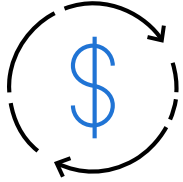AWS RESEARCH ROADSHOW — 4 APRIL 2023

Inspiration session: Innovating with AI/ML

AWS Education & Research Team

Nikiforos Botis - nbotis@amazon.com
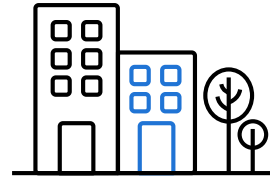
The reach of ML is growing

INCREASED SPENDING

By 2026, global spending on artificial intelligence will reach $300 billion growing 4.2 faster than average IT spend

IDC

"IDC Press Release, Worldwide Spending on AI-Centric Systems Will Pass $300 Billion by 2026, According to IDC, September 2022,"
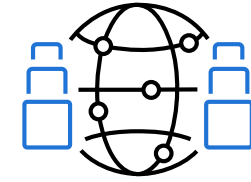https://www.idc.com/getdoc.jsp?containerId=prUS49670322

AI ACROSS SEVERAL BUSINESS UNITS

Gartner survey reveals 80% of executives think automation can be applied to any business decision

Gartner

Gartner, Press Release
https://bit.ly/3is16a2

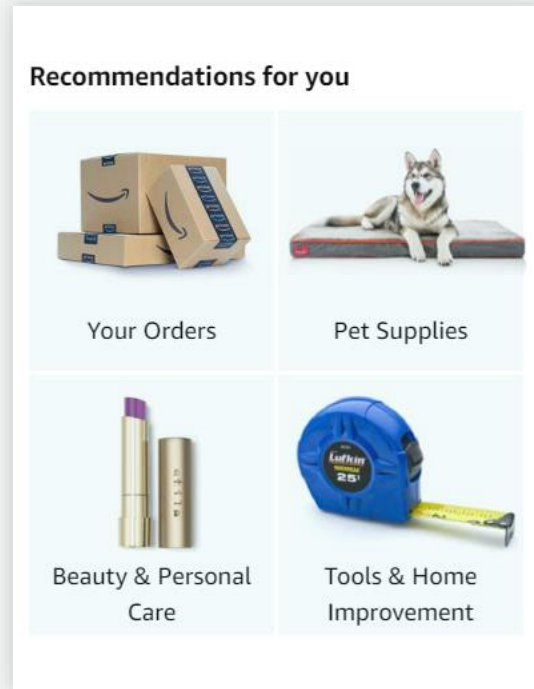## AI IS CRITICAL TO SUCCESS

94% of business leaders surveyed say AI is critical to their success

Deloitte

Deloitte, "State of AI in the Enterprise,"
https://bit.ly/3XiNOLR

Amazon machine learning innovation at scale

4,000 products per minute sold on Amazon.com
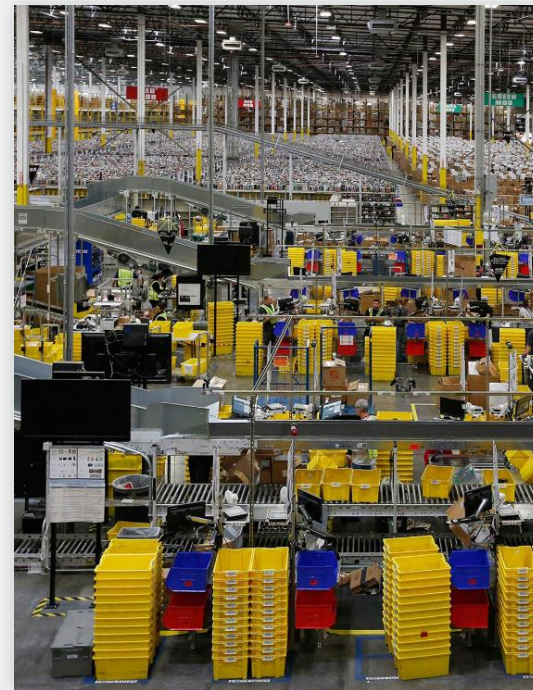
1.6 million packages every day

Billions of Alexa interactions each week

First Prime Air delivery on **December 7, 2016**

Enhance customer experience

Better and faster decision-making

Improve business operations

New products and services

Contact center intelligence	Intelligent search	Intelligent document processing	
Chatbots and virtual assistants	Forecasting	Content moderation	
Personalization	Lead scoring	Predictive Maintenance	New use cases
Identity verification	Churn prediction	Machine translation	
Fraud prevention		MLOps	
		Coding & DevOps	

Enhance customer experience | **Better and faster decision-making** | **Improve business operations** | **New products and services**

| Healthcare & life sciences	Industrial & manufacturing	Financial services	Retail	Media & entertainment
Reduce costs, improve patient outcomes, and enable the discovery and manufacturing of new drugs	Implement predictive maintenance, improve workplace safety, automate visual inspection, and optimize supply chains	Detect fraud, process documents faster, personalize recommendations, and perform credit decisioning	Personalize product recommendations, forecast demand, and predict fraud	Accelerate content creation and curation, increase customer engagement, and improve monetization
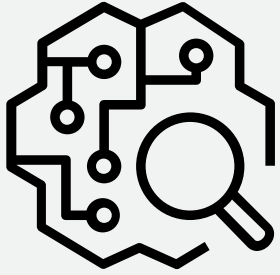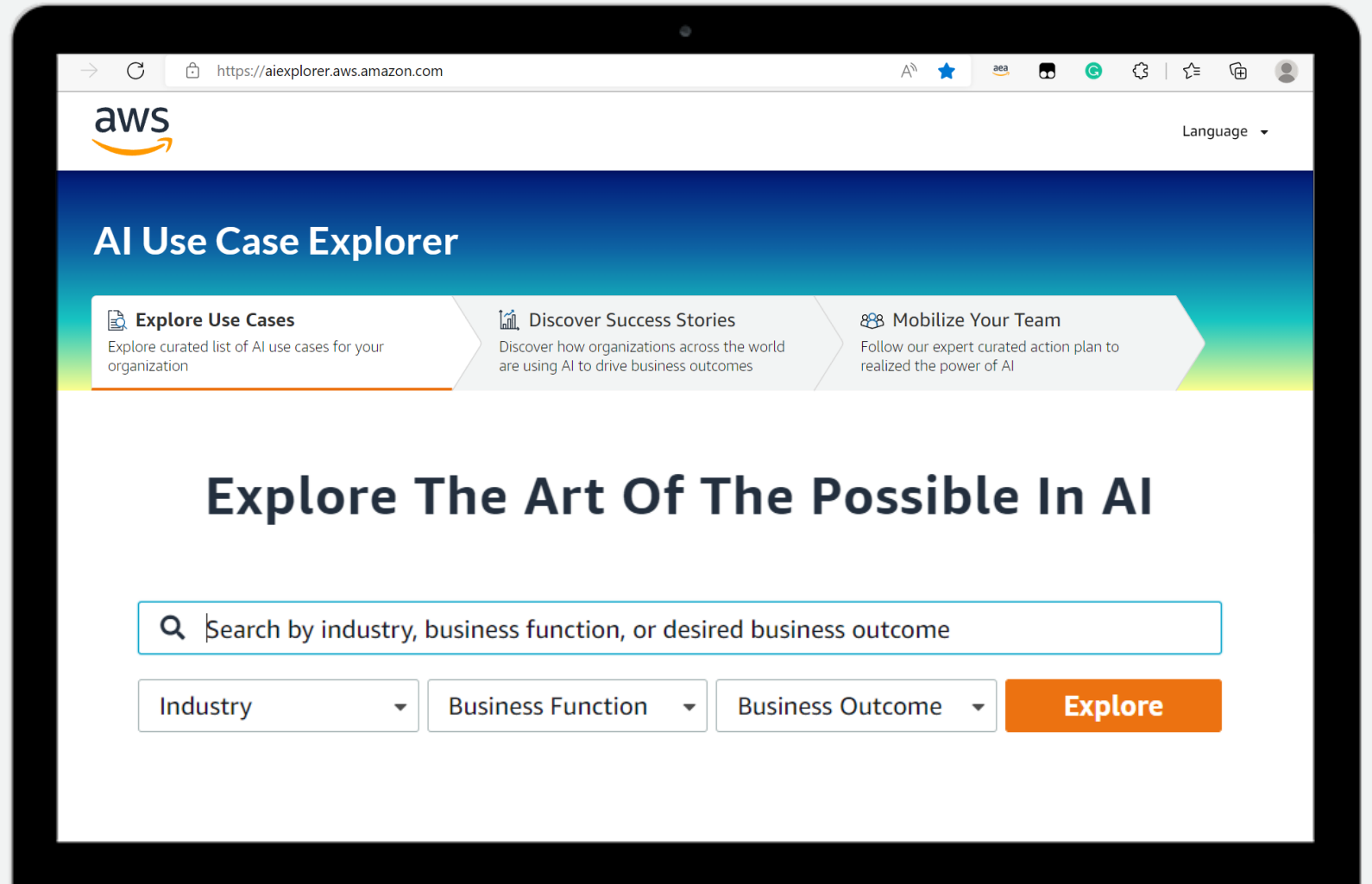
With the growth of AI comes the recognition that we must all use it **responsibly**

Our commitment to develop AI and machine learning in a responsible way is integral to our approach
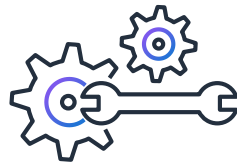
Transform responsible AI from theory to practice
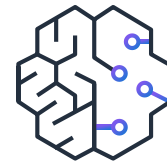
Integrate responsible AI into the end-to-end ML lifecycle

Nurture and educate a more diverse generation of leaders in ML

Advance the science behind responsible AI

The AWS AI/ML stack

BROADEST AND MOST COMPLETE SET OF MACHINE LEARNING CAPABILITIES

AI SERVICES

SPECIALIZED

BUSINESS PROCESSES
Amazon Personalize
Amazon Forecast
Amazon Fraud Detector
Amazon Lookout for Metrics

SEARCH
Amazon Kendra

CONVERSATION
Amazon Lex
Amazon Transcribe Call Analytics
Contact Lens
Voice ID

CODE + DEVOPS
Amazon CodeGuru
Amazon CodeWhisperer
Amazon DevOps Guru

INDUSTRIAL
Amazon Monitron
Amazon Lookout for Equipment
Amazon Lookout for Vision
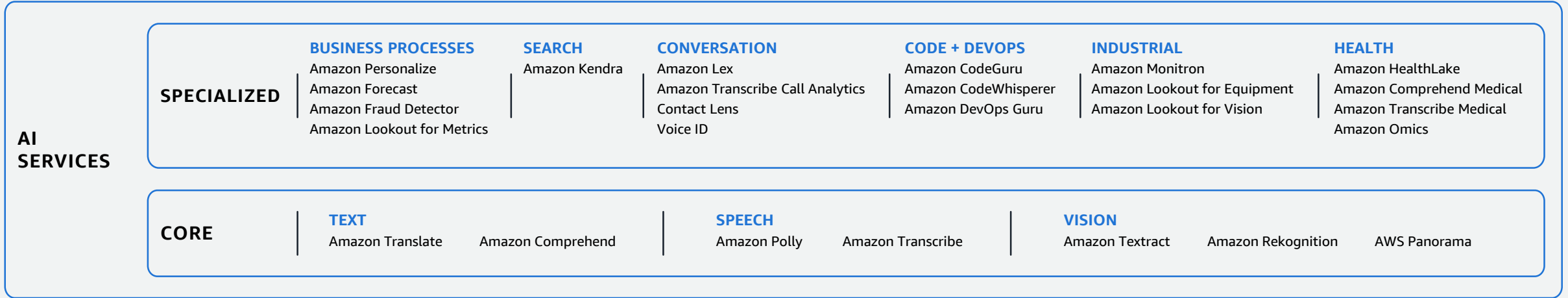
HEALTH
Amazon HealthLake
Amazon Comprehend Medical
Amazon Transcribe Medical
Amazon Omics

CORE

TEXT
Amazon Translate   Amazon Comprehend

SPEECH
Amazon Polly   Amazon Transcribe

VISION
Amazon Textract   Amazon Rekognition   AWS Panorama

AMAZON SAGEMAKER

CANVAS
No-code ML for business analysts

STUDIO LAB
Learn ML

GROUND TRUTH
Label data

STUDIO IDE

Prepare data

Store features

Geospatial ML

Build with notebooks

Train models

Tune parameters

Deploy in production

Manage and monitor

EDGE MANAGER
Manage edge devices

------- **CI/CD | GOVERNANCE | RESPONSIBLE ML** -------

ML FRAMEWORKS & INFRASTRUCTURE

PyTorch, Apache MXNet, TensorFlow | Amazon EC2 | CPUs | GPUs | AWS Inferentia | AWS Trainium | Habana Gaudi | FPGA

Inf1 instances are built from the ground up by AWS to provide high performance, cost-effective inference

https://aws.amazon.com/ec2/instance-types/inf1

AWS Nitro

AWS Inferentia

AWS Custom 2nd Gen Intel Xeon Scalable Processors

100Gbps Networking

HIGH PERFORMANCE   **LOW COST**
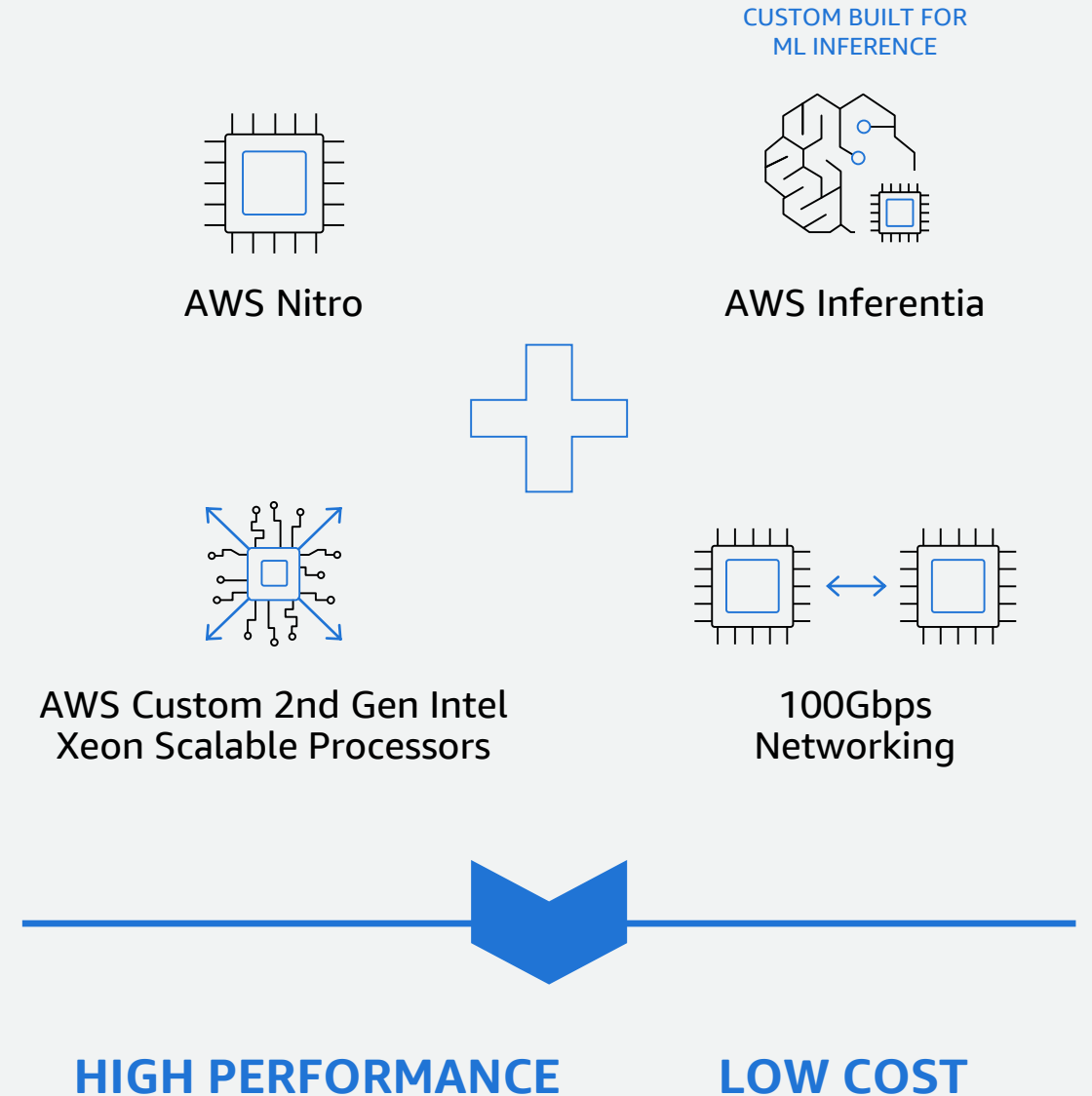
Integrated Workbench

IDE designed specifically for ML, data preparation, experiment management, and pipelines

Managed Infrastructure

Designed for ultra low latency and high throughput; automatic scaling, and distributed training
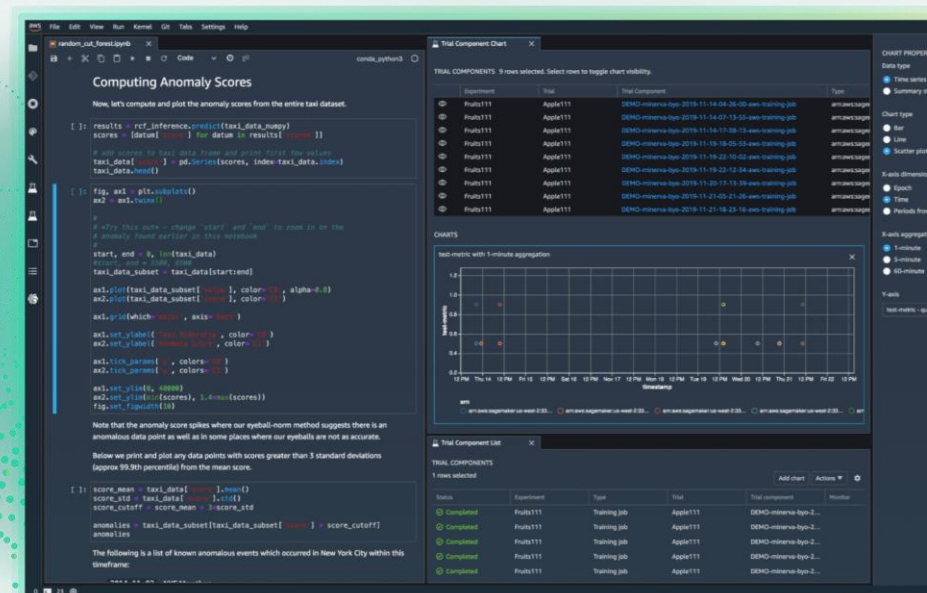
Managed Tooling

Purpose-built from the ground up to work together incl. Autopilot, collaboration, notebooks, experiments, debugger, and model monitor
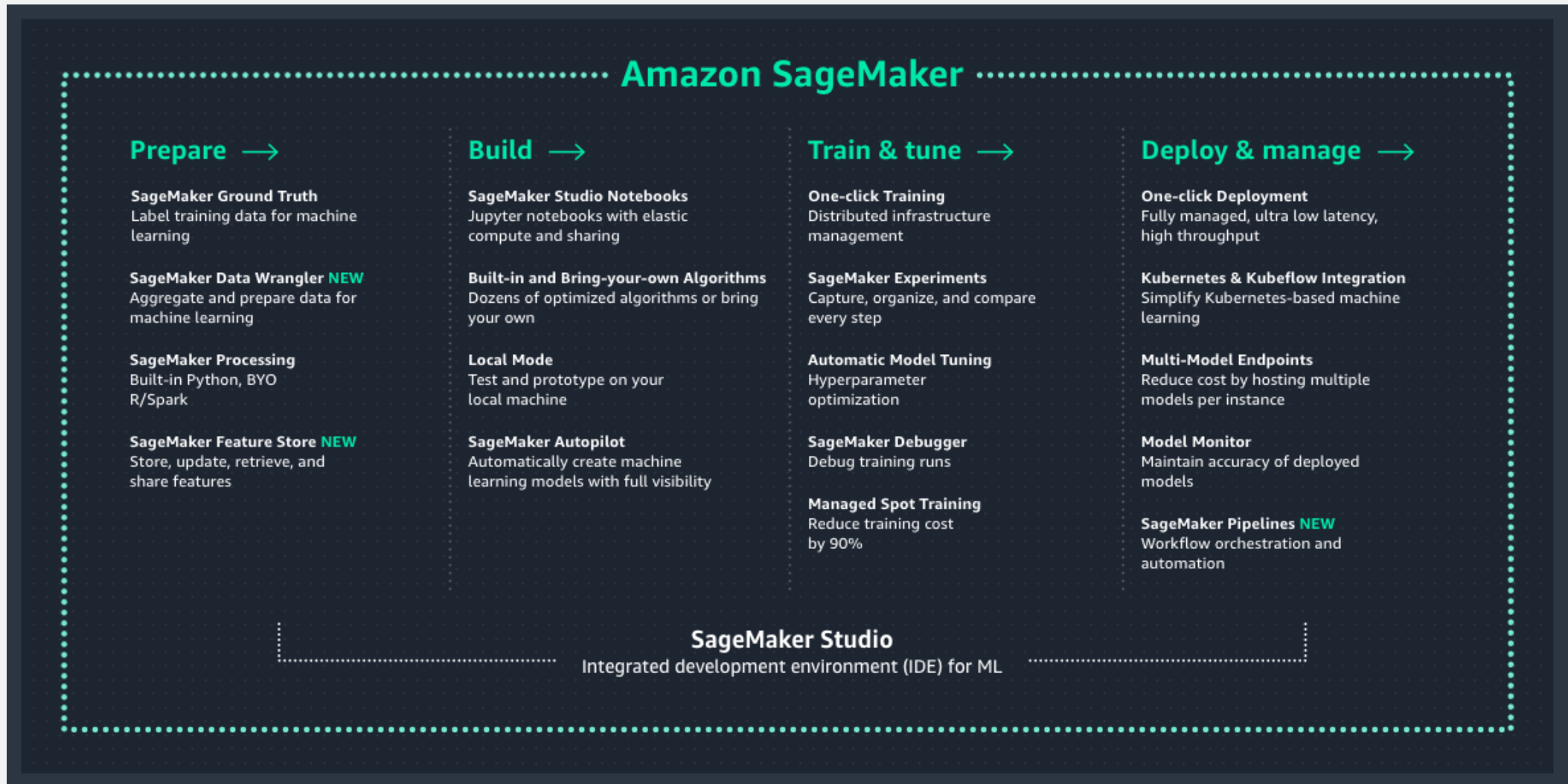
- https://aws.amazon.com/sagemaker

Amazon SageMaker

End-to-End ML service

Amazon SageMaker overview

Amazon SageMaker

Prepare →

SageMaker Ground Truth
Label training data for machine learning

SageMaker Data Wrangler NEW
Aggregate and prepare data for machine learning

SageMaker Processing
Built-in Python, BYO R/Spark

SageMaker Feature Store NEW
Store, update, retrieve, and share features

Build →

SageMaker Studio Notebooks
Jupyter notebooks with elastic compute and sharing

Built-in and Bring-your-own Algorithms
Dozens of optimized algorithms or bring your own

Local Mode
Test and prototype on your local machine

SageMaker Autopilot
Automatically create machine learning models with full visibility

Train & tune →

One-click Training
Distributed infrastructure management

SageMaker Experiments
Capture, organize, and compare every step

Automatic Model Tuning
Hyperparameter optimization

SageMaker Debugger
Debug training runs

Managed Spot Training
Reduce training cost by 90%

Deploy & manage →

One-click Deployment
Fully managed, ultra low latency, high throughput

Kubernetes & Kubeflow Integration
Simplify Kubernetes-based machine learning

Multi-Model Endpoints
Reduce cost by hosting multiple models per instance

Model Monitor
Maintain accuracy of deployed models

SageMaker Pipelines NEW
Workflow orchestration and automation

SageMaker Studio
Integrated development environment (IDE) for ML

No-code ML tools

ML requires a lot more builders

2x

the growth of any other emerging job role

+74%

annual growth in the past 4 years

aws machine learning

Business analysts require a different set of tools to use ML

PROGRAMMING TOOLS &
OPEN-SOURCE FRAMEWORKS

DATA SCIENTIST
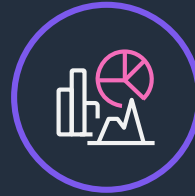
SOFTWARE ENGINEERING

DATA WRANGLING

DATA VISUALIZATION

Amazon SageMaker Canvas

**Generate ML predictions –
no code required**

Quickly access and prepare
data sources for ML

AutoML built in to generate
accurate predictions

Share ML models with data science teams

Purpose-built data preparation tools

Structured
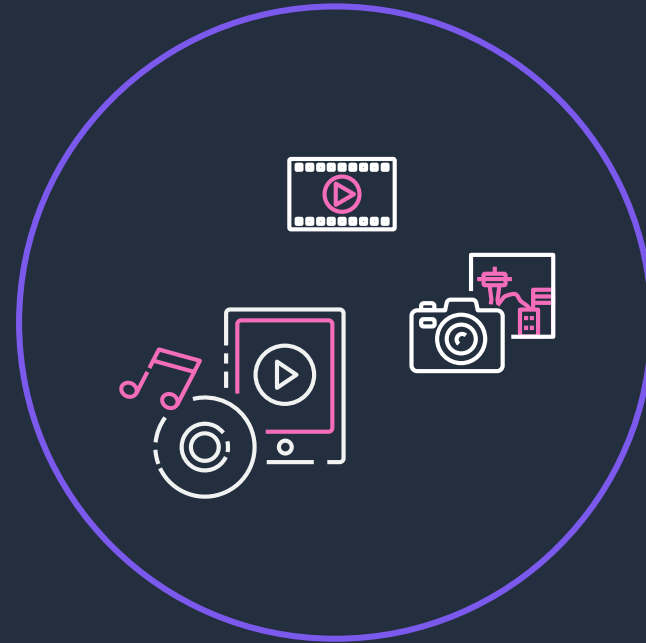
Unstructured

Amazon SageMaker Data Wrangler

No-code data preparation

Single visual interface for common data prep techniques

Select data from multiple sources

300+ built-in transformations to prepare data without writing code

Amazon SageMaker Feature Store

Securely store, discover, and share features for ML

- Online and off-line
- Millisecond latency
- Consistent features
- Visual search
- Sharing and collaboration

Amazon SageMaker Ground Truth Plus

Deliver high-quality training datasets fast and reduce data labeling costs

Increase data quality through ML-powered data labeling

Access expert data labelers
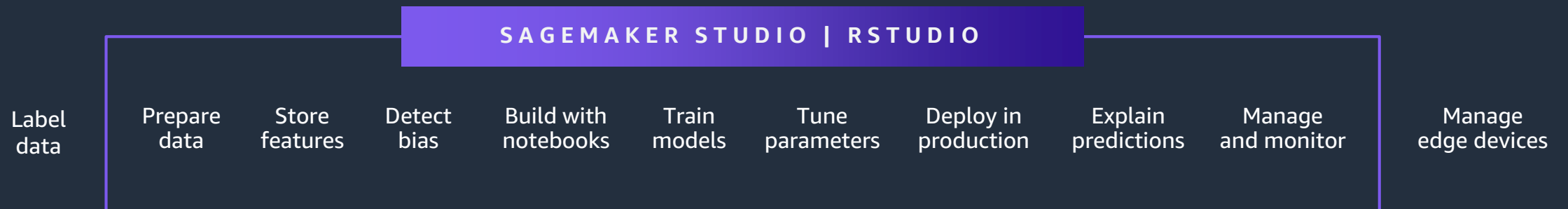
Reduce data labeling costs with assistive labeling features

Improve operational efficiency by reviewing project metrics

Make data labeling accessible to data operations and program managers

aws machine learning

Integrated ML tools in a single interface

# Amazon SageMaker brings tools for every step of the ML lifecycle under one unified visual user interface

SAGEMAKER STUDIO | RSTUDIO

| Label data | Prepare data | Store features | Detect bias | Build with notebooks | Train models | Tune parameters | Deploy in production | Explain predictions | Manage and monitor | Manage edge devices |

aws machine learning

Build ML models

Fully managed shareable notebooks on Amazon EC2

Fully managed, sharable Jupyter notebooks
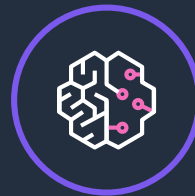Run notebooks on elastic compute resources
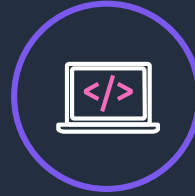
Built-in algorithms
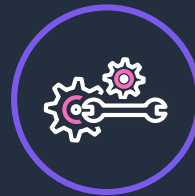15 built-in algorithms available in prebuilt container images
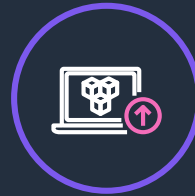
Prebuilt solutions and open-source models
Over 150 popular open-source models

AutoML
Automatically create ML models with full visibility

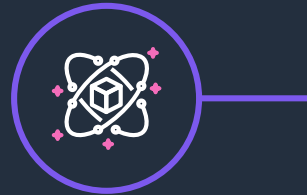**Support for major frameworks and toolkits**
Optimized for popular deep learning (DL) frameworks such as TensorFlow, PyTorch, Apache MXNet, and Hugging Face

Amazon SageMaker Studio Notebook

Perform data engineering, analytics, and ML workflows in one notebook

Connect with Amazon EMR, Amazon S3, and more

Interactively access, transform, and analyze a wide range of data

Build, train, and deploy models using your preferred framework

Train ML models

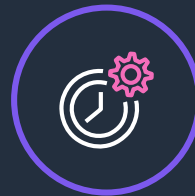**Fast and cost-effective ML model training**

Experiment management and model tuning
Save weeks of effort by automatically tracking training runs and tuning hyperparameters

Debug and profile training runs
Use real-time metrics to correct performance problems

Distributed training
Complete distributed training up to 40% faster

Training compiler
Accelerate training times by up to 50% through more efficient use of GPUs

Managed spot training
Reduce the costs of training by up to 90%

Deploy ML models

Fully managed deployment for inference at scale

Wide selection of infrastructure
70+ instance types with varying levels of compute and memory to meet the needs of every use case

Single-digit millisecond overhead latency
For use cases requiring real-time responses

Asynchronous inference
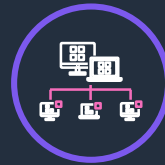Supports large models with long-running processing times

Cost-effective deployment
Multi-model/multi-container endpoints, serverless inference, and elastic scaling

Built-in integration for MLOps
ML workflows, CI/CD, lineage tracking, and catalog

Automatic deployment recommendations
Optimal instance type/count and container parameters, and fully managed load testing

aws machine learning

Support the responsible use of ML throughout the model lifecycle

Build
Perform bias analysis during exploratory data analysis

Train
Conduct bias and explainability analysis after training

Deploy
Explain individual inferences from models in production

Monitor
Validate bias and relative feature importance over time

Built-in
MLOps capabilities

Operational challenges with managing the ML lifecycle

Manual iterative processes slow down ML innovation

Difficult to scale the number of models in production
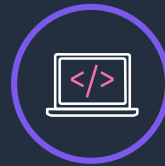
CI/CD for ML requires writing custom code

Compliance requirements are difficult to meet

Amazon SageMaker MLOps

Streamline the ML lifecycle

Automate ML workflows to scale model development

Build CI/CD pipelines for ML to accelerate model deployment

Catalog model versions, metadata, metrics, and approvals for traceability and reusability

Track lineage for troubleshooting and compliance

Maintain accuracy of predictions after models are deployed

Enhance governance and security

Amazon SageMaker key benefits

End-to-end ML service

Democratize ML innovation
Empower more groups of people, including business analysts

Prepare data at scale
Access, label, and process structured and unstructured data

Accelerate the ML lifecycle
Reduce training time from hours to minutes

Streamline ML processes
Automate and standardize MLOps processes

PROBLEM

3 TB+ data, 1,500+ hours play time per week

Needed ML solution for real-time stats

Lean team, no data science expertise

SOLUTION

Next Gen Stats (NGS):

Live data streamed to AWS from RFID tags on players and embedded in game ball

Data processed in 100+ steps in under 1 second

Real-time predictions

Output stats published through APIs and on-screen graphics

IMPACT

Quickly launched 20+ stats

Sports announcers get interesting data points to engage fans

aws machine learning

PROBLEM

Develops computer vision models for autonomous driving

Needed to reduce training times to accelerate model development

SOLUTION

Amazon SageMaker data parallelism library to speed up training jobs

Amazon SageMaker Debugger to monitor, trace, and analyze performance issues

Performance optimization with Amazon ML Solutions Lab

IMPACT

10x faster model training using SageMaker data parallelism library and ML Solutions Lab's code optimization

Hyundai Motor Company can now focus on data preparation and accelerate the development cycle

HYUNDAI MOTOR COMPANY

HYUNDAI

aws machine learning

PROBLEM

Ingests thousands of data points per day from the American secondary mortgage market

Wanted to leverage data more efficiently by implementing ML into its tools

SOLUTION

Used Amazon SageMaker to create ML models that help evaluate everything from loan risk assessment to property value analysis for its Multifamily line of business

IMPACT

Helped the company improve its ability to identify applicants that are likely to default from 3.5% to 48%

Provides transparency in the home loan process for lenders, investors, and analysts

THE FEDERAL NATIONAL MORTGAGE ASSOCIATION (FANNIE MAE)

Fannie Mae®

aws machine learning

PROBLEM

Wanted to speed up all stages of research and development, particularly the review of tissue samples

Labeling data was time consuming and tedious

SOLUTION

Worked with the Amazon ML Solutions Lab to integrate Amazon SageMaker Ground Truth to automate tedious portions of work

Helps annotate, collect, and classify ML model training data quickly

IMPACT

Reduces time humans spend cataloging samples by at least 50%

Accelerates the drug research process and the introduction of medicines to the market

Increases the pace of research by freeing up scientists

ASTRAZENECA PLC

AstraZeneca

aws machine learning
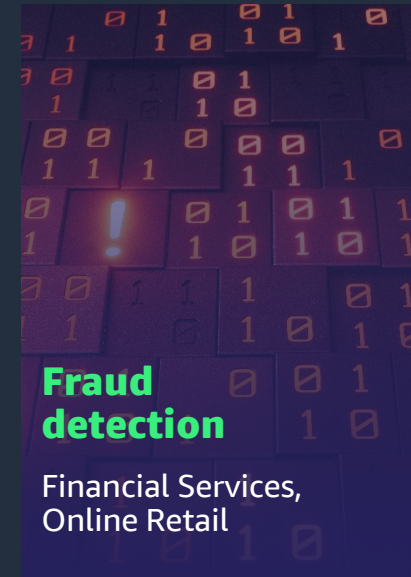
Amazon SageMaker

Common use cases

Predictive maintenance

Manufacturing, Automotive, IoT

Demand forecasting

Retail, Consumer Goods, Manufacturing

Fraud detection

Financial Services, Online Retail

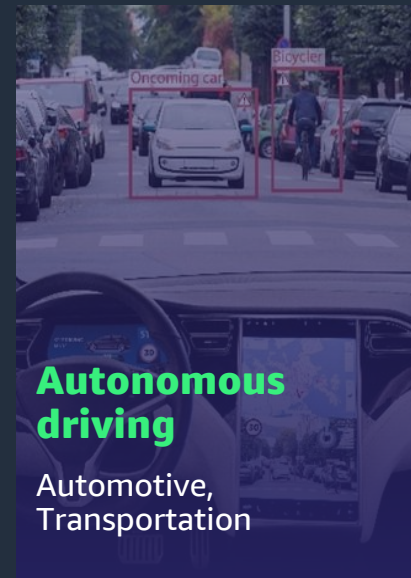**Credit risk prediction**

Financial Services, Retail

Extract and analyze data from documents

Healthcare, Legal, Media/Ent, Education

Computer vision

Healthcare, Pharma, Manufacturing

Autonomous driving

Automotive, Transportation

Personalized recommendations

Media and Entertainment, Retail, Education

Churn prediction

Retail, Education, Software and Internet

aws machine learning

Amazon SageMaker

Getting started

Business analysts

Get started with Amazon SageMaker Canvas

No-code ML

Data scientists

Onboard through Amazon SageMaker Studio

Single IDE for full-code ML

MLOps engineers

Implement MLOps with Amazon SageMaker Pipelines

Workflow automation and CI/CD

Learn and experiment with machine learning

Amazon SageMaker Studio Lab (Preview)

Free, no-configuration development environment

All you need is an email address to get started

As many CPU and GPU sessions as needed

Automatically save session data

Models ready for production

Industrial AI

AI and ML services for asset-intensive industry use cases

AWS Panorama Appliance
hardware appliance to add computer vision to existing onsite cameras

AWS Panorama Device SDK
build new cameras and devices that run computer vision applications at the edge

Amazon Monitron
end-to-end system for equipment monitoring to detect abnormal machine behavior and enable predictive maintenance

Amazon Lookout for Equipment
detect abnormal machine behavior using existing industrial sensor data

Amazon Lookout for Vision
spot defects and anomalies in manufacturing using computer vision

aws

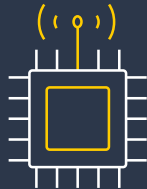# More relevant enterprise search with Amazon Kendra

DOMAIN EXPERTISE

Eliminate information silos

20+ NATIVE CONNECTORS

Simplify and accelerate integration

ML-OPTIMIZED INDEX

Find more precise answers

SIMPLE API TO INTEGRATE SEARCH

Easily develop search applications

NATURAL LANGUAGE SEARCH

Contextual search results

INCREMENTAL LEARNING THROUGH FEEDBACK

Up-to-date relevant answers

aws machine learning

Find More Fraud Faster with Amazon Fraud Detector

Enhance fraud detection with ML

Any level of ML expertise can build ML fraud models

ML boost from Amazon experience, enrichments

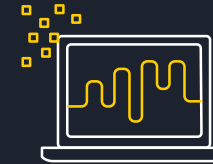Fewer false positives, manual reviews

Fraud staff self-service to address threats faster

Lower TCO, faster TTV

aws machine learning

Amazon CodeGuru to build and run high-performing software

WRITE + REVIEW

Built-in code reviews with intelligent recommendations

BUILD + TEST

Detect and optimize the expensive lines of code pre-prod

DEPLOY

MEASURE

Easily identify application inefficiencies in production environment

IMPROVE

aws machine learning

Amazon DevOps Guru

Next-gen DevOps service to improve application availability with ML-powered insights

Automatically detect operational issues with contextual insights

Reduce noise and alarm fatigue by automatically grouping related issues

Easily scale and maintain availability without manually updating alarms

Automatic configuration as you grow– no ML expertise required

THOMSON REUTERS

aws machine learning

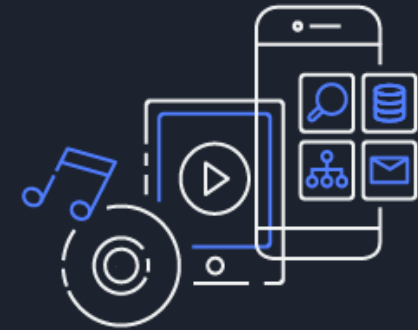# Improving customer experience with Amazon Personalize

Deliver high quality recommendations

Adapt to changes in customer intent in real time

Train a recommendation model with a few clicks

Generate recommendations for almost any product or content

aws machine learning

AWS DeepRacer

Getting started with reinforcement learning

Innovate faster with the most comprehensive set of AI/ML services

AWS ADVANTAGE

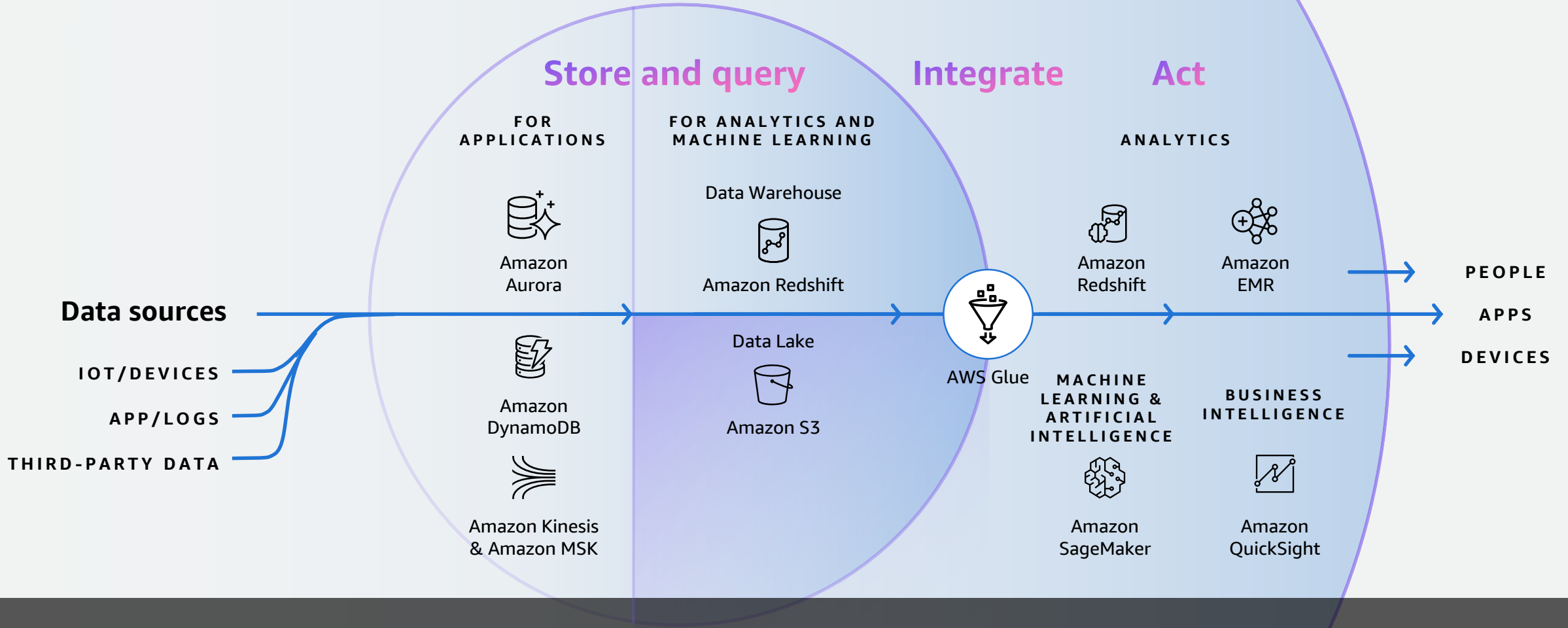|	
AI powered use cases	Hundreds of pre-built algorithms, models, and solutions for common use cases and industries Faster time to value with 23 pre-trained AI Services
Democratize access to ML	Empower any business analyst to make predictions without writing code with Amazon SageMaker AI Services to easily add AI capabilities to your applications, no ML skills required
Scale ML	Build, train, and deploy ML models for any use case with Amazon SageMaker Harness the exponential growth of ML models with foundation models on AWS
Grow ML skills	Learn more with AWS Training, AWS DeepRacer, AI & ML Scholarship Program, and AWS Machine Learning University Accelerate your journey with AWS ML Embark Program or Amazon ML Solutions Lab

Building an end-to-end data strategy

Store and query — Integrate — Act

Store and query

FOR APPLICATIONS

FOR ANALYTICS AND MACHINE LEARNING

Integrate

Act

ANALYTICS

Data Warehouse

Amazon Aurora

Amazon Redshift

Amazon Redshift

Amazon EMR

Data sources

Data Lake

AWS Glue

PEOPLE

APPS

IOT/DEVICES

Amazon DynamoDB

Amazon S3

DEVICES

APP/LOGS

MACHINE LEARNING & ARTIFICIAL INTELLIGENCE

BUSINESS INTELLIGENCE

THIRD-PARTY DATA

Amazon Kinesis & Amazon MSK

Amazon SageMaker

Amazon QuickSight

Catalog and govern | AWS Lake Formation, Amazon DataZone

Next steps

TRAINING

In-person, online and certification programs
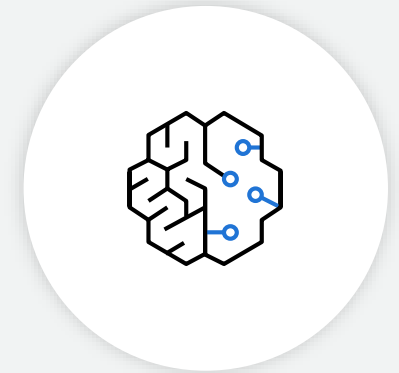
AWS deep devices

IDEATION AND PROOF OF CONCEPT

ML Solutions Lab

AWS ML Embark Program

AI SERVICES, NO ML SKILLS REQUIRED

Leverage our AI services to add AI capabilities to your business applications

BUILD YOUR OWN MODEL

Build your own model using SageMaker or SageMaker Canvas

Thank you!

AWS Education & Research Team

Nikiforos Botis - nbotis@amazon.com