AWS State, Local, and Education Learning Days

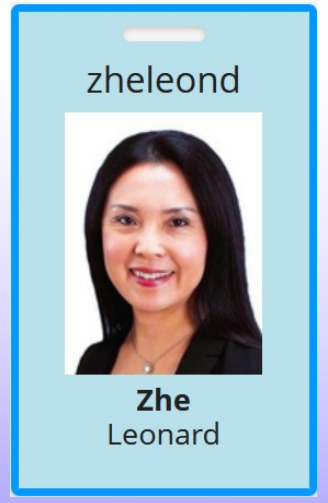
Phoenix

aws Learning Days
State, Local, and Education

AWS Data Foundations for AI in public sector

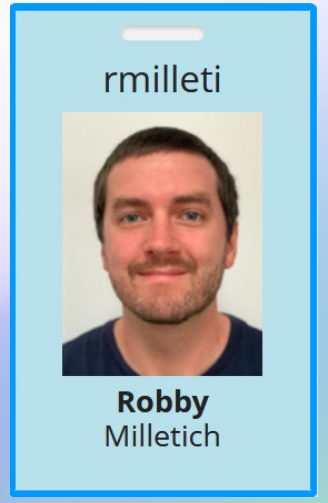
Zhe Leonard

Senior Analytics Specialist
Amazon Web Services
zheleond@amazon.com



Robby Milletich

Senior Applied Scientist
Amazon Web Services
rmilleti@amazon.com



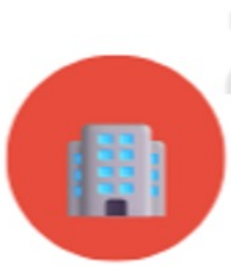
aws Learning Days
State, Local, and Education

Agenda

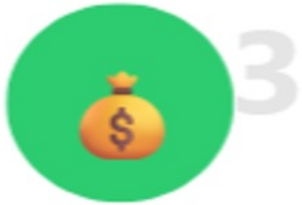
1. AI Trend, Challenges, Key Considerations
2. Freedom to Invent
3. Maximize Value
4. Trusted Data for Trusted AI

The Three Key AI Trends

AI Exposing Data Challenges



AI Becoming Institutionalized



Demand for Measurable ROI

Enterprises are doubling down on agents

33%

of enterprise software apps will include agentic AI by 2028, up from less than 1% in 2024.

Gartner, “Top strategic Technology Trends for 2025,” October 2024

15%

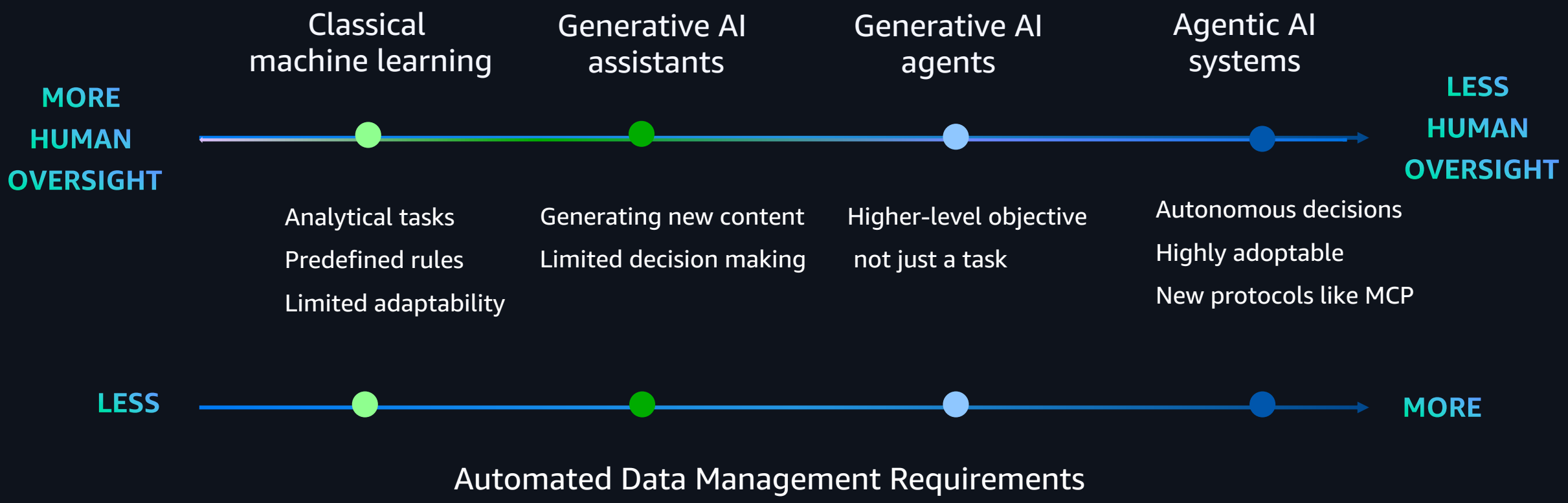
of day-to-day work decisions will be made autonomously through agentic AI by 2028.

Gartner, “Top Strategic Technology Trends: Agentic AI—the Evolution of Experience” February 2025

What are AI Agents?

Autonomous software system leverage AI to reason, plan, and complete tasks on behalf of humans or systems

AI gets more autonomous with Agents



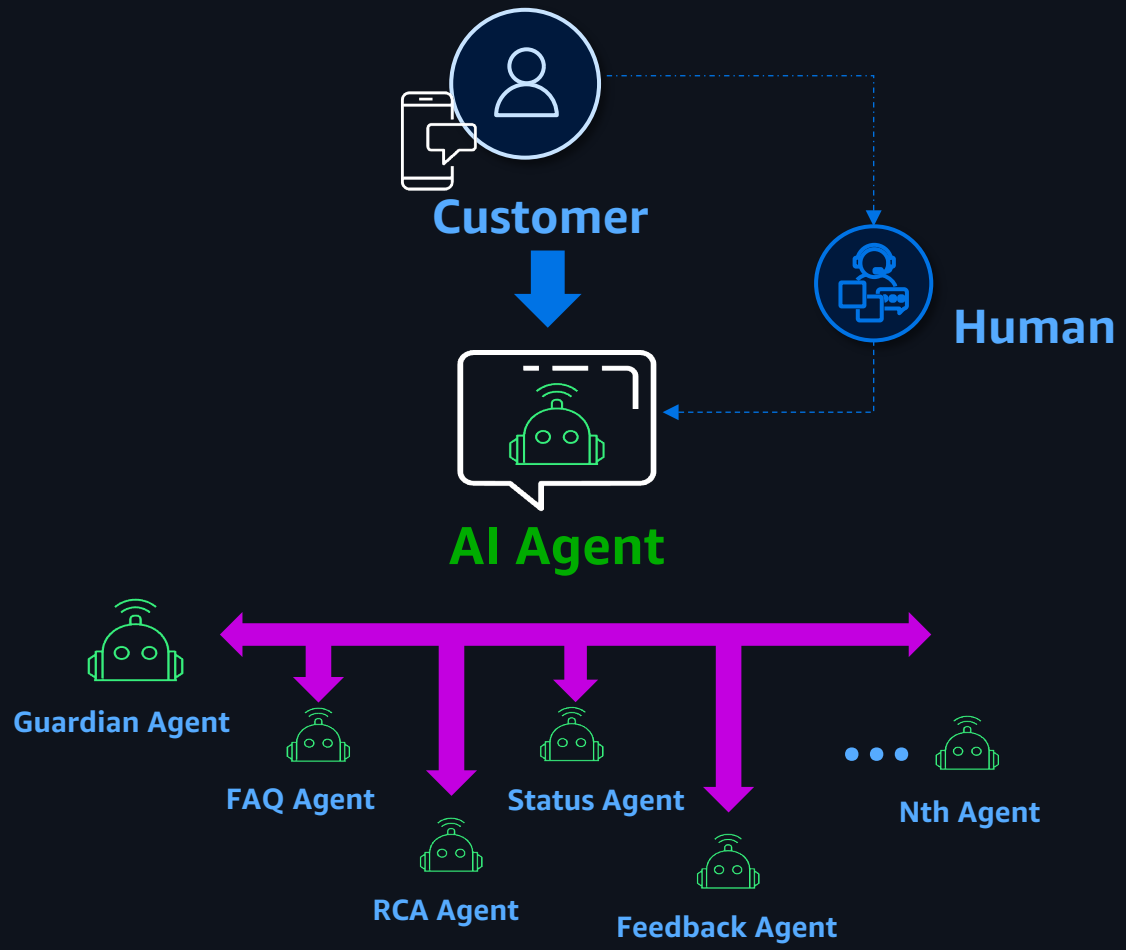
Experimentation is easy, enterprise scale, and monetization is hard

Over

60%

of AI projects will be canceled by the end of 2026, due to enterprise data is not AI-ready

(source: Gartner, June 2025)



Data is the foundation for your AI

AI Application and agent orchestration

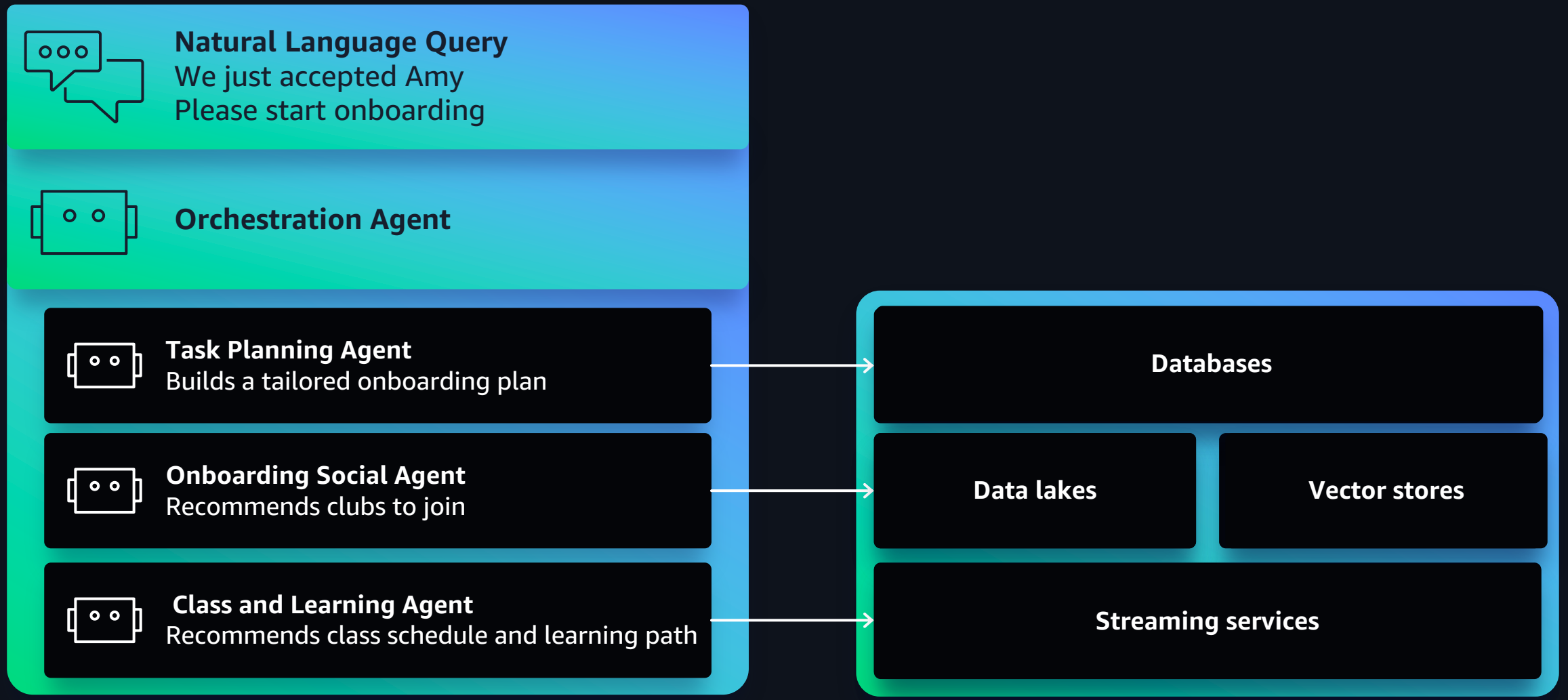
Data Foundation

DATA STORAGE& MANAGEMENT

DATA PROCESSING & TRANSFORMATION

DATA CATALOG & GOVERNANCE

How data supports your AI applications



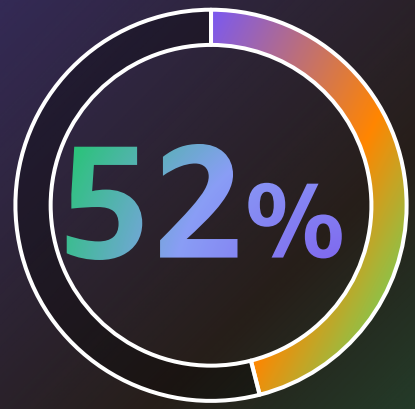
Building Trust in AI starts with your data foundation & people

Accuracy

Open and secure access

Reliability and auditability

People and process



of CDOs view data foundation not being ready for AI

Source: McKinsey State of the AI 2025 report.

Data makes or breaks your AI success

Lead with AWS – your trusted data partner since the dawn of the cloud

Freedom to invent

Value

Trusted data for trusted AI

aws

19+

years of innovation

Millions

of customers innovating with data and AI

Exabytes

of data for machine learning

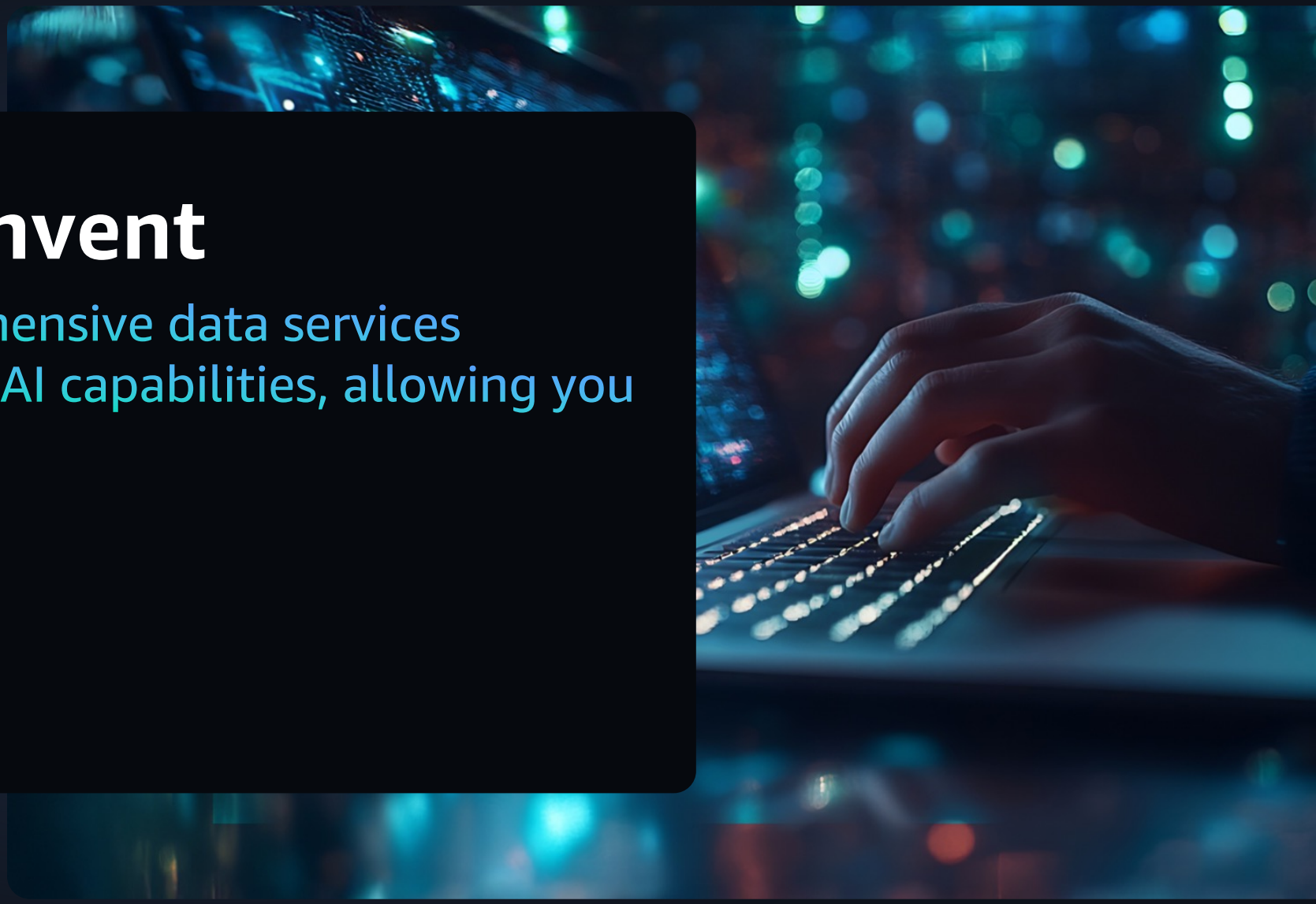
aws

Agenda

1. AI Trend, Challenges, Key Considerations
2. Freedom to Invent
3. Maximize Value
4. Trusted Data for Trusted AI

Freedom to invent

AWS provides comprehensive data services extended with agentic AI capabilities, allowing you to innovate freely



AWS comprehensive data services are AI ready

Built-in vector support

Cost-efficient vector store at massive scale

Performant databases

Fast-creation
Instant scale
Single-digit millisecond latency

End-to-end streaming pipelines

Real-time
Auto-scale

Intelligent processing

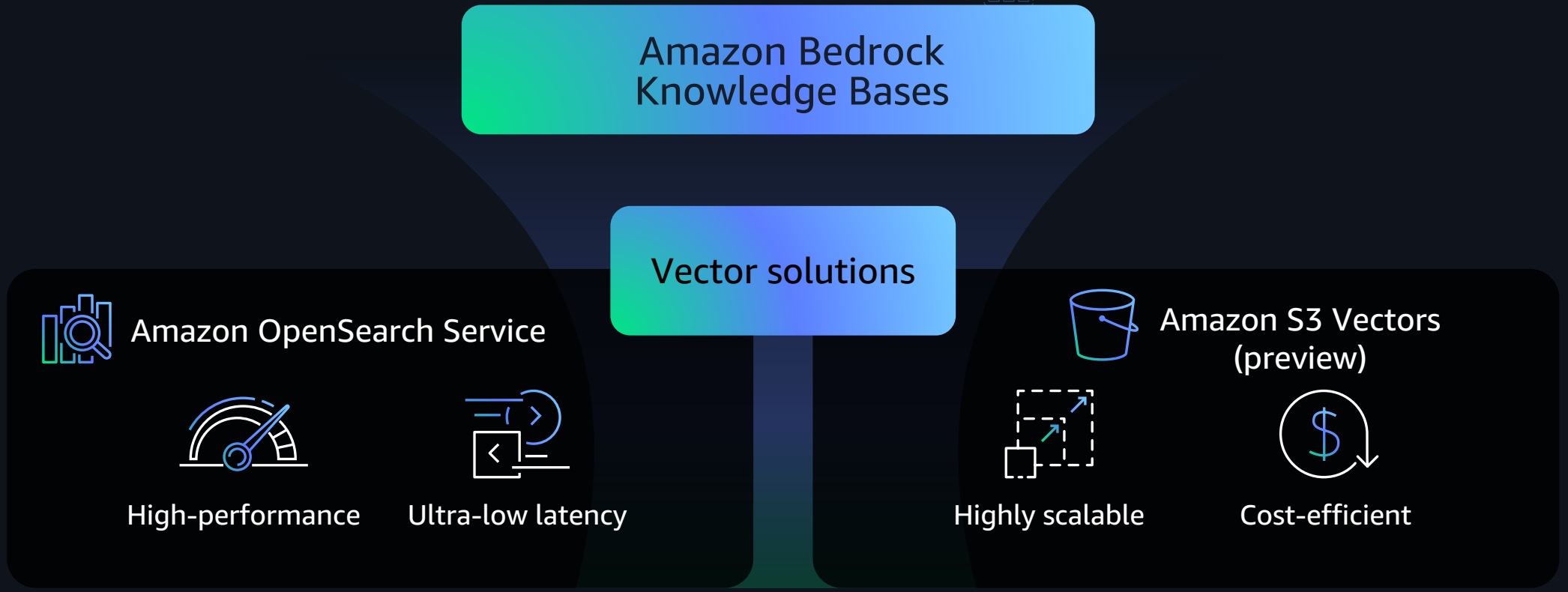
Simplified data integration
Flexible support

Unmatched price-performance

Effortless scale
ZeroETL integrations
AI-boosted

Easily extend your data architecture to support genAI and agentic AI

Vector solutions for all your AI use cases



How might we make this even better for our customers?

THE NEXT GENERATION OF

Amazon SageMaker

The center for all your data, analytics, and AI

Unified Studio

SQL analytics
Amazon Redshift

Data processing
Amazon EMR
AWS Glue
Amazon Athena

Model development
Amazon SageMaker AI

Gen AI App development
Amazon Bedrock

COMING SOON

Streaming
Amazon MSK
Amazon Kinesis

COMING SOON

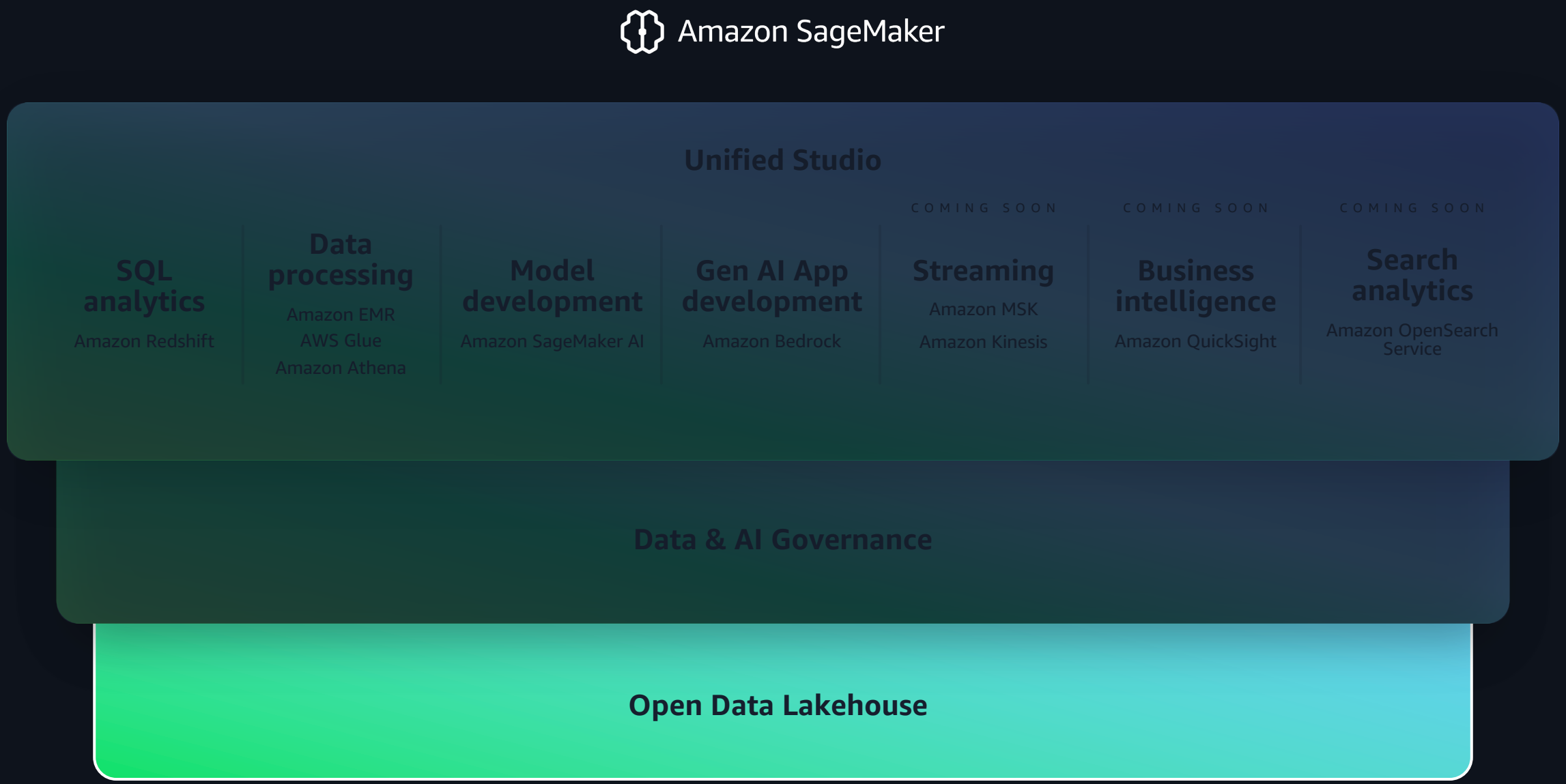
Business intelligence
Amazon QuickSight

COMING SOON

Search analytics
Amazon OpenSearch Service

Data & AI Governance

Open Data Lakehouse



Unified Studio

Data & AI Governance

Apache Iceberg compatible compute engines/3rd party applications

Iceberg APIs

LAKEHOUSE

Iceberg APIs

Amazon S3
data lakes

Amazon Redshift
data warehouses

Fine-grained access control

Zero-ETL integrations

Aurora	DynamoDB	RDS	Streaming data – MSK, Kinesis	OpenSearch vector data
ServiceNow	Salesforce	Zoho CRM	Instagram Ads	
SAP	Salesforce Pardot	Facebook Ads	Zendesk	

Federated query connectors + 100s of AWS Glue connectors

query

query

On-premise systems and 3rd party applications

S3 Tables

Fully managed Apache Iceberg tables in S3

Built In - Gen AI for Data Transformation

Bedrock

01 Generative AI powered database queries

02 Generative AI for data analytics

03 Generative AI powered data mapping

04 Generative AI for data relationship discovery

05 Generative AI for metadata extraction

Università degli Studi di Padova

Wolters Kluwer

Harvard Business School

Institutional GenAI assistant

Moodle plug-in for course-specific chatbots and an official university GenAI website

Solution supports institution chatbot, document analysis, eLearning, historical media assets

Expand to 100,000 users to serve the entire university community

Corporate performance management powered by AI

Generative AI supports decision-making for CFO and management

Wolters Kluwer uses AWS Generative AI in “CCH Tagetik expert solution”

Benefits include enhanced analytics, financial planning, and compliance-reporting

Data management with AI-powered unified platform

Unified Data Platform: Integrated previously siloed systems into a single platform

AI for curriculum and alumni insights and created a collaborative data science environment

The project included automated publication mining and staff training in data analytics and ML Ops

aws

Customer	Battelle
Industry	Nonprofit
Use Case	RFP Assistant
Country	United States

Customer Profile:
Battelle is the world's largest independent nonprofit applied science and technology organization, leveraging 90 years of innovation to solve complex scientific challenges for a diverse range of clients, from startups to multinational corporations and government agencies.

Problem

- Battelle frequently responds to complex Requests for Proposals (RFPs) from government agencies and institutions, requiring the comprehensive extraction of proposal requirements in order to respond quickly and completely.
- Currently, Battelle's proposal managers manually review RFPs for requirements, a time-consuming process that can take over a week to complete.
- Previous attempts to automate this process using GPT-4 faced challenges accurately distinguishing true requirements from false positives.

Solution

Battelle collaborated with the AWS Generative AI Innovation Center to:

- Develop an AI tool to automatically extract relevant requirements and evaluation factors from RFPs
- Generate an initial draft compliance matrix to reduce manual effort and increase completeness.

- *Services Used*: Amazon Bedrock, Amazon Textract, Amazon S3

Impact

- ***Improved accuracy:*** Identifying true requirements and evaluation criteria from RFPs, with a low false positive rate.
- ***Time savings:*** Reducing the time to generate compliance matrices from days down to minutes.
- ***Scalable, extensible solution:*** Flexibility to handle increasing RFP volumes and evolve to new RFP formats.

aws

Customer	MDRC
Industry	Nonprofit
Use Case	Research Chatbot
Country	United States

Customer Profile:
MDRC is a US nonprofit social policy research organization committed to improving the lives of people with low incomes. They design promising new interventions, evaluate existing programs, and provide technical assistance to build better programs.

Problem

- MDRC conducts rigorous studies on programs and policies affecting low-income populations, disseminates findings to policymakers and practitioners, and works directly with programs to improve effectiveness.
- MDRC wanted to develop a question-answering system using their research publications as a knowledge base, with the primary use case being a research assistant that can answer questions and provide citations from their work to accelerate research analysis.

Solution

MDRC partnered with the Generative AI Innovation Center to build a virtual research assistant that:

- Understands complex queries, retrieve information from MDRC's publications and provides well-cited responses tailored to research needs.
- Establishes a secure, extensible framework to expand knowledge bases to other research domains.
- *Services Used*: Amazon Bedrock, Amazon OpenSearch, Amazon S3, Amazon ECS

Impact

- ***Accelerated analysis and decision making:*** The AI assistant enables MDRC's stakeholders to quickly find relevant insights and citations from the organization's extensive research publications.
- ***Increased Scalability:*** As MDRC's research scope expands, the solution can grow alongside it, ensuring that researchers always have access to the most up-to date information.
- ***Secure AWS-native integration:*** The solution seamlessly integrates with MDRC's existing AWS infrastructure, ensuring data privacy, security, and governance while providing scalability and cost-efficiency.

Customer	Federal Agency
Industry	US Government
Use Case	Investigation Support
Country	United States

Problem

- Analysts have access to 100s of databases. Each database has 100s of tables with unique, complicated, and unintuitive schemas.
- It is very difficult for new analysts to productively use all of this data. Tenured analysts primarily use a few datasets they know well and use long hand-crafted queries.
- Given the steep learning curve of new data sources, relevant datasets go unused and institutional knowledge remains largely unavailable to junior analysts.

Solution

- Human-in-the-loop agentic system that utilizes customer APIs and data sources to:
  - (1) On-board databases for improved downstream performance
  - (2) Triage queries to appropriate database and table
  - (3) Build SQL queries from the analytic questions
  - (4) Describe SQL query plan for human approval,
  - (5) Incorporate human feedback to better align query
  - (6) Run query and retrieve data
- *Services Used*:
  - Amazon Bedrock
  - AWS Lambda

Impact

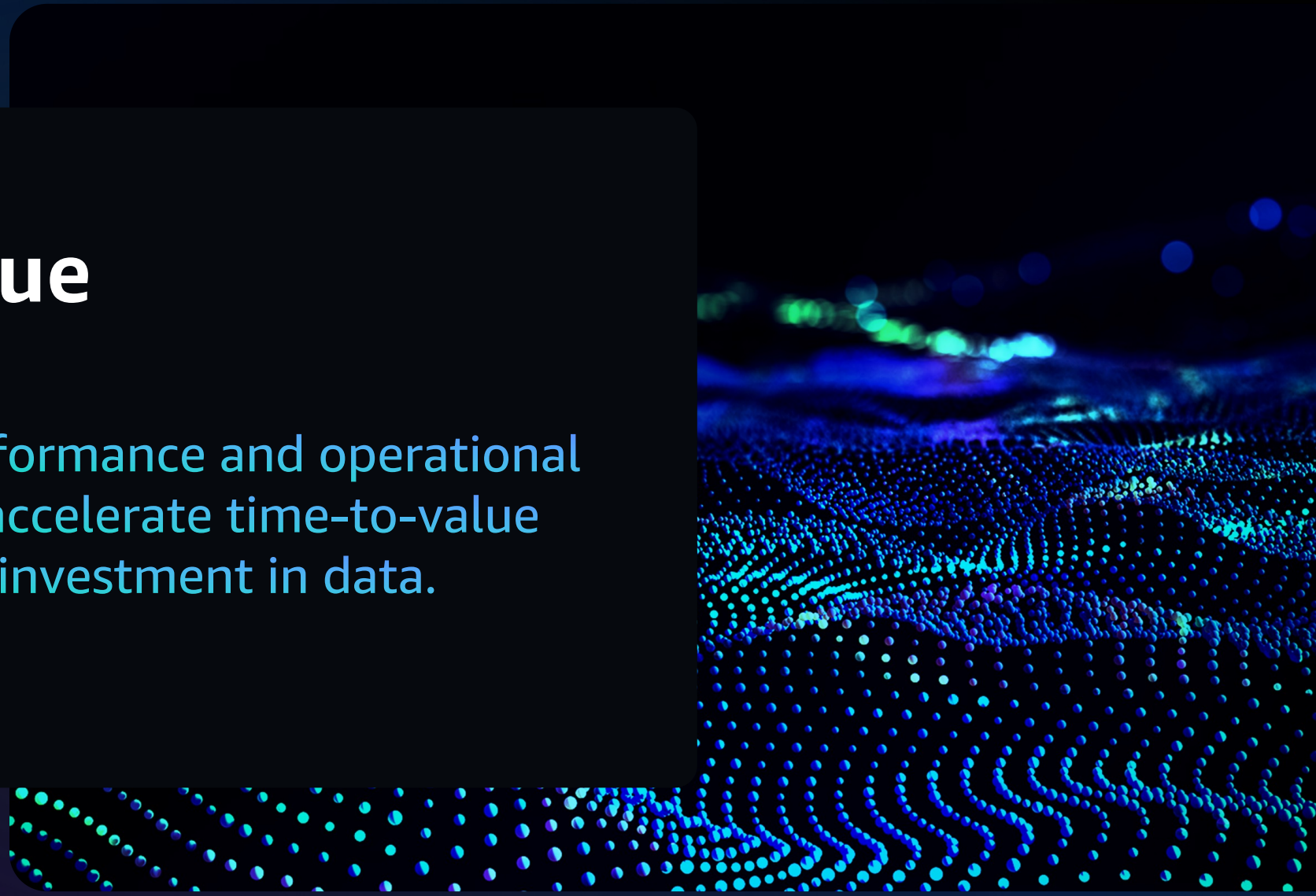
- ***Increased scalability.***
- ***Increased operational efficiencies.***
- ***Cost optimization.***

Agenda

1. AI Trend, Challenges, Key Considerations
2. Freedom to Invent
3. Maximize Value
4. Trusted Data for Trusted AI

Maximize value

We optimize price-performance and operational efficiency, so you can accelerate time-to-value and future-proof your investment in data.



Customer	State Workforce Agency – Midwest
Industry	State Government
Use Case	Unemployment Adjudication Support
Country	United States

Problem

- Adjudicators still sift through long claimant and employer filings by hand, so building a clear case summary takes hours and delays decisions.
- When new evidence arrives they reread entire packets because search is shallow, leading to inconsistent follow-up requests and back-and-forth with employers.
- Aged cases pile up, overtime spikes, and audits flag gaps in policy documentation—proof the current manual process can't scale.

Solution

- Human-in-the-loop agentic system that orchestrates a case review workspace to:
  - (1) Ingest claimant, employer, and policy documents into a unified evidence graph
  - (2) Auto-summarize facts, tag policy precedents, and surface conflicts for adjudicator review
  - (3) Draft consistent claimant/ employer outreach with required next steps prefilled
  - (4) Capture adjudicator edits and feedback to continuously refine prompts and guardrails
- *Services Used*:
  - Amazon Bedrock, Amazon Textract, Amazon OpenSearch Service

Impact

- ***Accelerated time-to-determination.***
- ***Consistent policy application and audit readiness.***
- ***Reduced adjudicator workload per case.***

BENCHMARK ANALYTICS®

Customer	Benchmark Analytics
Industry	Public Sector
Use Case	Data/Virtual Assistant
Country	United States

Customer Profile:

Benchmark Analytics' mission is to transform police force management through data science. In collaboration with esteemed research partners, Benchmark has developed an all-in-one solution to advance police force management, a groundbreaking early intervention system and proactive officer support platform.

Problem

- Benchmark Analytics equips law enforcement agencies with a data-driven early intervention system to understand key data related to officer performance.
- Law enforcement agencies can only access data via reporting tools and SQL queries, costing time and training up front for simple database commands.
- Due to the important nature of the problem space, database systems have to be be carefully designed to ensure data integrity and accuracy, while being easy to use for non-technical experts.

Solution

- The AWS GAIIC team in partnership with Benchmark Analytics, developed an AI solution that can answer questions with natural language about officer performance and behavior, by automatically generating database queries.
- This solution avoids answering questions about data it doesn't have access to, providing a stronger level of security than generic systems built with large language models (LLMs).
- A testing system was designed via LLMs so new questions can be generated and automatically tested against answers for further robustness.

Impact

- ***Data-driven, conversational AI:*** Proved concept of a chat agent which can retrieve and access event data, accurately answer relevant questions, and avoid answering questions the system has no data for.
- ***LLM efficacy***: Proved efficacy of Bedrock as an offering with integrated models and specialized prompting-developed solution outperforming single LLM in Q/A use case.
- ***Scalability***: Enabled path forward for deployment of the system by providing a testing system to generate further questions, this solution can be validated and deployed at scale.

aws

CPP Investments

Customer	CPP Investments
Industry	Finance
Use Case	Investment Analyst Assistant
Country	NAMER - Canada

Customer Profile:

Canada Pension Plan (CPP) Investments is one of the fastest growing institutional investors in the world. With current assets under management valued in excess of $500 billion, CPP Investments is a professional investment management organization that globally invests the funds of the Canada Pension Plan to help ensure long-term sustainability. CPP Investments invests in all major asset classes, including public equity, private equity, real estate, infrastructure and fixed-income instruments,

Problem

- CPP Investments is a data-driven investment group, leveraging cutting edge analytical techniques such as machine-learning, NLP, Generative AI and the use of multimodal datasets to enable investment decisions through proprietary research.
- CPP Investments built a prototype AI-powered market research assistant to help the Investment Science team with research. It utilized out-of-the box capabilities of available LLMs without task/domain specific customization.
- The prototype had high latency for text generation especially for analytical questions with high false negative rate making it less reliable for analysts.

Solution

CPP Investments partnered with the Generative AI Innovation Center to develop a robust and scalable investment assistant solution that:

- Takes in a user query, expands/transforms it and compares it against all relevant document chunks from the search index
- The retrieved chunks are reranked and further filtered with finance domain specific reranker and fed to Bedrock LLMs for answer generation.
- *Services Used*: Amazon Bedrock, Amazon SageMaker, Amazon OpenSearch.

Impact

- ***Reduced latency and costs:*** The solution generates answer within 10 sec from user input at *~1 cent per query.*
- ***Increased reliability***: accurate answering of single/multi doc, numerical, text and analytical queries on different financial documents.
- ***Scalability***: Built on managed services with serverless, autoscaling features.
- ***Quantitative metrics***: 33% improvement over CPP Investments' prior solution and 10% improvement over SOTA benchmark.
- ***More robust customer experience***: includes citation, custom metrics, automated metadata generation.

aws

coursera

Customer	Coursera
Industry	EdTech
Use Case	AI-powered Grading Assistant
Country	United States

Customer Profile:

Coursera is a global online learning platform that offers anyone, anywhere access to online courses and degrees from world-class universities and companies. They serve a vast global user base by providing a range of learning opportunities—from hands-on projects and courses to job-ready certificates and degree programs. Coursera's mission is to make high-quality education accessible to everyone, thereby empowering people with the knowledge and skills to pursue their passions and advance their careers.

Problem

- Coursera seeks consistent and high-quality feedback for assignments across their massive global user base.
- Feedback from peers and staff varied dramatically in usefulness and rigor, impacting the overall learning experience and effectiveness for students.
- Coursera sought to build a solution that provides insightful, structured feedback and fair evaluations to students on their work, while also making the assessment workflow more efficient for their growing online learning platform.

Solution

- Coursera collaborated with the AWS GenAIIC to develop an AI-powered assignment assistant to supplement human grading and provide meaningful guidance to learners in a scalable way.
- The solution involved building a Retrieval Augmented Generation (RAG) pipeline to generate feedback and grades for learner submissions
- An evaluation framework was also implemented to assess the quality of generated feedback and grades.
- *Services Used*: Amazon Bedrock, Amazon OpenSearch, Amazon SageMaker, Amazon S3

Impact

- ***Increased scalability***: By leveraging AWS services, the solution can scale to handle a massive volume of learner submissions and course materials across Coursera's global platform.
- ***Increased flexibility***: Amazon Bedrock provides numerous choices for embedding and language models, which gives Coursera great flexibility to experiment with and switch to different models on the fly.
- ***Increased reliability***: The evaluation framework allows Coursera to reliably test and monitor the performance of the AI grading tool during initial development and long-term usage.

aws

The University of Texas at Austin

Customer	The University of Texas at Austin
Industry	Higher Education
Use Case	Virtual Instructional Designer & AI Tutor
Country	United States

Customer Profile:
The University of Texas at Austin, founded in 1883, is a prestigious public research university and flagship institution of the UT System. With over 52,000 students, it ranks among the top 40 universities globally and offers top national programs across 19 colleges and schools.

Problem

- UT Austin seeks to enhance student learning by developing AI-powered chatbots tailored to course content and pedagogical approaches.
- Manual creation of personalized learning tools is time-consuming and limited in scale across diverse courses.
- The university aims to leverage generative AI to enhance student engagement, customize, and empower instructors to enhance teaching practices.

Solution

UT Austin collaborated with the Generative AI Innovation Center to:

- Develop a scalable chatbot creation tool allowing professors to provide customized learning capabilities for courses.
- Enables instructors to create multiple chatbots with different knowledge bases for an interactive, Socratic tutoring experience.
- *Services Used*: Amazon ECS, Amazon Textract, AWS Lambda, Application Load Balancer, Amazon Bedrock, Amazon OpenSearch

Impact

- Increased scalability: The solution has the potential to roll out to over 30,000 students in 2025, significantly expanding the reach of personalized learning.
- Enhanced learning experience: Empowers students and instructors through personalized learning tools and improving student engagement.
- Improved efficiency: The application is powered by Amazon Bedrock for easy and secure use of large language models at scale.

INSTRUCTURE

Customer	Instructure
Industry	EdTech
Use Case	Content Alignment
Country	United States

Customer Profile:

Instructure is a prominent educational technology company revolutionizing the way we learn. Founded in 2008, the company's flagship product, Canvas, has become one of the most widely adopted learning management systems (LMS) in the world.

Problem

- Instructure faced a significant challenge in efficiently aligning diverse educational content to various state and national standards.
- This process was time-consuming and complex due to the vast number of standards and semantic discrepancies between the educational content and the language used in the standards.
- Developing an automated system that could accurately understand educational content and align it with existing standards across multiple subjects, grade levels, and standards bodies proved to be a difficult task.

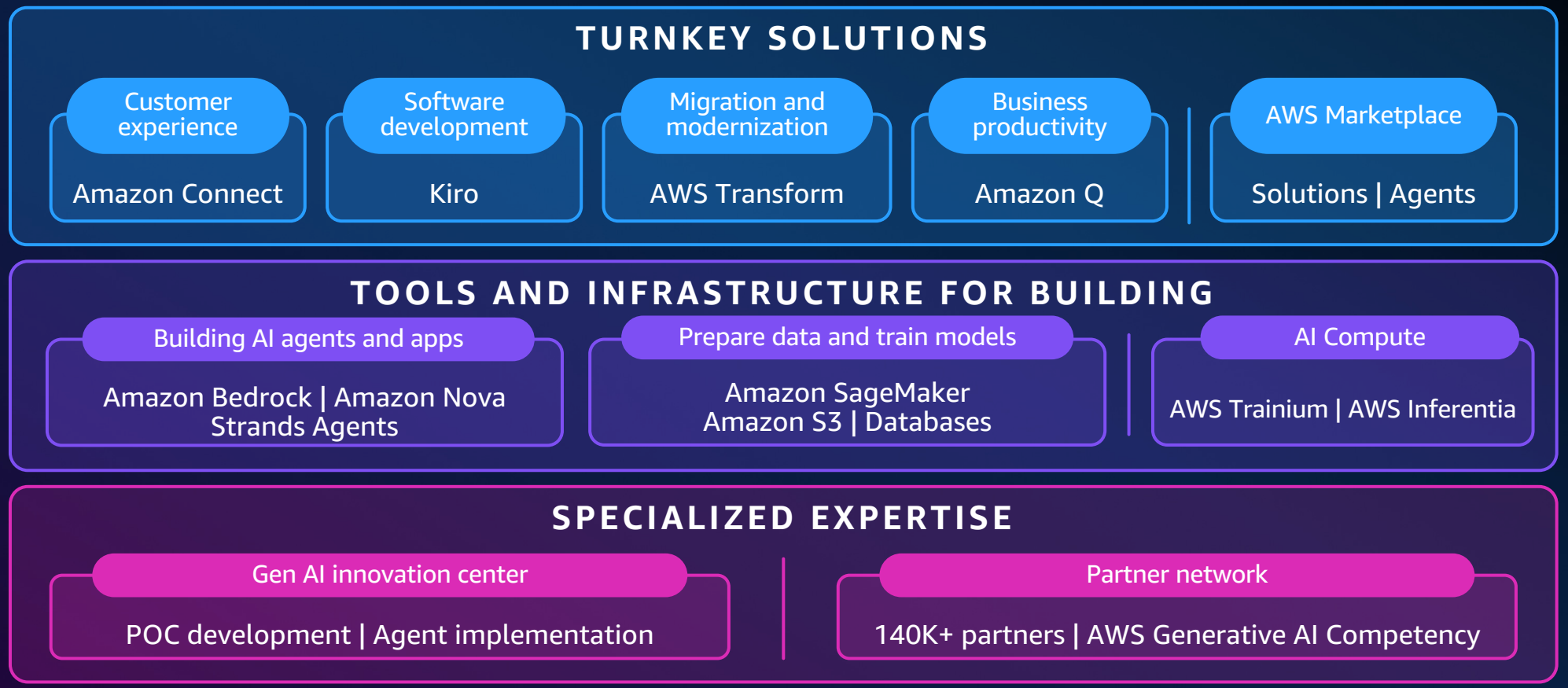
Solution

- To address this challenge, Instructure partnered with the Generative AI Innovation Center to develop a solution using:
- Novel guided chain-of-thought approach: provides the model with prompts to enable a greater understanding of how to align educational content to standards.
- Metadata enrichment: Provides additional signals for the model to more accurately understand the content
- Bloom's Taxonomy levels and Webb's Depth of Knowledge framework: allows the model to map educational content to standards based on the cognitive skills and complexity involved.
- *Services Used*: Amazon Bedrock, Amazon Sagemaker

Impact

- ***Increased scalability.*** By automating the standards process, Instructure is able to rapidly expand across various markets and jurisdictions.
- ***Increased operational efficiencies.*** New solution is poised to improve SME workflows and save time needed.
- ***Cost optimization.*** This solution allow the team to optimize cost and performance across multiple educational topics and standards groups — this was prohibitively expensive with prior solutions.

Everything you need to innovate with AI



Deliver value with managed data services

Reduced operational costs, accelerated value creation

Optimize costs in the background while you focus on innovation

Choice for all your workload cost profiles

Unparalleled high performance and availability at global scale

Amazon Bedrock AgentCore Gateway

Agents

MCP Client

Bedrock AgentCore Gateway

- MCP
- Tools creation
- Tools search
- Inbound and Outbound Auth

APIs, tools, resources

SageMaker MCP

Databases

AWS Lambda/APIs

AgentCore Identity

CloudWatch

Agenda

1. AI Trend, Challenges, Key Considerations
2. Freedom to Invent
3. Maximize Value
4. Trusted Data for Trusted AI

Trusted data for trusted AI

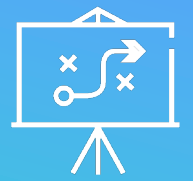
Build secure and reliable data foundations for accurate and trusted AI applications at enterprise scale.

We build trust with customers via...

DATA QUALITY AND LINEAGE

RELIABLE INFRASTRUCTURE

OBSERVABILITY AND RISK MITIGATION



BEST-IN-CLASS GOVERANCE AND GUARDRAILS

SECURITY AND COMPLIANCE FOR AN AGENTIC AI ERA

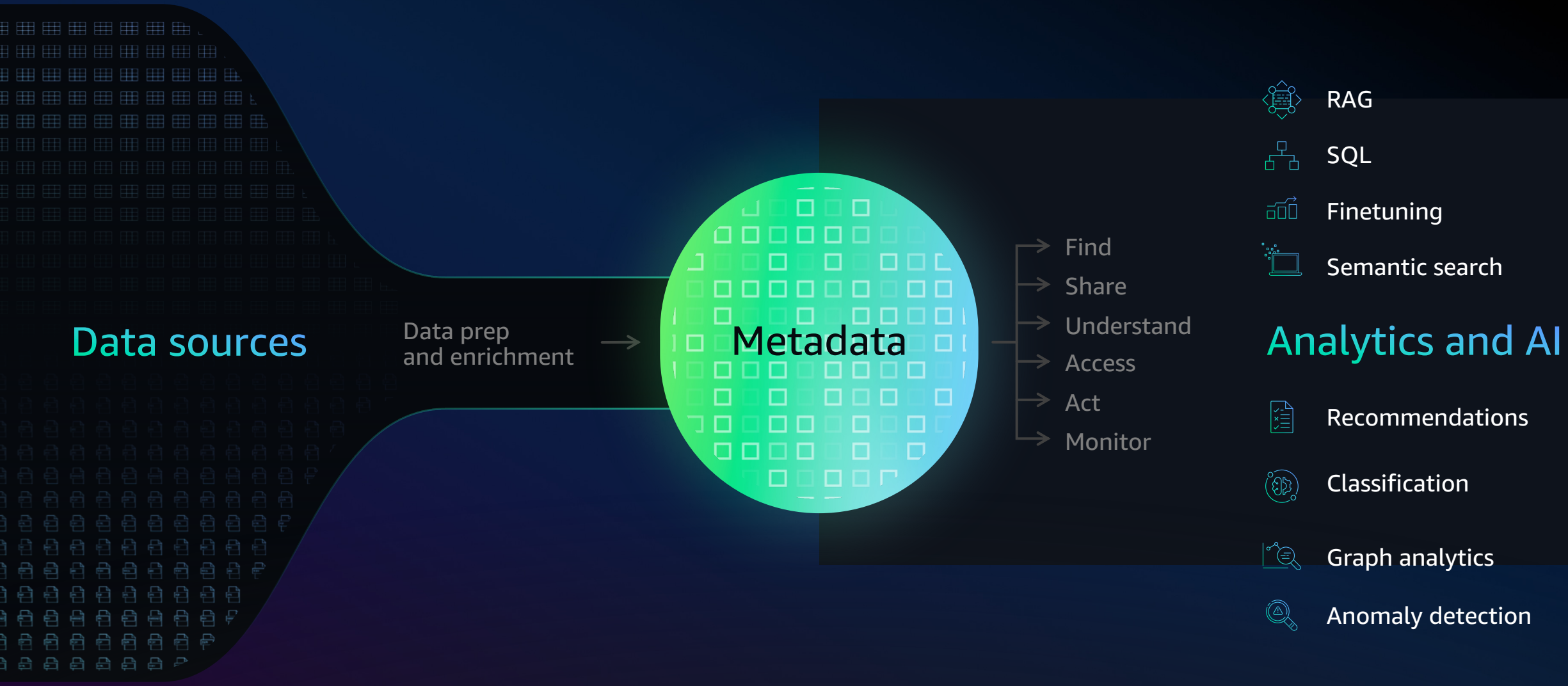
Structured

Data sources

Unstructured

Data streaming and processing

Cleaning | Enriching | Generating | Business context | Transformation | Aggregation



Unified Studio

SQL analytics
Amazon Redshift

Data processing
Amazon EMR
AWS Glue
Amazon Athena

Model development
Amazon SageMaker AI

Gen AI App development
Amazon Bedrock

COMING SOON

Streaming
Amazon MSK
Amazon Kinesis

COMING SOON

Business intelligence
Amazon QuickSight

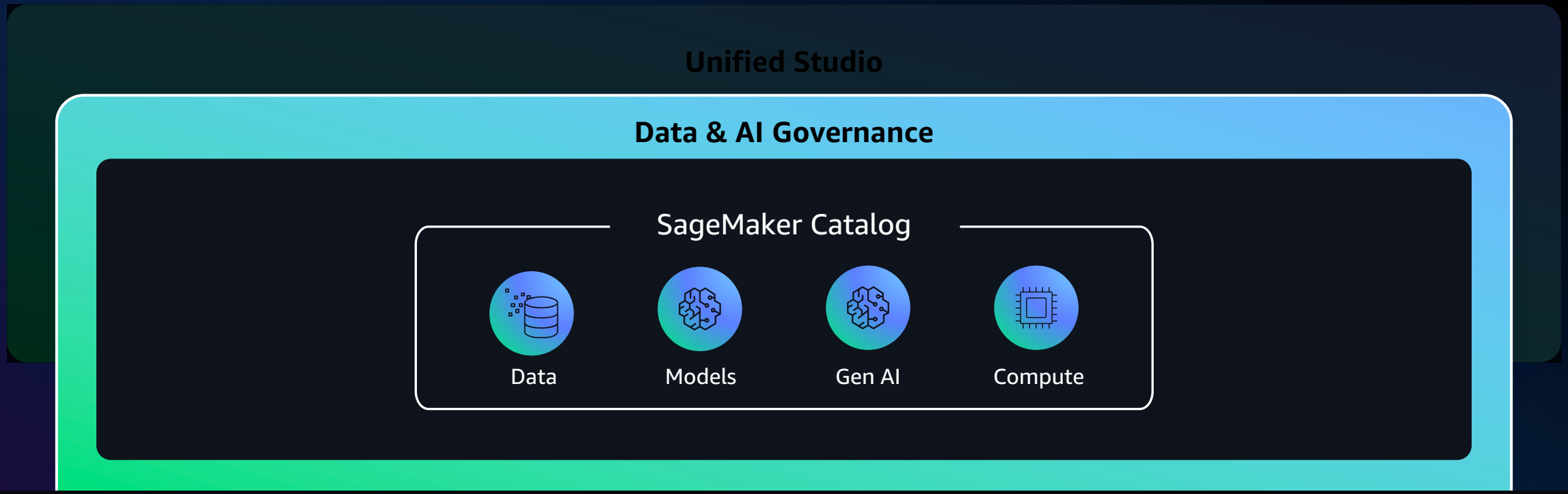
COMING SOON

Search analytics
Amazon OpenSearch Service

Data & AI Governance

Open Data Lakehouse

Amazon SageMaker



Centralized metadata repository

Auto-ingestion of technical metadata

AI-generated business context

Multi-Modal Data

Data discovery and sharing

AWS SageMaker Catalog

DISCOVER, GOVERN, AND COLLABORATE ON DATA AND AI SECURELY, WITH A UNIFIED CATALOG.

SageMaker Catalog

Discover

Automate data discovery and cataloging with ML and GenAI. Provide data quality and lineage

Govern

Govern data access with fine grained controls and publish/subscribe workflows

Collaborate

Connect people and data through shared tools to drive business insights

Different Data Governance needs

Type of user

Role

Technical producer

- File formats
- Schema definition
- Partitioning
- Data ownership
- Data classification

Technical consumer

- Schema metadata
- Data location and format
- Partitioning information
- Versioning
- Size

Business producer

- Field definitions with business context
- Data classification
- Asset documentation
- Data quality indicators
- Access control

Business consumer

- Data lineage
- Data categories
- Data freshness
- Data quality indicators
- Access and usage guidelines

Build trusted AI with built-in governance

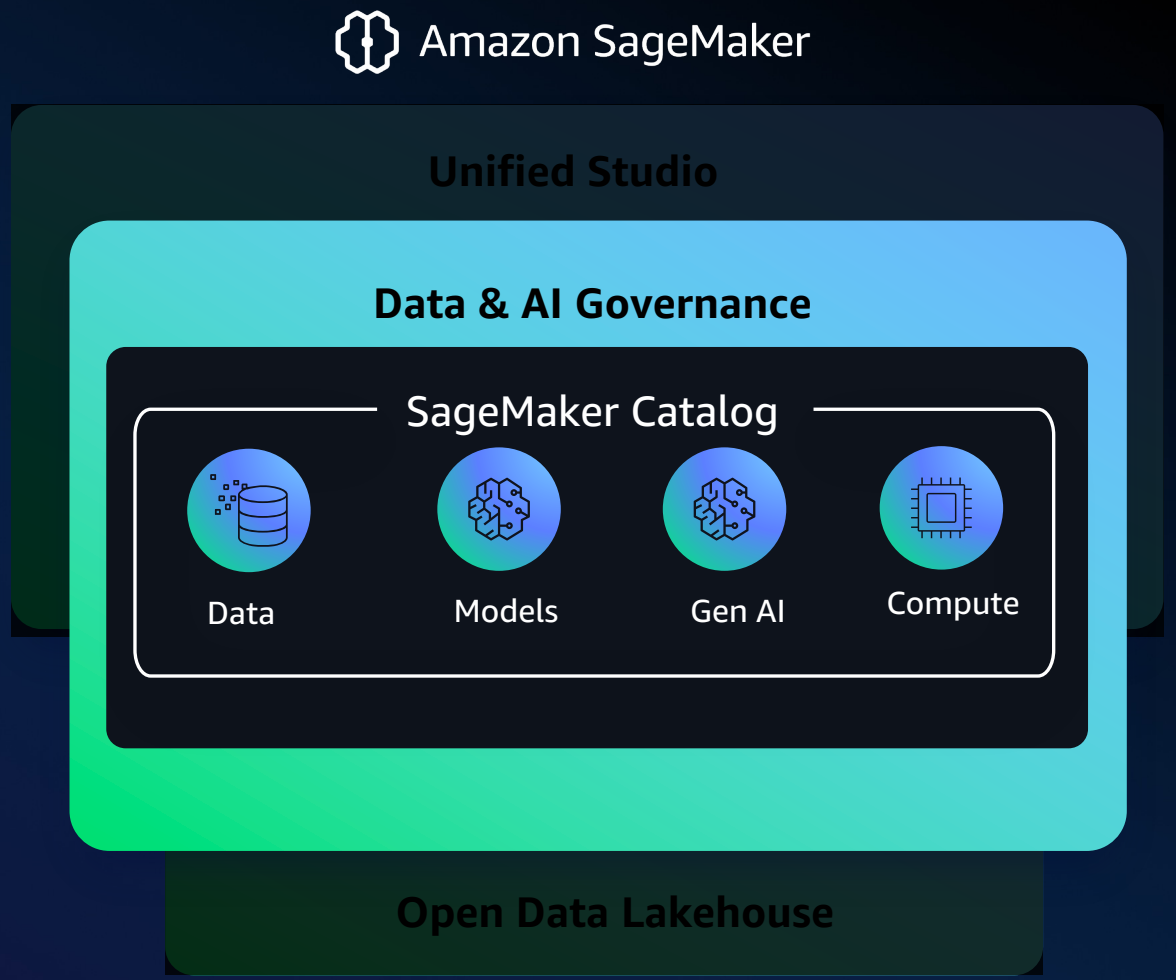
Data quality and classification

Data & ML lineage

Publish and subscribe to workflows

Fine-grained access controls

Cost logging



Customer	Modern state government software leader for nation's cities, counties, and state agencies
Industry	State Government Tech Vendor
Use Case	State Government IT and Application Support
Country	United States

Problem

- Bottlenecks and dependencies
- Limited Self-Service Capabilities
- Data quality and governance issues
- Scalability challenges

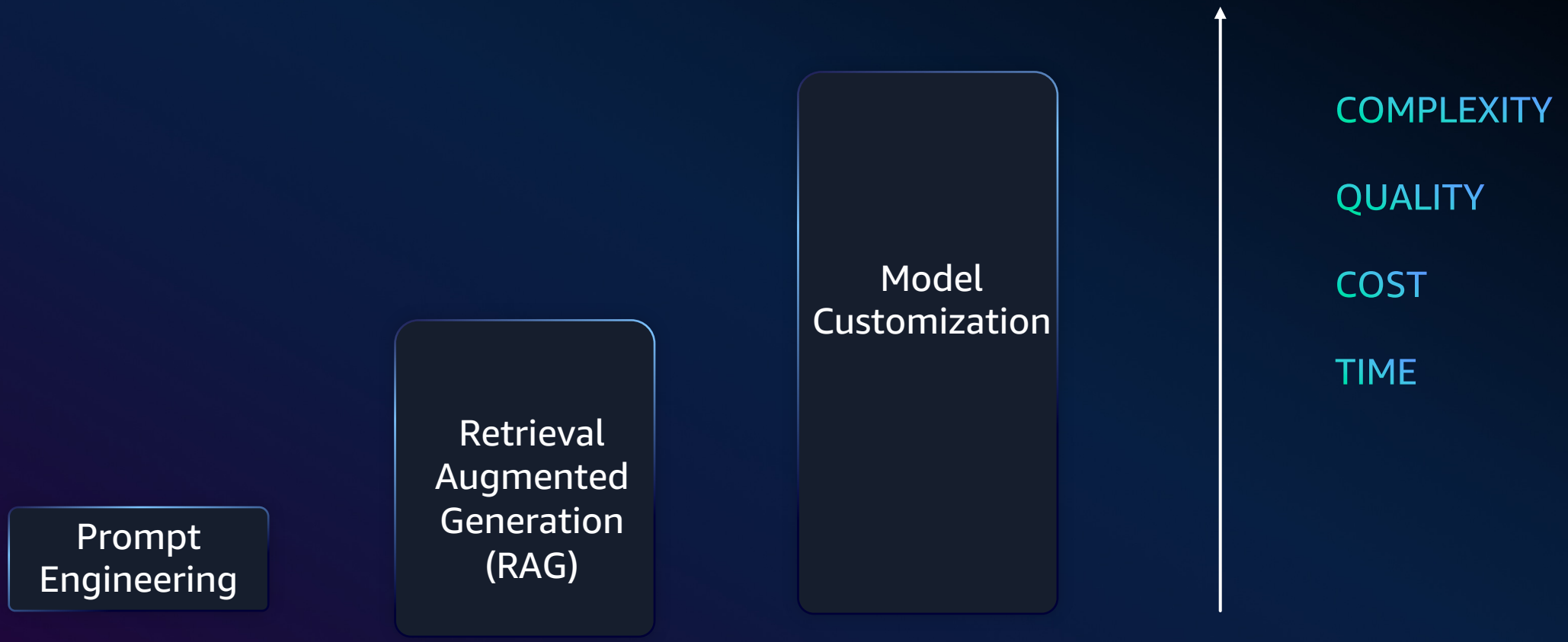
Solution

- SageMaker Catalog, SageMaker LakeHouse (Redshift, Iceberg table on S3), DMS, Amazon Managed Workflows for Apache Airflow (MWAA), Glue; SageMaker Unified Studio (later phase)

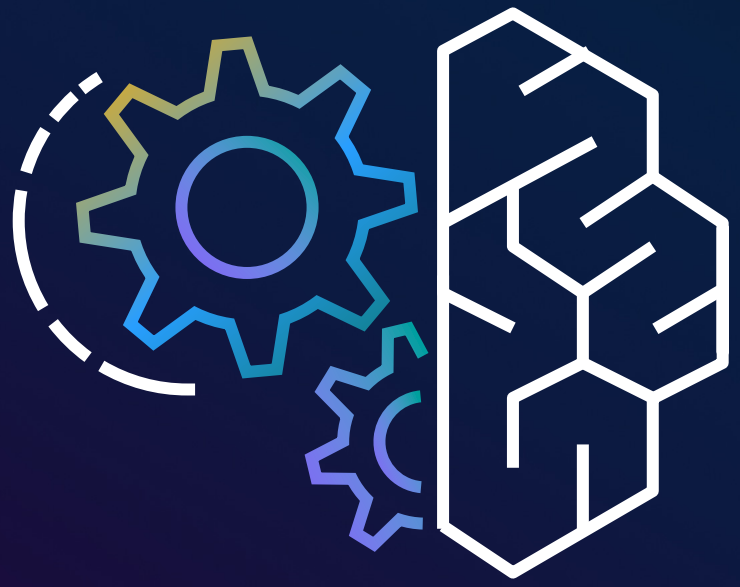
Impact

- ***Demonstrates successful alignment of technical capabilities with business objectives, establishing a foundation for scalable, organization-wide data mesh implementation***
- ***Two additional business units will adopt the solution in early 2026***

Common approaches to improve AI accuracy



Considerations for improving AI accuracy



Responsible AI is the practice of designing, developing, and using AI technology with the goal of maximizing benefits and minimizing risks.

At AWS, we define responsible AI using a core set of dimensions that we assess and update over time as AI technology evolves.

AWS Responsible AI dimensions

Controllability

Having mechanisms to monitor and steer AI system behavior

Privacy & Security

Appropriately obtaining, using and protecting data and models

Safety

Preventing harmful system output and misuse

Fairness

Considering impacts on different groups of stakeholders

Veracity & Robustness

Achieving correct system outputs, even with unexpected or adversarial inputs

Explainability

Understanding and evaluating outputs generated by an AI system

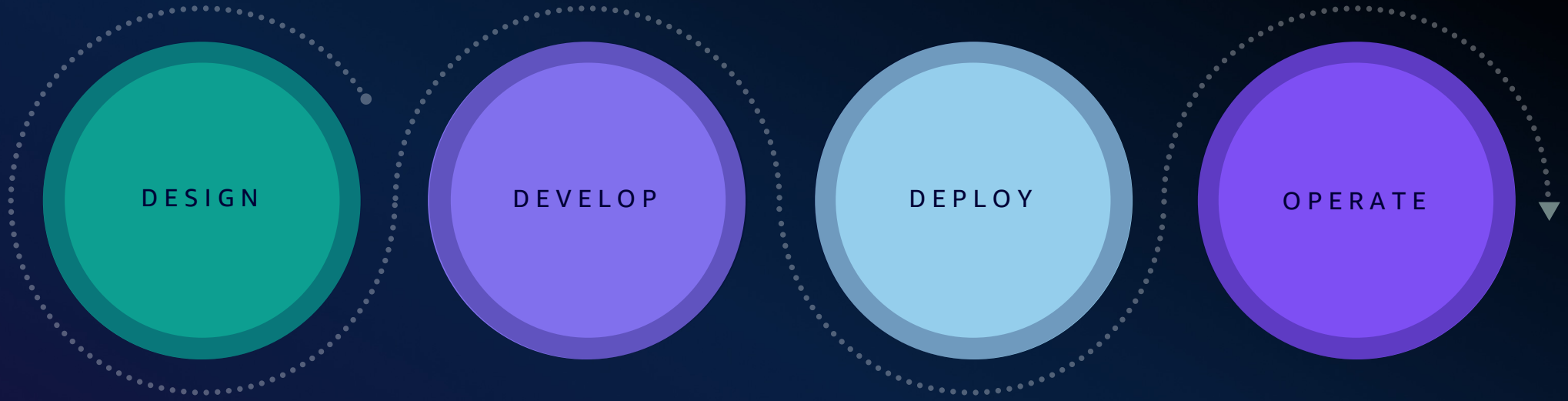
Transparency

Enabling stakeholders to make informed choices about their engagement with an AI system

Governance

Incorporating best practices into the AI supply chain, including providers and deployers

AWS support across the AI lifecycle



Amazon Partner Network & AWS Solution Architects

AWS Generative AI Innovation Center

AWS Audit Manager, AWS Artifact, AWS Config

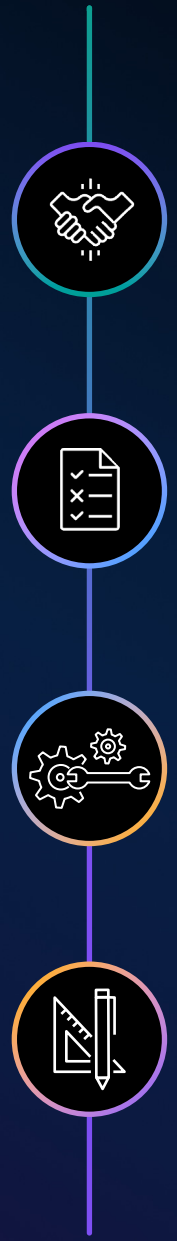
Amazon SageMaker AI (Data Wrangler & Ground Truth & Clarify)

Amazon Bedrock (Knowledge Bases, Guardrails, Evaluations, LLM-as-a-judge)

Amazon SageMaker ML Governance tools (e.g., Model Registry)

Amazon SageMaker Model Monitor

Responsible AI:
Best practices



Define use cases—the more specific & narrow, the better

Assess risk on a (use) case-by-case basis

Iterate across the AI lifecycle

Test, test again, and then test again

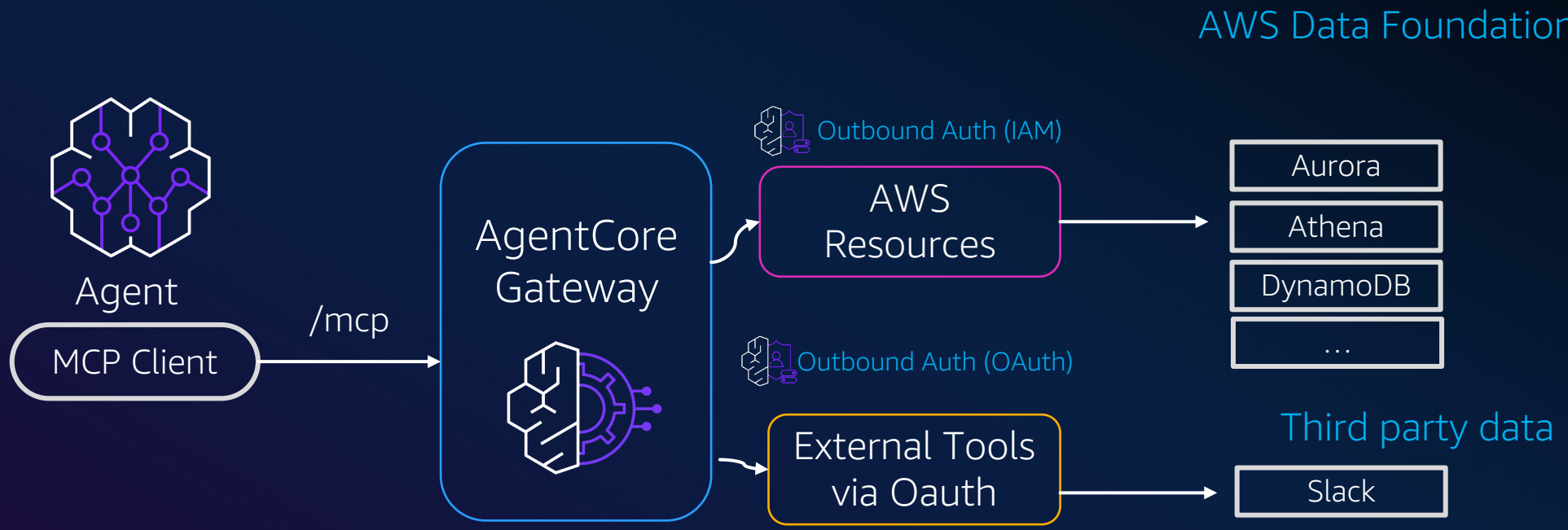
Amazon Bedrock Guardrails

Evaluate prompts and model responses for agents, knowledge bases, FMs in Amazon Bedrock, and self-managed or third-party FMs

- Configure thresholds to filter undesirable and potentially harmful text and image content, jailbreaks, and prompt attacks
- Identify, correct, and explain factual claims in responses using Automated Reasoning checks
- Define and disallow denied topics with short natural language descriptions
- Remove personally identifiable information (PII) and sensitive information in generative AI applications
- Define a set of words to detect and block in user inputs and model responses
- Filter hallucinations by detecting grounded ness and relevance of model responses based on context

aws

Data Access Flow via AgentCore Gateway & Identity



Full observability for end-to-end visibility

Monitor data quality, lineage, usage, and model drift

Gen AI driven debugging, quality, and root cause analysis

Multi-layered observability with AWS CloudWatch and CloudTrail

Open telemetry with integration with 3rd party tools

Data Ingestion | Processing | Transformation | Data Stores & Knowledgebases | AI & Agent Applications

Your data is your competitive advantage. It should stay that way.

 Security

 Resilience

 Governance

Well Architected guidance Recovery Replication Audits Data classification Incidence response

Multi-az configurations Encryption at rest and in transit Object versioning Row level and column level security

IAM policies 143 security certifications No sharing customer data trained for LLMs with model providers Backup and recovery

Data protection, privacy, and resilience

AWS is there with you, no matter where you are in your journey

Jumpstart your data and AI flywheel

AWS Data-Driven Everything

Guide generative AI use cases

AWS Generative AI Innovation Center

Transform your business

AWS Partner

Thank you!

Please complete the session survey in the mobile app

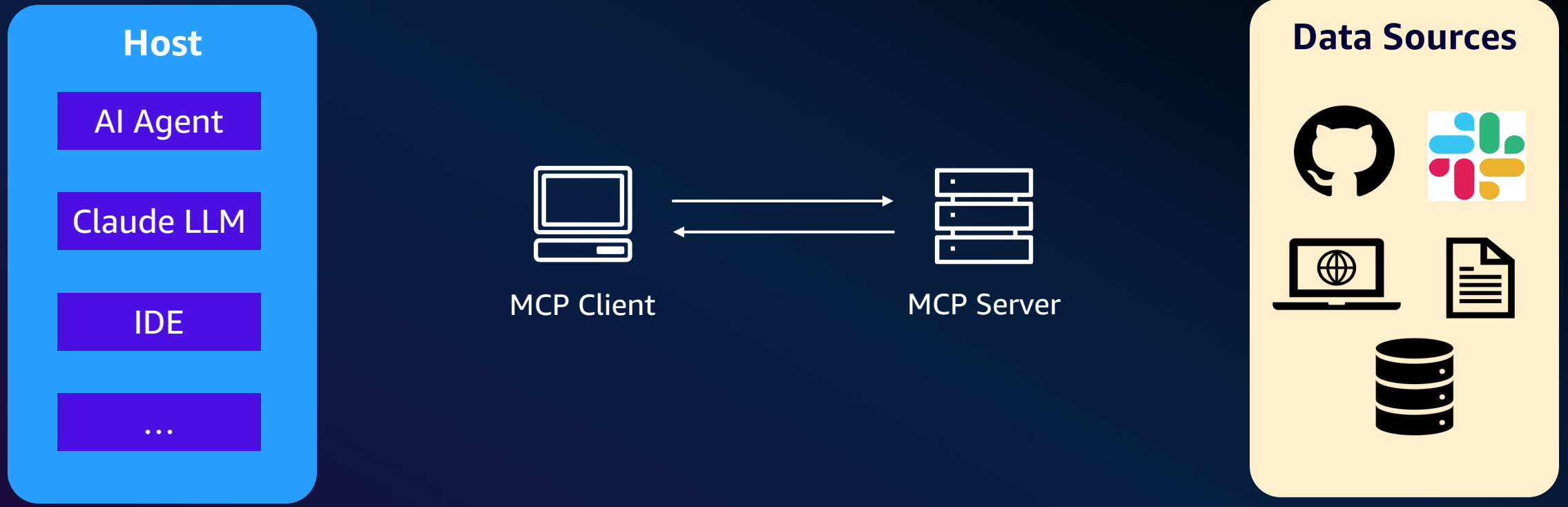
What is MCP?

MCP (Model Contextual Protocol)

- Introducing by Anthropic on Nov 2024 as open-source project
- It is proposed to be a new standard for connecting AI assistants to the systems where data lives – think of databases, content repositories, business applications
- At its core, MCP uses a client-server architecture that enables developers to expose their data through lightweight MCP servers while building AI application as MCP clients that connect to these servers

Reference: https://www.anthropic.com/news/model-context-protocol

MCP main components



1. **Host:** A program, AI tool, or AI agents requiring access to data through MCP
2. **MCP Client:** Protocol clients maintaining one-to-one connections with servers
3. **MCP Server:** Lightweight programs exposing capabilities through standard MCP
4. **Data sources:** both local (databases, files) and remote services (APIs) that MCP server can access

Why is MCP important

Standardization

Integration with External Data Sources

Enhanced Functionality

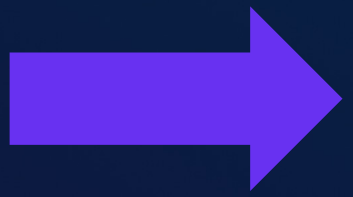
Flexibility and Scalability

The old way of doing GenAI application …

Tools for agents

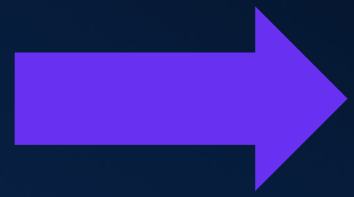
Internet search

Database or knowledge base



The code for each tool is being configured into Lambda function

AWS Lambda

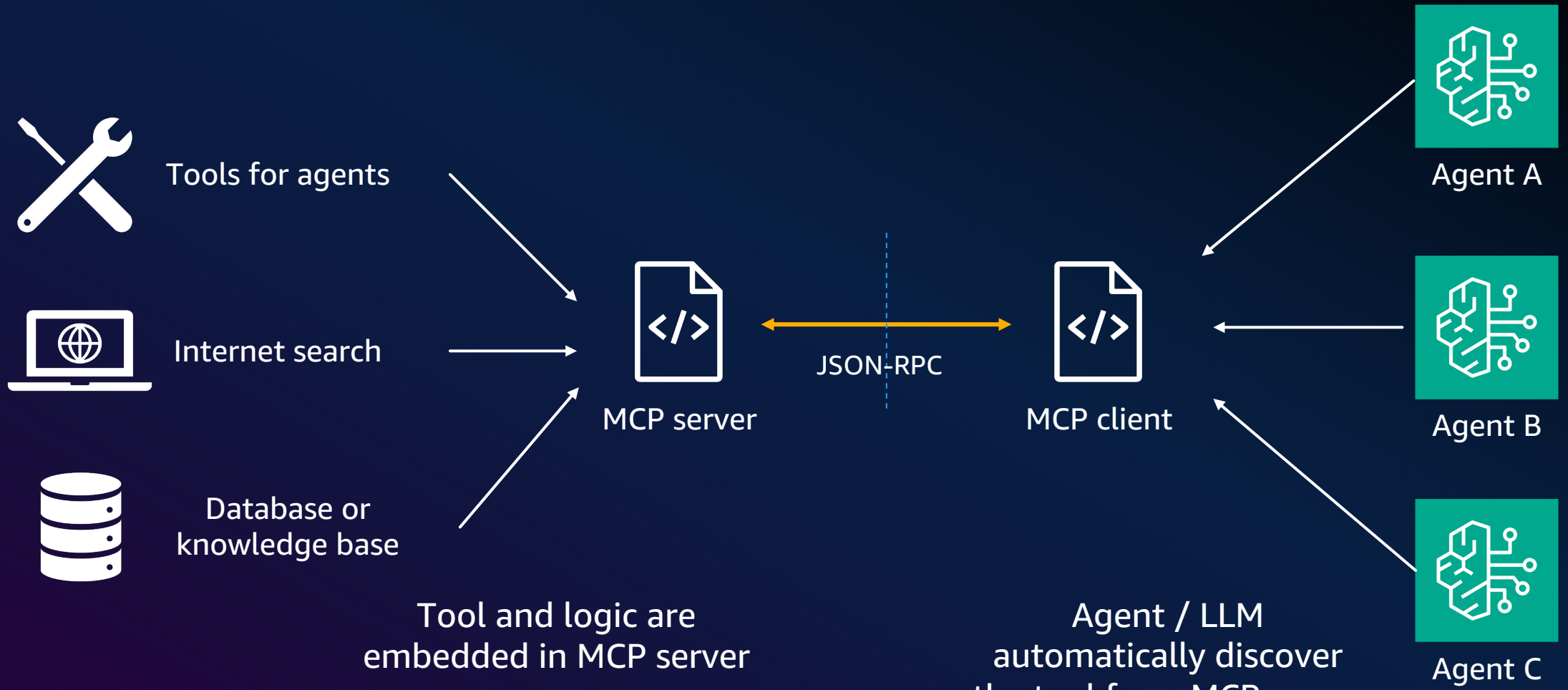


Associated the function to Amazon Bedrock Agent

Bedrock agent

Now, you can see that the tools for agent are embedding in Lambda function
-- think of **scaling**, and how you **manage** the tools --

… and now with MCP



MCP Request Flow