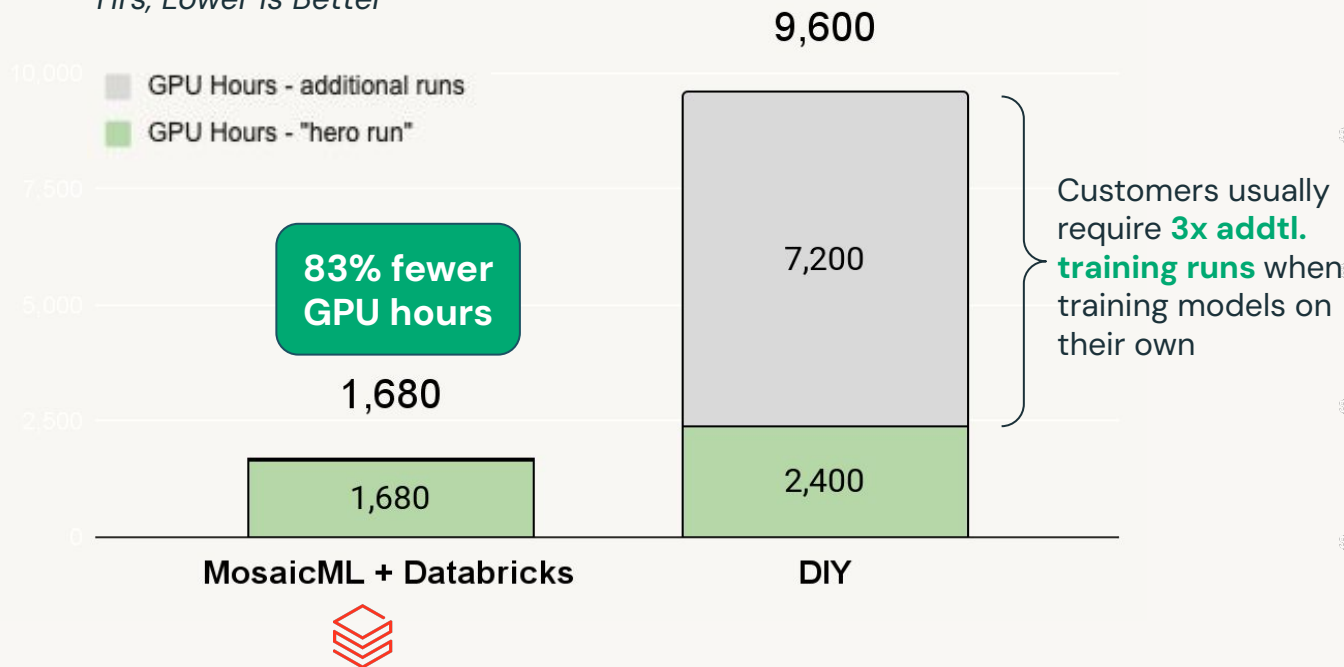
1

2

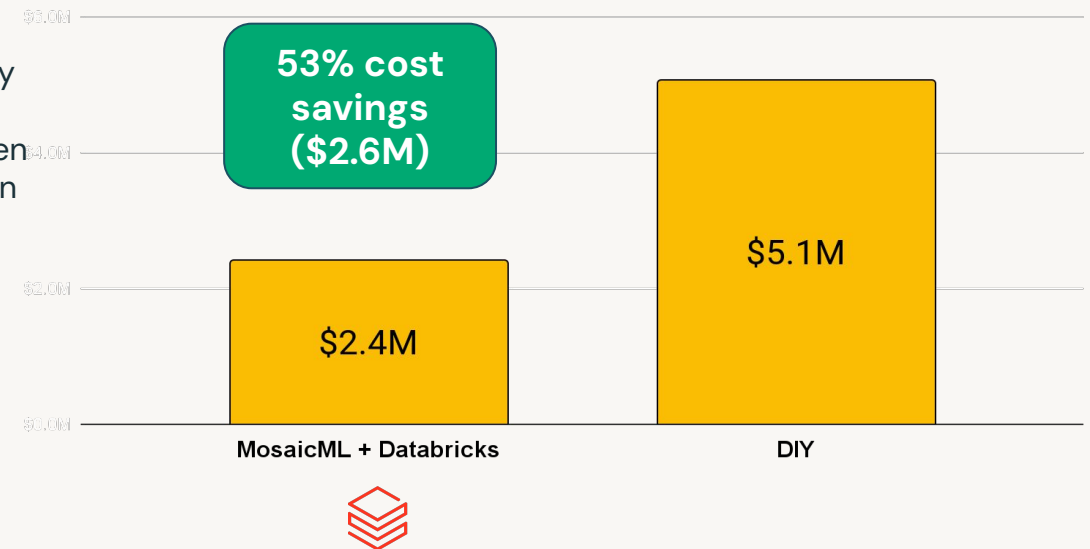
Fewer training runs and higher GPU utilization...

...also translates to cost savings

GPU hours required for model training
Hrs, Lower is Better



Model training GPU costs compared to DIY
Est. based on on-demand pricing¹



- Databricks is more **efficient** at processing GPUs, and requires **fewer training runs** compared to DIY, resulting in **~83% fewer GPU hours** for model training
- This translates to **52% cost savings (\$2.6M)**, in addition to **faster time to market**

1: Spot prices for DIY taken from [Lambdalabs.com](https://www.lambdalabs.com)
Source: [Databricks and Factset GenAI Value Model](#)

