



AWS for Data and AI

Dr. Chomchana “Chom” Trevai
Head of AI and Machine Learning AWS
ASEAN

The **data-driven** organization

Culture

Align business and technology leaders

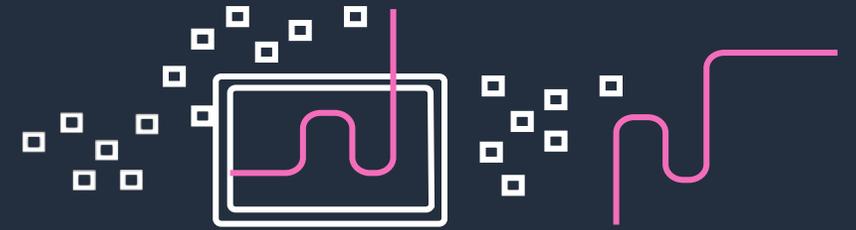
People and process

Build the right organization and process model

Technology

Empower builders with an end-to-end data strategy

AWS data services help accelerate innovation



PRICE PERFORMANCE AT SCALE

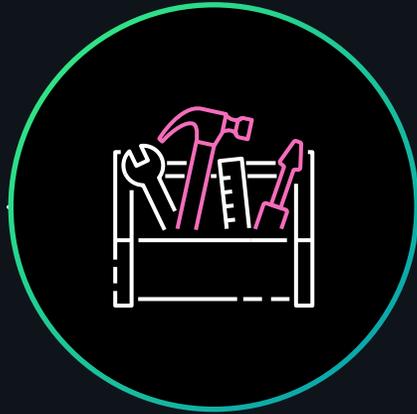


SECURE AND RELIABLE



REMOVE HEAVY LIFTING

AWS is the best cloud to unlock the value of data



Comprehensive

Tools for any
data use case



Integrated

Easily connect to all
your data



Governed

End-to-end
governance

End-to-end data foundation



Comprehensive



Integrated



Governed

Data sources

IOT/
DEVICES

APP/
LOGS

THIRD-PARTY
DATA

Integrate

Store, query, analyze

DATABASES



Amazon Aurora



Amazon DynamoDB



Amazon Kinesis
& Amazon MSK

DATA WAREHOUSE



Amazon Redshift

DATA LAKE



Amazon S3

BIG DATA



Amazon EMR

Act

BUSINESS INTELLIGENCE



Amazon QuickSight

MACHINE LEARNING



Amazon SageMaker

GENERATIVE AI



Amazon Bedrock



PEOPLE



APPS



DEVICES

Catalog and govern | AWS Lake Formation, Amazon DataZone



Data integration gives you a full picture of your business



Comprehensive



Integrated



Governed



AWS enable governance capabilities



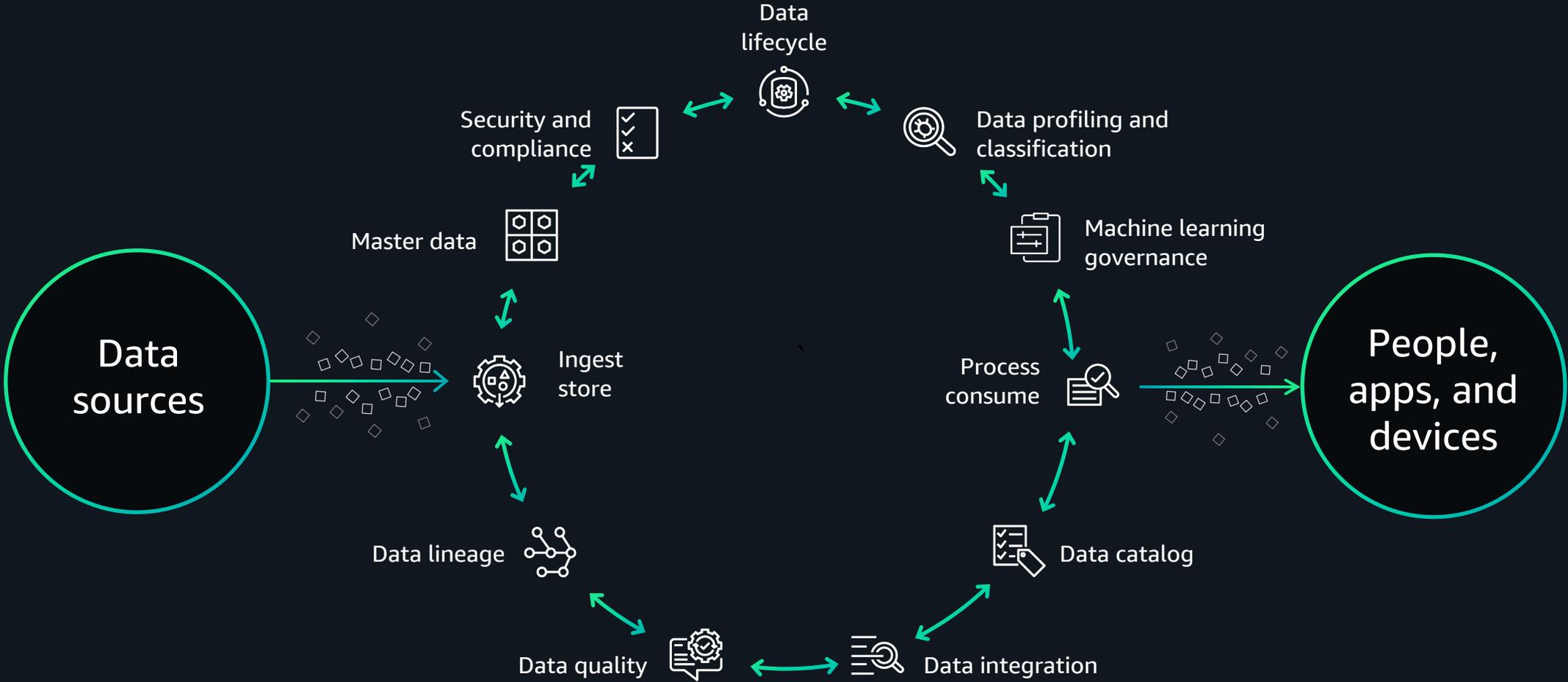
Comprehensive



Integrated



Governed



New era of AI and Machine Learning



Artificial intelligence (AI)

Any technique that allows computers to mimic human intelligence using logic, if-then statements, and machine learning



Machine learning (ML)

A subset of AI that uses machines to search for patterns in data to build logic models automatically



Deep learning (DL)

A subset of ML composed of deeply multi-layered neural networks that perform tasks like speech and image recognition



Generative AI

Powered by large models that are pretrained on vast corpuses of data and commonly referred to as **foundation models (FMs)**

What you need



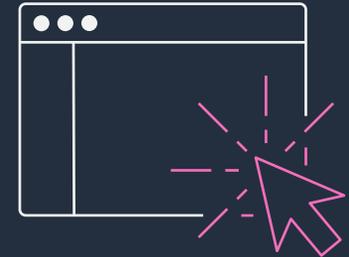
EASE OF USE

Choose a base FM and customize it securely and quickly to build differentiated apps



COST EFFECTIVENESS

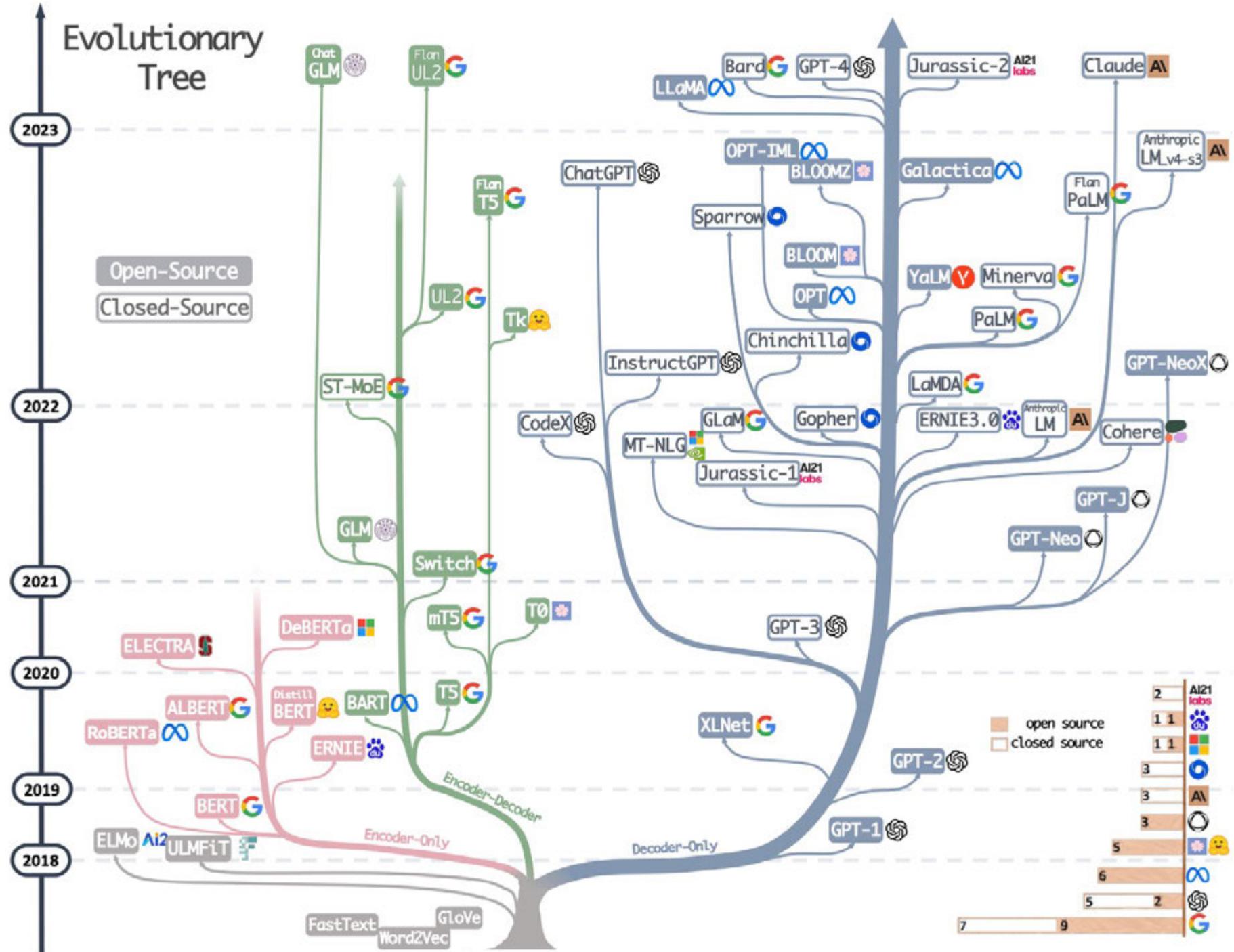
Manage infrastructure to ensure optimal performance and cost



FASTER TIME TO VALUE

Get started with generative AI services for common business tasks

FMs as Commodity



Source: <https://www.sequoiacap.com/article/generative-ai-a-creative-new-world/>

Source: JingFeng et al. "Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond", arXiv:2304.13712 (2023)



Generative AI enables new and improved use cases



New generative AI use cases

Generative AI capabilities enable completely new use cases (e.g., Content creation)



Improved AI use cases

Combines traditional ML and deep learning with generative AI capabilities to enable improved AI use cases (e.g., Call transcript summarization)



High-value non-generative AI use cases

Traditional AI use cases that do not require nor benefit from generative AI capabilities (e.g., Condition monitoring)

The AWS AI/ML Stack

Consumer

AI SERVICES

Generative AI

SCALING GENERATIVE AI
Amazon Bedrock

AMAZON BUILT FOUNDATIONAL MODEL
Amazon Titan

Specialized AI Services

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

SEARCH
Amazon Kendra

CONVERSATION
Amazon Lex
Contact Lens
Voice ID
Amazon Transcribe Call Analytics

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
Panorama Amazon Rekognition AWS

Tuner

AMAZON SAGEMAKER

JUMPSTART
Use pre-built models in 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

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

Provider

ML FRAMEWORKS & INFRASTRUCTURE

PyTorch, TensorFlow

Amazon EC2

CPUs

GPUs

AWS Inferentia

AWS Trainium

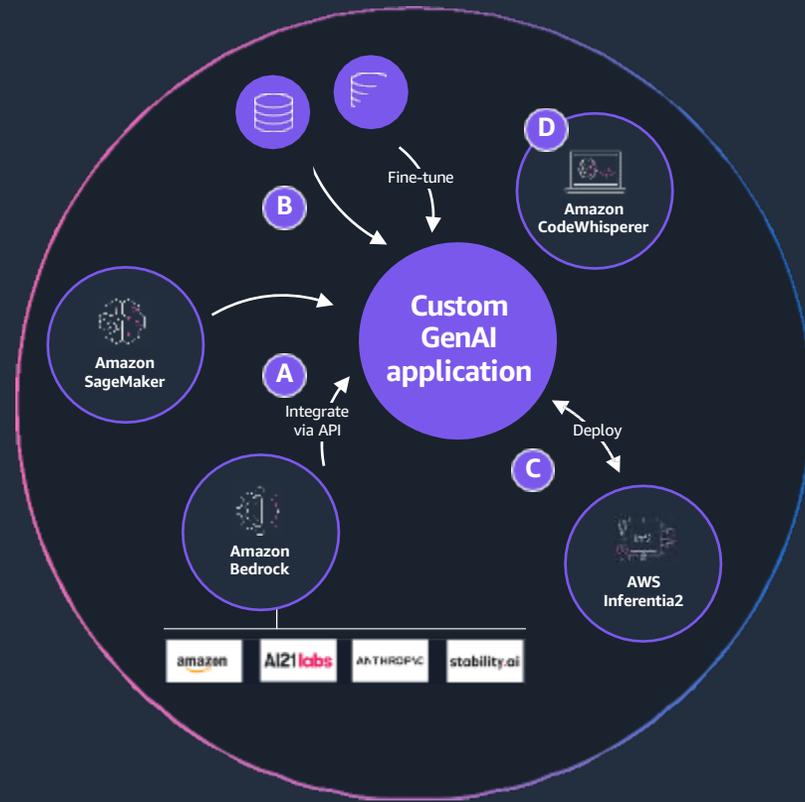
Habana Gaudi

FPGA



AWS GenAI solution suite helps you address these challenges and accelerate time to business value

Seamlessly deploy tailored GenAI models on AWS...



- A Integrate pretrained models**
Select from a variety of pretrained foundation models and integrate them into your GenAI application via API
- B Tailor models to your need**
Fine-tune models to address company-specific use cases by providing a small number of labelled examples
- C Efficiently operate your solution**
Run your models on purpose-built Instantia2 compute instances, driving improved price performance
- D Accelerate solution development**
Leverage our CodeWhisperer solution supporting developers with real-time code suggestions and vulnerability detection

... and drive real business impact

Integrate, customize and deploy foundational models using controls with the familiar depth, breadth and security of AWS services



~57%

Avg. productivity increase of developers supported by CodeWhisperer co-pilot



4

Providers of FMs enabling integration of their models via API (incl. Amazon Titan)



Up to 10x

Team productivity increase for work on ML models through purpose-built tools



Up to 40%

Better price performance of running LLMs on Inferentia2 instances compared to EC2



Could we spark your interest?

We would be delighted to discuss your GenAI journey and how AWS can partner with you

Feel free to reach out to your key contacts

Thank you!



Dr. Chomchana Trevai

Head of AI and Machine
Learning ASEAN

cttrevai@amazon.com