



## LAKIREDDY BALI REDDY COLLEGE OF ENGINEERING

(An Autonomous Institution since 2010)

Accredited by NAAC with 'A' Grade & NBA (Under Tier - I),

An ISO 21001:2018, 14001:2015, 50001:2018 Certified Institution

Approved by AICTE, New Delhi and Affiliated to JNTUK, Kakinada

L.B. REDDY NAGAR, MYLAVARAM, NTR DIST., A.P.-521 230.

[hodcse@lbrce.ac.in](mailto:hodcse@lbrce.ac.in), [cseoffice@lbrce.ac.in](mailto:cseoffice@lbrce.ac.in), Phone: 08659-222 933, Fax: 08659-222931

### DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

# A Case Study on Amazon SageMaker AI



Submitted by

P. Saranya Durga  
22761A05A9

Submitted to

M. Gayathri  
Assistant Professor

## **Introduction:**

- An e-commerce company faced challenges in providing accurate product recommendations to its customers.
- Their old recommendation system was slow, required manual model training, and couldn't handle large datasets.
- To solve this, the company decided to use **Amazon SageMaker**, a fully managed machine learning service.
- So by using this service we can build, train, and deploy AI models for any use case with fully managed infrastructure, tools, and workflows.

## **Why SageMaker AI?**

### **1. All-in-One ML Platform**

- Build, train, and deploy AI models in one place
- No need to manage servers or separate tools

### **2. Fast and Low-Cost**

- Uses optimized cloud resources
- Only pay for what you use
- Trains models faster and cheaper

### **3. Easy Tools for Developers**

- Ready-to-use notebooks
- Debuggers and profilers to improve models

### **4. Strong Security & Governance**

- Secure access control
- Full visibility of your ML projects
- Meets compliance and audit requirements

### **5. Supports Foundation Models (FMs)**

- Create your own large AI models
- Fine-tune and retrain big models easily

### **6. Access to Many Pretrained Models**

- Use hundreds of ready-made AI models
- Deploy them in just a few steps

### **7. Scales Automatically**

- Handles small to very large workloads
- Provides reliable, high-performance deployment.

## **Benefits of SageMaker AI :**

- Help more people innovate with AI through a choice of tools—IDEs for data scientists and no-code interface for business analysts.
- Get assistance throughout the entire AI model development journey from data preparation and model training, to model deployment.
- Fully managed, scalable infrastructure.
- SageMaker AI helps **AI model developers, data scientists, and ML engineers** build FMs from scratch, evaluate and customize FMs with advanced techniques, and deploy FMs with fine-grained controls for AI use cases that have stringent requirements on accuracy, latency, and cost.

## **Working of Amazon SageMaker AI:**

### **1. Data Preparation:**

- Data is first collected and stored in **Amazon S3**.
- SageMaker tools like **Data Wrangler** and **Notebooks** are used to clean the data, remove errors, and create useful features for training.

### **2. Model Building:**

- Developers use **SageMaker Studio** to create or choose a model.
- They can use:
  - 1) Built-in algorithms
  - 2) Popular ML frameworks (TensorFlow, PyTorch)
  - 3) Pretrained models or foundation models
- This makes model creation faster and easier.

### **3. Model Training**

- SageMaker automatically provides the required compute power.
- It trains the model at high speed, performs **hyperparameter tuning**, and uses **Debugger** to track training performance.

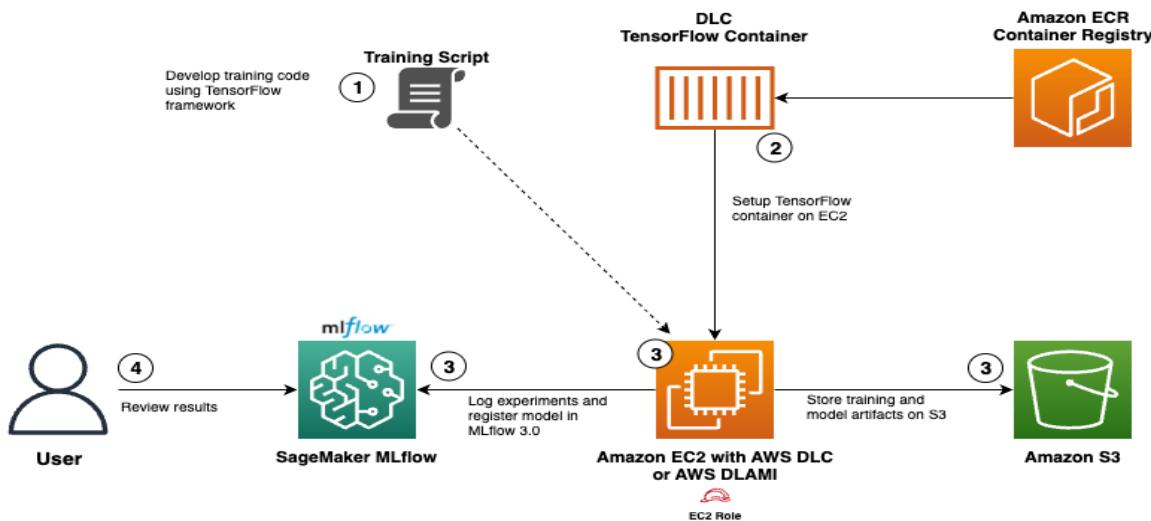
### **4. Model Deployment**

- When training is complete, the model can be deployed in a few clicks.
- SageMaker offers:
  - 1) Real-time endpoints for live predictions

- 2) Batch transform for large offline jobs
- 3) Serverless deployment for low-cost usage

## 5. Monitoring

- After deployment, SageMaker keeps monitoring the model with Model Monitor.
- It checks accuracy, data drift, and system health, ensuring the model stays reliable.



## SageMaker AI pricing and AWS Free Tier:

The SageMaker AI AWS Free Tier offers a **2-month free trial** that provides 250 hours per month of t2.medium or t3.medium notebook usage, 50 hours per month of m4.xlarge or m5.xlarge for training, and 125 hours per month of m4.xlarge or m5.xlarge for hosting.

## Unified Support for Popular AI Frameworks and Programming Languages:

- Python
- Jupyter Notebook
- Tensorflow
- PyTorch
- Hugging Face
- R Langauge
- Scikit learn

## Customers:

- **AI21 Labs** – AI21 labs accelerates generative AI model adoption with SageMaker AI
- **AT&T** – AT&T cybersecurity helps business improve threat detection and response using SageMaker AI
- **LG** – LG AI research scales foundation model development using SageMaker AI
- **Stability.ai** – Stability AI uses SageMaker AI to build generative AI

- **Workday** - Workday Accelerates Generative AI & ML Product Development Using Amazon SageMaker

### **Conclusion:**

- Amazon SageMaker AI provides a powerful and efficient platform for developing machine learning solutions at scale.
- By bringing together data preparation, model building, training, deployment, and monitoring into a single managed environment, SageMaker greatly simplifies the entire AI lifecycle.
- It reduces development time, lowers operational costs, and ensures consistent model performance through automated tools and governance features.
- Overall, SageMaker enables organizations to build smarter applications faster and supports innovation with reliable, scalable, and production-ready AI workflows.