



INNOVATE

AI/ML EDITION

From POC to production: Strategies for achieving machine learning at scale

Alex McClure
Principal Machine Learning Business Development Manager
AWS

Agenda

Session overview

What does Machine Learning at scale look like?

7 steps toward operating at scale

Technology and tools

How to get started

Resources to help

Topics we will be . . .

Covering

- Principles and key strategies
- Technology capabilities
- High-level system architectures

Not covering

- Code and APIs
- 400-level technical content

Machine Learning at scale

ML builders and business stakeholders
Stay agile



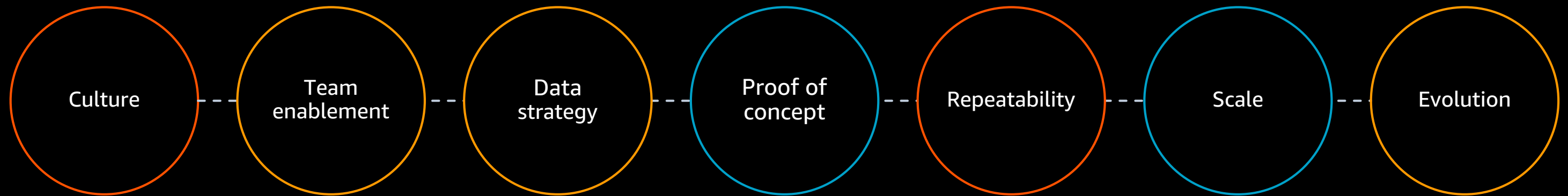
Innovate with the speed and agility of AWS
Self-service access
Experiment fast
Respond quickly to change

Cloud IT
Governance and operational readiness



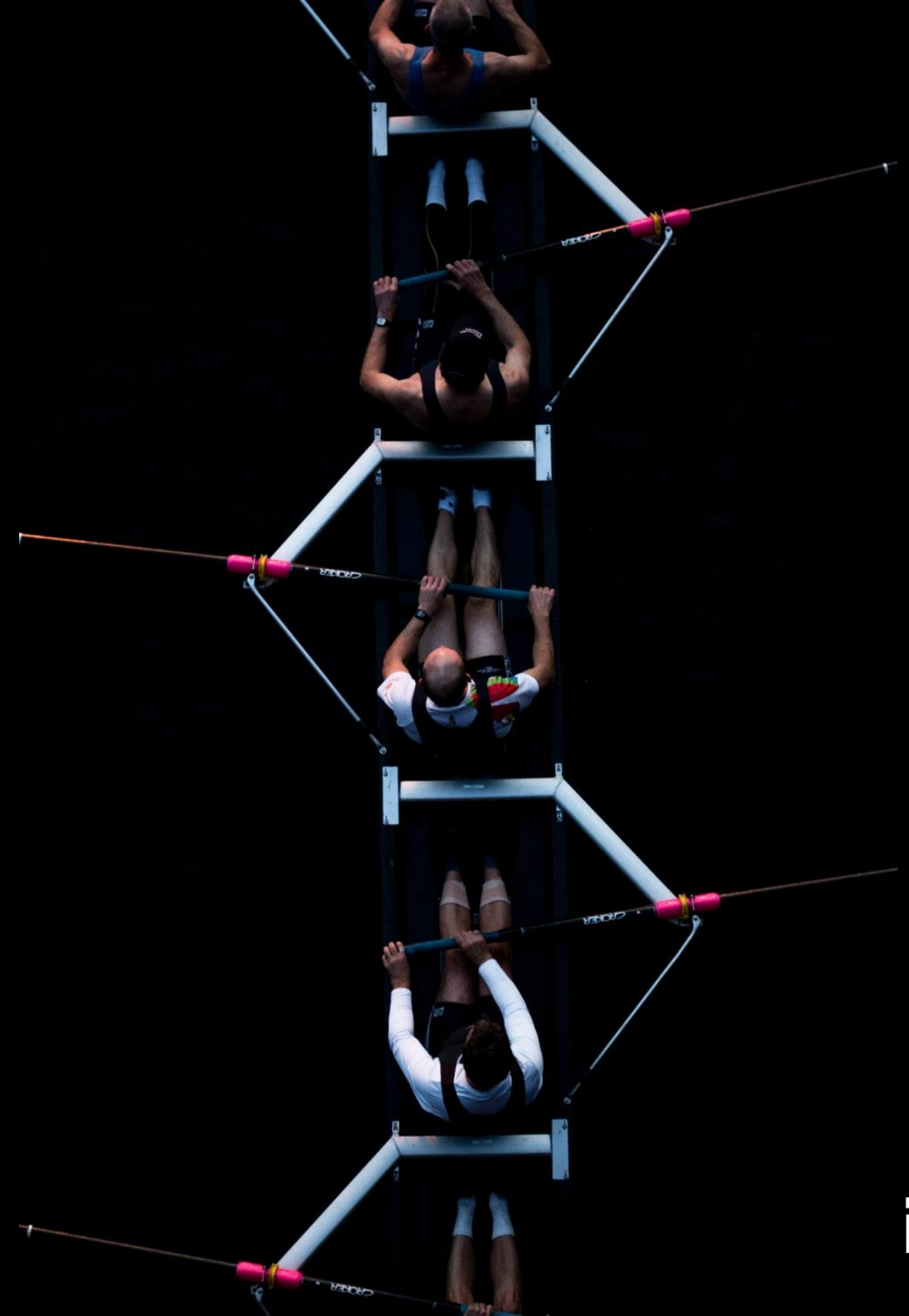
Govern and enable with central controls
Security
Compliance
Operations
Spend management

Achieving ML at scale is a journey



Establishing culture

- Executive alignment on expectations
- Focused on the right projects
- Prioritize security and governance
- Investment in developing teams
- Collaborate and share knowledge
- Embrace experimentation
- Establish repeatable MLOps practices
- Measure results against business outcomes
- Celebrate success



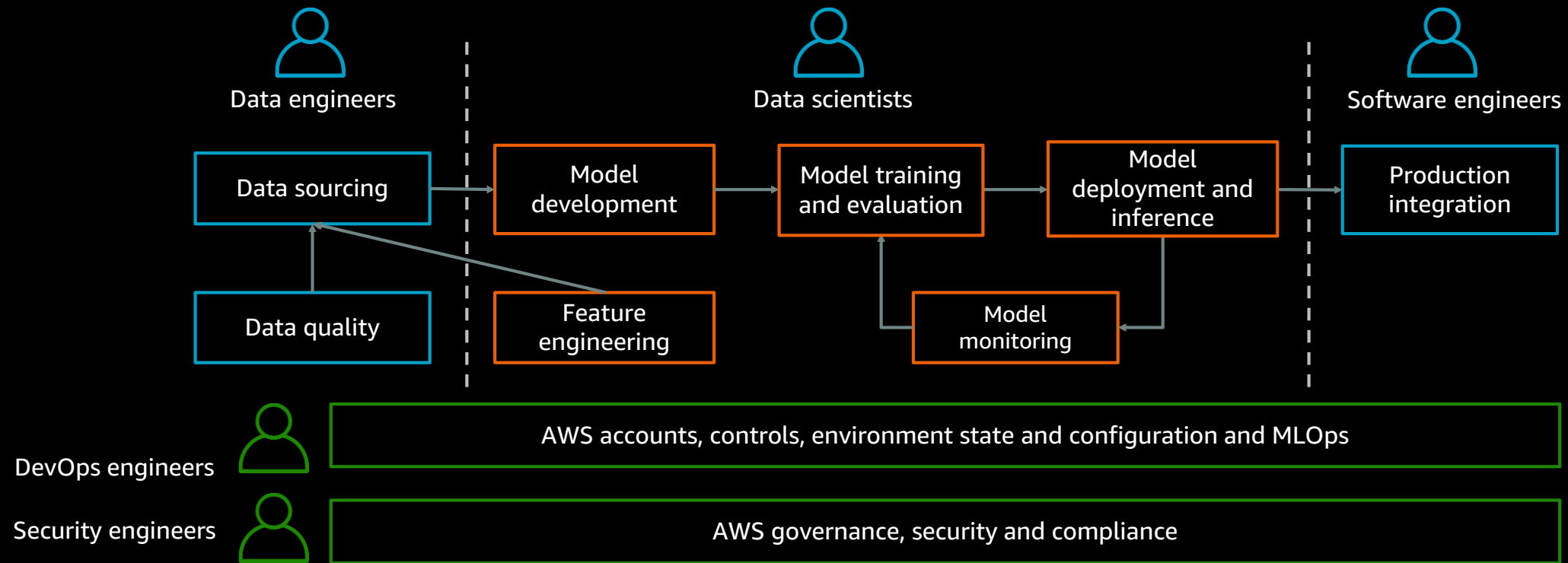
Enabling teams

Establishing culture **Enabling teams**



Enabling teams

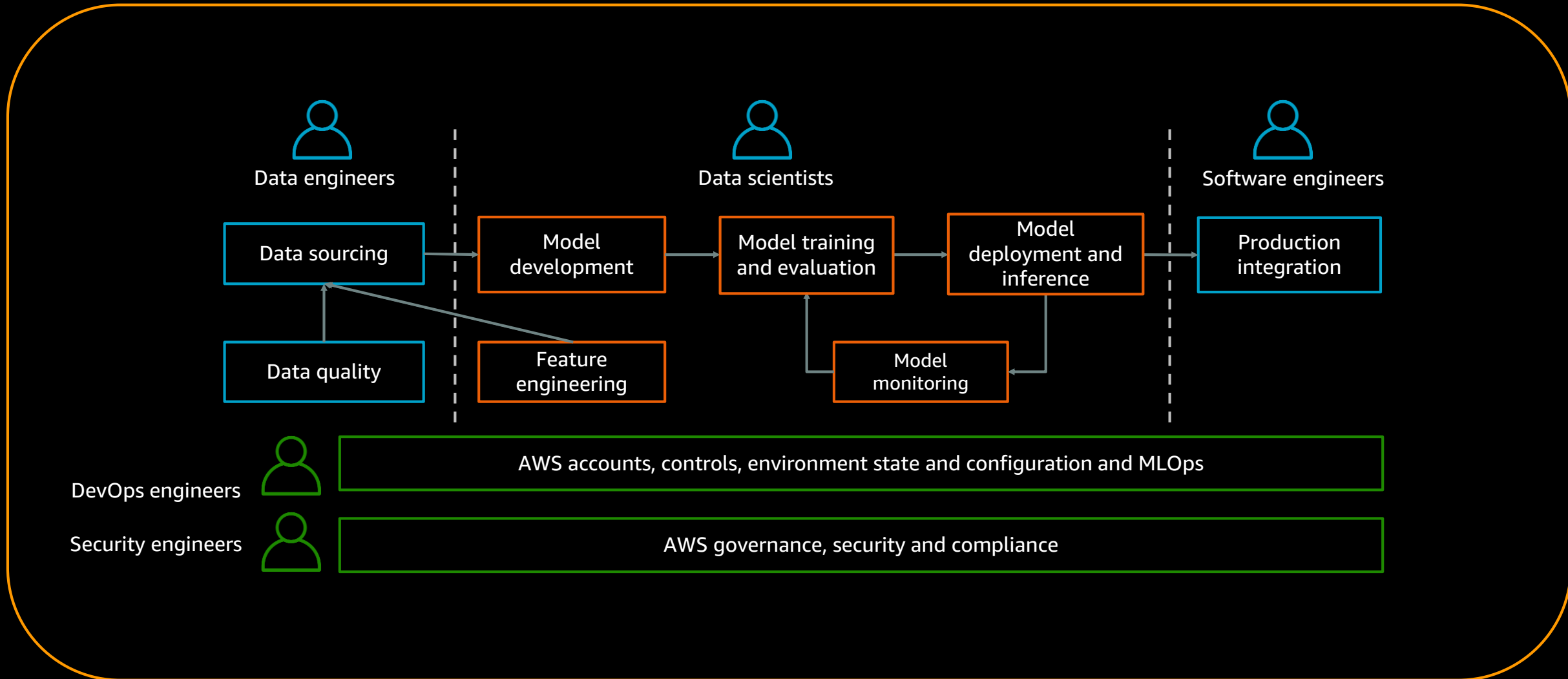
Establishing culture **Enabling teams**



Enabling teams

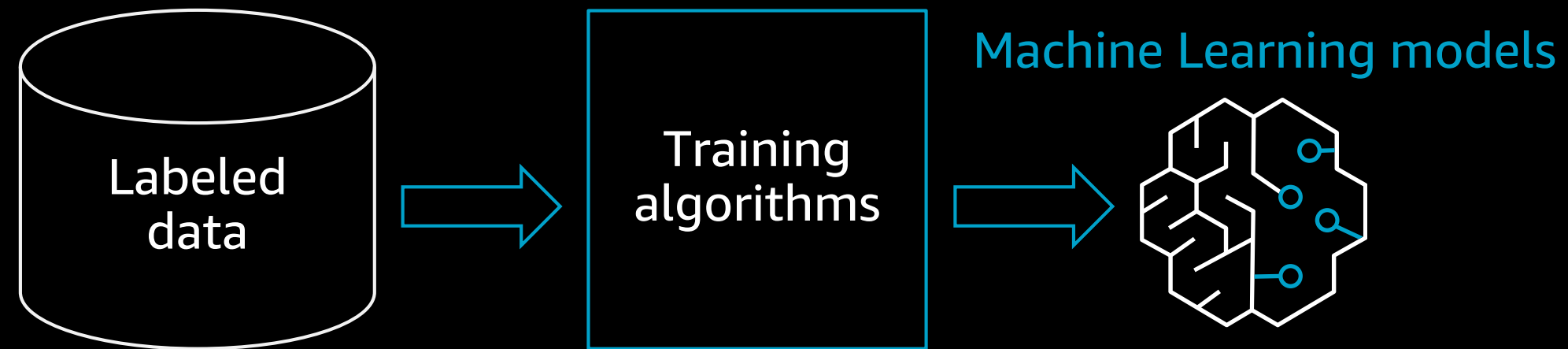
Establishing culture **Enabling teams**

Domain experts



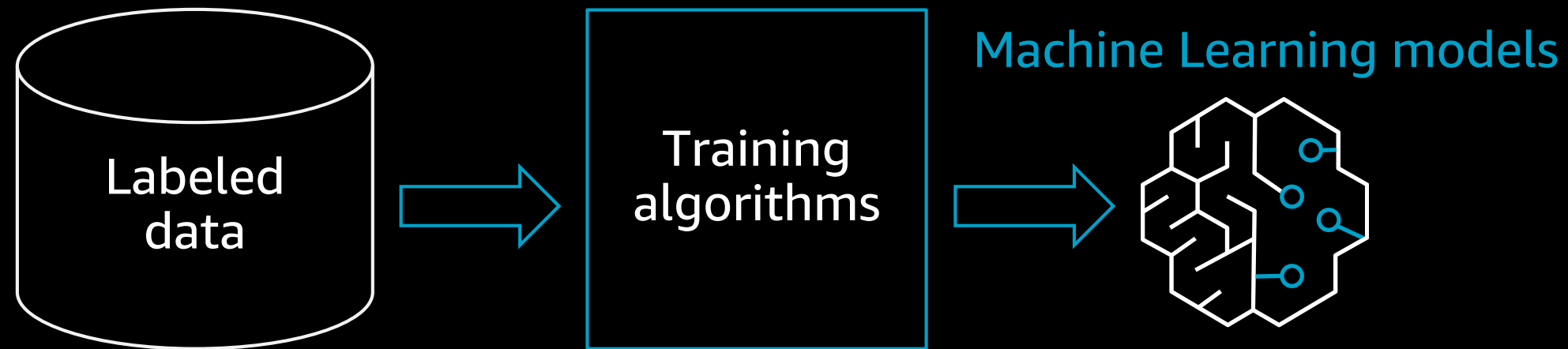
Data strategy

Establishing culture Enabling teams **Data strategy**



Data strategy

Establishing culture Enabling teams **Data strategy**



Common security approaches

Data science sandbox

Data is copied from production to a new environment

Easier to prevent accidents in production

Requires additional administration when destroying data

Establish a data platform account









Team members given "read" access to production data

Reduces the amount of places to administer data

Accidents can lead to disruptions in production

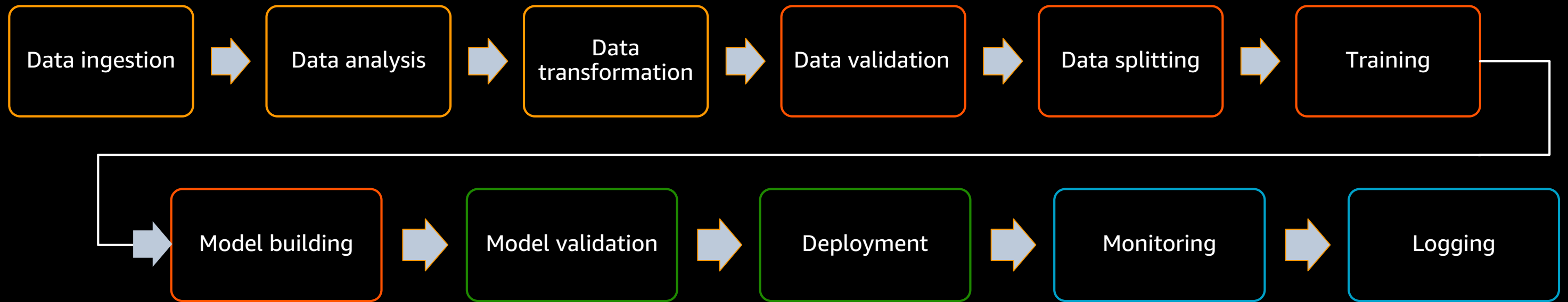
Proof of concept

Establishing culture Enabling teams Data strategy **Proof of concept**

-  What is the business challenge that this ML workload is trying to solve?
-  How good are you currently at solving this problem?
-  What is the business value that this solution creates?
-  Who would use this solution and how?
-  What metric will you use to measure success of this solution?
-  Have you established your core team and processes?
-  Do you have the right data? Is it already labeled or can it be labeled?
-  Can this project be completed within pre-determined timelines and budgets?

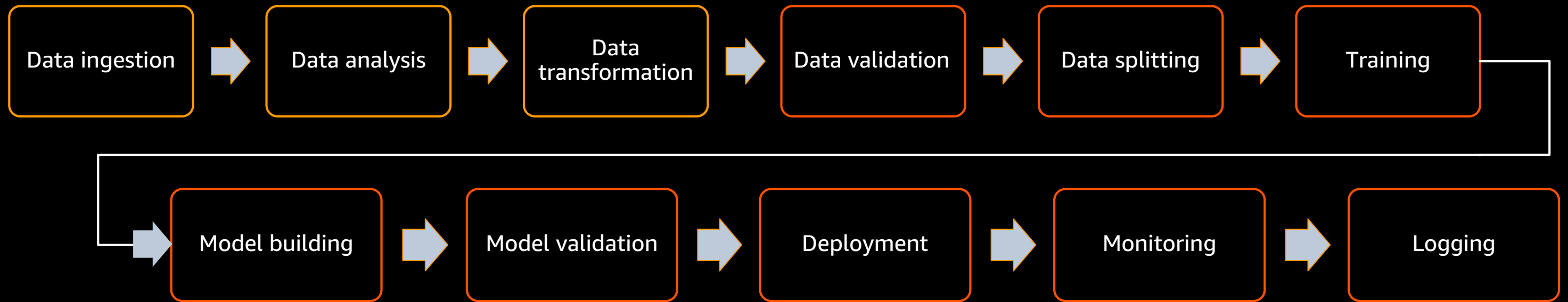
Repeatability

Establishing culture Enabling teams Data strategy Proof of concept **Repeatability**



Repeatability

Establishing culture Enabling teams Data strategy Proof of concept **Repeatability**



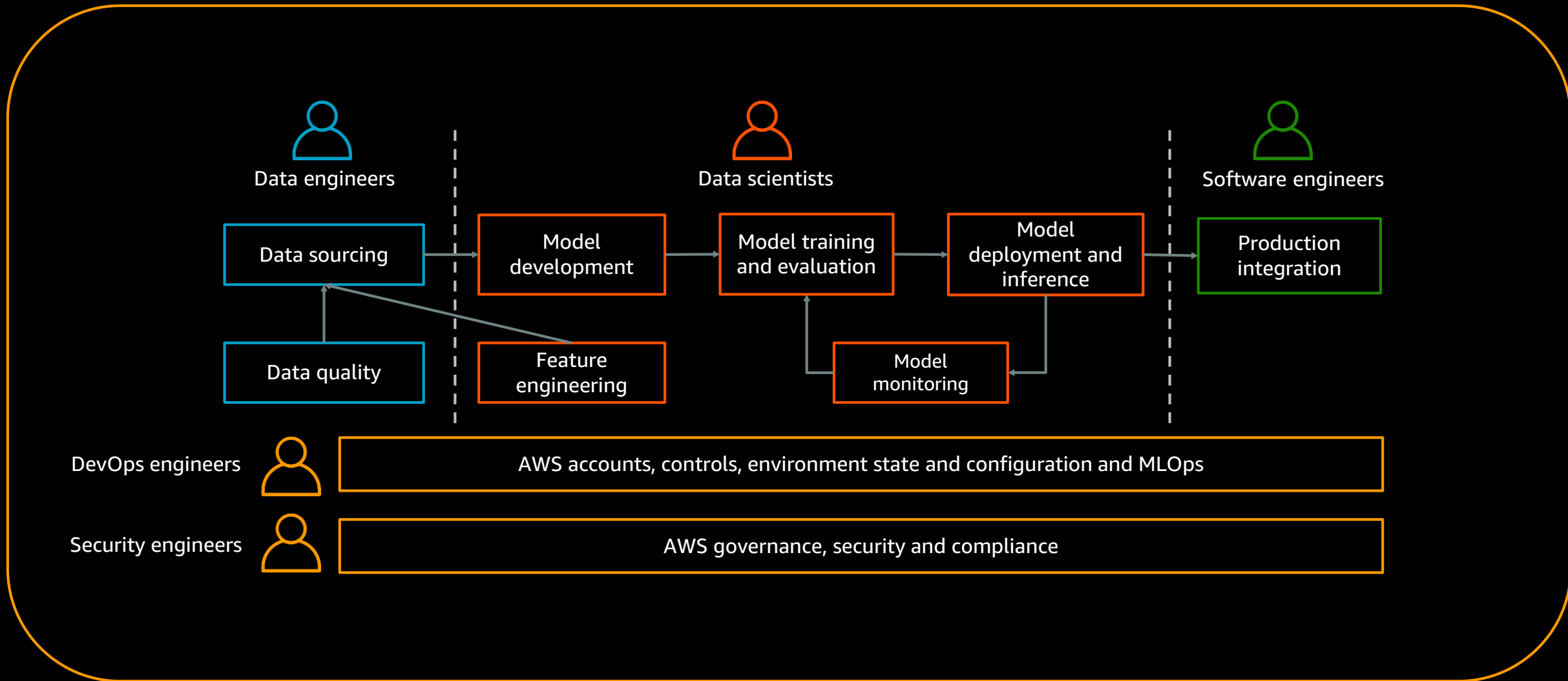
Automation techniques can eliminate manual hand-offs between teams

- Utilize AWS Step Functions and AWS CloudFormation templates to define your pipelines as code
- MLOps and CI / CD (continuous integration and continuous deployment) enable full traceability and lineage tracking of model training jobs and deployments
- Amazon SageMaker Debugger and Amazon SageMaker Model Monitor can be used to troubleshoot operational and model performance issues

Scale

Single Project Team

Domain experts



Scale

Multiple Teams Supported by a Machine Learning Center of Excellence (COE)

Team 1



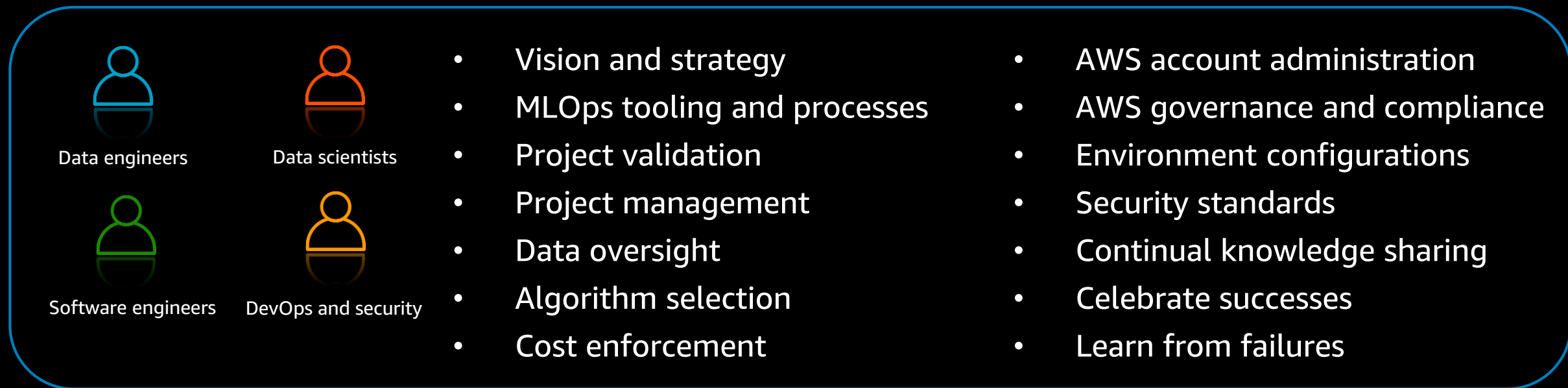
Team 2



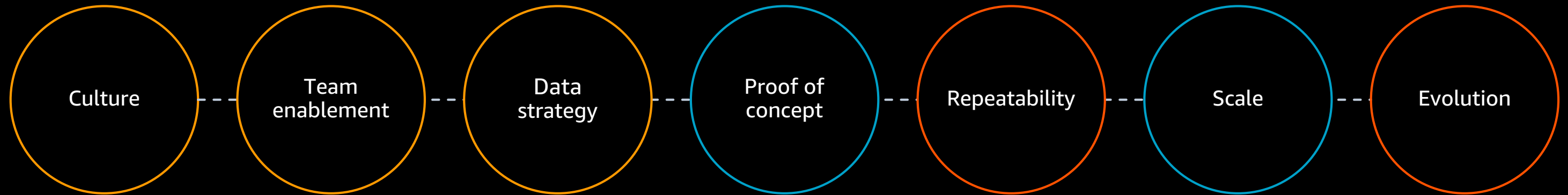
Team 3



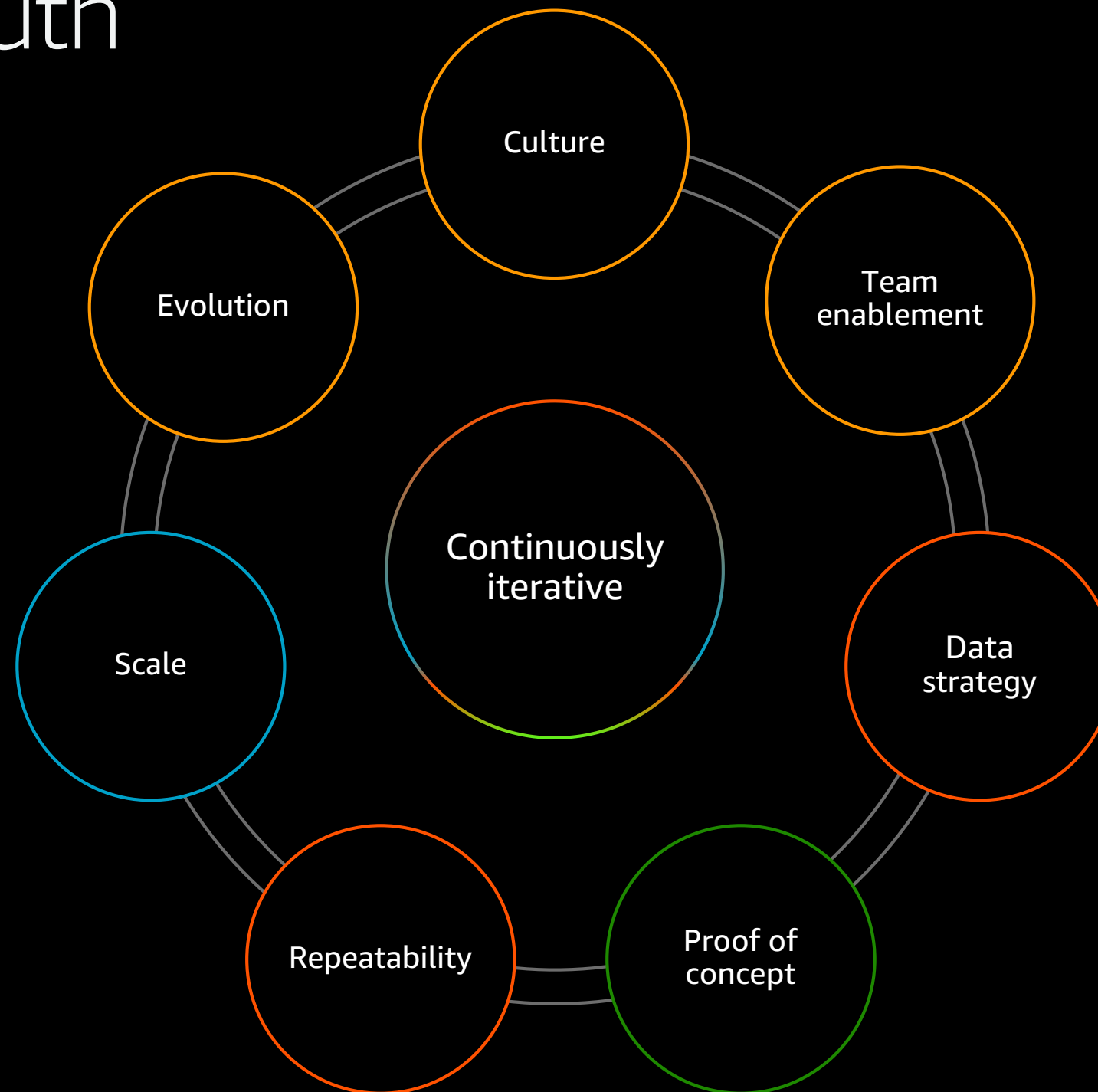
Machine Learning Center of Excellence with representation from your project teams



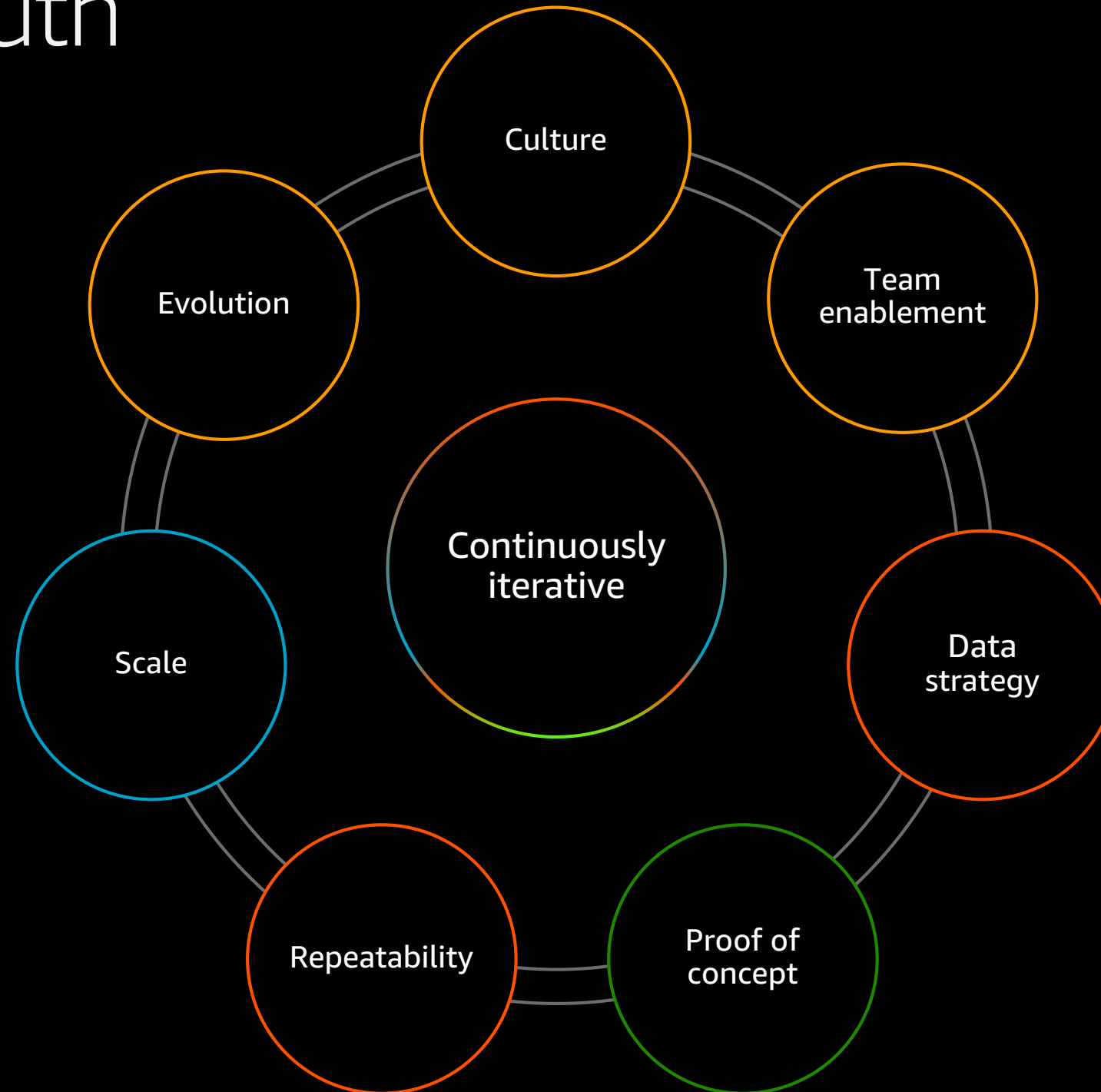
Completing the journey



A hidden truth



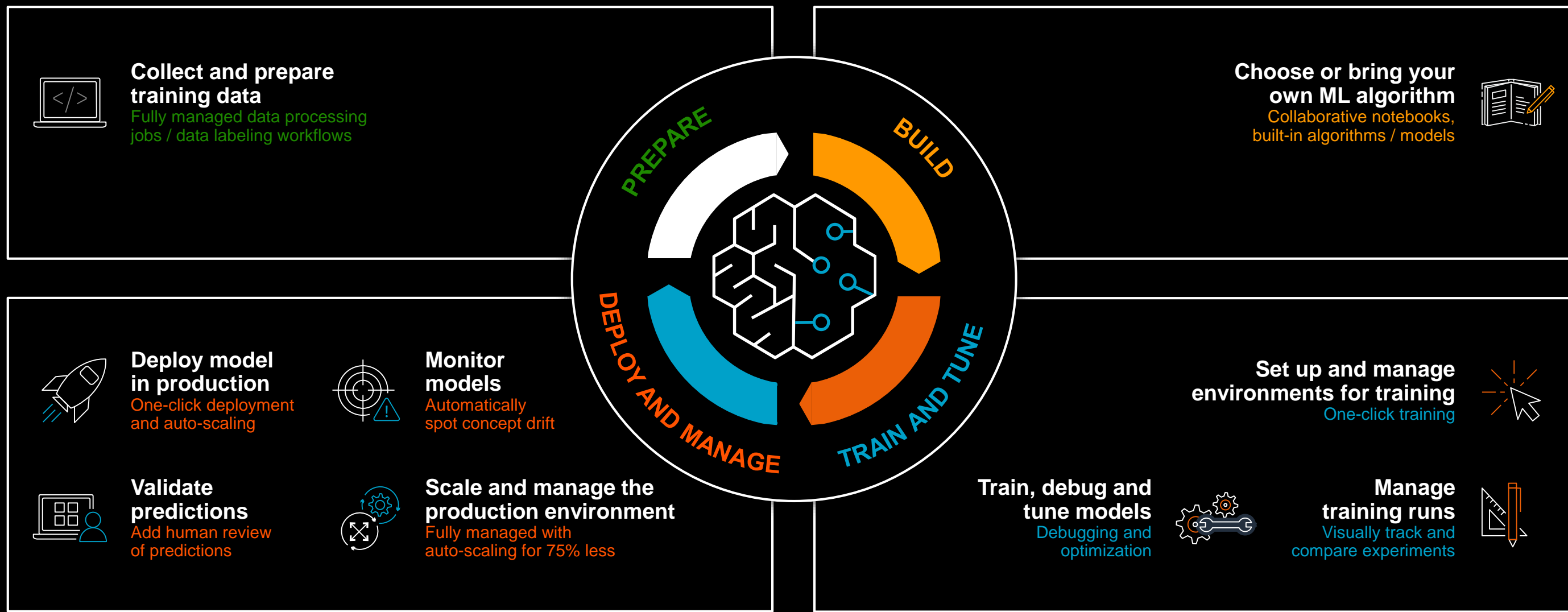
A hidden truth



Suggested techniques

- Enable data scientists to safely deploy to production as needed
- A/B tests in production
- Shadow deployments

Amazon SageMaker helps you build, train and deploy models



“With Amazon SageMaker, we can accelerate our Artificial Intelligence initiatives at scale by building and deploying our algorithms on the platform. We will create novel large-scale machine learning and AI algorithms and deploy them on this platform to solve complex problems that can power prosperity for our customers.”

intuit

Ashok Srivastava
Chief Data Officer, Intuit

“Amazon SageMaker allows GE Healthcare to access powerful artificial intelligence tools and services to advance improved patient care. The scalability of Amazon SageMaker, and its ability to integrate with native AWS services, adds enormous value for us.”

GE Healthcare

Sharath Pasupunuti
AI Engineering Leader,
GE Healthcare

Tens of thousands of customers using Amazon SageMaker



Best practices for achieving ML at scale

Team enablement

- AWS Working Backwards session
- Ensure executive alignment
- Define your culture
- Invest in training
- Integrate domain experts
- Establish collaborative platforms – cloud, studio, source control, etc.

Data strategy

- Data Lab Workshop
- Balance security with accessibility
- Label data with Amazon SageMaker Ground Truth
- Gather actual event outcomes; compare predictions vs. actuals using Amazon SageMaker Model Monitor

Proof of concept

- ML Workload Design Workshop
- Focus on business value
- Set measurable business goals
- Establish cross-functional core team
- POCs should be time bound

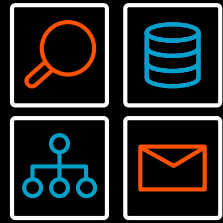
Repeatability

- MLOps Workshop
- ML Modernization Workshop
- Create pipelines to automate tasks
- Explore AWS Step Functions and the Data Science SDK
- Incorporate CI/CD practices and custom containers to increase reliability in building and deploying models

Scale

- ML CoE Workshop
- Run model training in Amazon SageMaker training jobs
- Utilize Amazon SageMaker Spot Instances to reduce training costs
- Improve effectiveness with A/B Testing and Shadow Deployments

AWS resources to help



Machine Learning Immersion Days



ML Workload Design Workshop

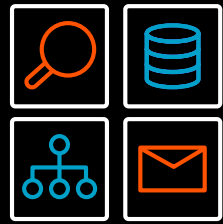


ML Solutions Lab and Embark



Partner Network for AI / ML

AWS resources to help



Machine Learning Immersion Days



ML Workload Design Workshop



ML Solutions Lab and Embark



Partner Network for AI / ML

accenture

Deloitte.

Provectus

ONICA
a rackspace technology company

quantiphi

slalom

ThoughtWorks®

REPLY

INAWISDOM ➤

Keeper

k a i n ● s

Thank you!



Please complete
the session survey