
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

INO104

AI/ML at Amazon.com

Laura Squier (she/her)
Americas AI/ML Specialist
Amazon Web Services

Choong Lee (he/him)
WW Leader – AWS Innovation Programs, Learning from Amazon
Amazon Web Services

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

"We intend to build the world's most customer-centric company. But there is no rest for the weary. I constantly remind our employees to be afraid, to wake up every morning terrified. Not of our competition, but of our customers.

We consider customers to be loyal to us – right up until the second that someone else offers them a better service."

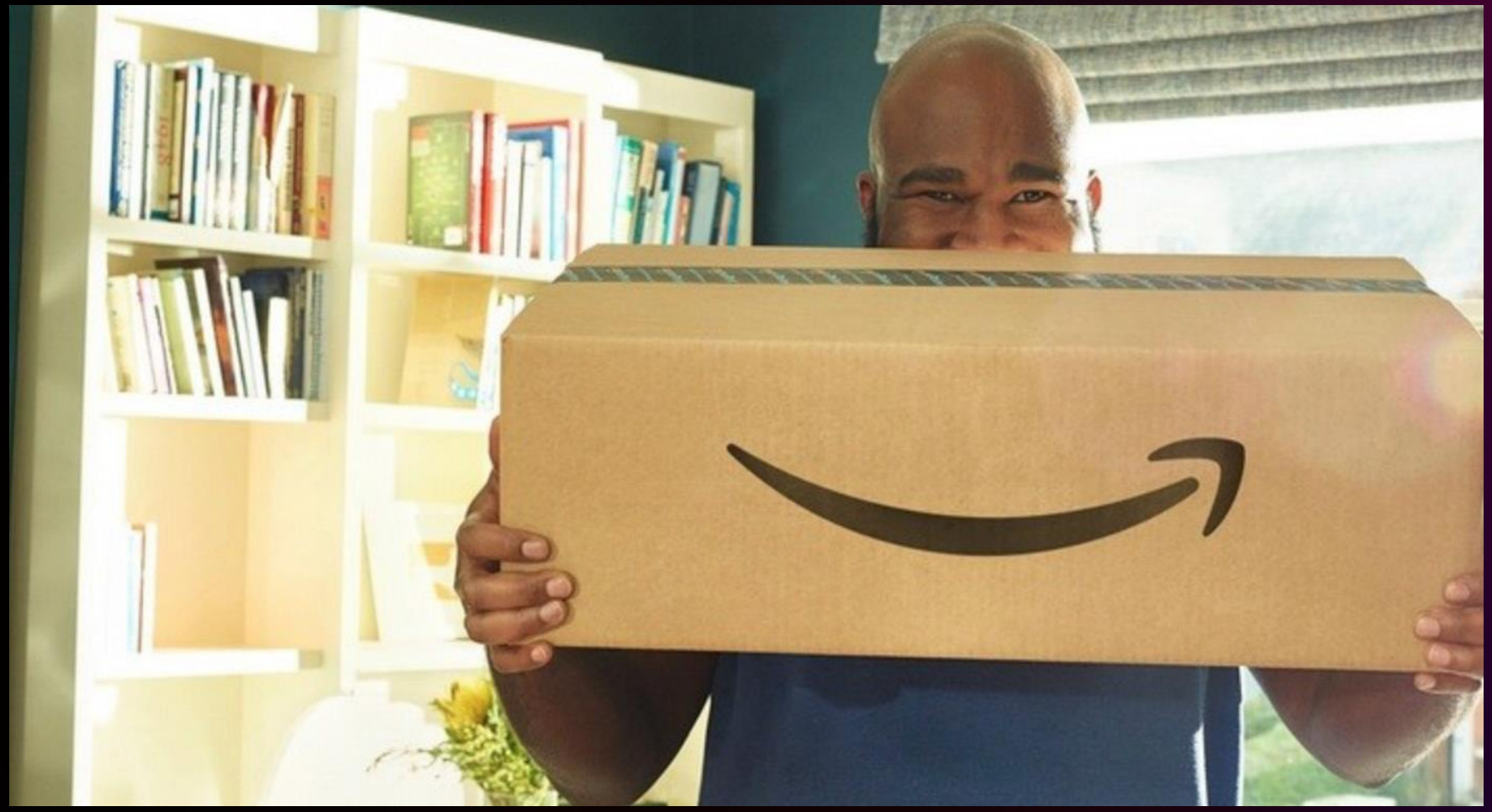
Jeff Bezos

Founder and Executive Chair, AWS

aws

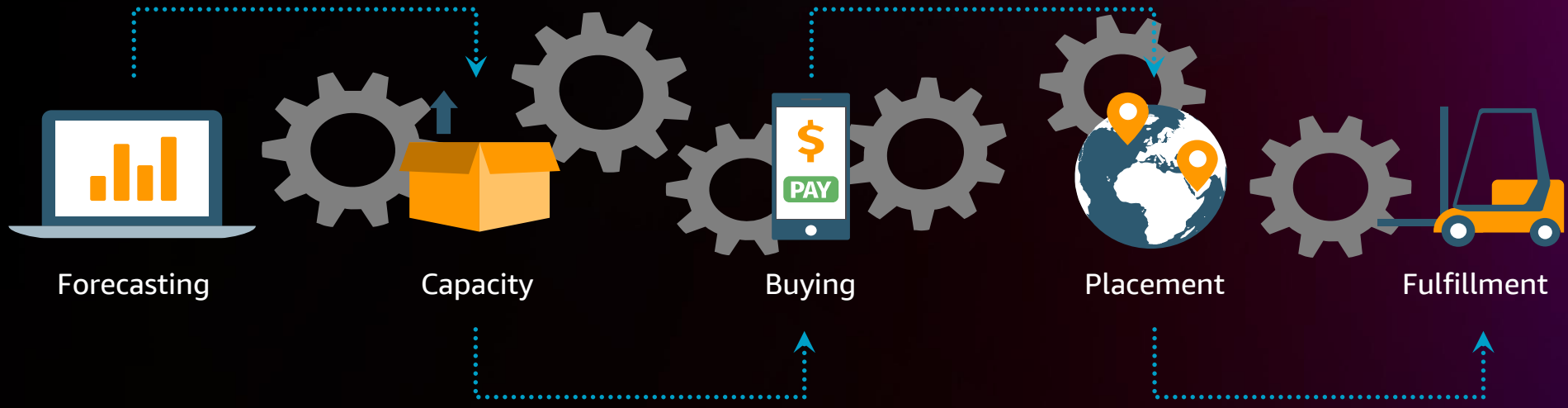
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer promise: Prime One-Day | Prime Now



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What powers the customer promise



400 million products
forecast every day

300+ fulfilment centers, 200K+ robots,
19K trailers, and 30K+ delivery vans

Accurate delivery promise

(1 billion+ packages shipped to Prime customers last year)

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How does Amazon.com get the right items in the right place . . . even before the customer clicks the “Buy” button?

Obvious





Not so obvious





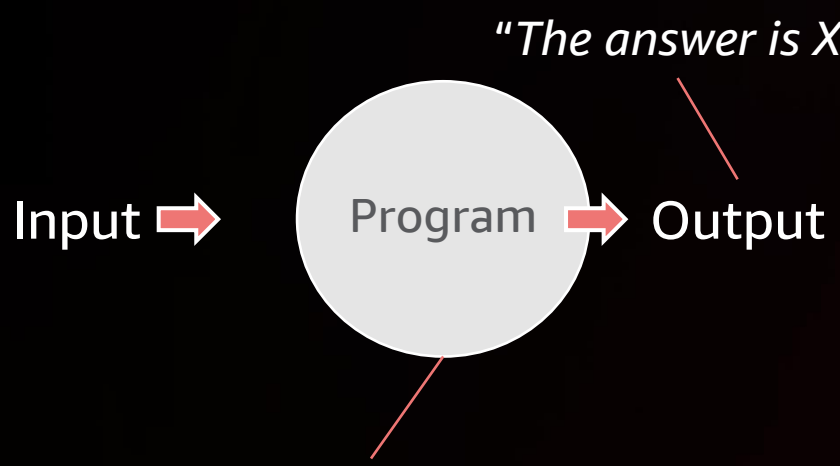


aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

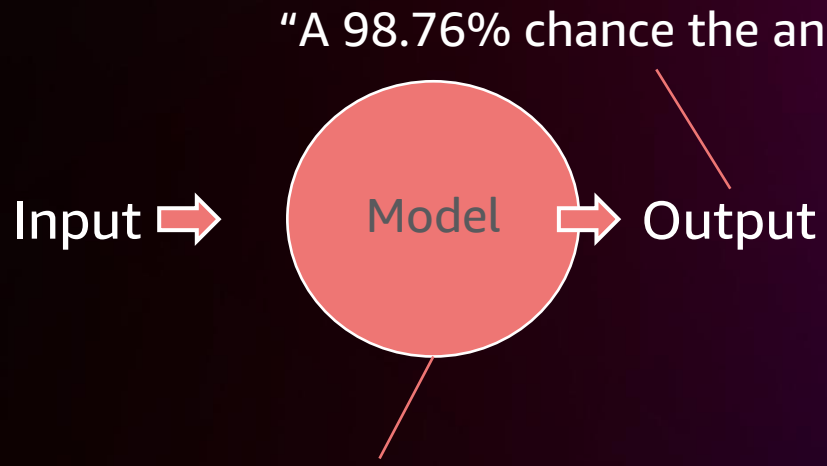
Forecasting models that learn from data

Traditional approach



A bunch of rules humans coded (for example, IF . . . THEN . . . ELSE); limited by what humans can observe and program

Machine learning



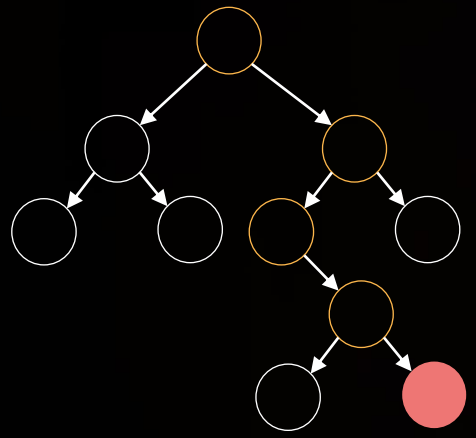
What the machine "learned" and generated after being trained by hundreds of millions of data points

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Predicting product demand at Amazon.com

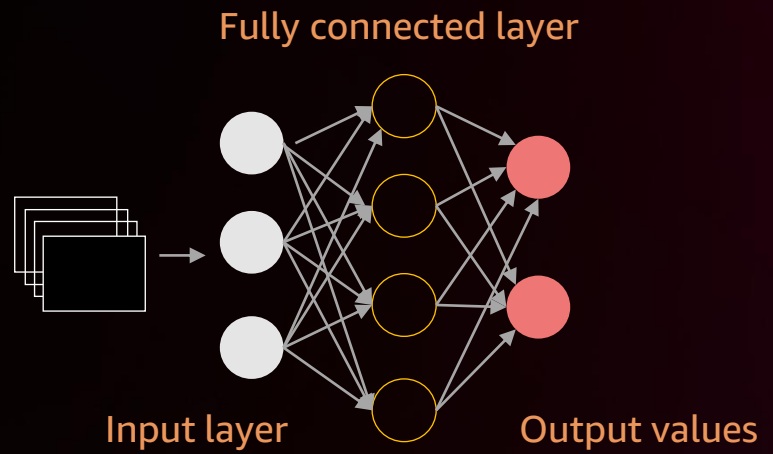
Circa 2007

Decision tree algorithm-driven model



Today

Convoluted neural networks



Billions of training data points

Powerful, non-linear relationships

Automated feature engineering

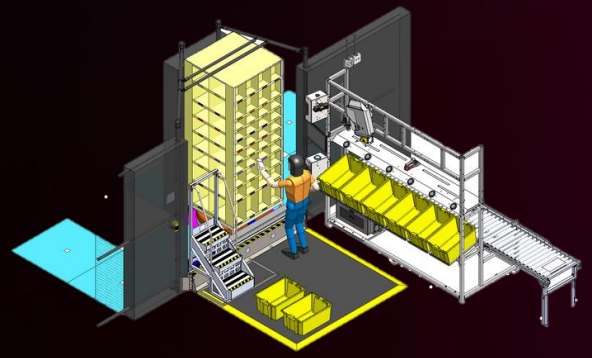
Forecast 400M+ items every day

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Package delivery:
A symphony of humans and robots







- Computer tracks millions of **randomly stowed items**
- A fleet of **robots bring storage mods** to humans
- Guided stow/pick activities through **Bin Vision System**

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What other customer experiences are transformed through ML (at scale)?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

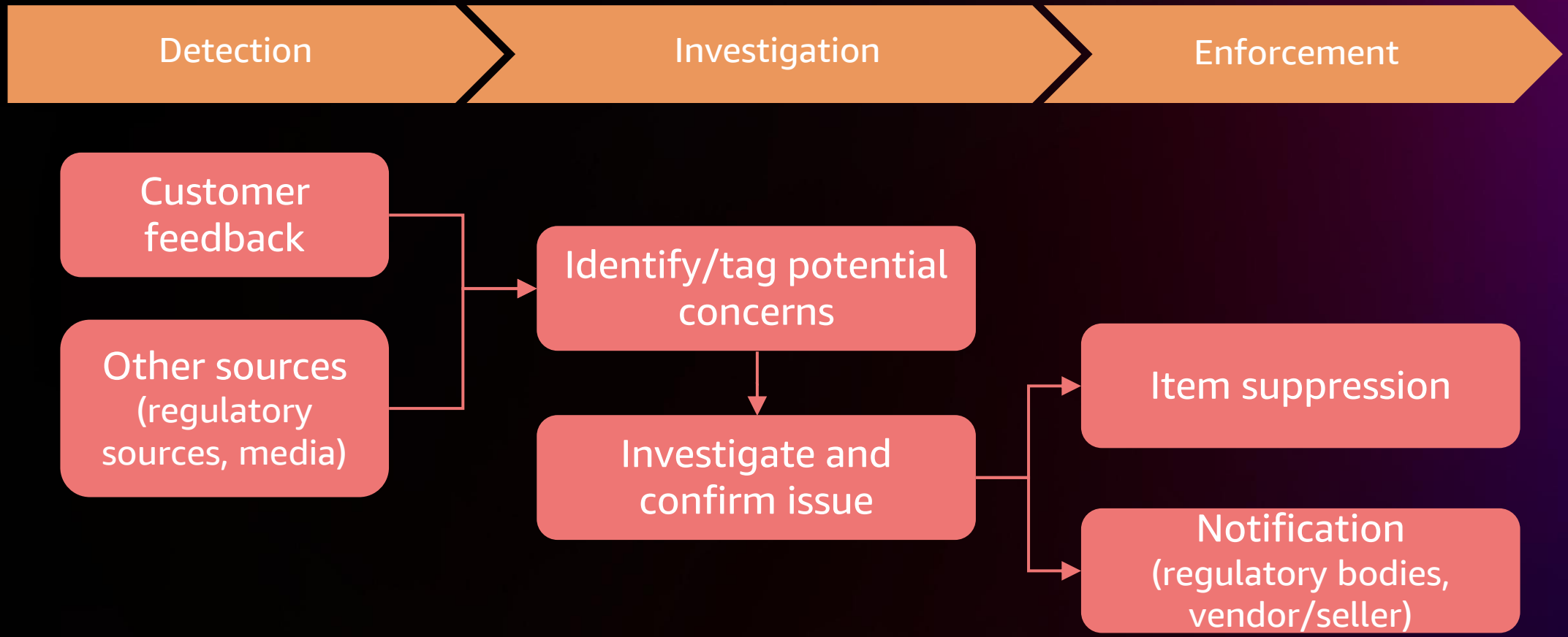
A difference in perspective





© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What are the limitations of this approach?



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Predicting safety risk before products ship



Billions of product and sales data updates

Hundreds of thousands of concerns a year

Predictive and highly accurate models

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why does shipping packaging matter?

Less of this . . .



. . . and (a lot) more of that



75% lighter
than similarly-sized boxes

40% less space
during shipping

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

ML model changing the game

Basic decision making rules

+

Visual inspection of (only) top products

Handle small fraction of products

Untapped data

Limits to what an individual can observe

Ability to process all products Amazon sells

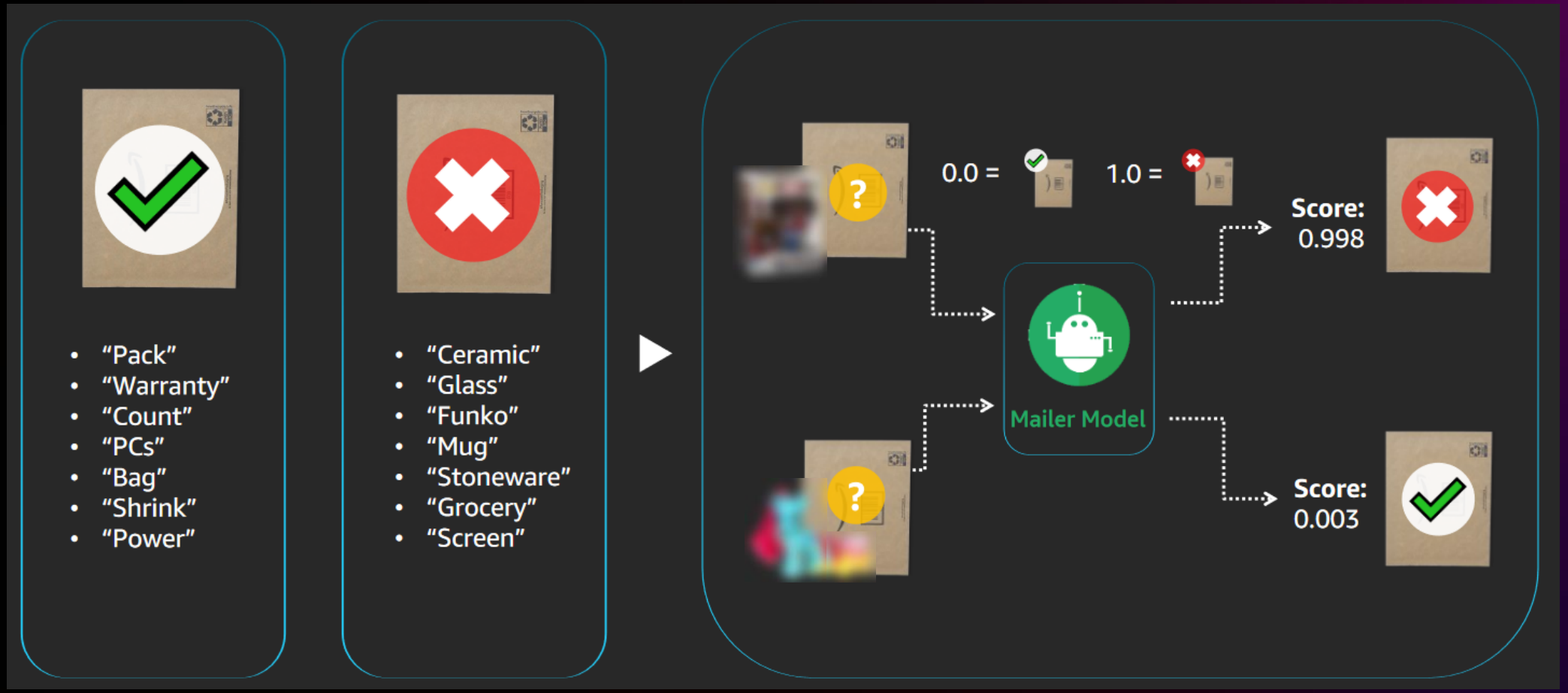
Apply complete set of intelligence generated in the delivery process

Handle non-obvious use cases (for example, collectible)

aws

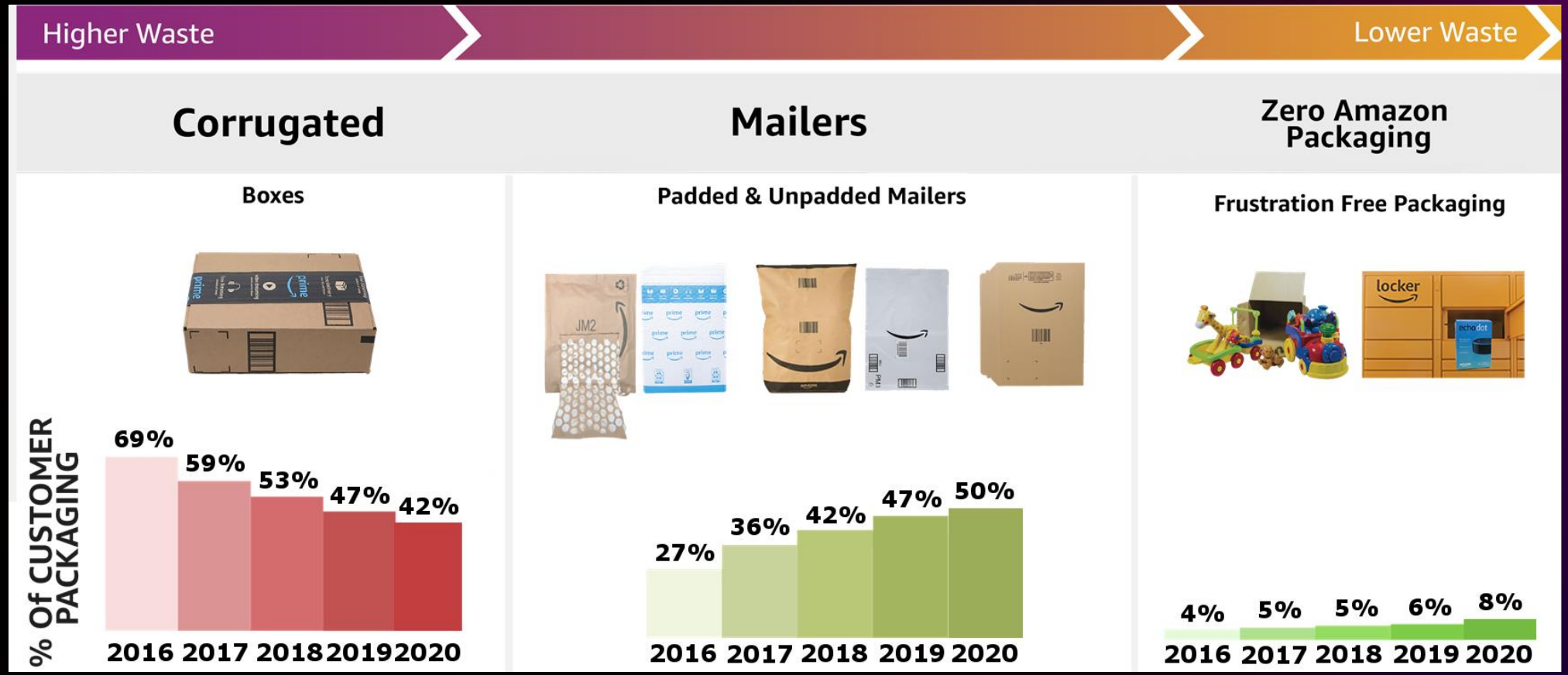
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is a collectible?



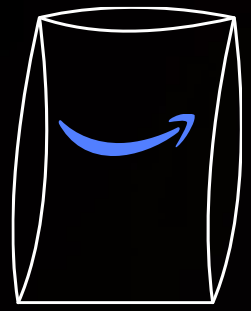
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

>50% of Amazon's shipments in mailers in 2020



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

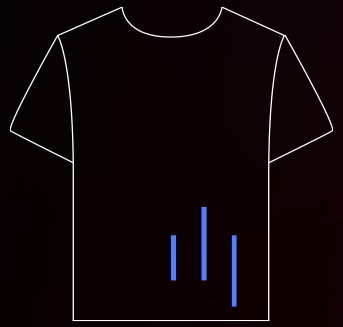
Team can launch new models in weeks



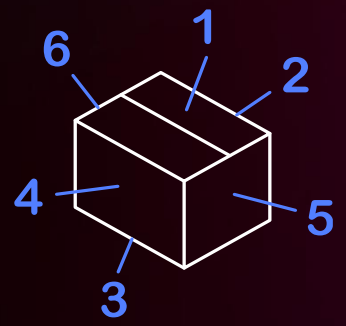
SUITABLE FOR PAPER BAG?



MIGHT LEAK?



FOLDABLE?



NO AMAZON PACKAGE NEEDED?



PRONE TO DAMAGE?

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Team can launch new models in weeks

Automation

Better, cheaper, and faster

Augmentation

Smarter, more convenient

Enrichment

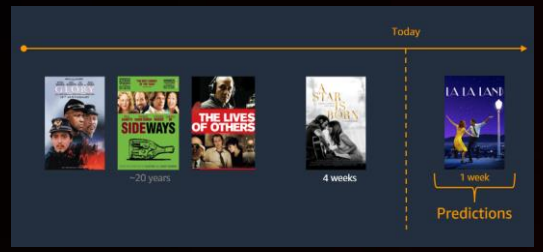
New features/services

Invention

New products and categories



Supply chain automation and operations



Personalized Amazon Video recommendation



"Style Snap" – find that style in the picture



No check out lines – ever

aws

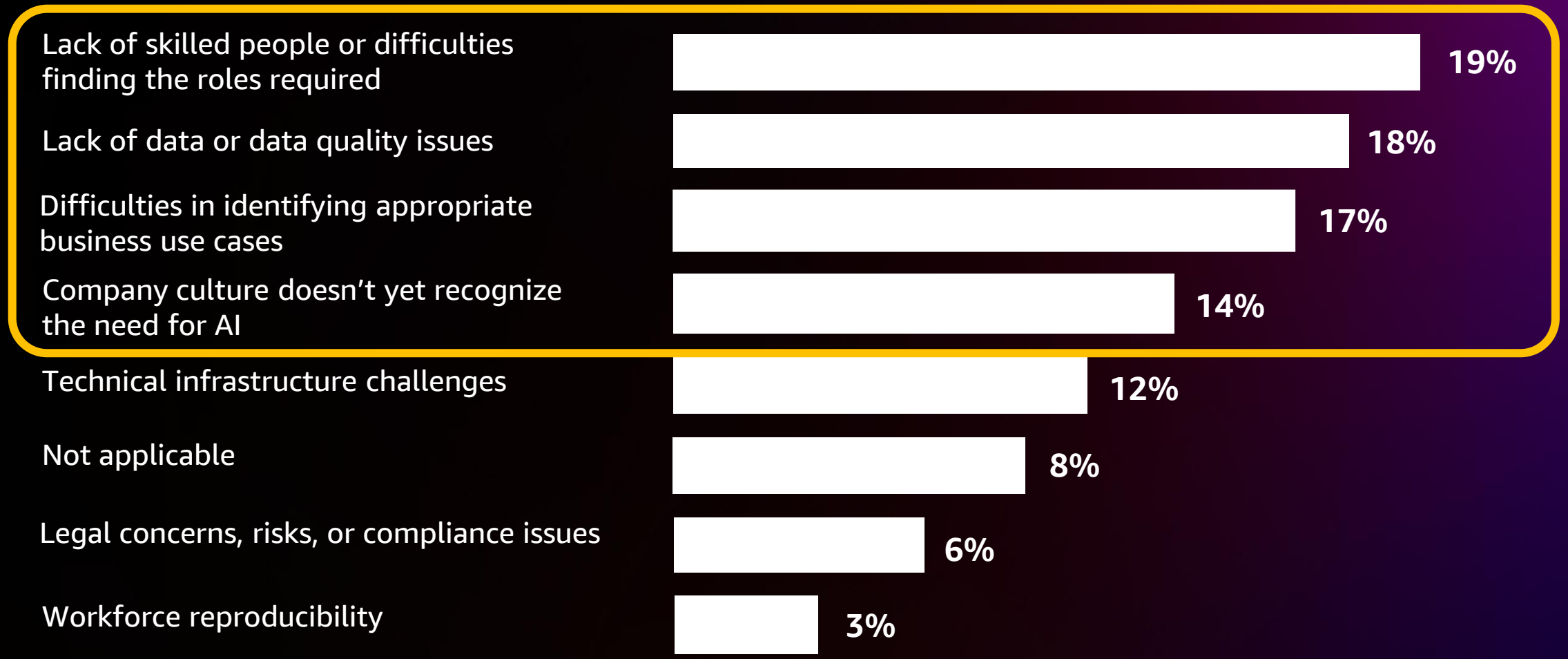
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What did Amazon learn was most important to get right in the ML journey?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

It’s (always) the basics

? What are the key challenges of AI/ML adoption?

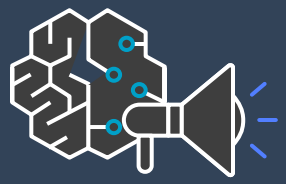


aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Source: O’Reilly AI Adoption in the Enterprise 2022

Solve for a system of enablers



ML-first mindset and culture



Enabling teams for the mission



Powerful data platform



Choosing the right first project



Scaling

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Start with one question



How will you use machine learning?

("We won't" isn't an acceptable answer)

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Break down the silos

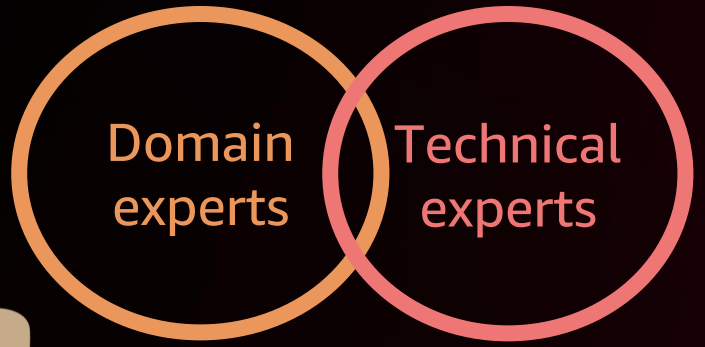




Product manager
Requirements/validations



Data scientist
Model development and testing





Data engineer
ML data foundation and performance



Research scientist
Cutting-edge theories and applications



Cloud architect
Model production and optimization

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Evolve teams as you grow and scale



Enabling teams for the mission

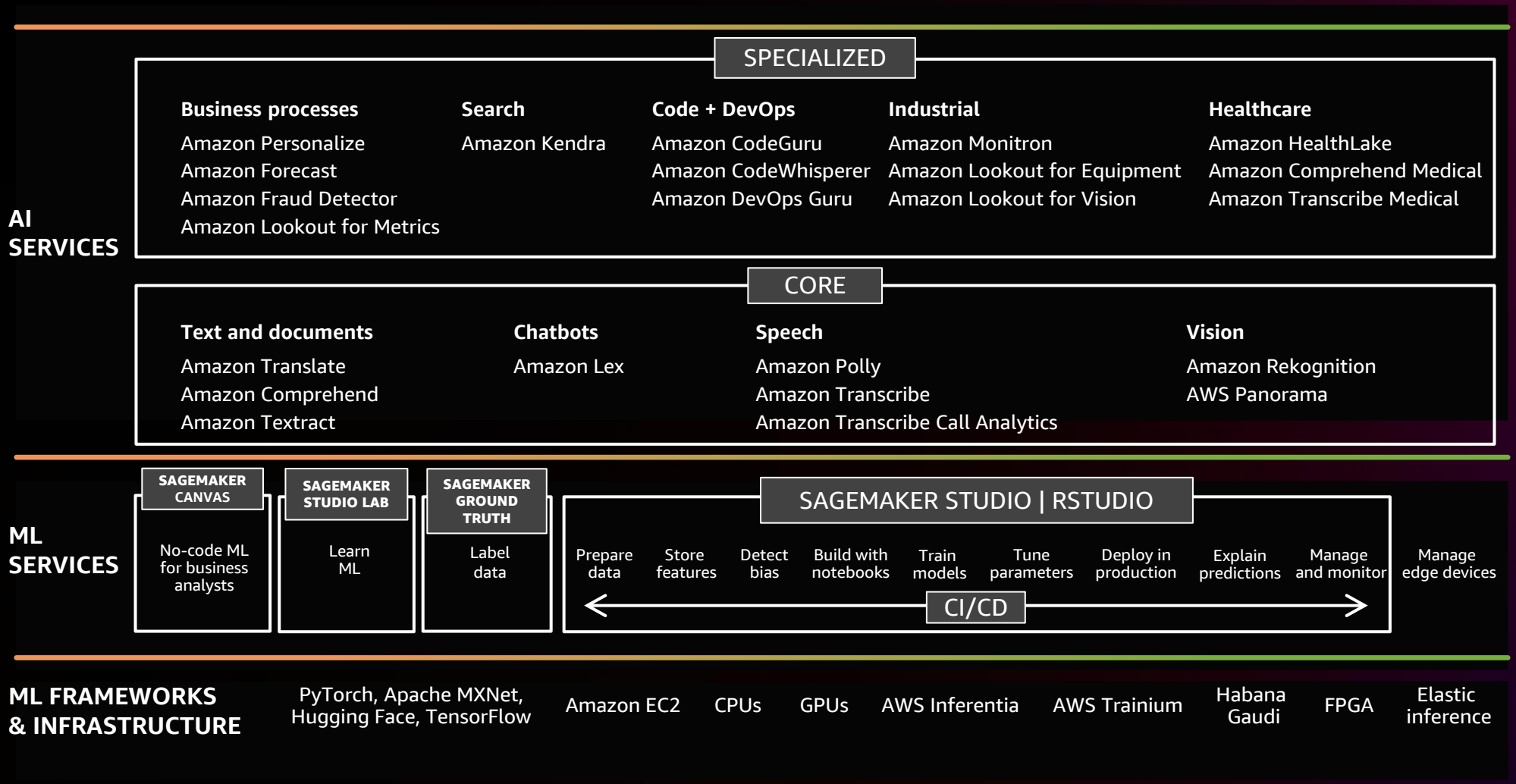
	How teams might start	Scaling within teams	Scaling across teams
Team level	Product manager	Product manager, Data scientist	Product manager, Data scientist
Function/ LoB level	Data engineer	Data engineer	Research scientist, Platform engineer, Data engineers
Enterprise level/ consultants	Data scientist	Platform engineers	Other specialist

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Apply the right tool for the job

Enabling teams for the mission

Example: AWS AI/ML service stacks



Simple, powerful and ready to use

Control and flexibility at the most granular levels

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data fuels ML-driven innovations



Amazon.com's big data marketplace

Innovation happens at the edge

Self-service discovery and subscription

Bring your own clusters/queries

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Speed and substance is critical

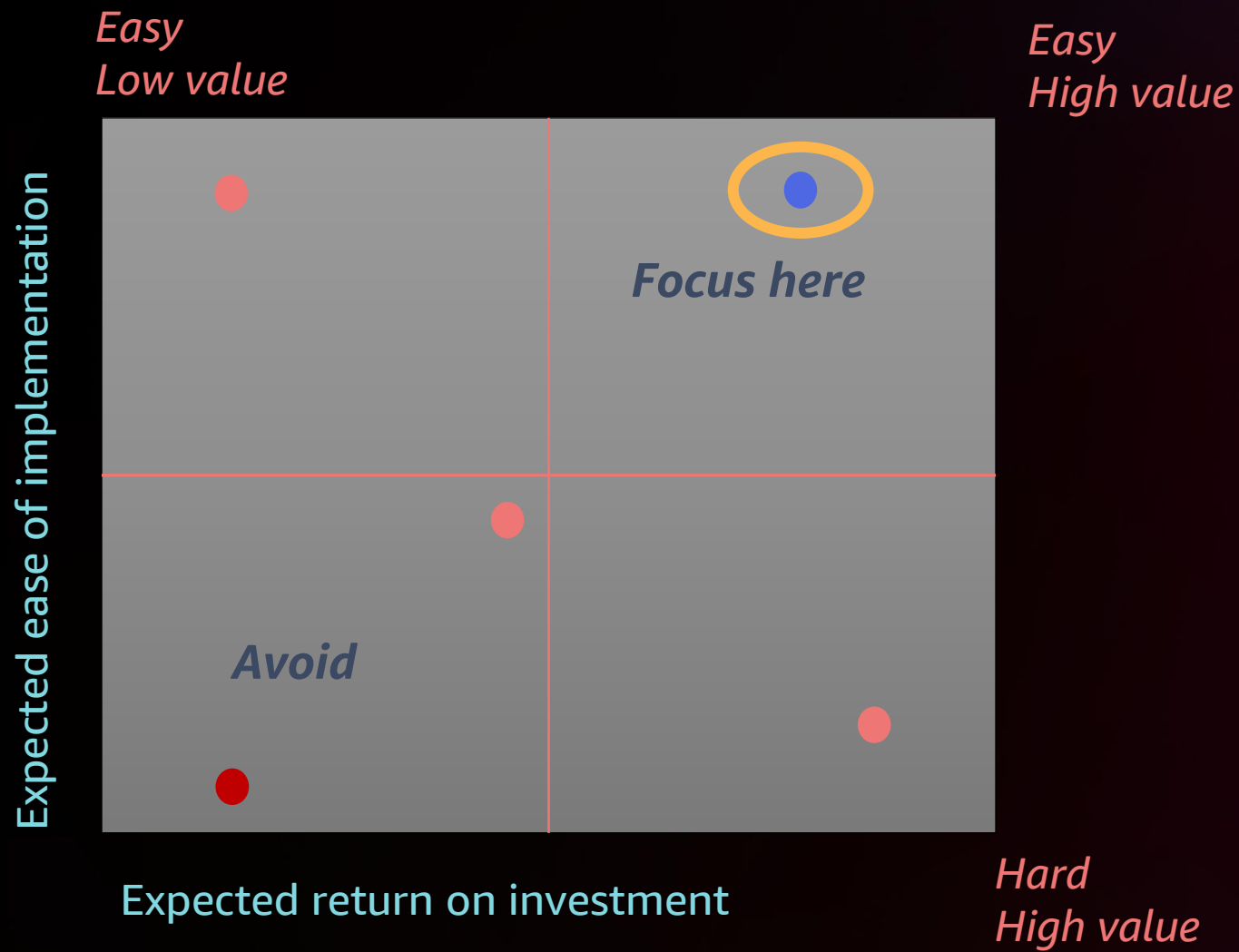


1. Does it solve a real and significant problem for your customer?
2. Where/how does machine learning unlock new capabilities?
3. Are there places where we already have a lot of untapped data?
4. Can we achieve success in the 6-10 months?
5. Is it important enough to get sustained attention and support?

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Prioritization matrix (example)

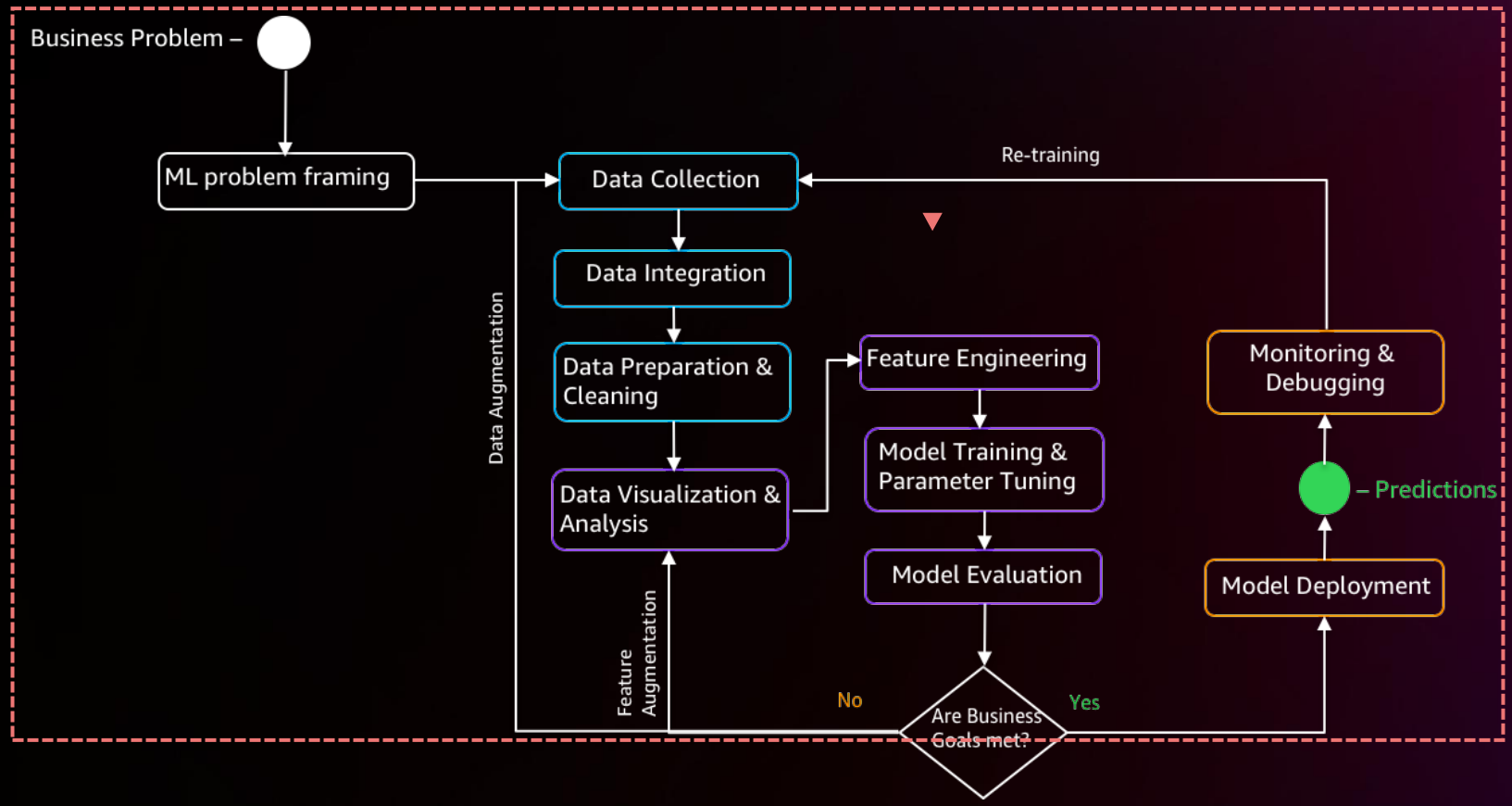




© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Disconnected tools and processes

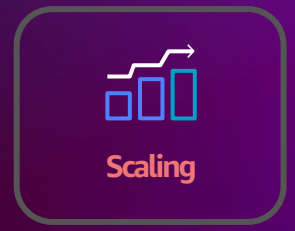




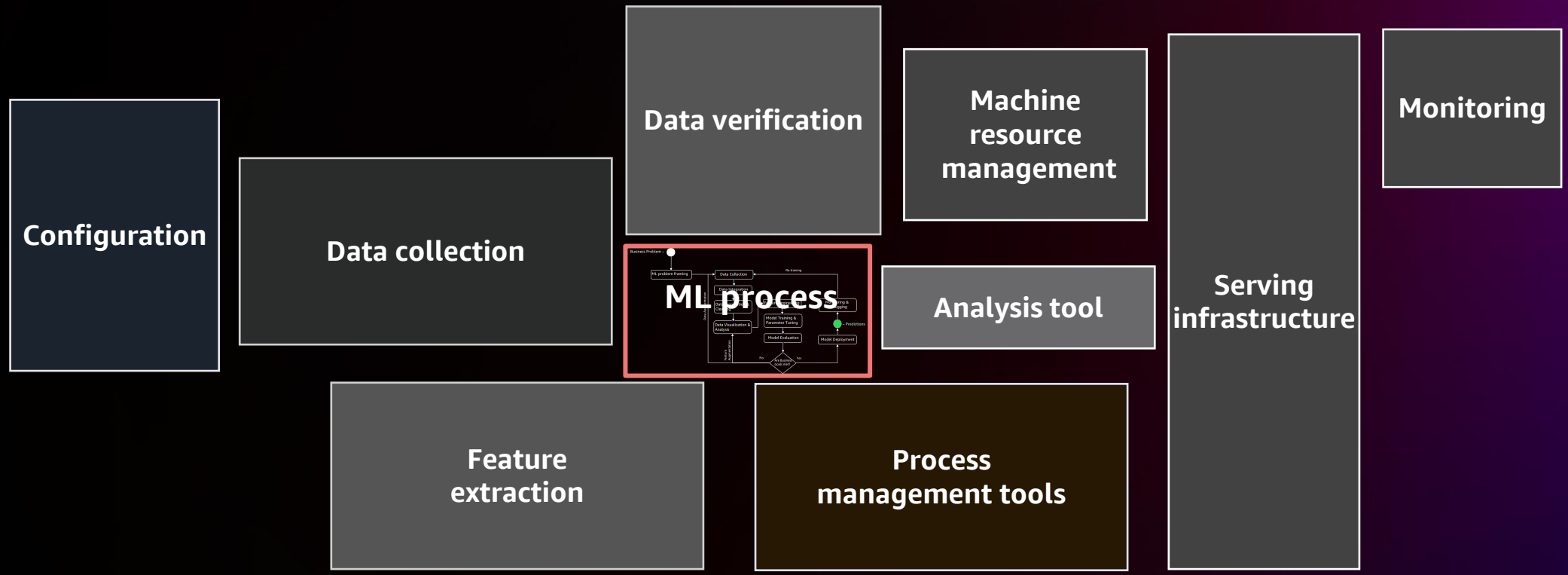
We need to accelerate this process

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

“Only a small fraction of real-world ML systems are composed of the ML code”



The overall deployment picture



Source: Hidden Technical Debt in Machine Learning Systems [D. Sculley, & al.] – 2015
https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Integrated tools and processes for ML operations



Discoverable and accessible data

Version control and artifact repositories

Continuous deployment orchestration for pipelines

Infrastructure for multiple environments and experiments

Model performance and assessment tracking

Model monitoring and refresh orchestration

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Summary

1. Amazon sees AI/ML as a table stake toward delivering better services
2. ML delivers differentiated impact and non-linear scaling for Amazon
3. Building and scaling ML usage at enterprise need requires clear strategies across the “system of enablers”
4. Creating the right way of working and providing teams with the right tools for the job is critical toward building and scaling a ML-driven innovation engine
5. Identify and remove blocks to the path to production

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Laura Squier
squierls@amazon.com

Choong Lee
choong@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.