



# Data governance in the age of generative AI

Mike Thomas

Solutions Architect, EDU  
zmthoma@amazon.com

Dan Buettner

Solutions Architect, EDU  
danjbu@amazon.com

“In the fullness of time - virtually every app in every process in every business in every industry will use generative AI.

Over the next few years, it’s going to become very common; just part of the fabric of what we do and how we build.”

**Dr. Matt Wood**

VP of Artificial Intelligence Products, Amazon Web Services



© 2024, Amazon Web Services, Inc. or its affiliates.

# Unlocking the potential of generative AI



## **Easiest way to build**

with Foundation Models



## **Flexibility to build**

your own Foundation Models



## **Ready to use Applications**

powered by generative AI



## **Data as a Foundation**

for generative AI

Your data is  
**your differentiator**



# Core elements of a data foundation



---

Comprehensive  
set of services



---

Integrated  
data sources



---

Governance  
strategy

# Key Governance GenAI Features

Data discoverability

Data privacy

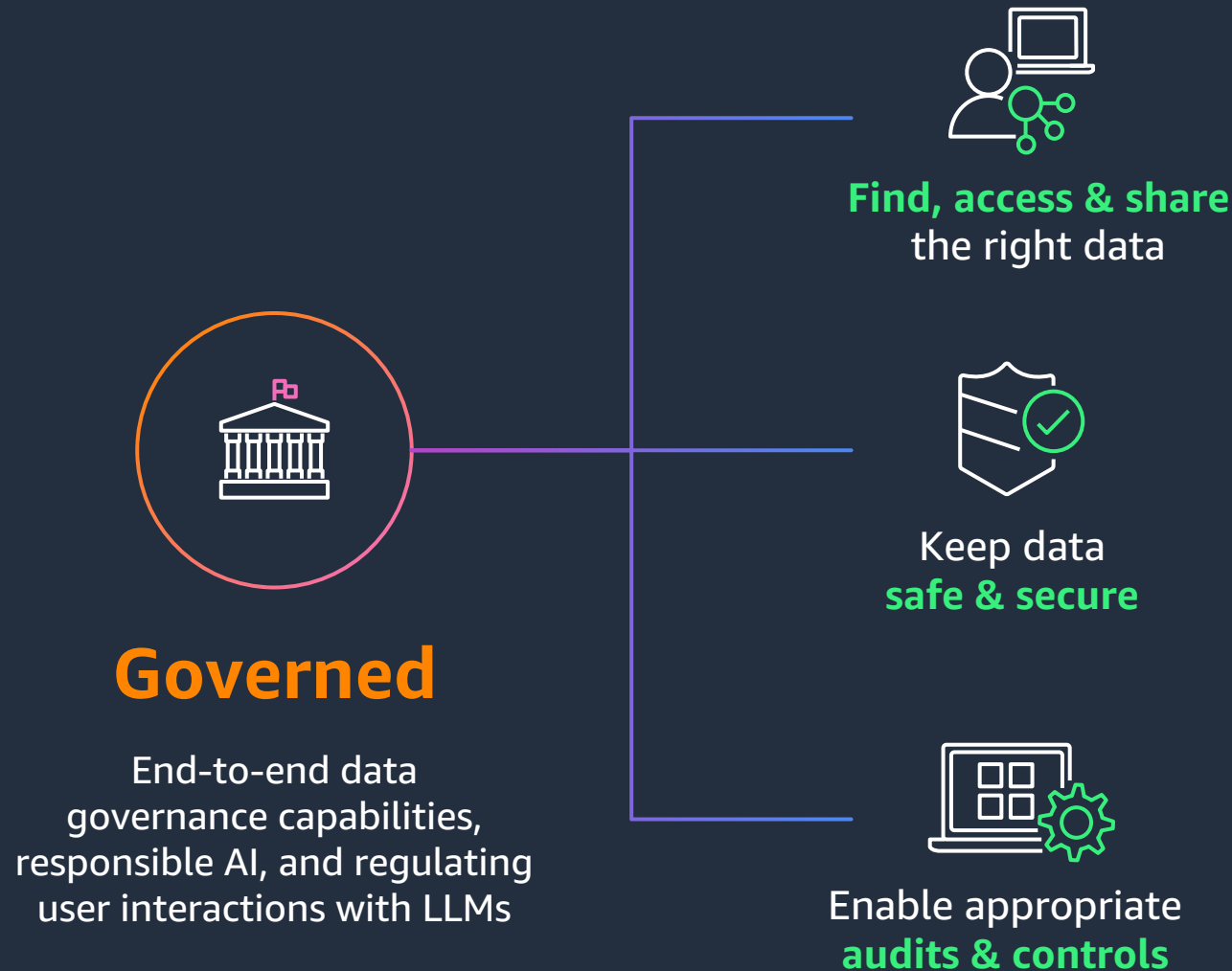
Data enrichment

Data quality

Vector store governance



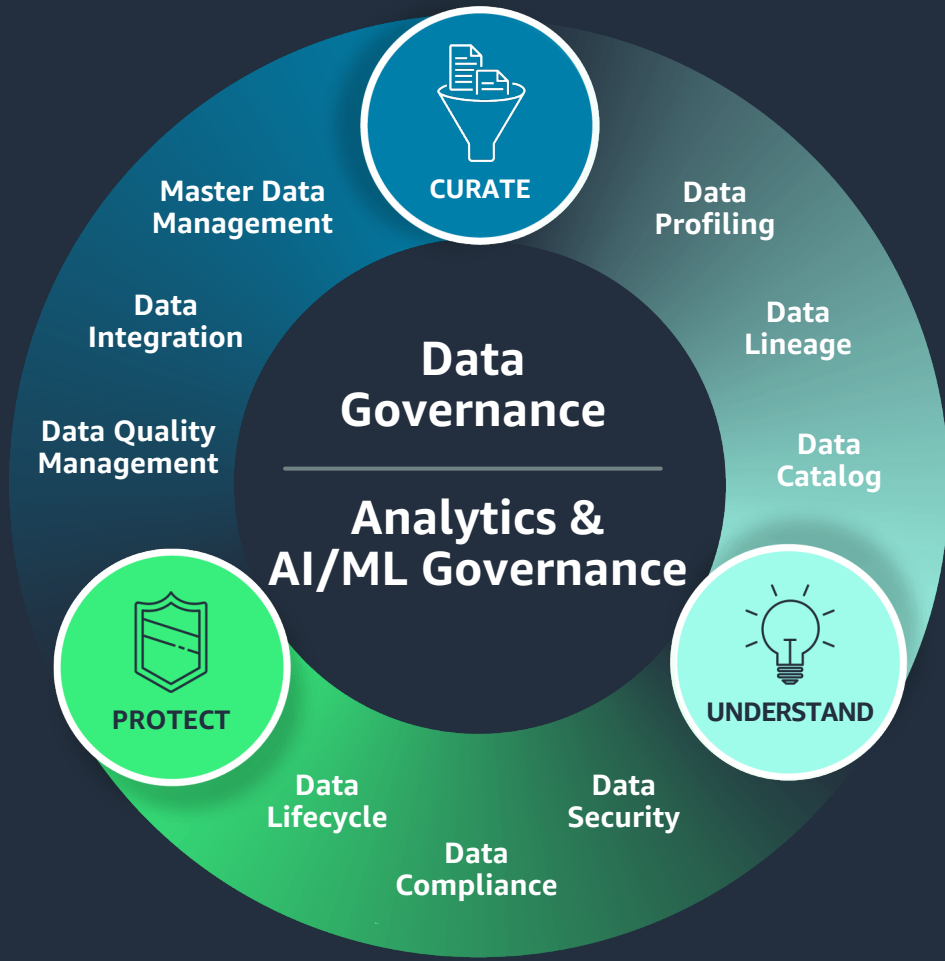
# Governance provides data guardrails to innovate faster with generative AI applications



*"80-90% of enterprise data is in unstructured formats (e.g. pdfs, emails, text files etc.) and is growing three times faster than structure data" - Gartner*

These formats need to be ingested, pre-processed and transformed into clean, curated formats that are ready for generative AI applications

# AWS and our partners enable governance capabilities



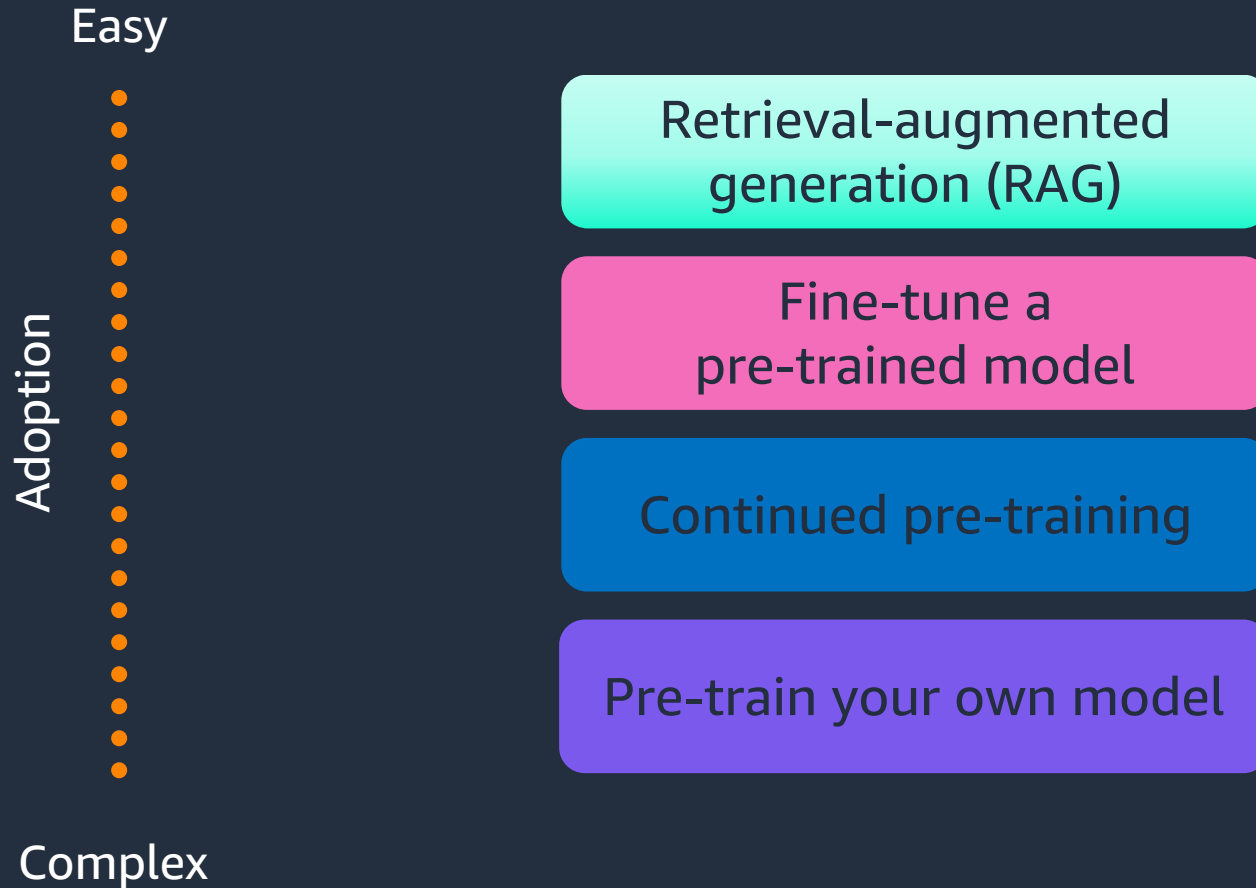
- **Security and Privacy:** Data must be governed with the right access and permissions controls implemented through GenAI applications
- **Lineage:** Businesses must track data lineage, which requires understanding the source (and accuracy/authenticity) of data used in AI models
- **Ownership:** It must be clear who has stewardship over data and is therefore responsible for managing, securing, sharing and addressing any concerns that arise on its usage



# WHITEBOARD



# The age of Generative AI



Popular approaches to building generative AI solutions

# Your data continues to be the differentiator



Enterprise Data



Retrieval-augmented  
generation (RAG)

Fine-tune a  
pre-trained model

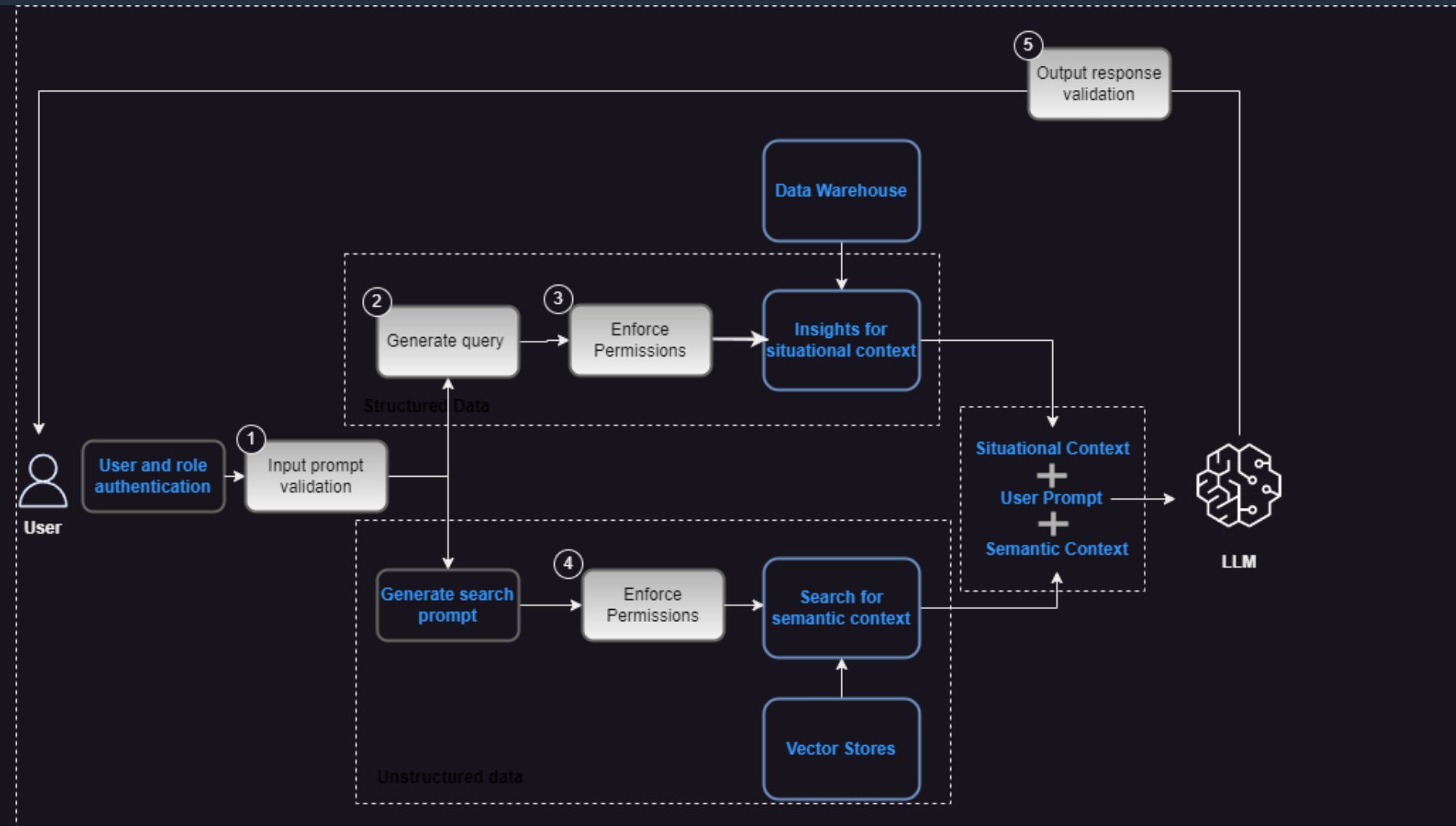
Continued pre-training

Pre-train your own model

# WHITEBOARD



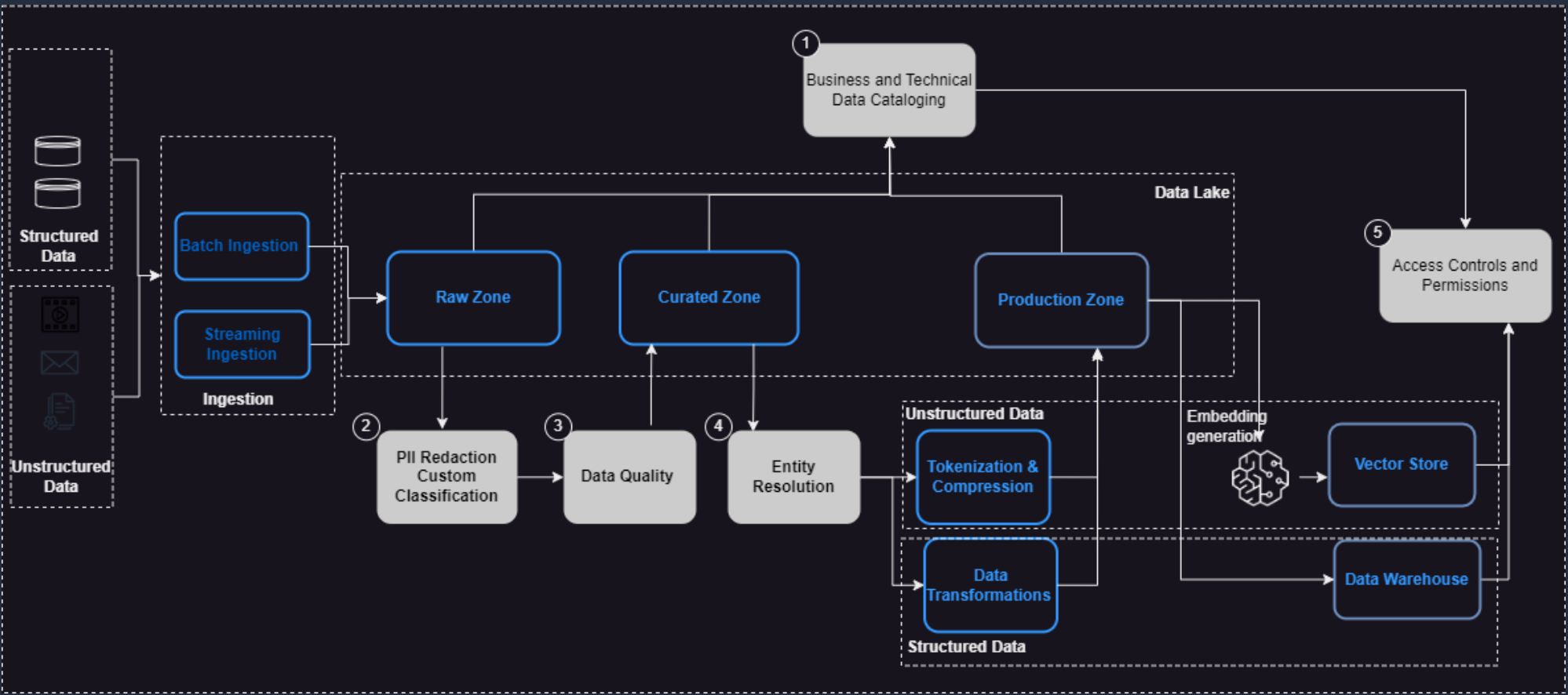
# Enforce data permissions and controls in the user workflows



## Data controls integrated into user workflows

- 1 Validate input prompt based on user permissions model
- 2 Generate query using the data catalog
- 3 Enforce FGAC (fine grained access controls) based on user permissions model
- 4 Enforce vector store access control policies based on user permissions model
- 5 Validate LLM output for data quality, bias, toxicity prior to response

# Expand the data lake to govern and manage unstructured data



## Governance of unstructured data

- 1 Enable discoverability through the Enterprise Catalog
- 2 Enforce data privacy controls with PII redaction
- 3 Monitor and flag data based on quality guardrails
- 4 Link with structured data entities to drive deeper context from data
- 5 Apply enterprise permissions with fine-grained access control policies

# Governance on AWS

CONTROL WHERE YOUR DATA SITS, WHO HAS ACCESS TO IT,  
AND WHAT CAN BE DONE WITH IT

USER WORKFLOWS  
(Real-time processes)

DATA ENGINEERING  
(Batch processes)



AWS identity and Access  
Management



Amazon Cognito



AWS Glue

Data discoverability and data quality



AWS Lake Formation

Enable centralized permissions, and  
fine-grained access controls



Amazon DataZone

Business catalog, discover, share, and  
govern data across the organization



Amazon SageMaker

Model monitoring and bias detection



Amazon Comprehend

PII identification and redaction



Amazon S3

Resource-based policies defined and  
enforced by object





# Thank you!

Mike Thomas

Solutions Architect, EDU  
zmthoma@amazon.com

Dan Buettner

Solutions Architect, EDU  
danjbu@amazon.com

Please take the survey:



Data & analytics track

Data Governance in the age of GenAI