aws

Digital Collections and Generative AI

Greg Ritter (he/him)

Strategy Lead for Teaching, Learning, & Libraries

AWS Higher Education

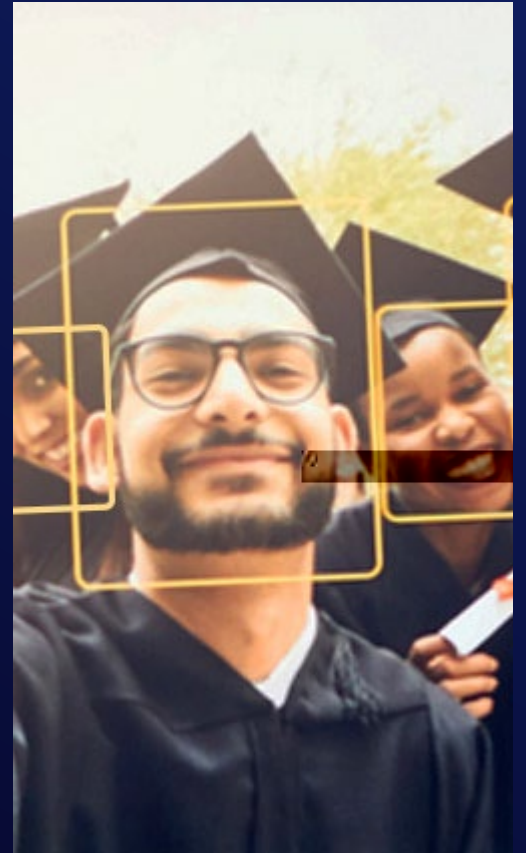# Transforming libraries with machine learning and AI

Reduce librarian workloads

- Connect the output of digitization directly to the cloud
- Automate library collection workflows
- Extract metadata with artificial intelligence
- Leverage "Human in the Loop" to assure integrity of results
- Create custom machine learning models tailored to your collections
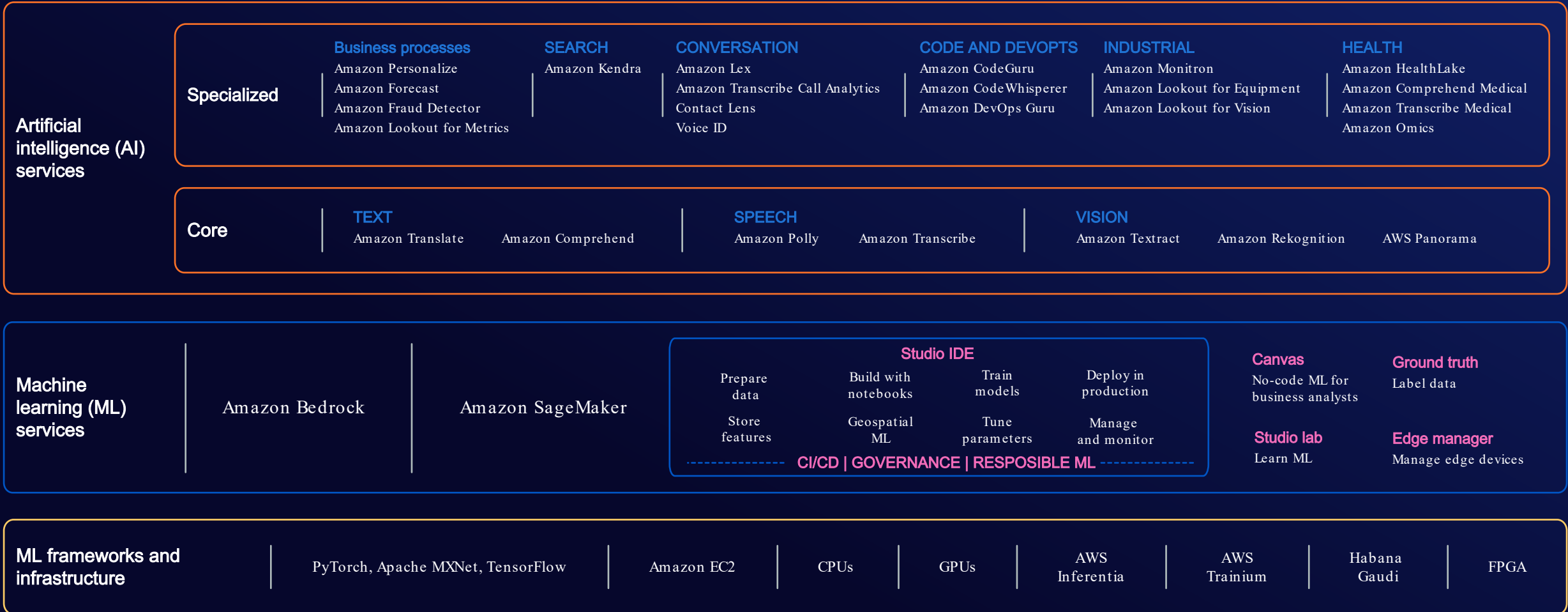

Create paths for discovery & scholarship

- Make the previously un-discoverable discoverable
- Increase access to collections through transcription and translation
- Create opportunities for expansive text mining and textual analysis
- Build AI-driven conversational interfaces to interact with collection

The AWS AI/ML stack
THE BROADEST AND MOST COMPLETE SET OF MACHINE LEARNING CAPABILITIES

Artificial intelligence (AI) services

Specialized

Business processes	**SEARCH**	**CONVERSATION**	**CODE AND DEVOPTS**	**INDUSTRIAL**	**HEALTH**
Amazon Personalize	Amazon Kendra	Amazon Lex	Amazon CodeGuru	Amazon Monitron	Amazon HealthLake
Amazon Forecast		Amazon Transcribe Call Analytics	Amazon CodeWhisperer	Amazon Lookout for Equipment	Amazon Comprehend Medical
Amazon Fraud Detector		Contact Lens	Amazon DevOps Guru	Amazon Lookout for Vision	Amazon Transcribe Medical
Amazon Lookout for Metrics		Voice ID			Amazon Omics

Core

TEXT		**SPEECH**		**VISION**		
Amazon Translate	Amazon Comprehend	Amazon Polly	Amazon Transcribe	Amazon Textract	Amazon Rekognition	AWS Panorama

Machine learning (ML) services

Amazon Bedrock   Amazon SageMaker

Studio IDE

Prepare data	Build with notebooks	Train models	Deploy in production
Store features	Geospatial ML	Tune parameters	Manage and monitor

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

Canvas
No-code ML for business analysts

Ground truth
Label data

Studio lab
Learn ML

Edge manager
Manage edge devices

ML frameworks and infrastructure

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

Flexibility

Cost-effective infrastructure

Amazon Bedrock

The easiest way to build and scale generative AI applications with foundation models

Choice of industry-leading models from AI21 Labs, Anthropic, Cohere, Mistral AI, Meta, Stability AI, and Amazon

Customize your models using your own library data

Enterprise-grade security and privacy

Amazon Bedrock
Keeps your data secure and private

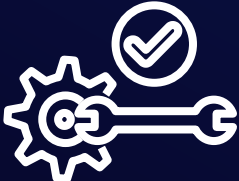None of the library's data is used to train the underlying models

All data is encrypted in transit and at rest and remains in the library's virtual private cloud.

Support for GDPR, SOC, ISO, CSA compliance and HIPAA eligibility

6

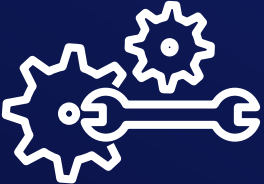# Customizing model responses for your needs

Fine tuning

PURPOSE

Maximizing accuracy
for **specific tasks**

DATA NEED

Small number
of labeled examples

Continued pre-training

PURPOSE

Maintaining model
accuracy for **your domain**

DATA NEED

Large number
of unlabeled datasets

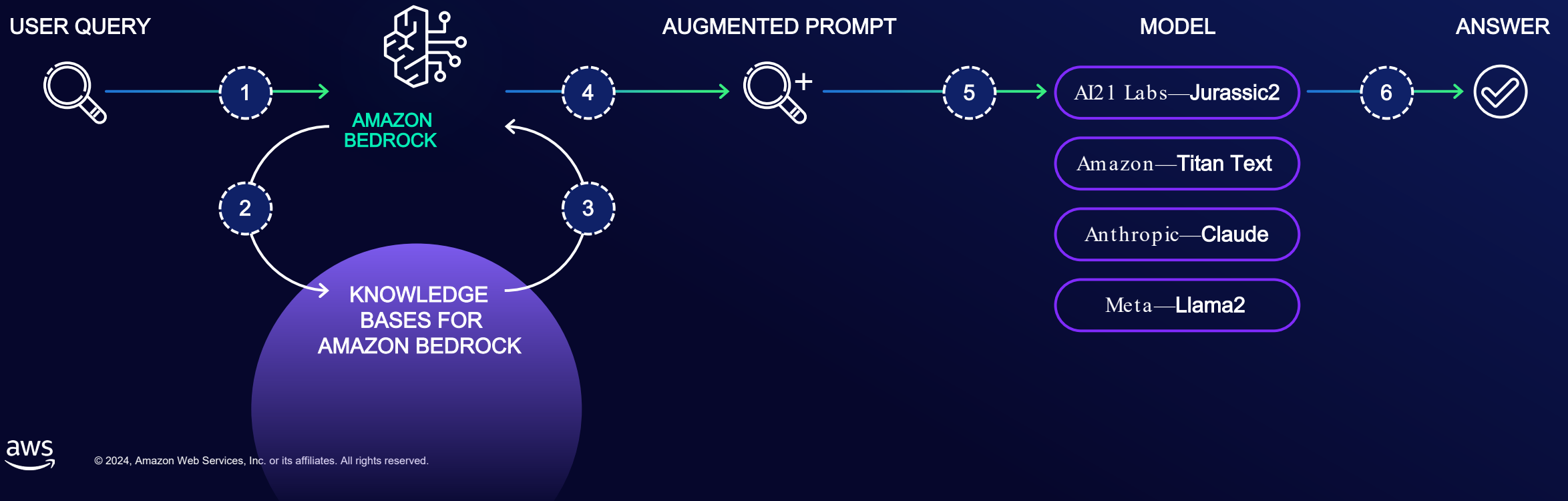# Knowledge bases for Amazon Bedrock
NATIVE SUPPORT FOR RETRIEVAL-AUGMENTED GENERATION (RAG)

Securely connect models to data sources to deliver more relevant responses

Fully managed RAG workflow including ingestion, retrieval, and augmentation

Built-in session context management for multi-turn conversations

Automatic citations with retrievals to improve transparency

USER QUERY

AMAZON BEDROCK

AUGMENTED PROMPT

MODEL

ANSWER

1

2

3

4

5

6

KNOWLEDGE BASES FOR AMAZON BEDROCK

AI21 Labs—Jurassic2

Amazon—Titan Text

Anthropic—Claude

Meta—Llama2

Key library use cases for generative AI

"Ask your collection" conversational interfaces

Topic identification & metadata generation

Resource description & content summary

Multi-disciplinary & multi-modal content matching

Pillars for the responsible use of AI

Value alignment
Systems should be designed and used in ways that align with company mission, social norms, and legal compliance

Privacy & security
Protects the quality and integrity of data used, its relevance, access, and processing

Inclusion
Inclusion of unique skills, experiences, perspectives, and cultural backgrounds

Fairness
Systems must be designed to minimize bias and promote inclusive representation

Training & education
Appropriate knowledge sharing and education to understand purpose, use, and impact

Accountability
Structured maintaining human involvement and responsibility for design, development, decision processes, and outcomes
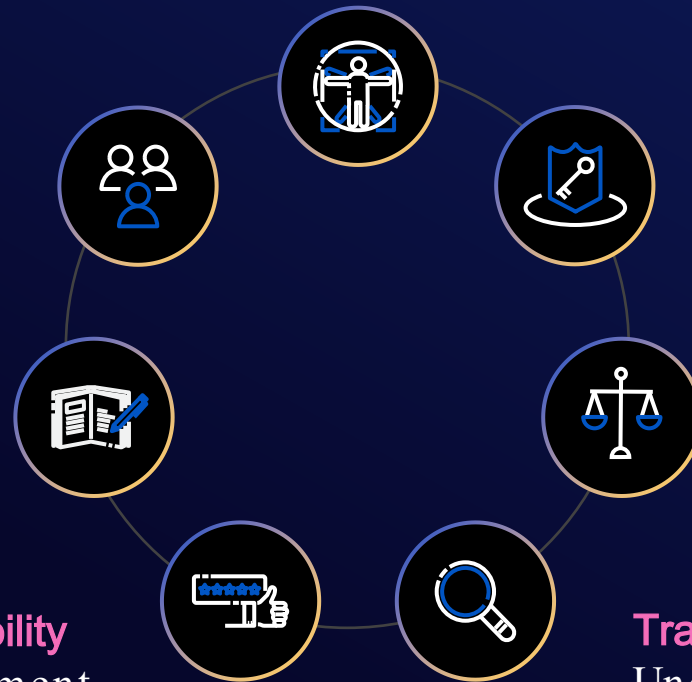
Transparency & explainability
Understanding how data is used, and how decisions and outcomes are made in a human understandable way

aws

Thank you!

Greg Ritter
gritter@amazon.com