Data Strategy for Generative AI Applications

Why Your Data is the Differentiator

Manuel Heinkel
Solutions Architect
AWS

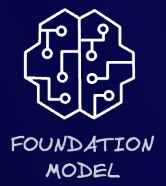
Puria Izady
Solutions Architect
AWS

Innovation can **transform industries**

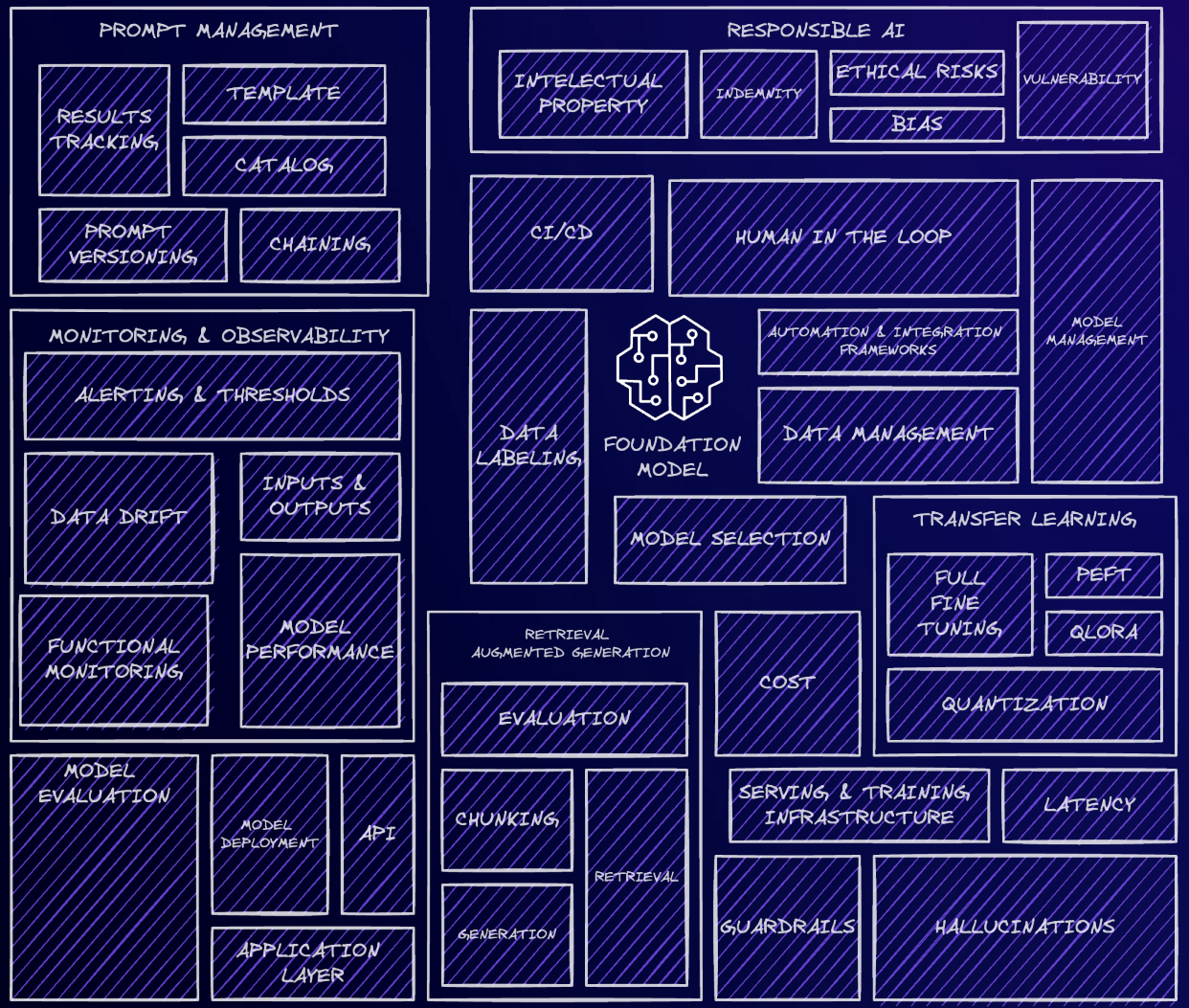

Generative AI
Application


Generative AI
Application
Data
Foundation
STORAGE
GOVERNANCE
& COMPLIANCE
DATABASES,
ANALYTICS,
& DATA LAKES
DATA INTEGRATION

Innovating beyond the foundation model

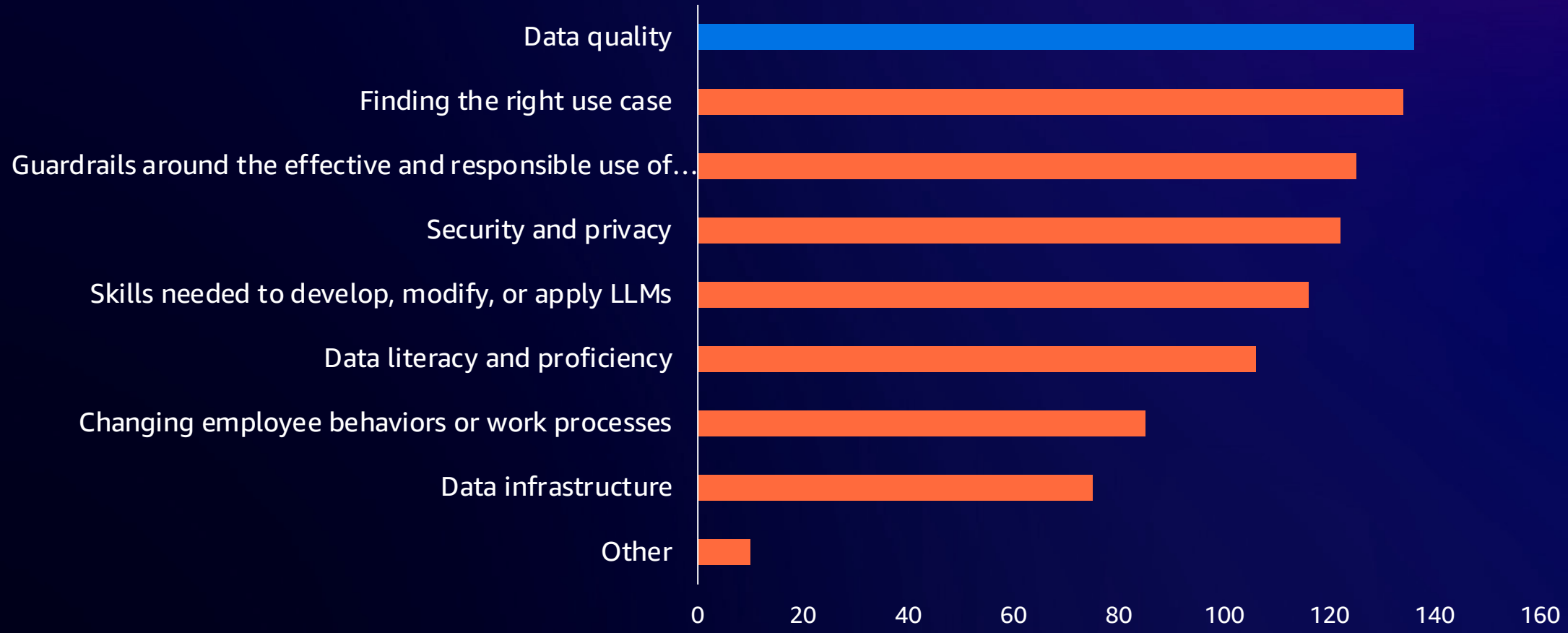


Innovating beyond the foundation model



Your data is the differentiator

Question: In your opinion what is the biggest challenge for your organization in realizing the potential of generative AI? (Select all that apply)



Generative AI core capabilities

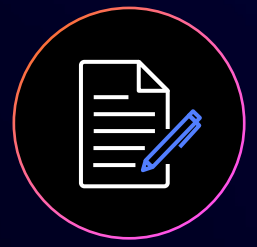
Capabilities

Generation

Question answering

Summarization

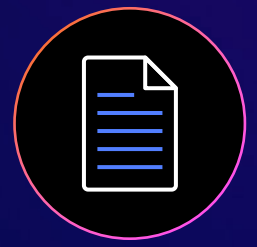
Translation



Correction

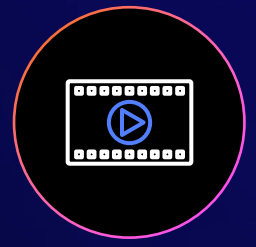
Classification

Input/Output



Text

Images



Video/Audio

Code

2023
The Year of POCs

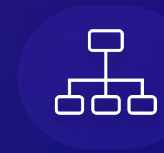
What is generative AI?

Is this secure?

Do I need to become a prompt engineer?

How do I choose a model?

Where do I get started?



What does this mean for my business?

What is a Foundation Model?

Which models should we try out?

What is FM?

What is a Large Language Model?

2024

The Year of Production

(FOR SOME)

How do I prioritize my projects?

How can I lower my costs?

How do I make this real?

What customization method should I use?

How I can I scale this?

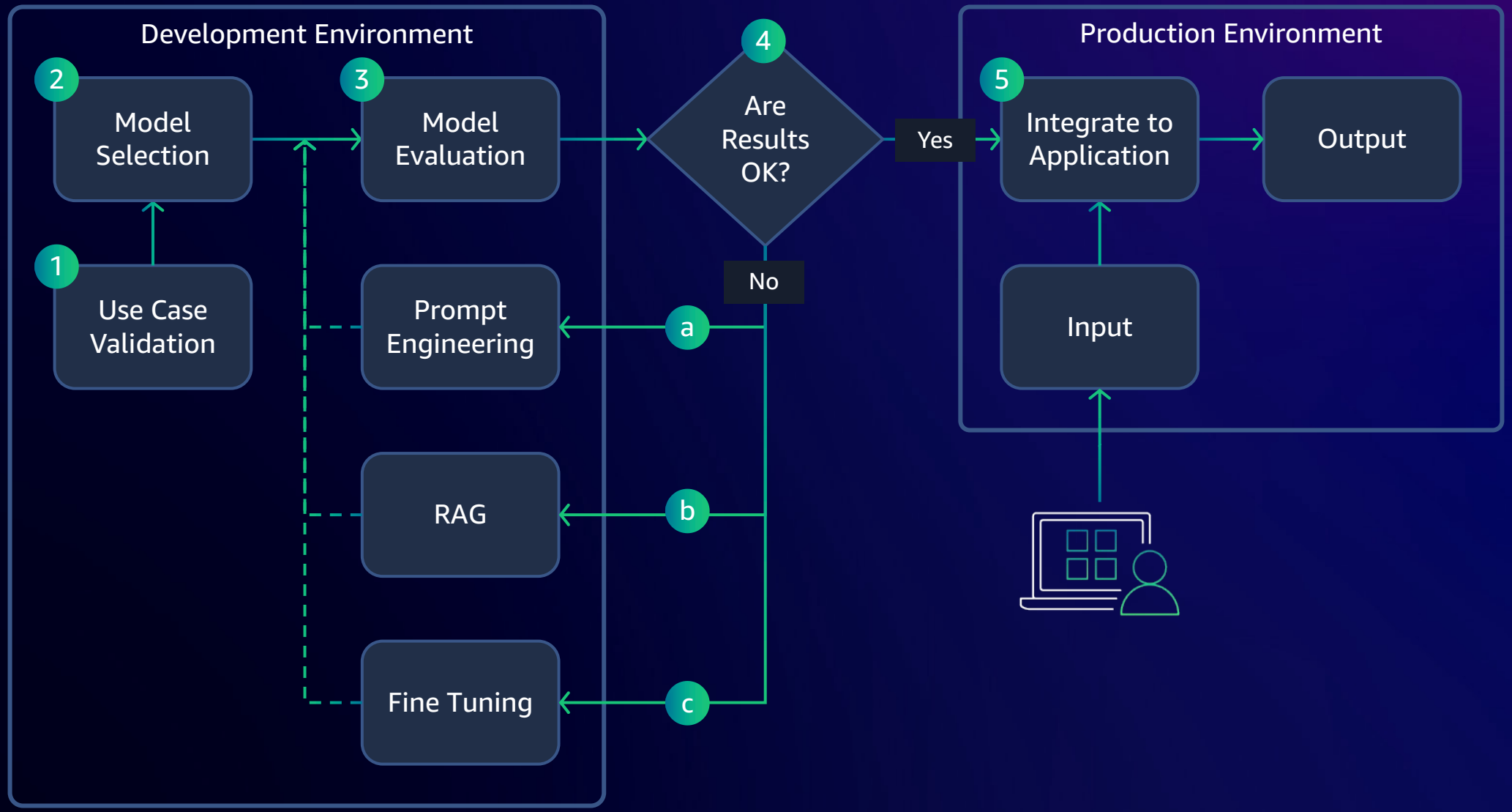
Which models should I use?

Should I train my own model?

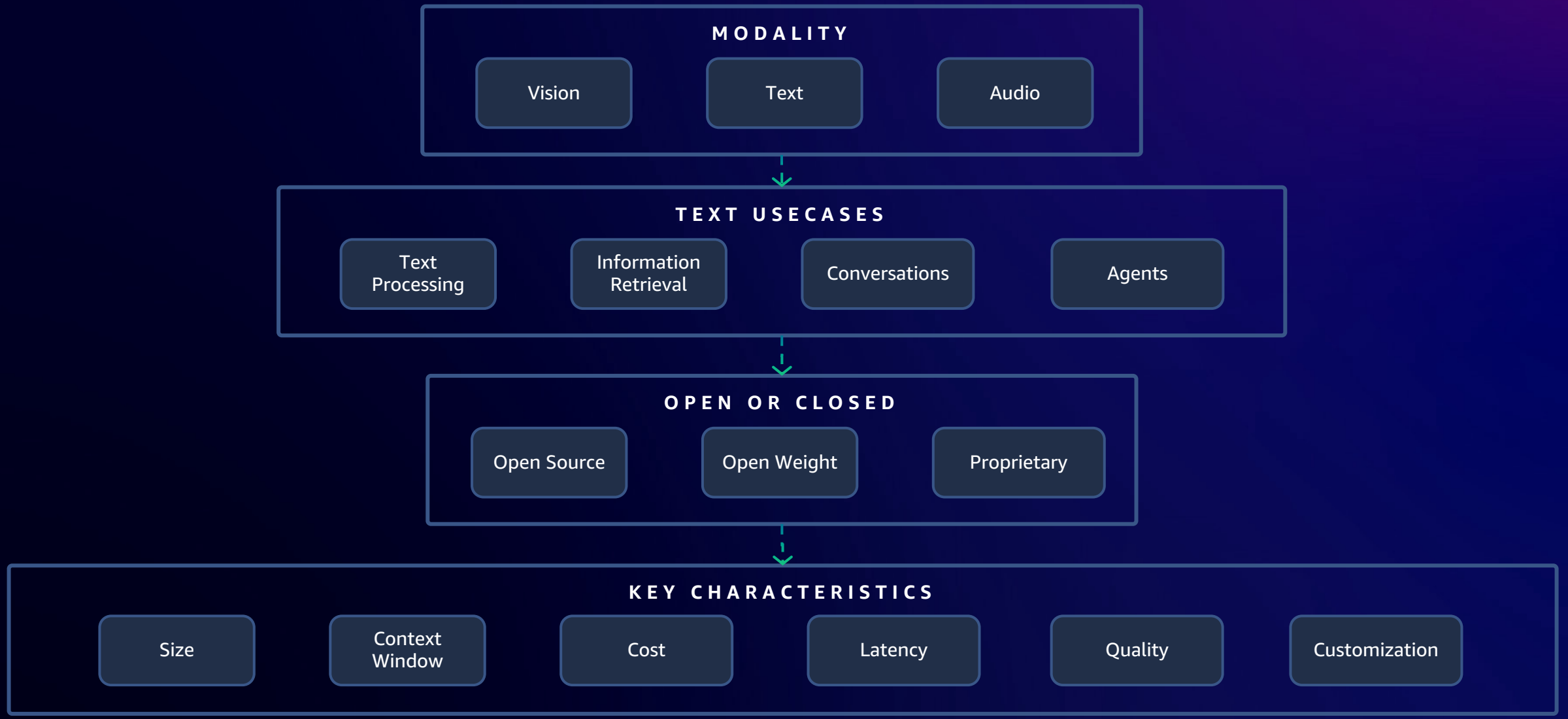
How do I manage risks?

How can we move faster?

Generative AI lifecycle



Foundation model selection framework



A more personal virtual agent

THE OBJECTIVE

An online travel agency wants to generate personalized travel itineraries

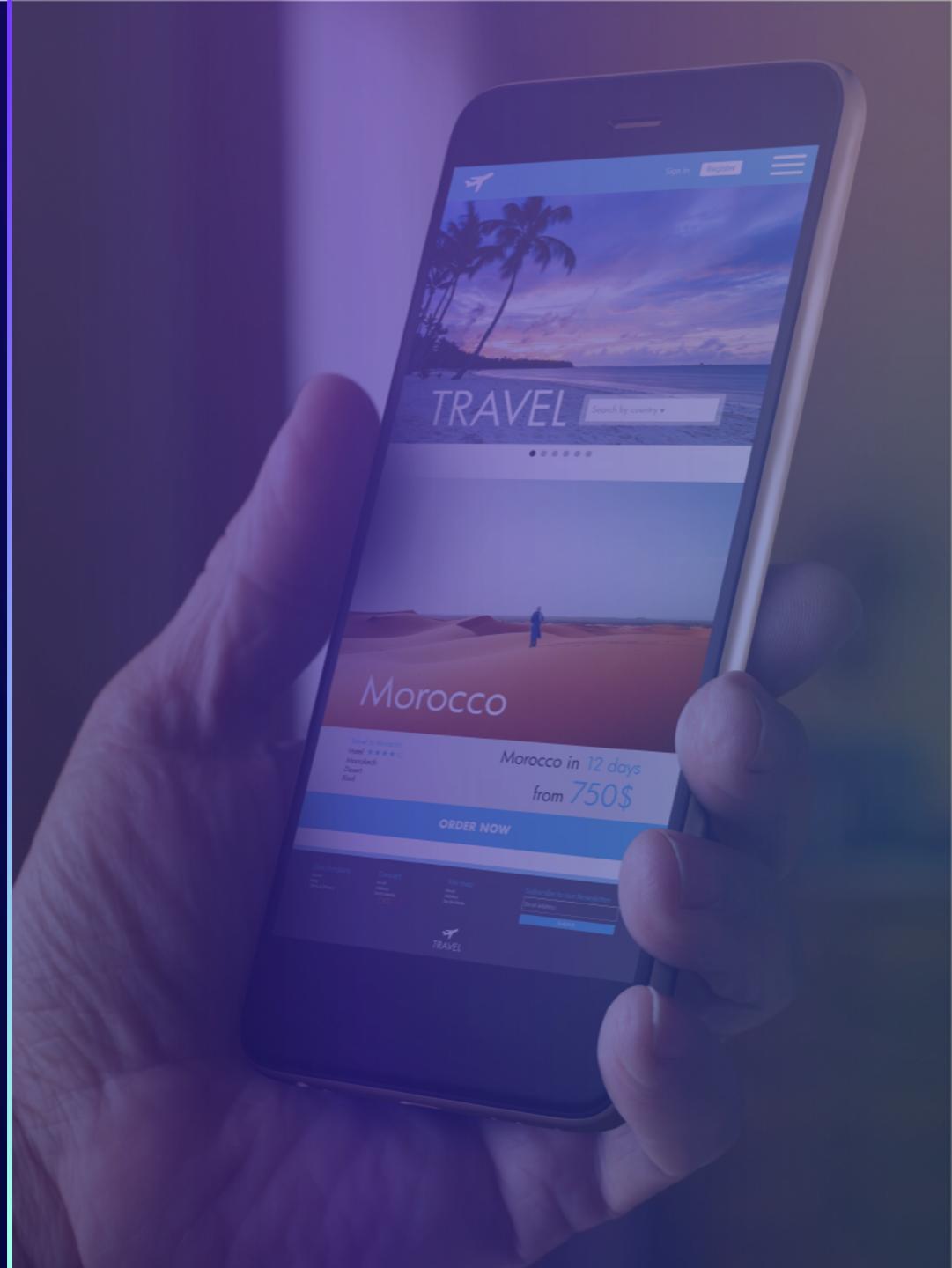
The data

CUSTOMER PROFILE

- Past trips
- Web history
- Travel preferences

COMPANY DATA

- Flight and hotel inventory
- Promotions
- Similar traveler details

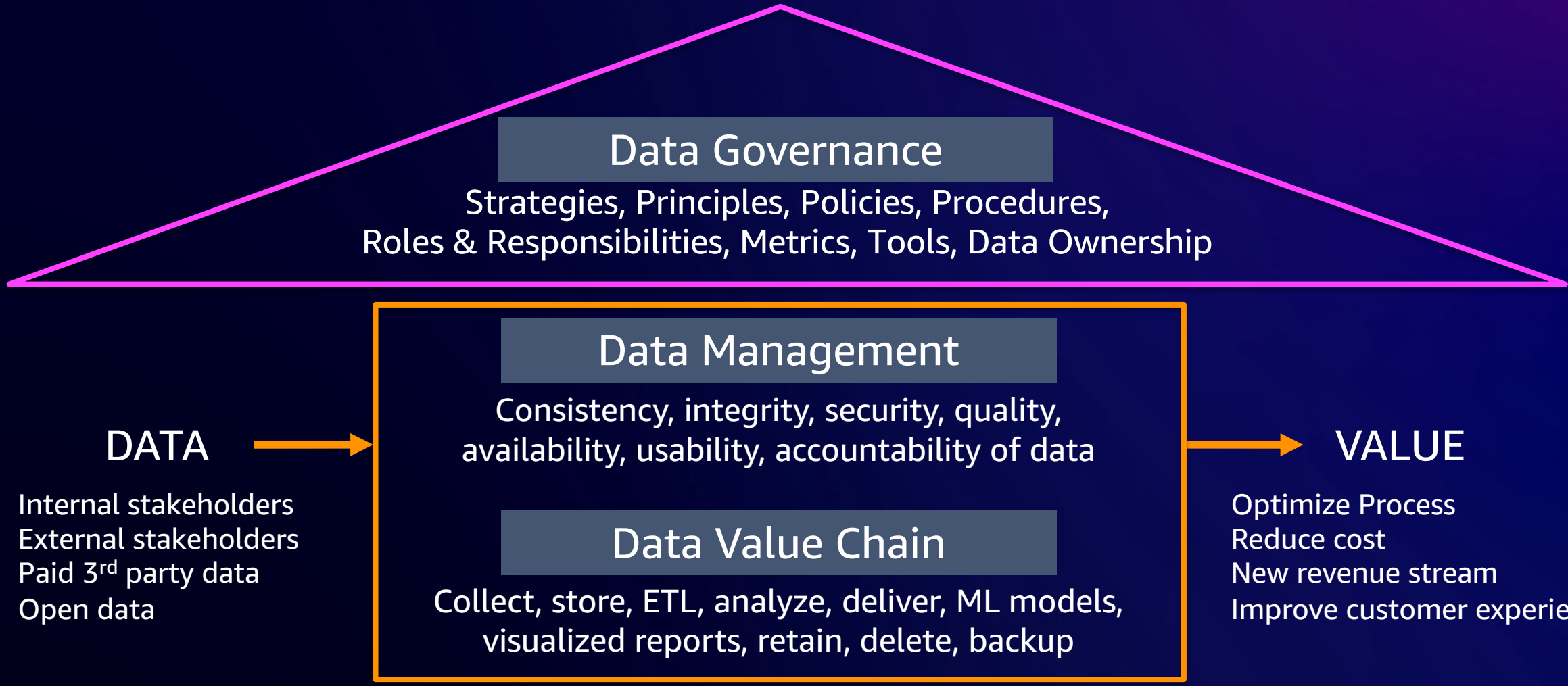


Data strategy

Value from data

DATA → ??? → VALUE

Data management



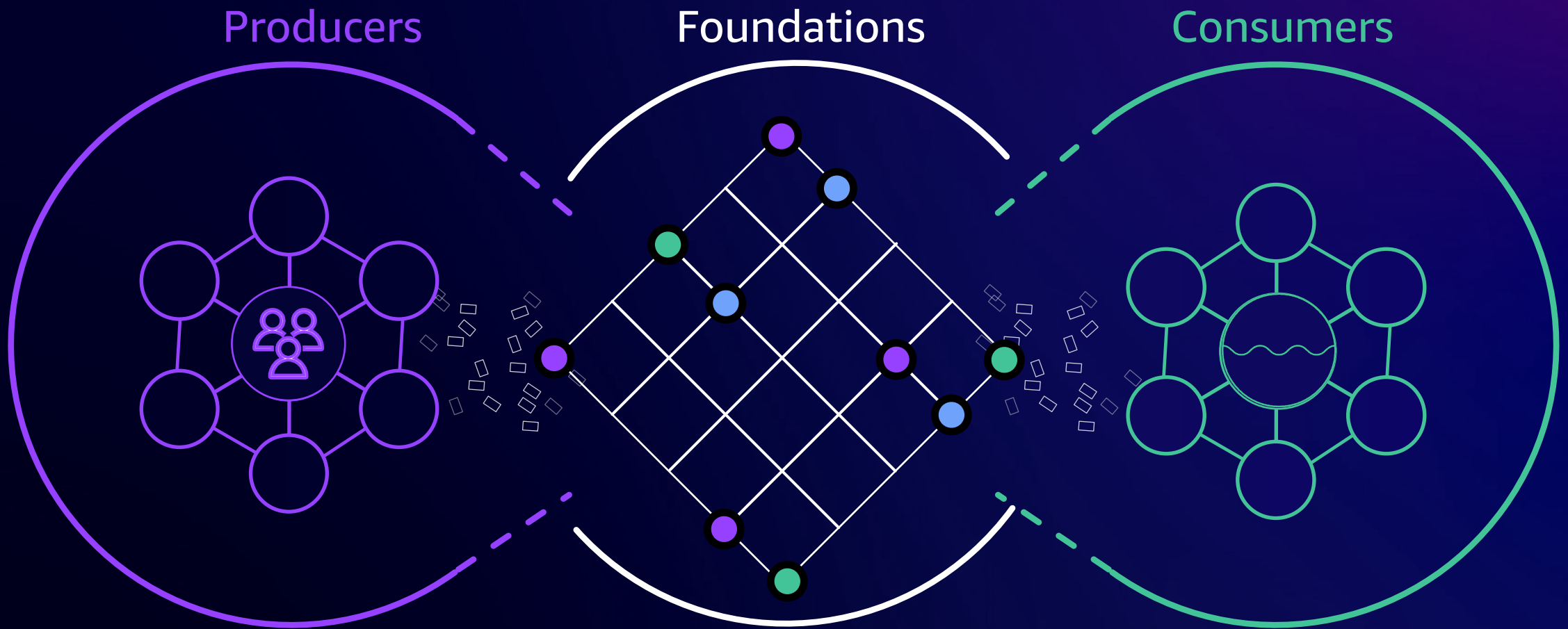
Deriving insights from data can be challenging

Data silos

People silos

Business silos

End-to-end data strategy



Domain experts

Build curated data products

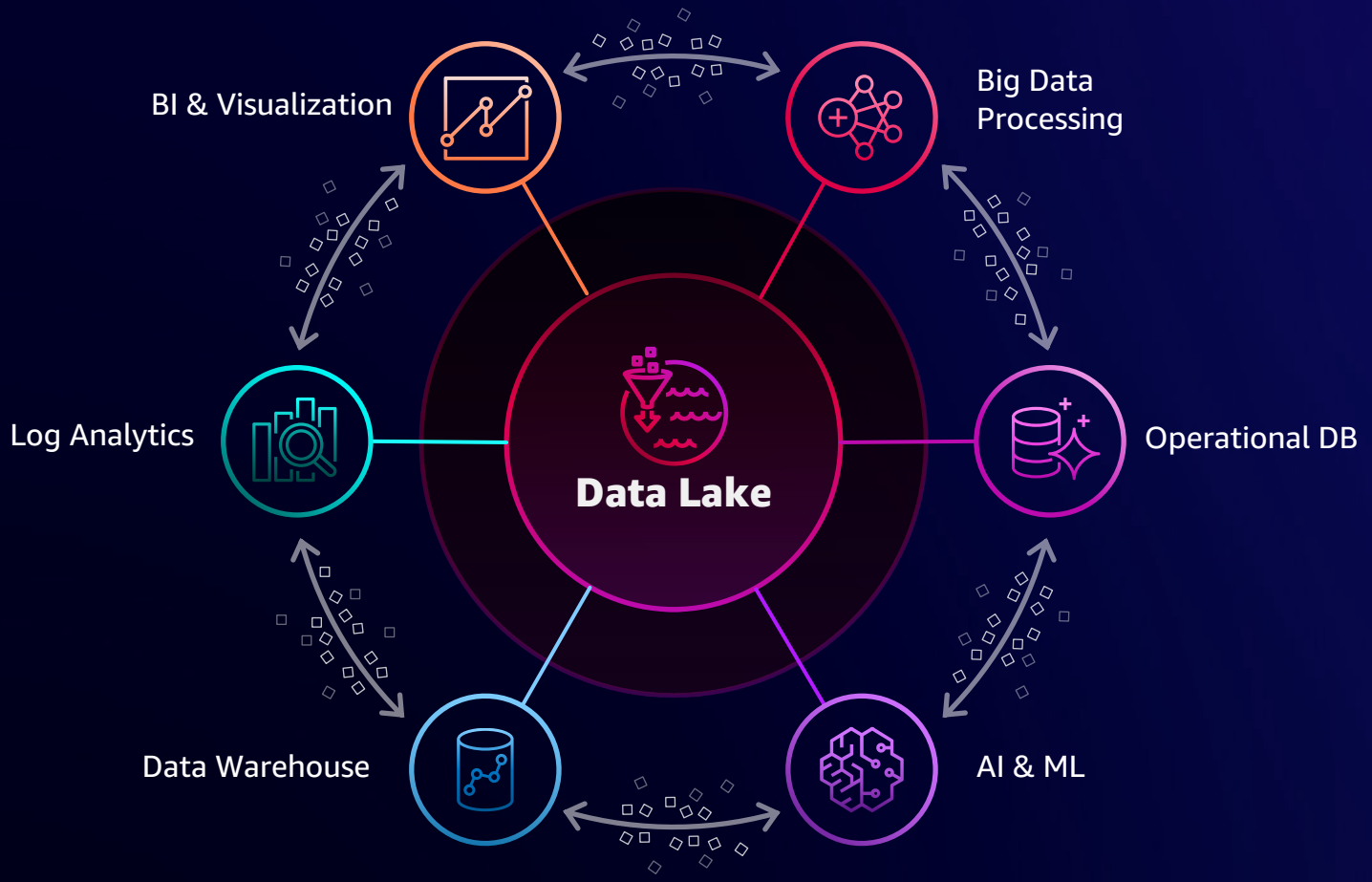
Central IT

Select and deploy tools

Business users

Drive business results

Modern Data Strategy



TECHNOLOGY

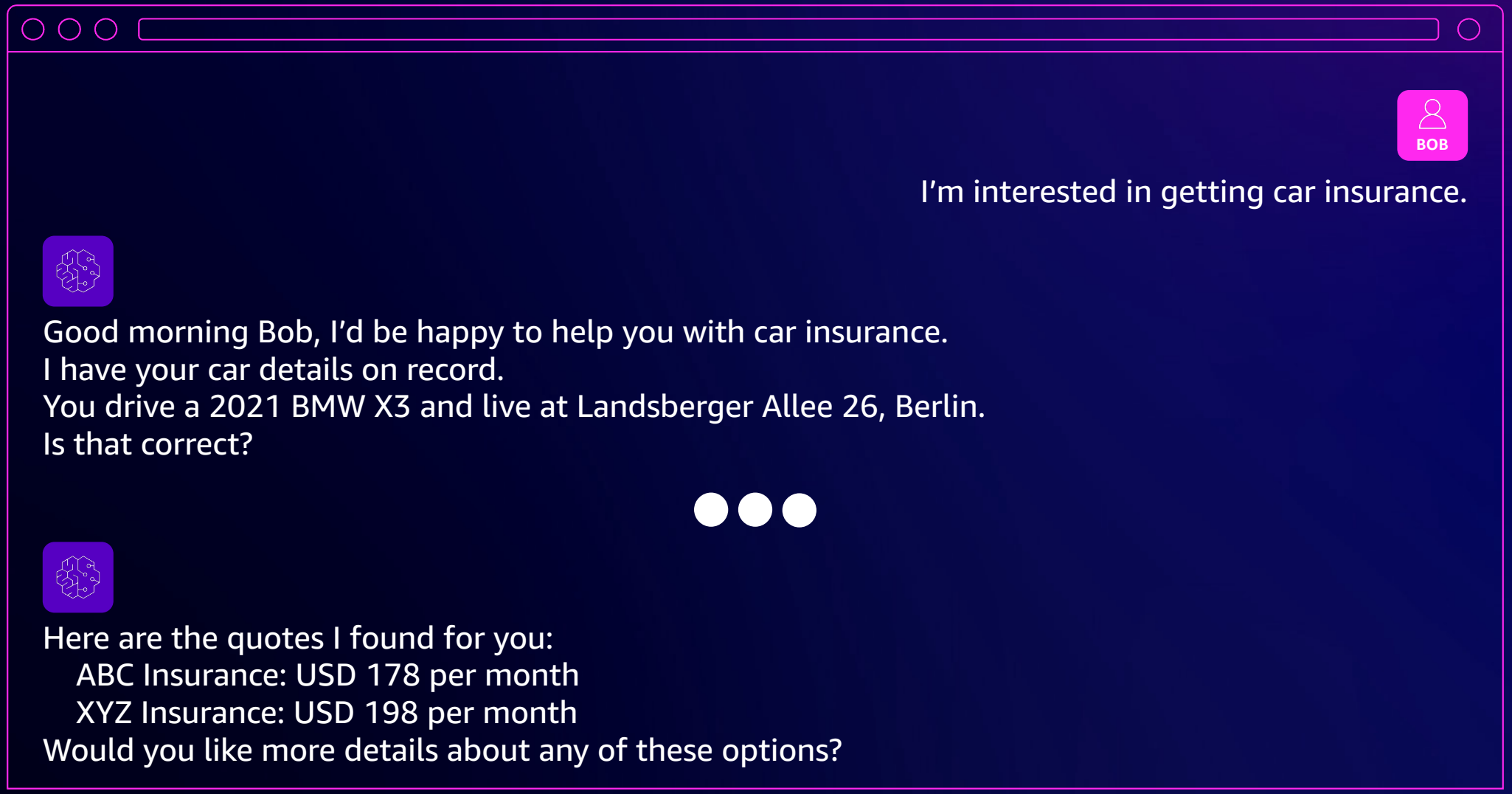
PEOPLE

PROCESS

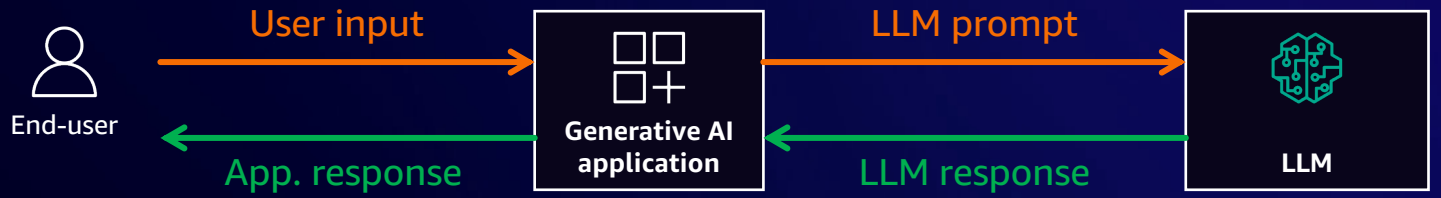
MINDSET

How to provide your data to generative AI applications?

Example generative AI interaction



Basic generative AI application



Prompt engineering behind the example

User input

“I’m interested in getting car insurance.”

Prompt engineering behind the example

User input

Prompt engineering
Using prompt template, context from structured data sources, and semantic context

```
You are a conversation agent for the
HorizonGuard insurance marketplace.
Responses to questions are written to
be helpful and inform in a unbiased
manner. Ask human to confirm
responses.
```

Instructions for the model

"I'm interested in getting car insurance."

Prompt engineering behind the example

User input

Prompt engineering
Using prompt template, context from structured data sources, and semantic context

"I'm interested in getting car insurance."

```
You are a conversation agent for the HorizonGuard insurance marketplace. Responses to questions are written to be helpful and inform in a unbiased manner. Ask human to confirm responses.

The human is Bob, lives at 123 Main St, Anytown, FL, USA and owns a 2021 BMW X3. Household members include Sarah (wife) and Jake (son).
```

Instructions for the model

Situational context

Prompt engineering behind the example

User input

Prompt engineering
Using prompt template, context from structured data sources, and semantic context

"I'm interested in getting car insurance."

```
You are a conversation agent for the HorizonGuard insurance marketplace. Responses to questions are written to be helpful and inform in a unbiased manner. Ask human to confirm responses.

The human is Bob, lives at 123 Main St, Anytown, FL, USA and owns a 2021 BMW X3. Household members include Sarah (wife) and Jake (son).

Auto insurance is one of the most used types of personal insurance. Most states require that you purchase some kind of insurance coverage to drive legally in the state. […]. Reference: https://content.naic.org/consumer/auto-insurance.htm

[…]
```

Instructions for the model

Situational context

Semantic context derived using RAG

Prompt engineering behind the example

User input

"I'm interested in getting car insurance."

Prompt engineering
Using prompt template, context from structured data sources, and semantic context

```
You are a conversation agent for the HorizonGuard insurance marketplace. Responses to questions are written to be helpful and inform in a unbiased manner. Ask human to confirm responses.

The human is Bob, lives at 123 Main St, Anytown, FL, USA and owns a 2021 BMW X3. Household members include Sarah (wife) and Jake (son).

Auto insurance is one of the most used types of personal insurance. Most states require that you purchase some kind of insurance coverage to drive legally in the state. […]. Reference: https://content.naic.org/consumer/auto-insurance.htm

[…]

I'm interested in getting car insurance.
```

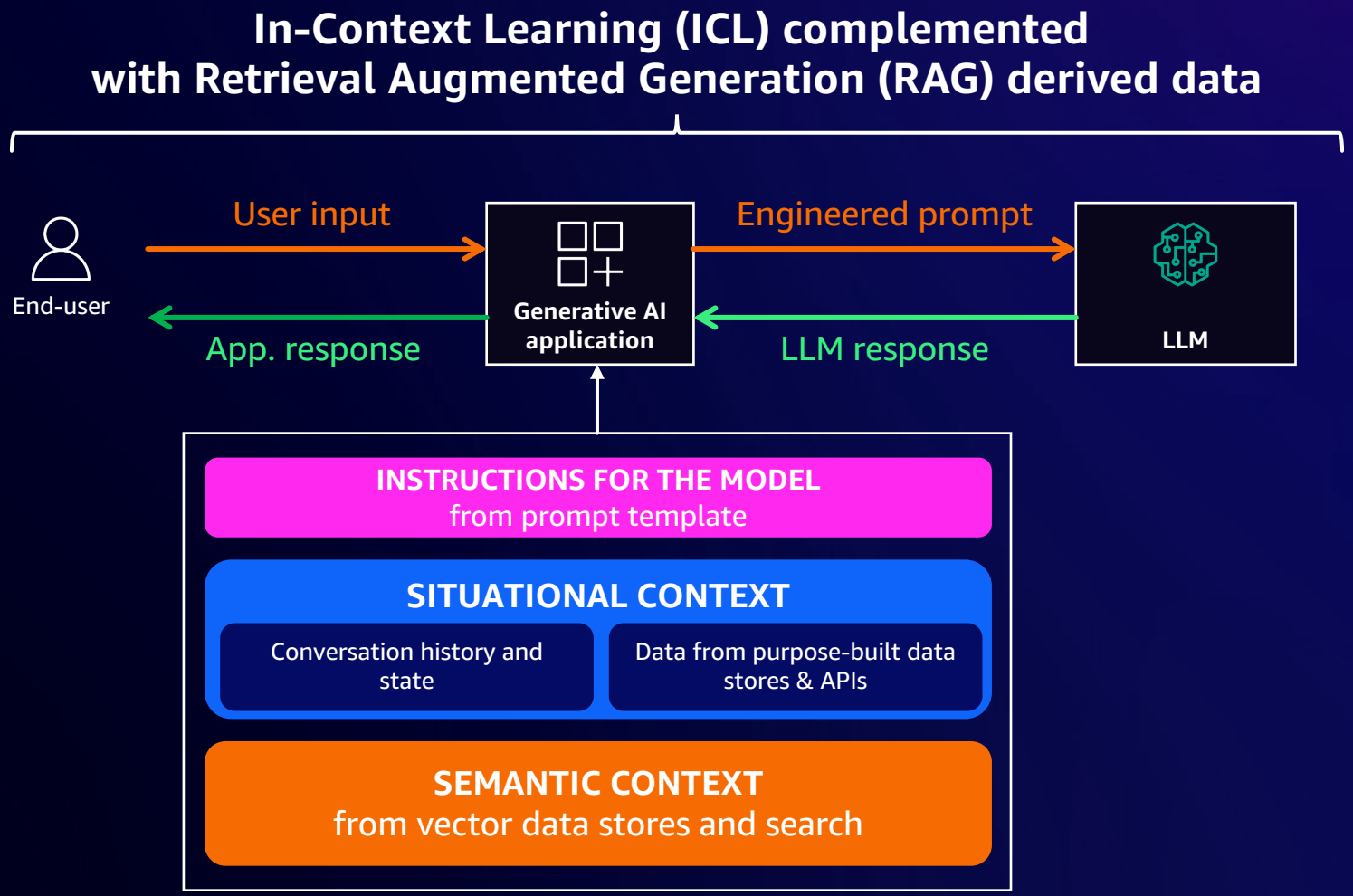
Instructions for the model

Situational context

Semantic context derived using RAG

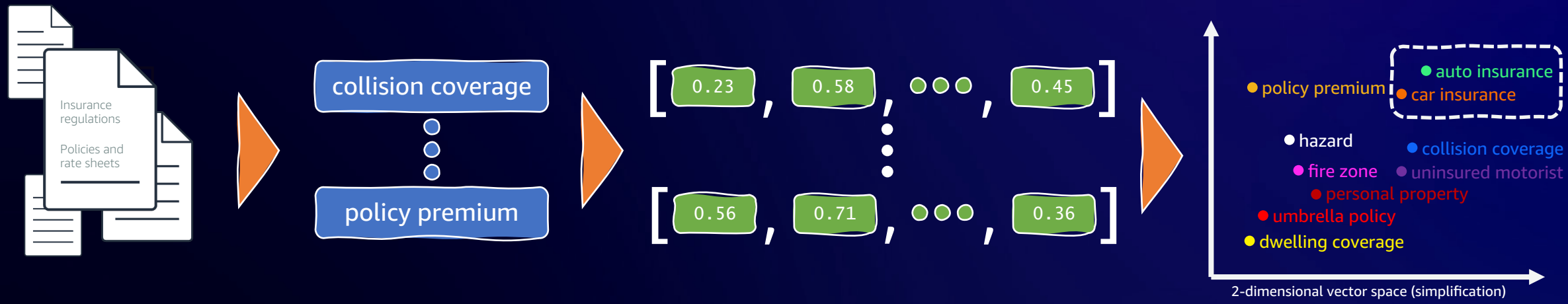
User prompt/question

Differentiating generative AI with your data



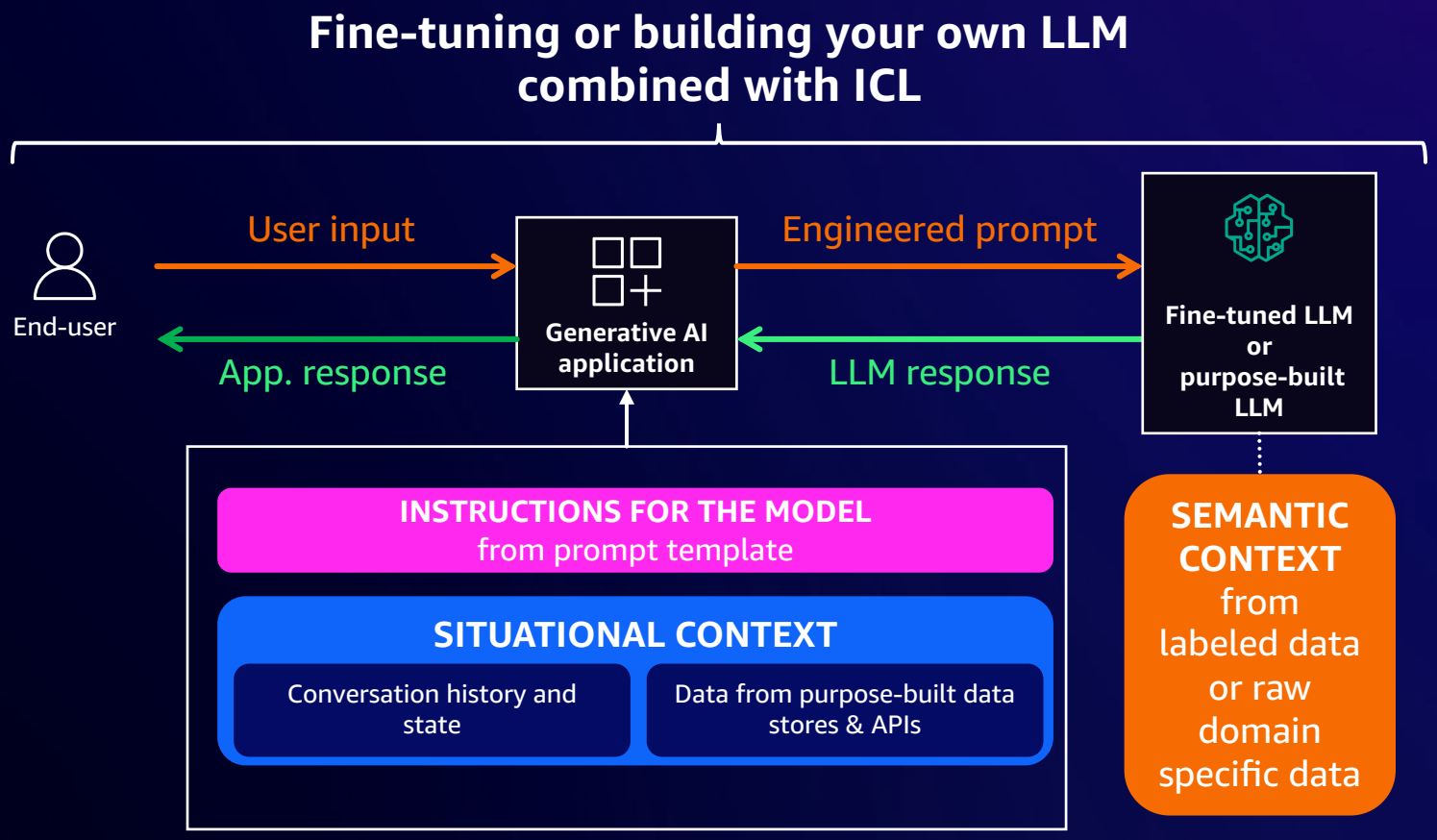
What are vector embeddings?

Source domain specific data → Tokenization → Vectorization → Store in vector datastore → Perform semantic similarity search → Include semantically similar context in prompt

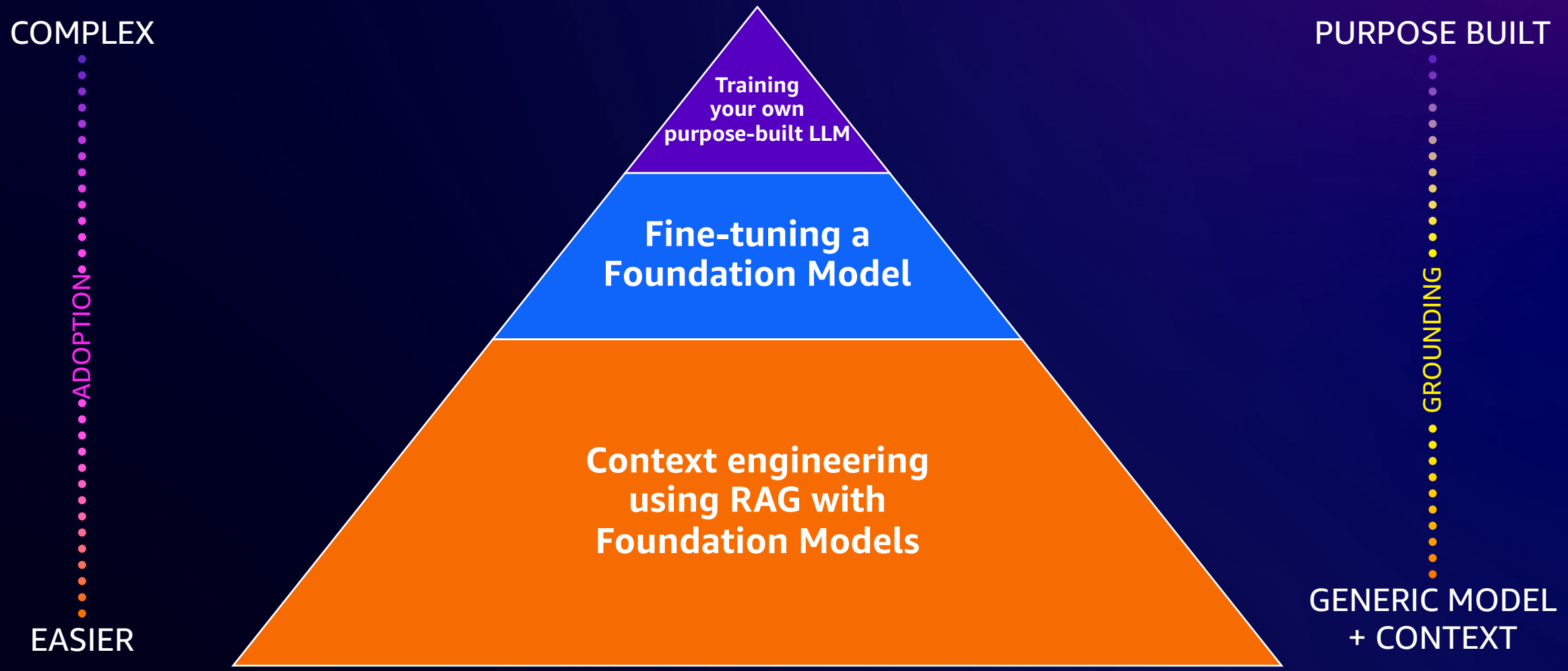


Embeddings: When vectors elements are semantic, used in Generative AI

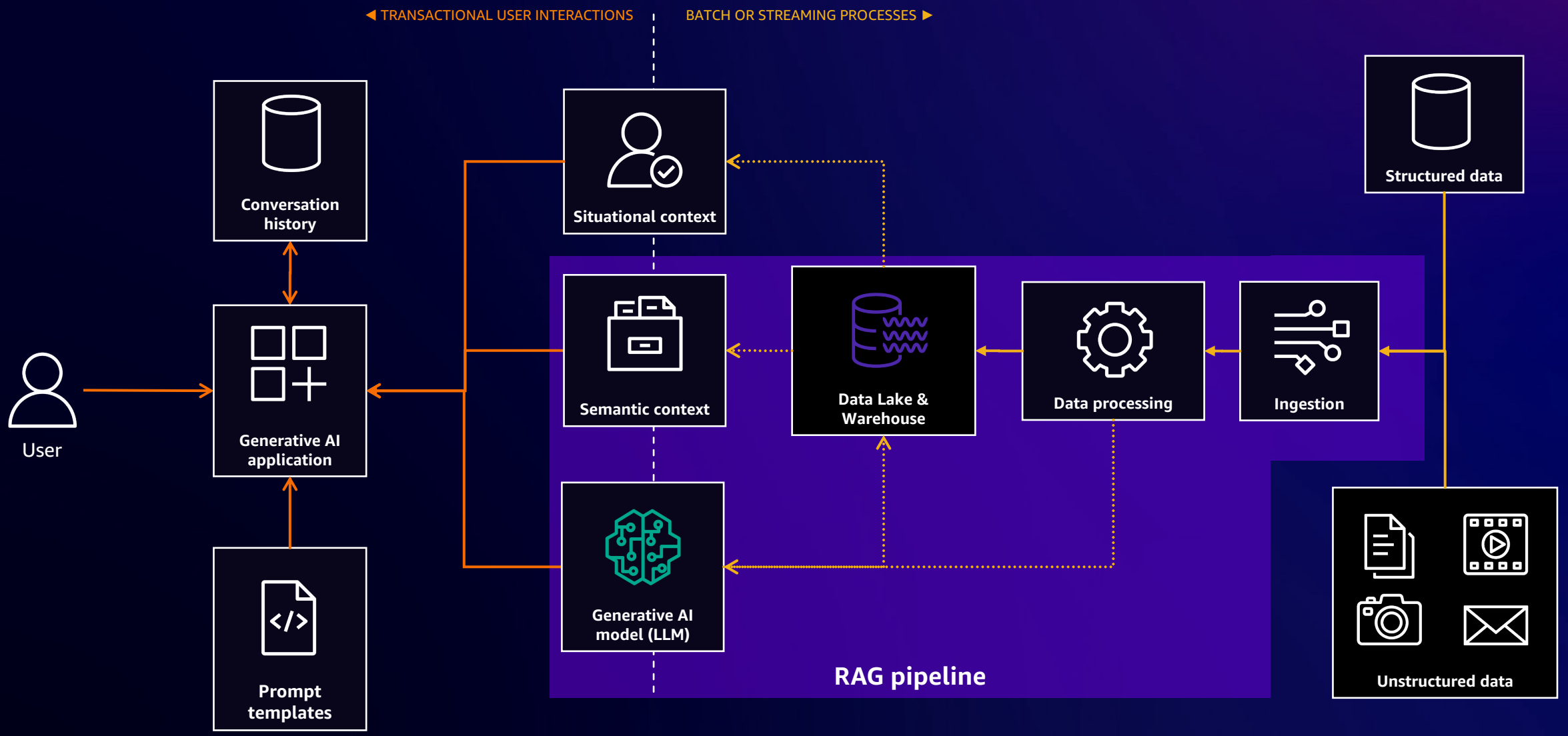
Moving the semantic context into the LLM



Generative AI implementation patterns



Data architecture for context engineering



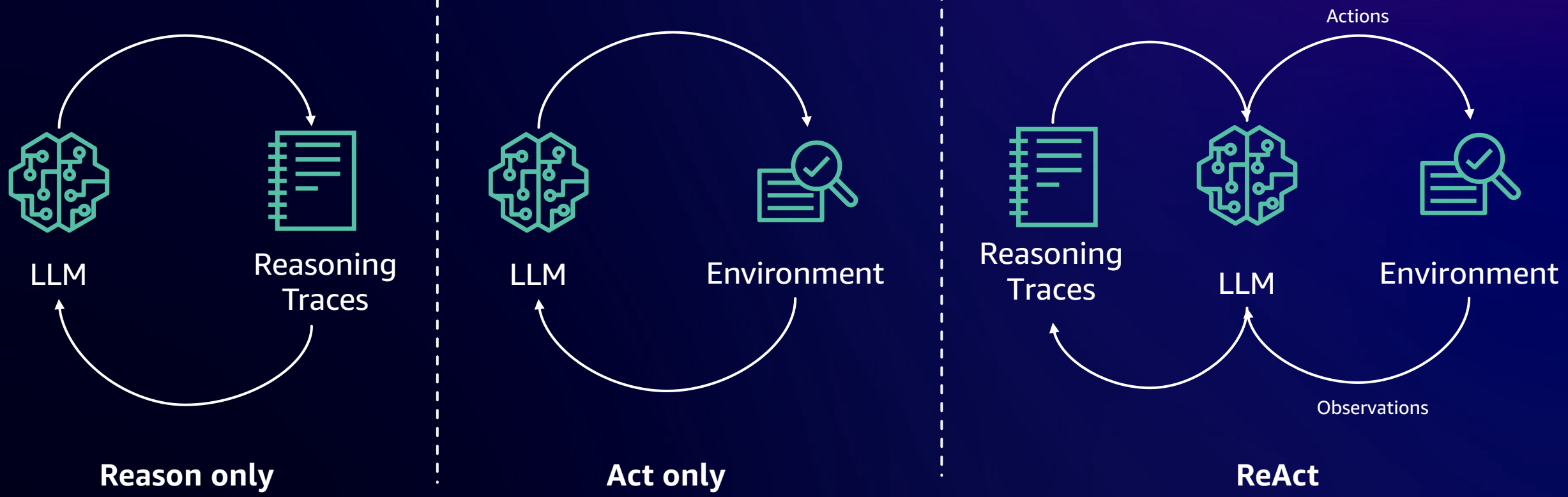
Hybrid architecture design pattern

Keep up with changes to your domain specific data sets and user profiles in near real-time.

Fine-tune your models with updated domain specific data sets occasionally.

Train a model using your domain specific data sets infrequently.

Reason and act – ReAct 🧠🛠️👀



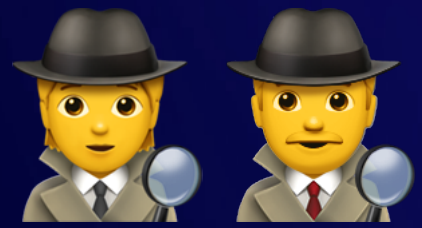
Use ReAct to create agents leveraging tools

Prompt

"How many active devices do we currently have in our fleet?"

"How do I configure your product if I want to use feature A & B at the same time?"

"Find our most popular item for January and start a 10% discount campaign for it.



ReAct Agent

Tools/API

Situational Context

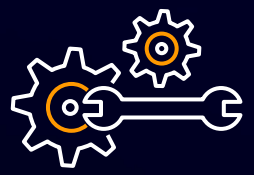
Knowledge

Semantic Context

Evolving data strategy for generative AI

Comprehensive

Manage both structured and unstructured data
Maintain both native and vector formats



Integrated

Unify from disparate data sources
Prepare data for model training, fine-tuning, vectorization

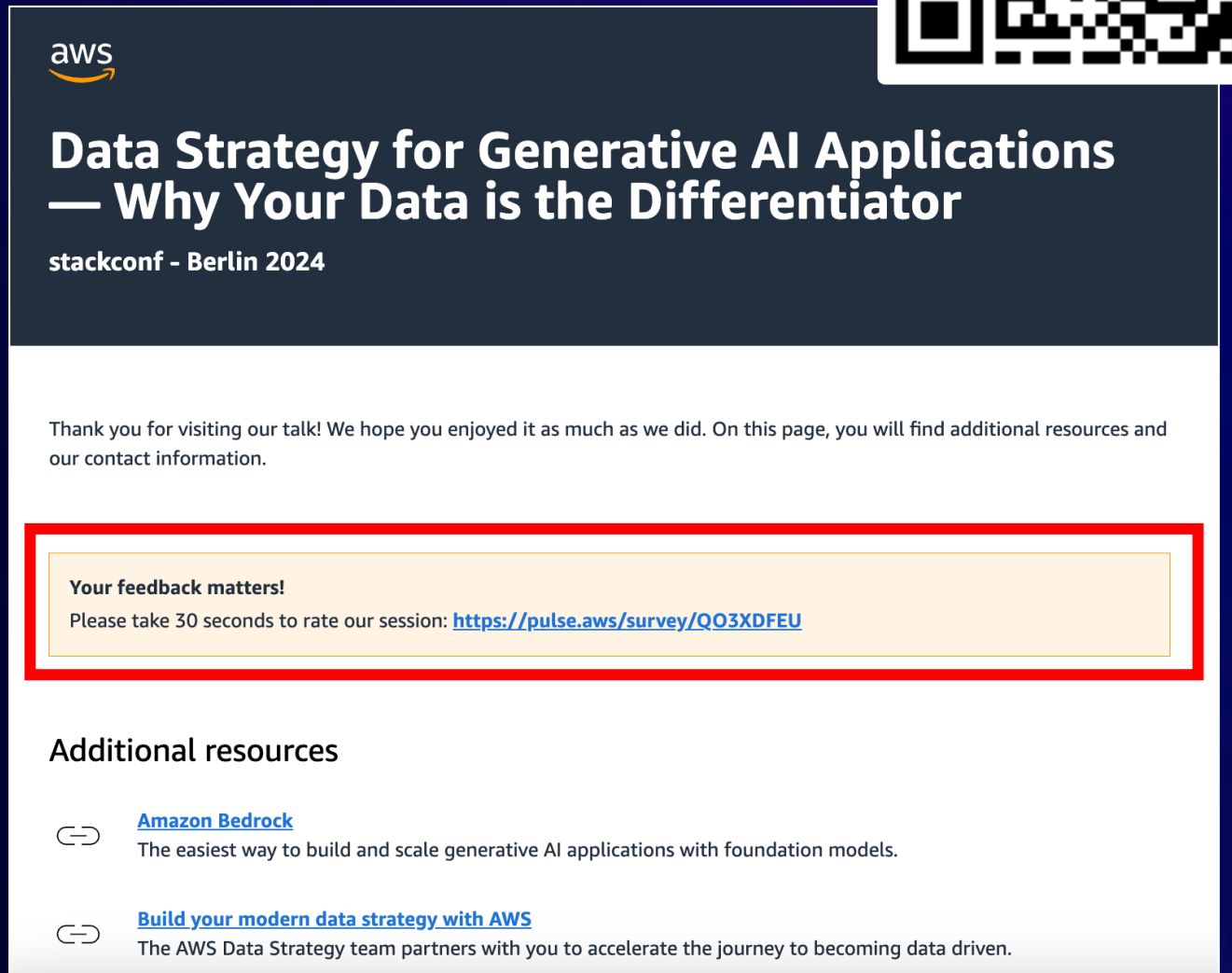
Governed

Expand data security and compliance to prompts and completions, data quality and Responsible AI

Thank you!

Manuel Heinkel
Solutions Architect
AWS

Puria Izady
Solutions Architect
AWS



aws

Data Strategy for Generative AI Applications — Why Your Data is the Differentiator

stackconf - Berlin 2024

Thank you for visiting our talk! We hope you enjoyed it as much as we did. On this page, you will find additional resources and our contact information.

Your feedback matters!
Please take 30 seconds to rate our session: https://pulse.aws/survey/QO3XDFEU

Additional resources

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

Build your modern data strategy with AWS
The AWS Data Strategy team partners with you to accelerate the journey to becoming data driven.