Aviva x AWS

How Aviva deployed enterprise MLOps to automate machine learning and drive £3 billion in annual claims efficiency.

Aviva at a Glance

Global Reach & Heritage

- Founded 1696—one of the world's oldest **financial services** organisations

- Operating across 16 countries

- Serving 33+ million customers worldwide

Core Offerings

- Life and general insurance

- Pensions and investment services

- **70+ active AI and ML** use cases in production

Business Problems

Time-Consuming Operational Tasks

Data scientists spend over **50% of their time** on operational tasks, diverting focus from innovative model development.

High On-Premise Infrastructure Costs

Inefficient resource utilization and **high maintenance expenses**, driving significant fixed costs.

Slow, Error-Prone Deployment

An estimated **50% of ML projects never reach production**, indicating significant deployment hurdles and wasted investment.

Massive Claims Volume & Cost

Aviva processes **400,000** annual claims, amounting to **3 billion EUR**, demanding enterprise-grade automation and governance for optimized workflows.

The solution:

AWS Enterprise MLOps Framework

Implement a modern, **cloud-native MLOps architecture** built on **four building blocks** designed for enterprise scale, security, and governance.

4 Building Blocks

1. Network Infrastructure

- 3 isolated AWS accounts: Development, Staging and Production
- new network infrastructure set up across all environments via VPCs and endpoints; zero-trust security and compliance built in.

2. Amazon SageMaker Studio

- Unified workspace for ML development, training, and deployment
- native collaboration, versioning, and monitoring.

3. Amazon Project Templates

- Automated ML pipelines enforcing best practices, CI/CD, security across environments for consistency at scale.

4. Seed Code

- Production-ready scaffolding, preprocessing scripts, and baseline pipeline code
- accelerating experimentation and standardising patterns

The Remedy Use Case:
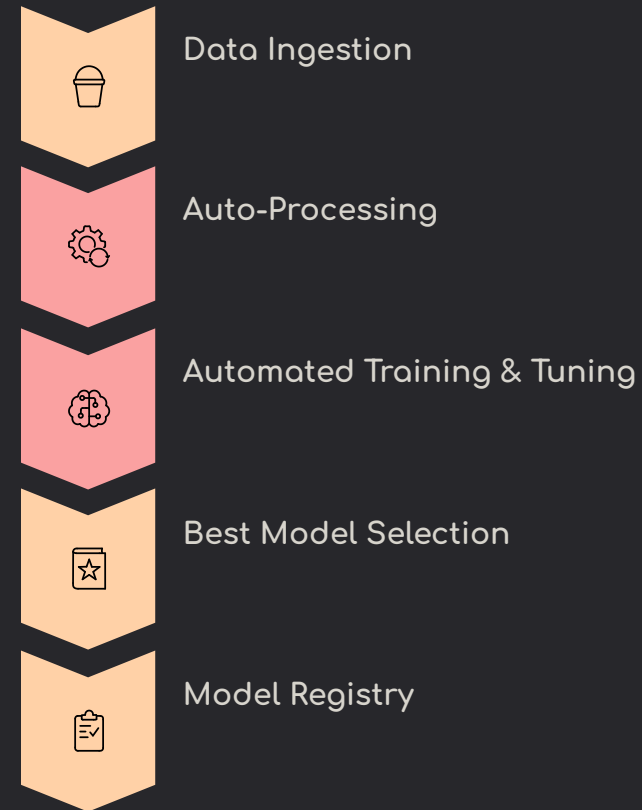## Complex Orchestration of 14 ML Models

- The Remedy process involves deciding **whether to repair or write off a damaged vehicle** — a **key decision** in Aviva's operations.

- Managing **14 different machine learning models** and business rules on outdated legacy systems caused **inefficiencies**.
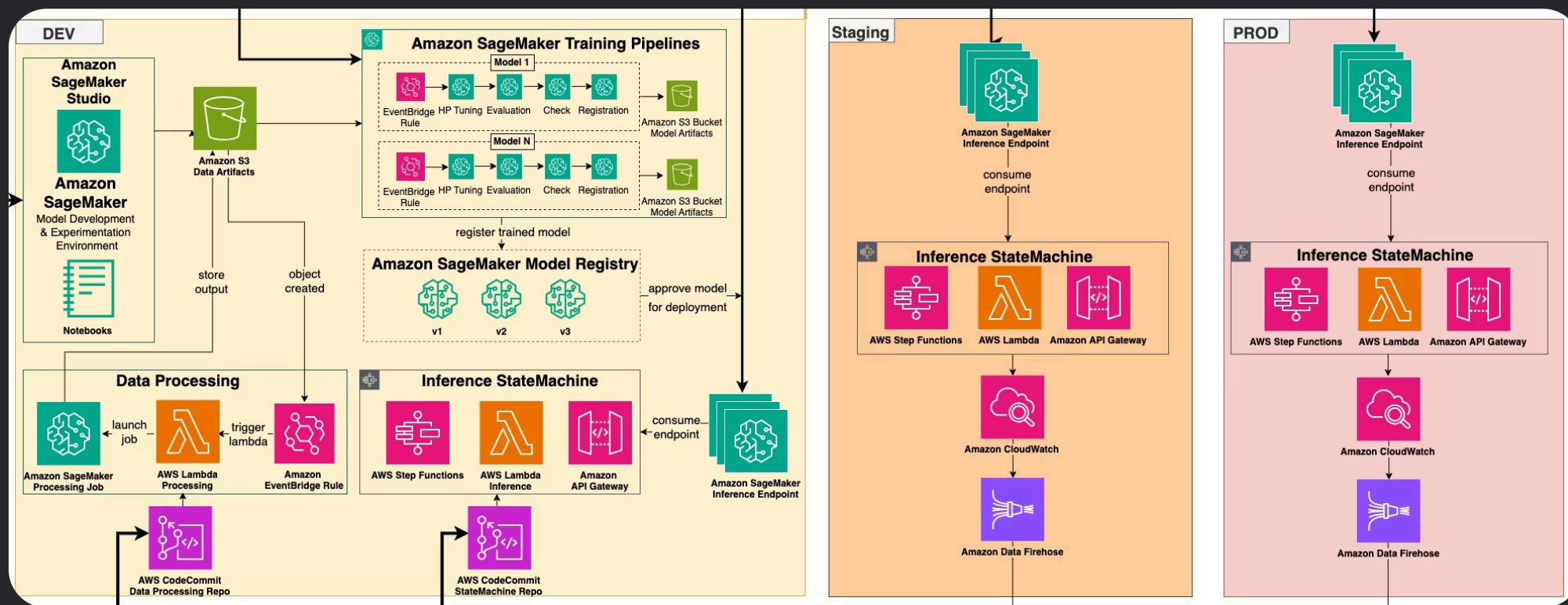
Transforming Model Development

Manual Development

Data Ingestion

Training & Tuning

Best Model Selection

Model Registry
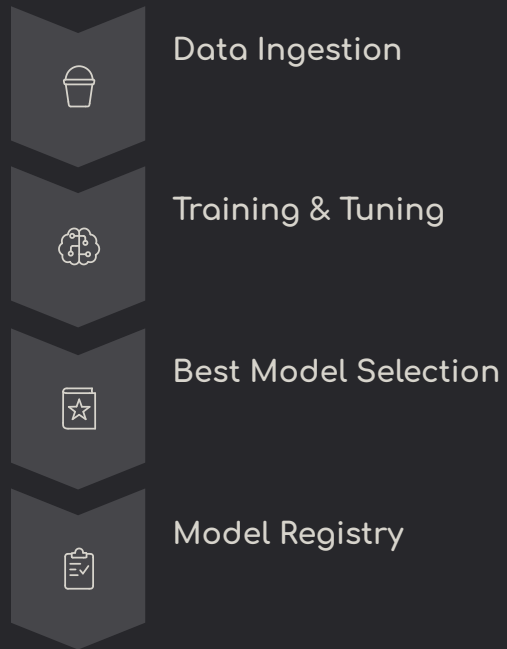
Automated Development Pipeline

Data Ingestion

Auto-Processing

Automated Training & Tuning

Best Model Selection

Model Registry

Made with GAMMA

Overview MLOps Architecture
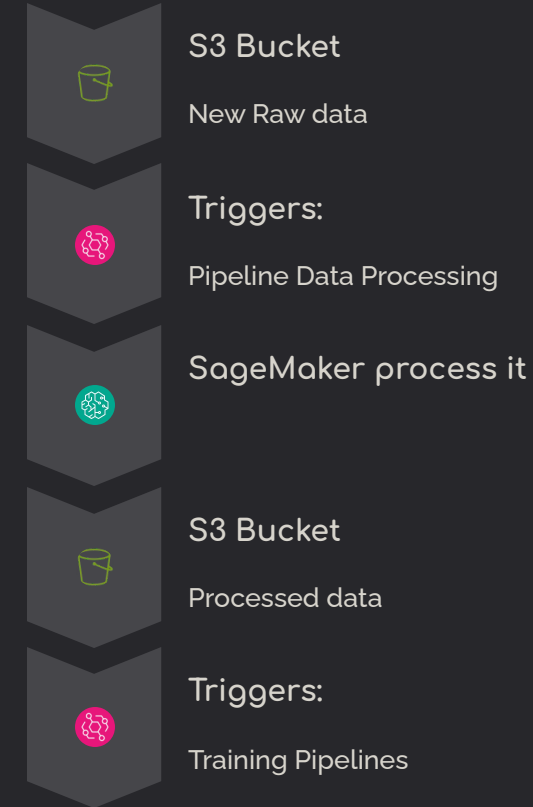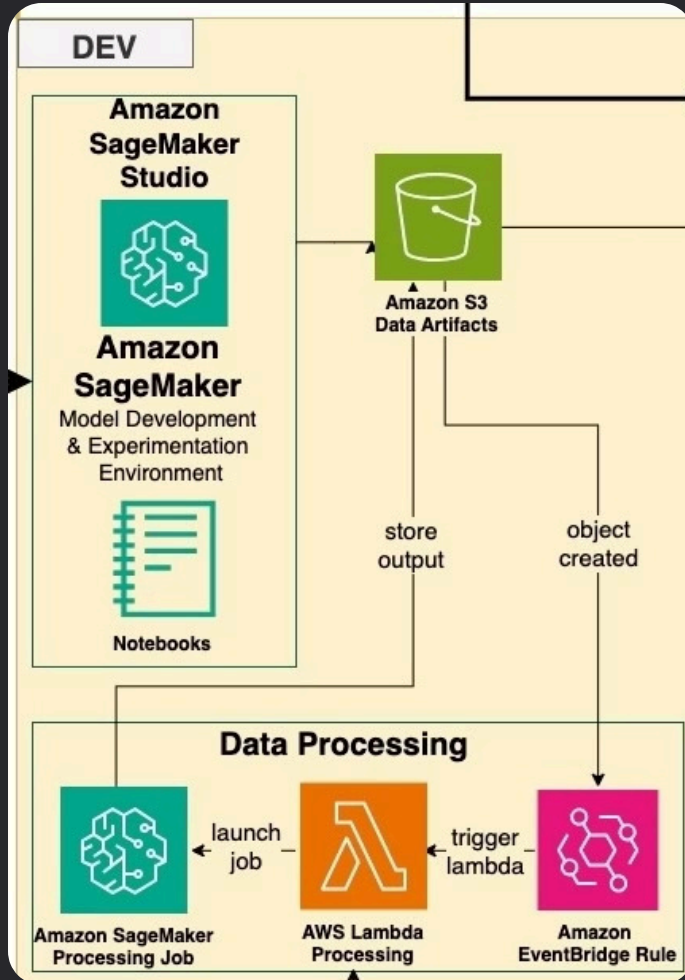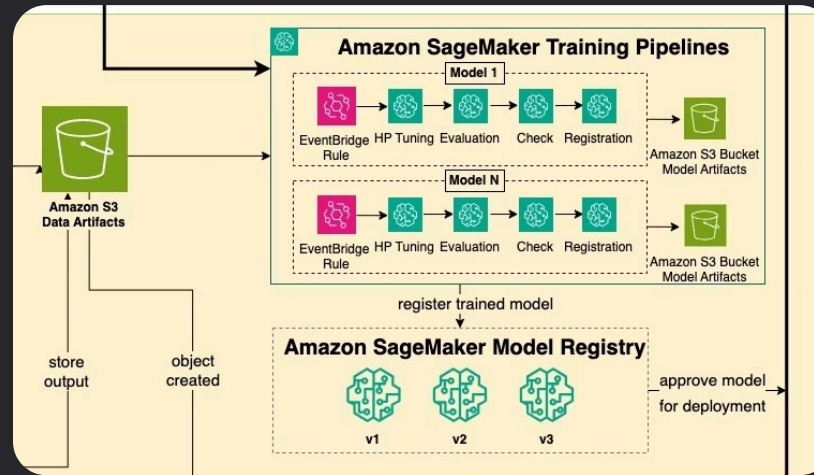
Data ingestion & Auto-Processing Pipeline
Development Account

Automated Training & Tuning Pipeline
Development Account

Amazon SageMaker Training Pipelines

Model 1: EventBridge Rule → HP Tuning → Evaluation → Check → Registration → Amazon S3 Bucket Model Artifacts

Model N: EventBridge Rule → HP Tuning → Evaluation → Check → Registration → Amazon S3 Bucket Model Artifacts

Amazon S3 Data Artifacts

store output | object created

register trained model

Amazon SageMaker Model Registry
v1 | v2 | v3

approve model for deployment

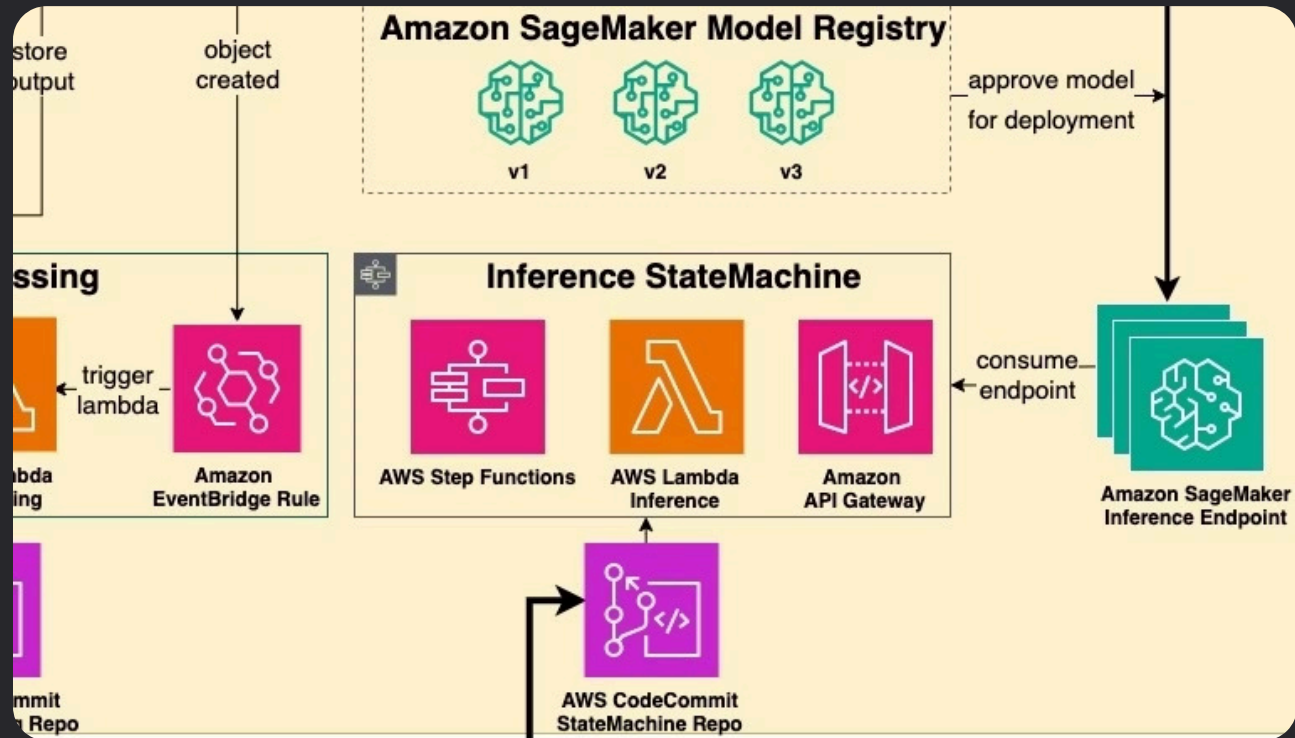| Training N Pipelines in parallel	Train + Hyper tuned	Evaluation	Registration

Model artifacts

Model Registry

Made with GAMMA

Inference Pipeline
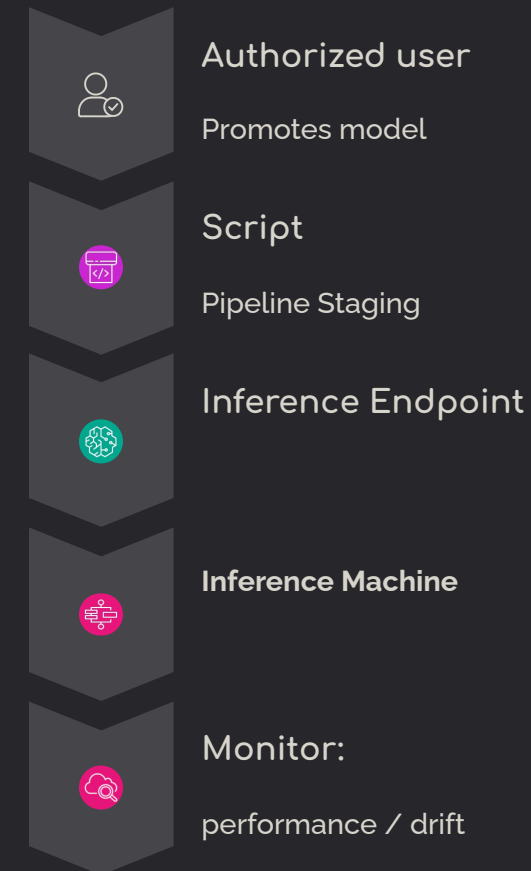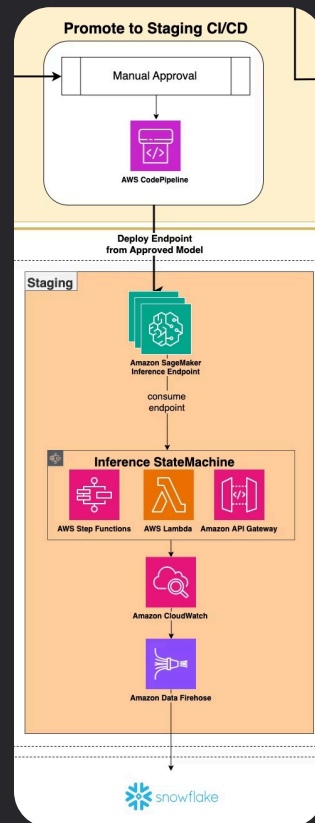## Development Account

Best model: approved   Inference Endpoint   Inference Machine

MLOps Architecture: *Staging*

Model approval and inference (i.e. beta testing)

Authorized user

Promotes model

Script

Pipeline Staging

Inference Endpoint
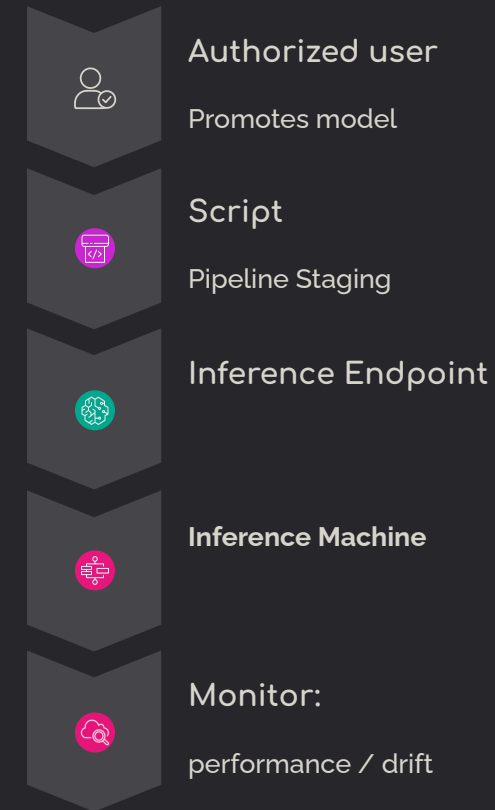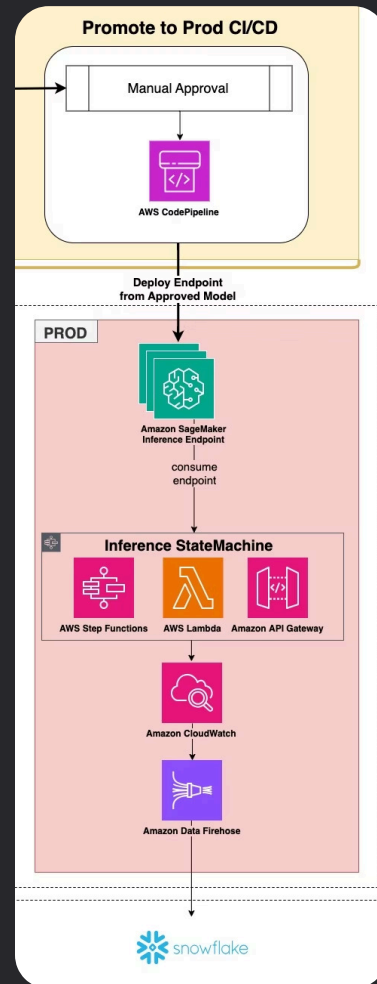
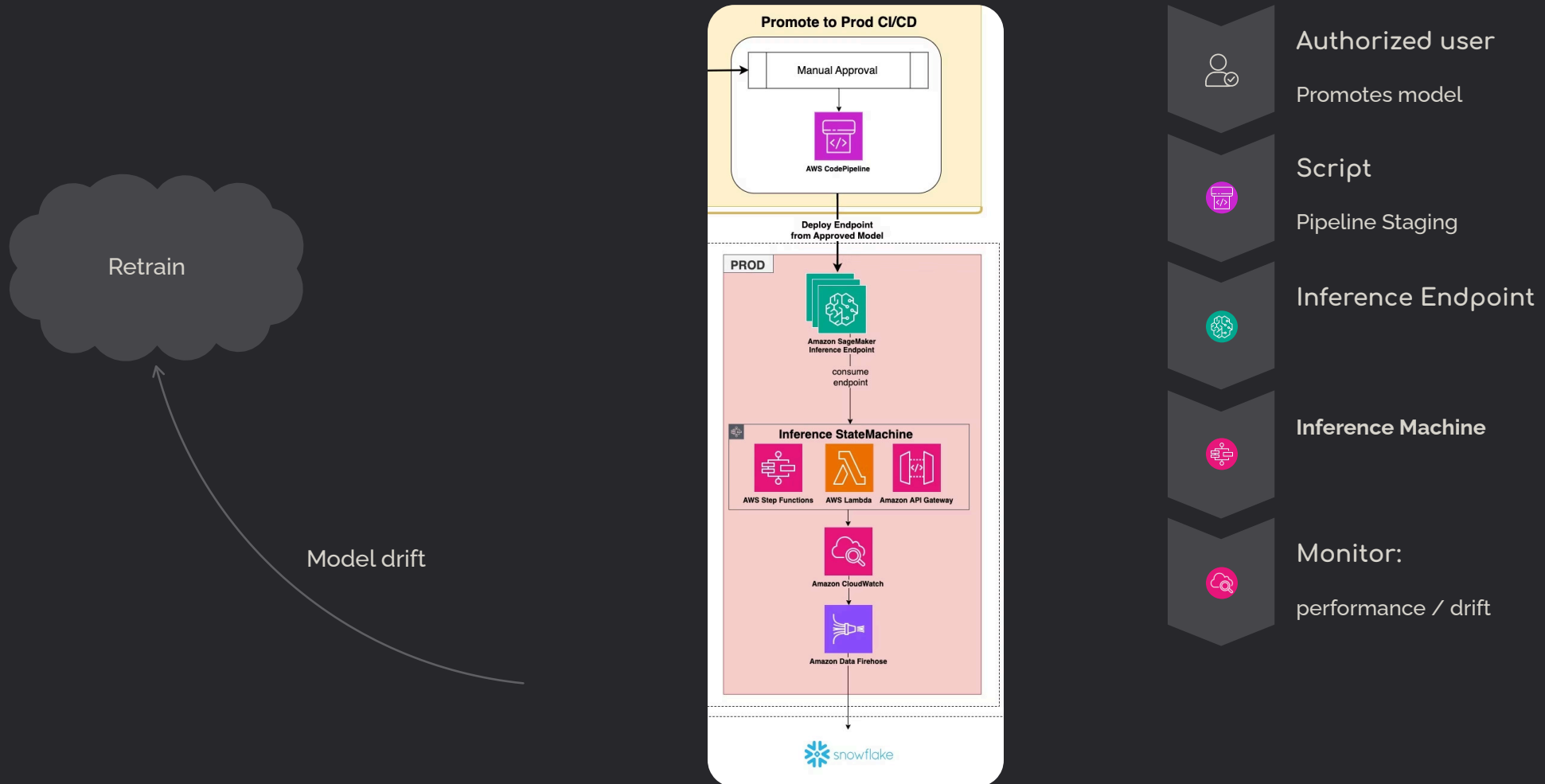**Inference Machine**

Monitor:

performance / drift

MLOps Architecture: *Production*

Model approval and inference deployment

MLOps Architecture: *Production*

Model approval and inference deployment

Retrain

Model drift

Promote to Prod CI/CD

Manual Approval

AWS CodePipeline

Deploy Endpoint
from Approved Model

PROD

Amazon SageMaker
Inference Endpoint

consume
endpoint

Inference StateMachine

AWS Step Functions   AWS Lambda   Amazon API Gateway

Amazon CloudWatch

Amazon Data Firehose

snowflake

Authorized user

Promotes model

Script

Pipeline Staging

Inference Endpoint

Inference Machine

Monitor:

performance / drift

Made with GAMMA

MLOps Architecture: Automated Pipeline

DEV

Amazon SageMaker Studio

Amazon SageMaker — Model Development & Experimentation Environment

Notebooks

Amazon S3 Data Artifacts

Amazon SageMaker Training Pipelines

Model 1: EventBridge Rule → HP Tuning → Evaluation → Check → Registration → Amazon S3 Bucket Model Artifacts

Model N: EventBridge Rule → HP Tuning → Evaluation → Check → Registration → Amazon S3 Bucket Model Artifacts

register trained model

Amazon SageMaker Model Registry — v1, v2, v3

store output / object created

approve model for deployment

Data Processing

Amazon SageMaker Processing Job ← launch job ← AWS Lambda Processing ← trigger lambda ← Amazon EventBridge Rule

AWS CodeCommit Data Processing Repo

Inference StateMachine

AWS Step Functions, AWS Lambda Inference, Amazon API Gateway

AWS CodeCommit StateMachine Repo

consume endpoint

Amazon SageMaker Inference Endpoint

Staging

Amazon SageMaker Inference Endpoint

consume endpoint

Inference StateMachine

AWS Step Functions, AWS Lambda, Amazon API Gateway

Amazon CloudWatch

Amazon Data Firehose

PROD

Amazon SageMaker Inference Endpoint

consume endpoint

Inference StateMachine

AWS Step Functions, AWS Lambda, Amazon API Gateway

Amazon CloudWatch

Amazon Data Firehose

Development   Staging   Production

Making AI Actionable for Claims Handlers

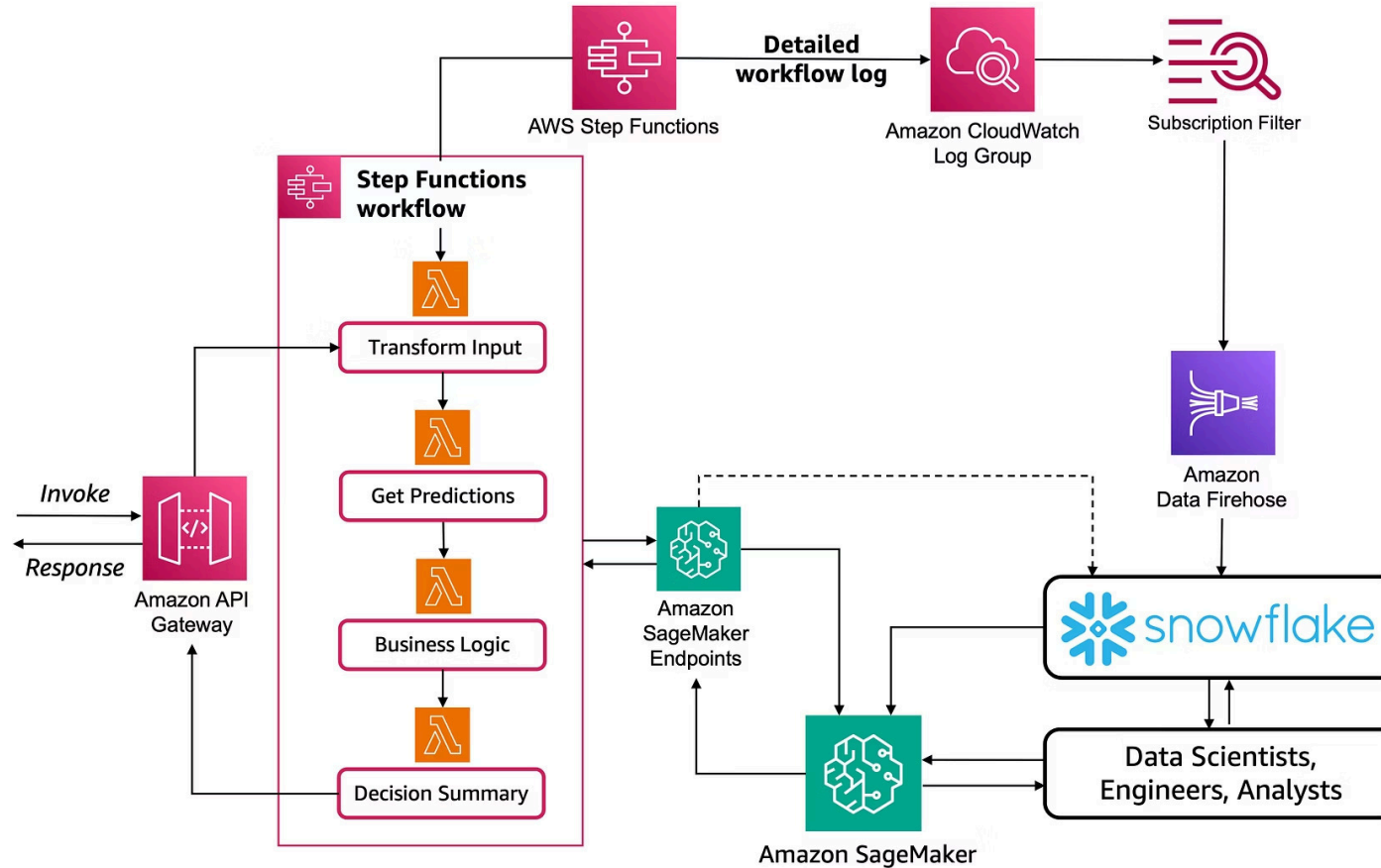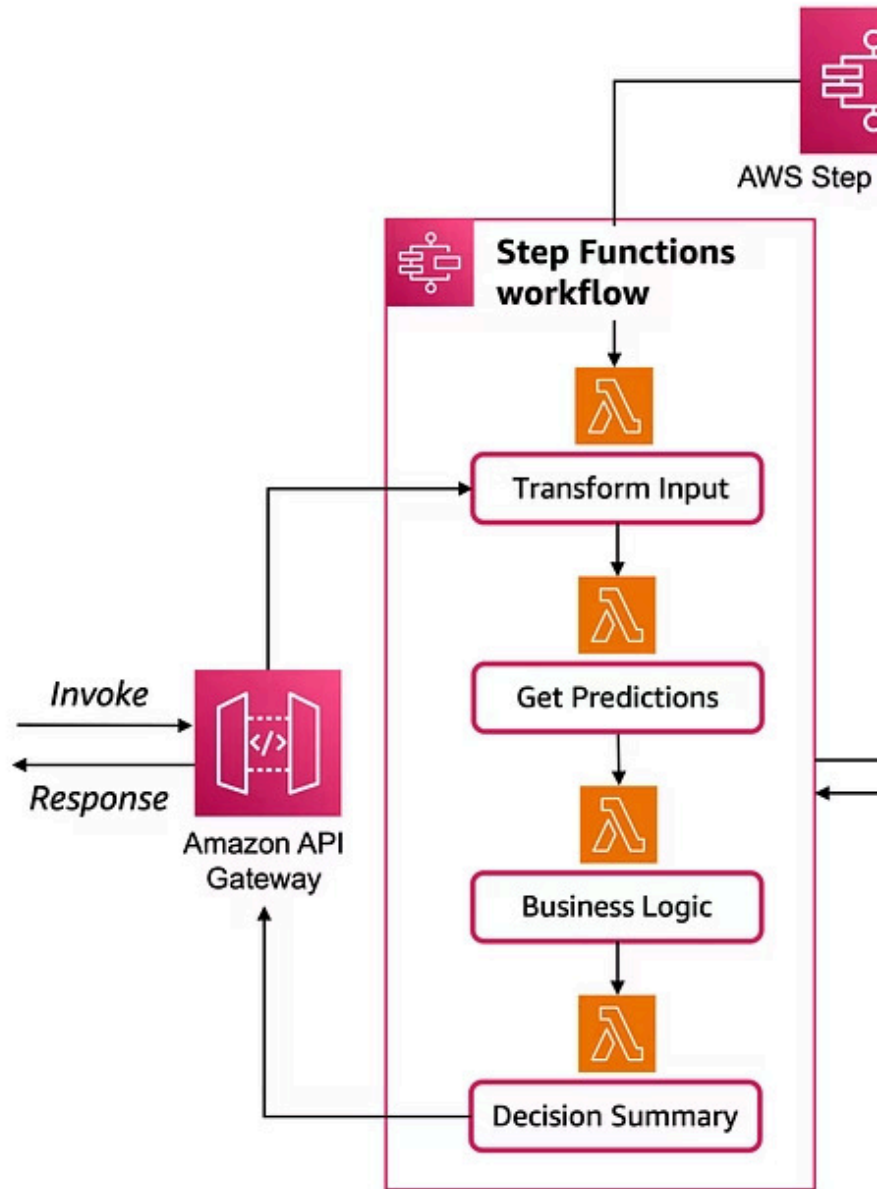**Car Insurance Claim**

Write Off
Remove the item from records.
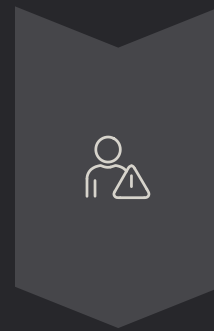
SEEK REPAIR
Vehicle value > Repair cost

Investigate Further
Examine to determine the issue.

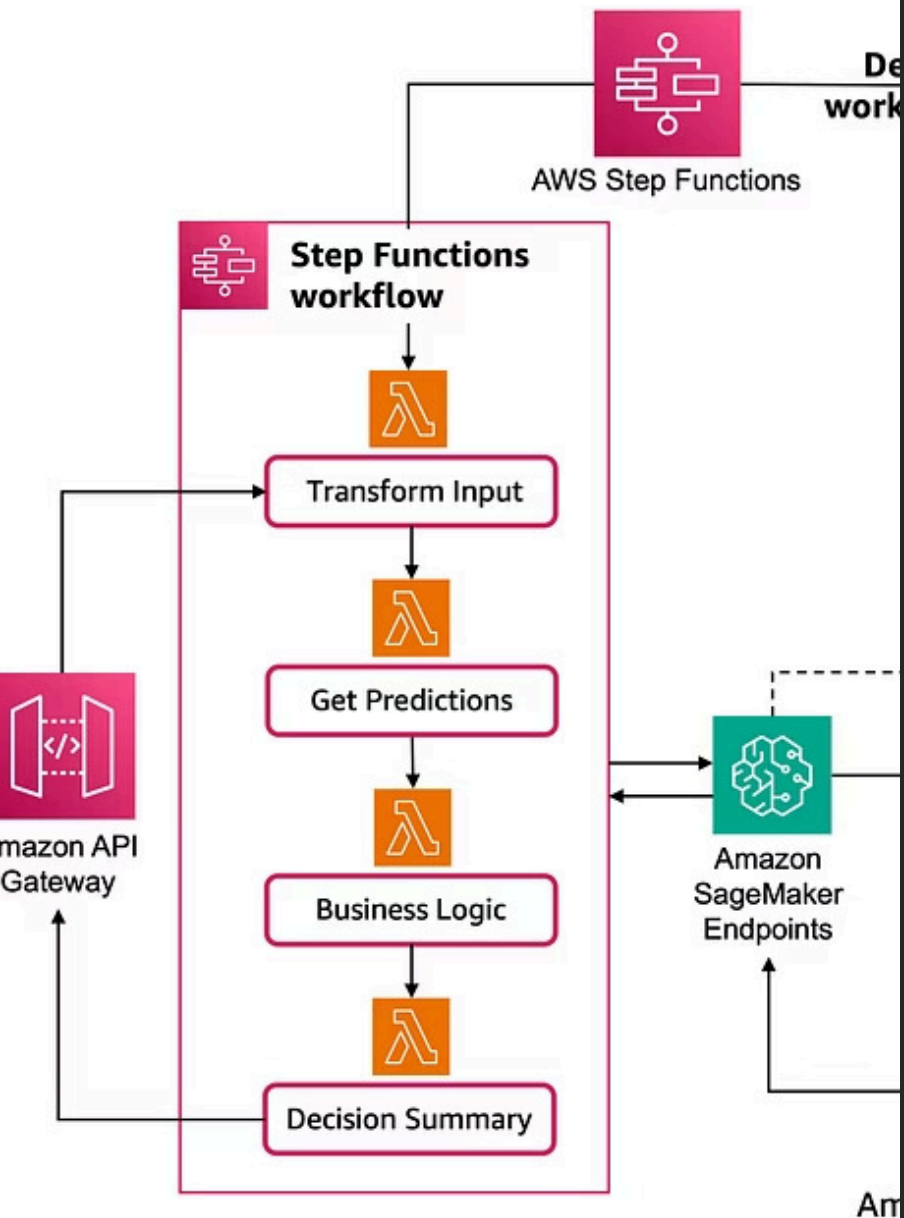# Production: Real-time Inference Pipeline

Production: Real-time Inference Pipeline

CLAIM INITIATION
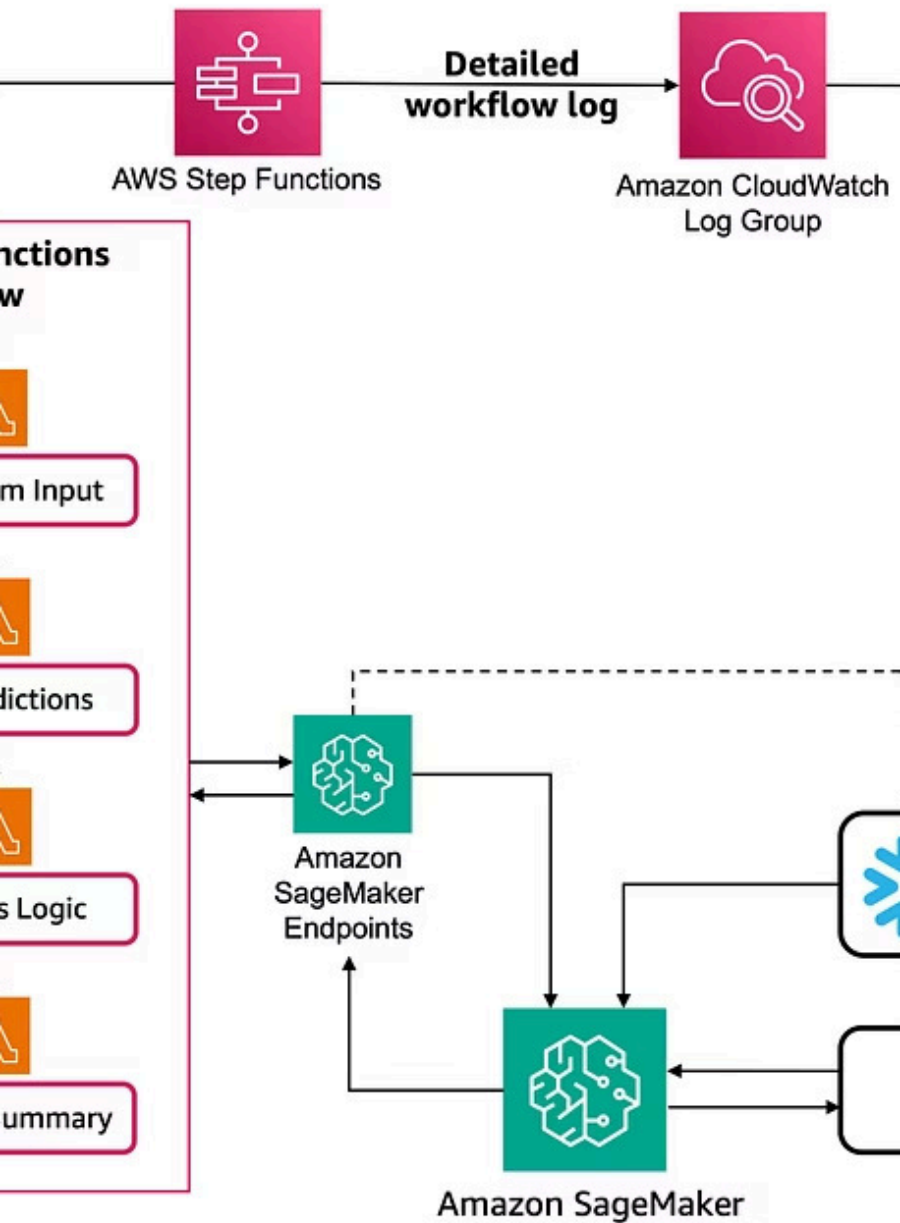
Claims Handler submits Request via API

Production: Real-time Inference Pipeline

CLAIM INITIATION

Claims Handler submits Request via API

DATA TRANSFORMATION

Feature Encoding

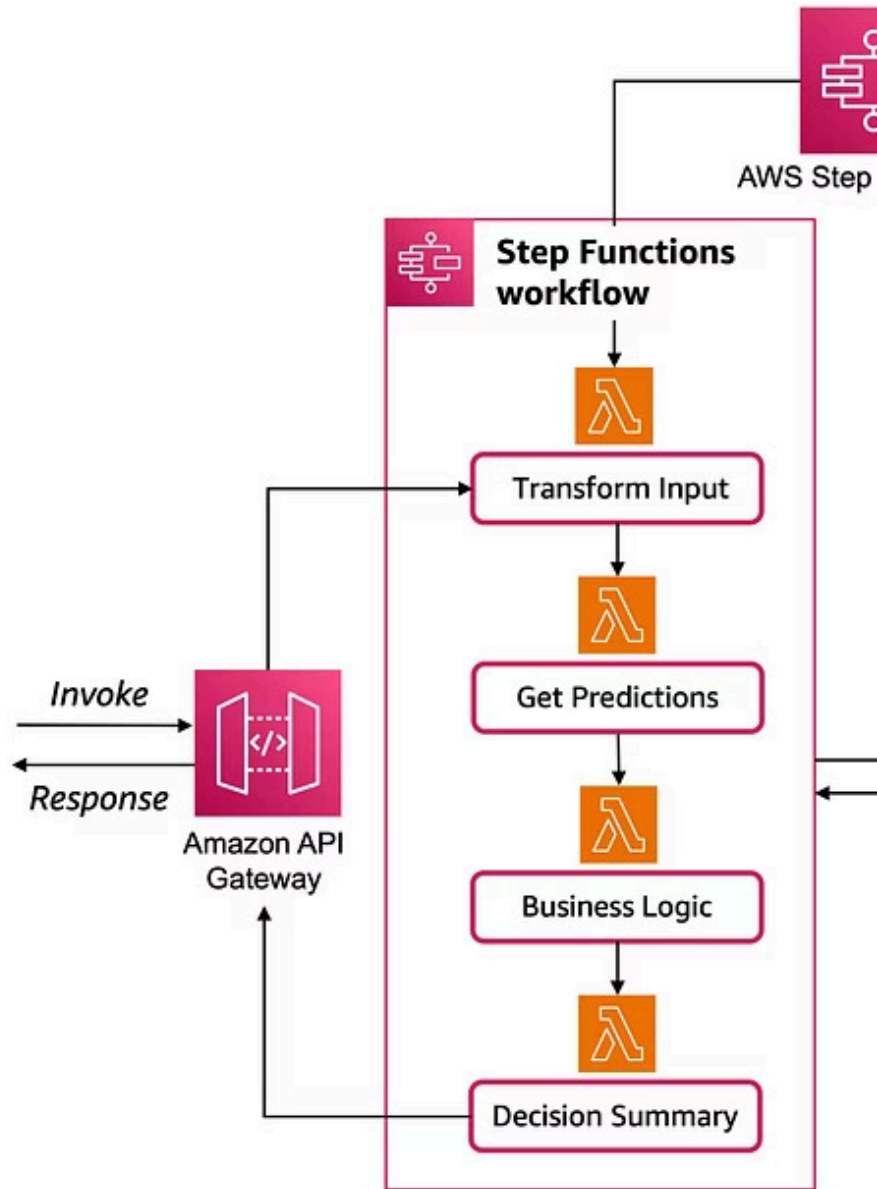Made with GAMMA

Production: Real-time Inference Pipeline

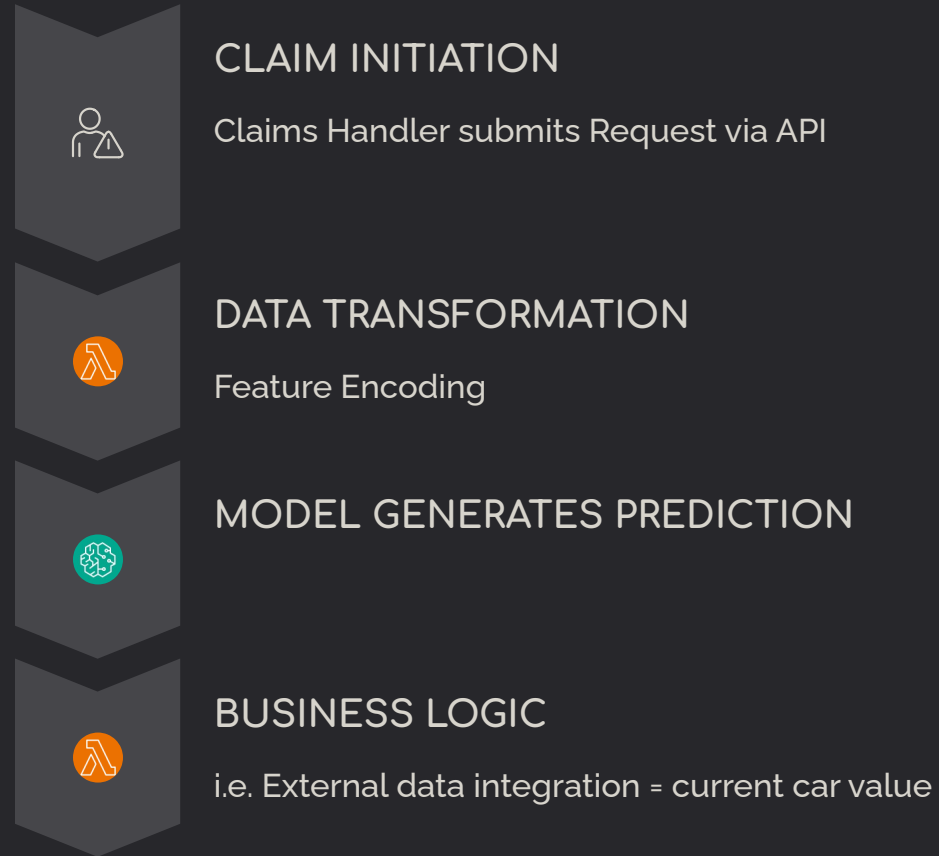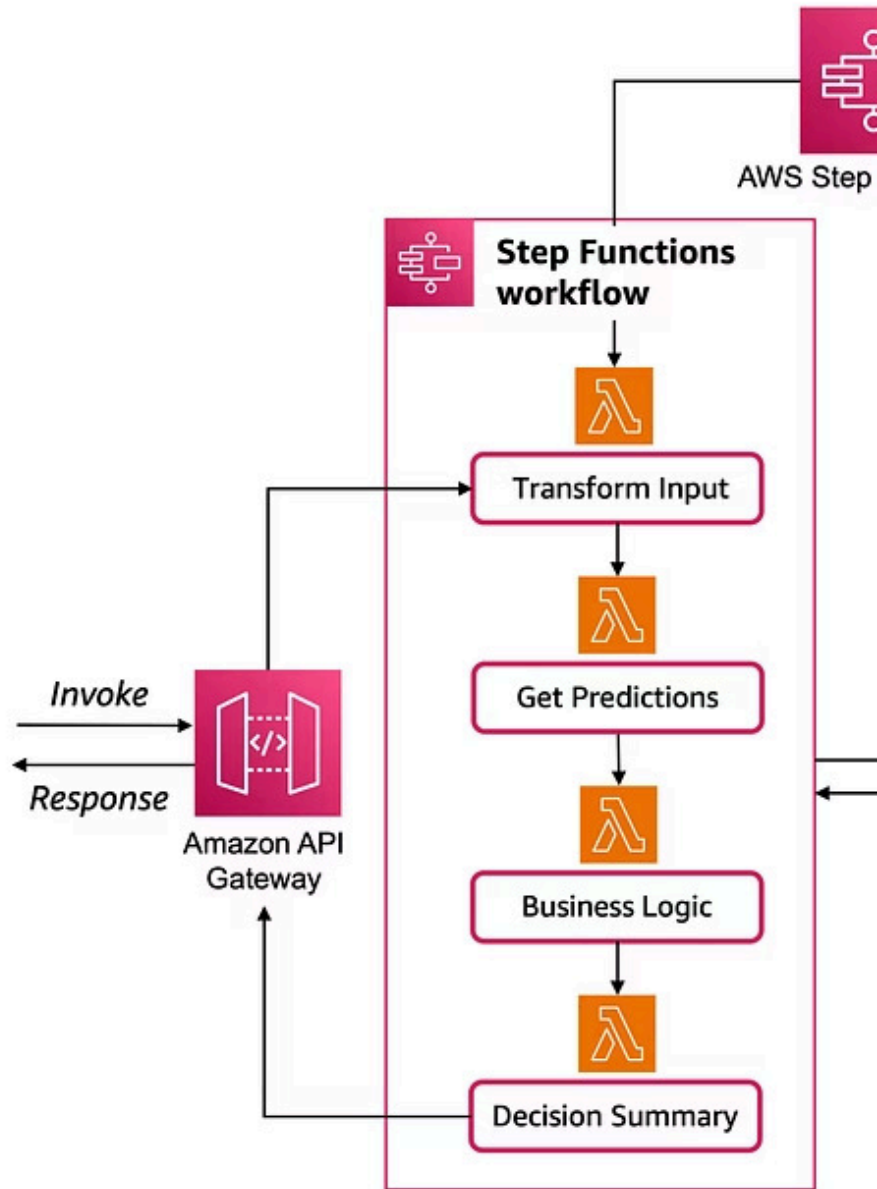**CLAIM INITIATION**

Claims Handler submits Request via API

DATA TRANSFORMATION

Feature Encoding

MODEL GENERATES PREDICTION

Made with GAMMA

Production: Real-time Inference Pipeline

CLAIM INITIATION

Claims Handler submits Request via API

DATA TRANSFORMATION

Feature Encoding

MODEL GENERATES PREDICTION

BUSINESS LOGIC

i.e. External data integration = current car value

Made with GAMMA

Production: Real-time Inference Pipeline

CLAIM INITIATION

Claims Handler submits Request via API

DATA TRANSFORMATION

Feature Encoding

MODEL GENERATES PREDICTION

BUSINESS LOGIC
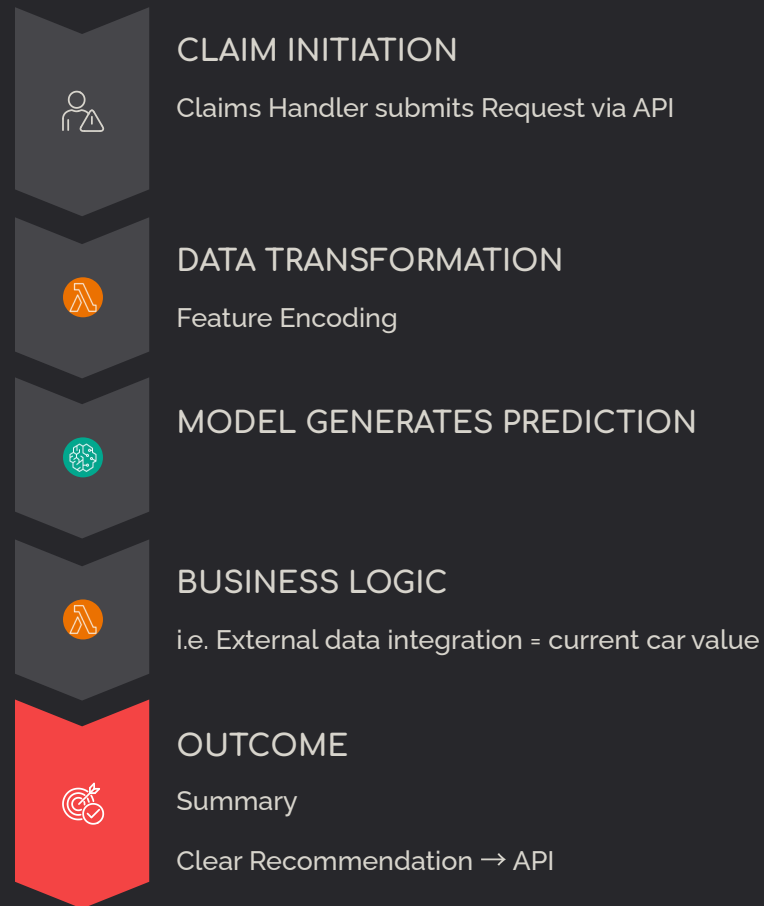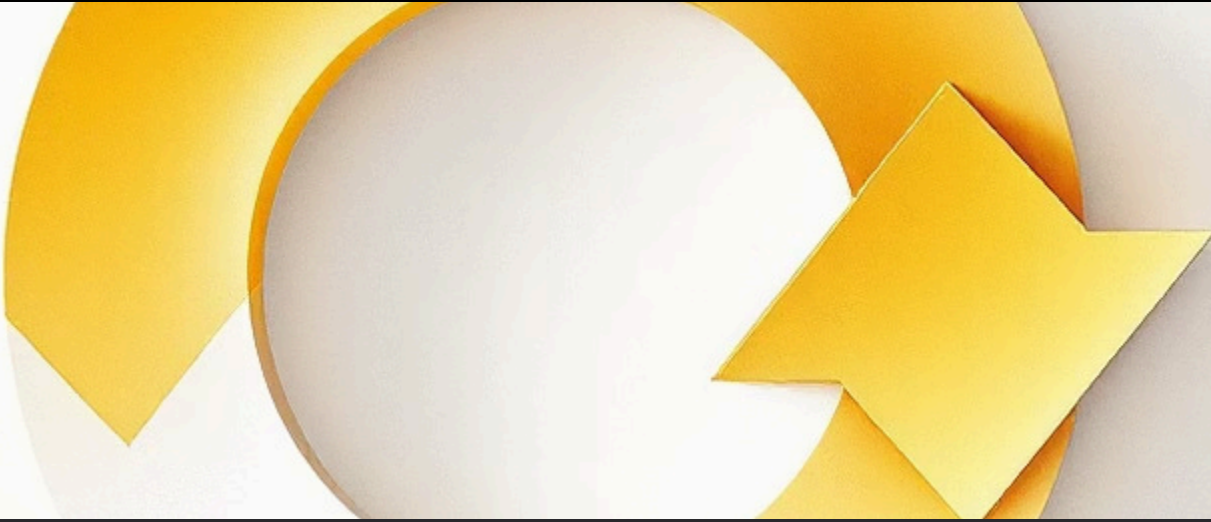
i.e. External data integration = current car value

OUTCOME

Summary

Clear Recommendation → API

Made with GAMMA

Full Automation, Maximum Efficiency

- Pipelines Trigger Pipelines

- Human-in-the-Loop for Governance

- Continuous Monitoring Triggers Retraining

Business Impact

1 MLOps Architecture Drives Innovation

Enterprise-grade infrastructure—network isolation, unified workspaces, automated pipelines, and seed code—shifts focus from operations to innovation.

2 Cost and Speed Gains Are Real

90% infrastructure cost reduction, **100x faster deployment cycles**, and **50% reduction in manual operational work** demonstrate tangible business value.

3 Governance at Scale Enables Growth

Standardised templates, centralised model governance, and reproducible workflows allow Aviva to confidently **scale from 70 active use cases to hundreds across the enterprise**.

Aviva MLOps Case Study Team

Guillermo Fiallo Montero   Fee Pieper