



SUMMIT  
ONLINE

# How to build smart applications without a machine learning background

Alex Casalboni

Developer Advocate

@alex\_casalboni

# Agenda

Why are we here?

What's new in the stack?

BYOD – Bring your own data

Building blocks approach

# Why are we here?

# Our mission at AWS

---

Put machine learning (ML) in the  
hands of every developer

# The AWS ML stack

## AI services

Vision			Speech			Text			NEW!	NEW!	NEW!	NEW!
Amazon Rekognition			Amazon Polly			Amazon Transcribe + Medical NEW			Amazon Kendra	Amazon Lex	Amazon Personalize	Amazon Forecast
						Amazon Comprehend + Medical						
						Amazon Translate						
						Amazon Textract						
												</

# What's new in the stack?

New!

# Amazon Kendra

Transform the way you search and interact with enterprise data  
– no ML experience required



# Amazon Kendra

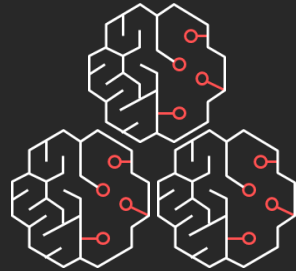
Rethinking enterprise search



Natural language  
queries



NLU and  
ML core



Domain  
expertise



Continuous  
improvement



Native  
connectors



Simple API  
and console  
experiences,  
code samples

Easy to find what you are looking for

Simple and quick to set up

# Amazon Kendra

## Connectors



Amazon S3



File systems (SMB)



Web crawler



Databases



SharePoint Online



SharePoint 2013, 2016, 2019



Box



Dropbox



Microsoft Exchange



OneDrive



Google Drive



Salesforce



Confluence



Jira



ServiceNow



Zendesk



Jive

...and more coming in 2020

1

# Create index [info](#)

## Index details [info](#)

Index name

hr-index

Maximum of 1000 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your AWS account in the region.

Description - *Optional*

Human resources index

IAM Role [info](#)

Amazon Kendra requires permissions to access your S3 bucket and CloudWatch logs. Choose an existing IAM role or let Kendra create a role for you here.

Choose an option



Encryption key - *Optional* [info](#)

By default, Kendra will encrypt your data automatically. Alternatively, you can provide your own KMS key.

Choose an option



Cancel

Create

2

# Add data source

## Name data source

Data source name

hr-data

Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

Description - *Optional*

Data for hr site

## Select connector type [info](#)

Choose connector type based on where your data resides.

Connectors enable your data source to sync with Kendra. [Learn more.](#)



Sharepoint



Intranet/Web



Amazon S3



Box



Salesforce

3



## Kendra Search console

Test search & Deploy

Where is the i



Where is the it help desk day1

Where is the it help desk nearest me

Where is the it support desk dopler

Where is the it support desk kumo

Where is the it support desk blackfoot

New!

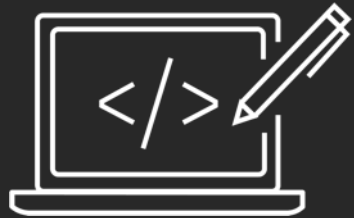
# Amazon CodeGuru

Automate code reviews and identify expensive lines of code  
– no ML experience required

# Amazon CodeGuru

ML helps build and run high-performing software

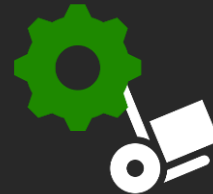
CodeGuru Reviewer



Write + Review



Build + Test



Deploy

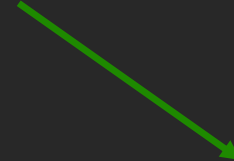
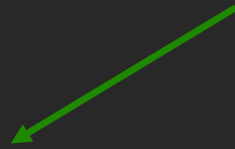


Measure



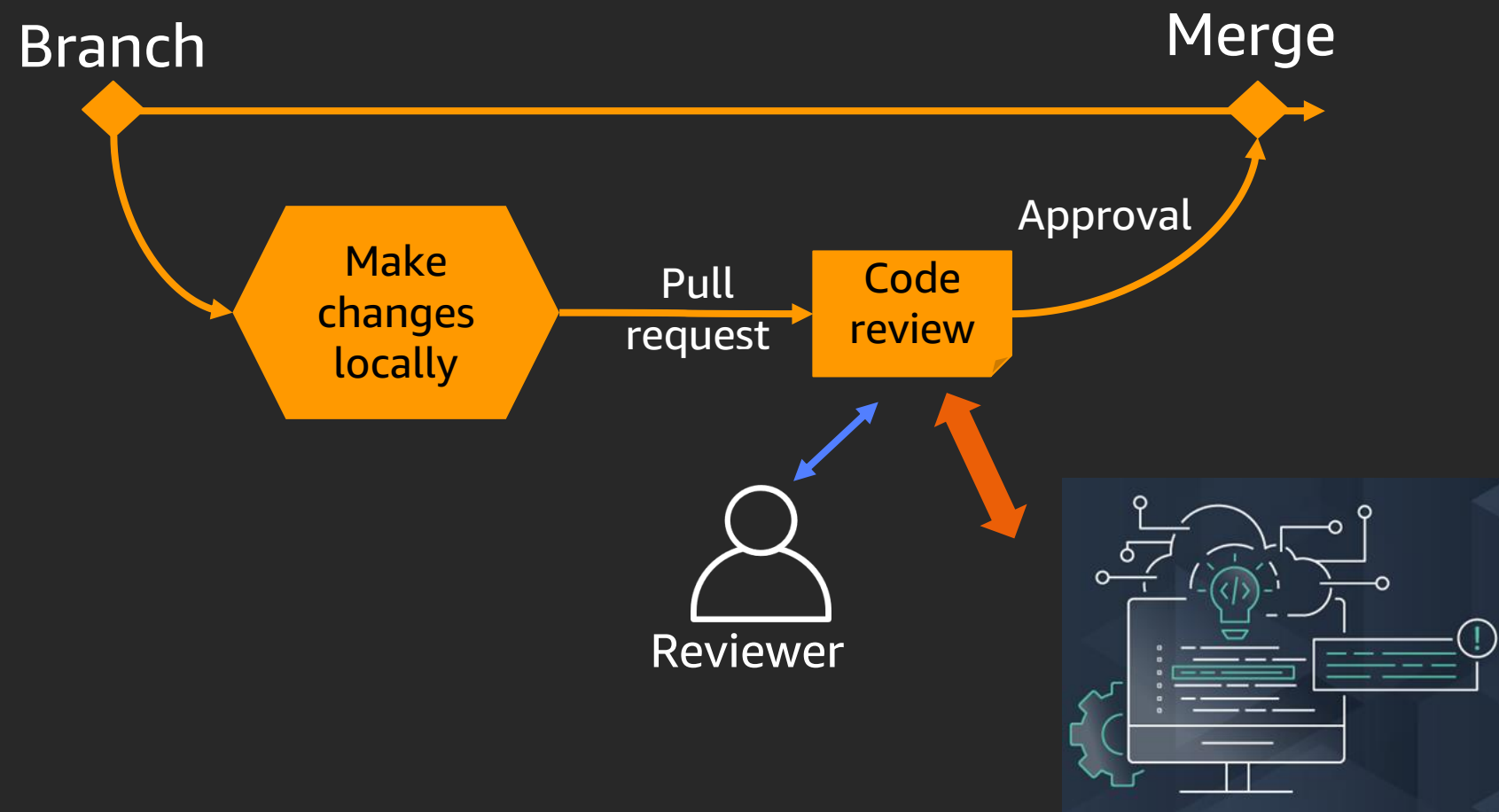
Improve

CodeGuru Profiler



# Amazon CodeGuru Reviewer

Flags critical defects and reliability issues in source code



Amazon CodeGuru Reviewer augments the human code review process, it does not replace it

Amazon CodeGuru Reviewer



# Code areas addressed by CodeGuru Reviewer

## **AWS best practices**

*Incorrect use results in performance issues (e.g., polling) or correctness and completeness issues (e.g., pagination)*

## **Concurrency**

*Incorrect use results in correctness issues (e.g., missing synchronization) or performance issues*

## **Resource leaks**

*Incorrect handling (e.g., not releasing database connection) results in slowdown and impacts availability*

## **Sensitive information leak**

*Leakage of sensitive information (e.g., logging of credit card number) leads to compliance issues*

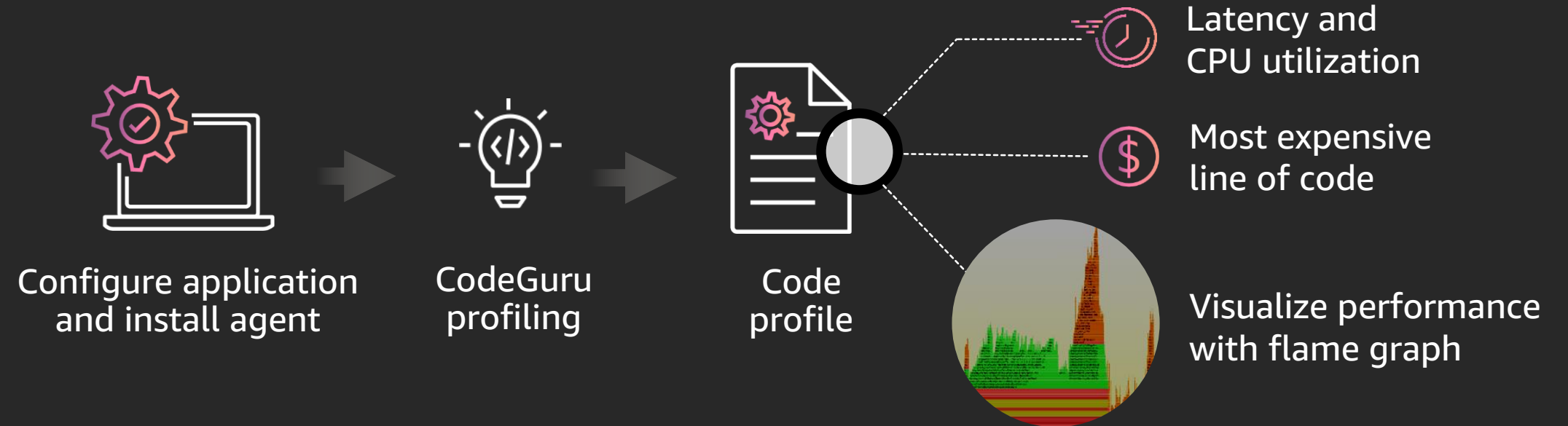
## **Code defects discovered by mining data**

*Correcting issues (e.g., not creating a client for each AWS Lambda invocation) improves code quality*

# CodeGuru Profiler

Find your most expensive lines of code

How does it work?



# Prime Day improvements with Amazon CodeGuru

Optimization of the most expensive line of  
code identified by Amazon CodeGuru

Prime Day 2017 vs. 2018

325%

Increase in  
CPU utilization

39%

Lower  
cost

New!

# Amazon Fraud Detector

Detect more online fraud faster  
– no ML experience required

# Amazon Fraud Detector

Automate online fraud detection



Pre-built fraud  
detection model  
templates



Automatic  
creation of  
custom fraud  
detection models



Models learn  
from past attempts  
to defraud  
Amazon.com



Amazon  
SageMaker  
integration



One interface to  
review past  
evaluations and  
detection logic

# Generating fraud predictions

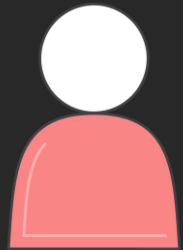
**Guest checkout: Purchase**

*IP: 1.23.123.123*

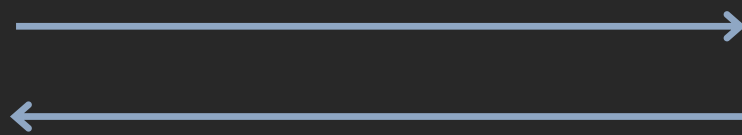
*Email: john@example.com*

*Payment: AnyCompany Bank*

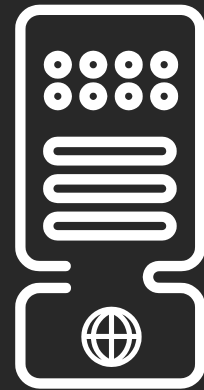
...



Customer



*Purchase approved*



Website

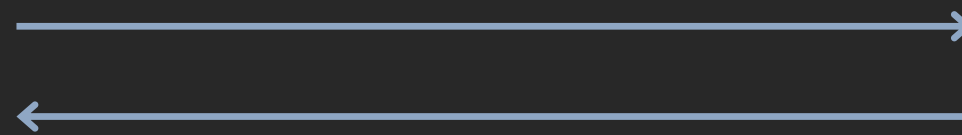
**Call service with:**

*IP: 1.23.123.123*

*Email: john@example.com*

*Payment: AnyCompany Bank*

...



**Fraud Detector returns:**

*Outcome: Approved*

*ML score: 160*

Amazon Fraud  
Detector



Customized  
fraud detection  
endpoint

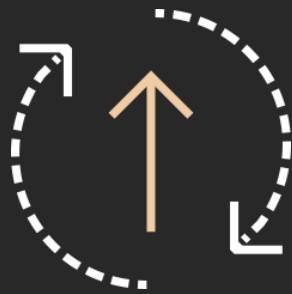
New!

# Amazon Augmented AI (Amazon A2I)

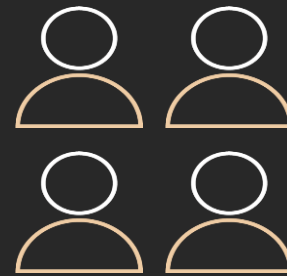
Implement human review of ML predictions  
– no ML experience required

# Amazon A2I

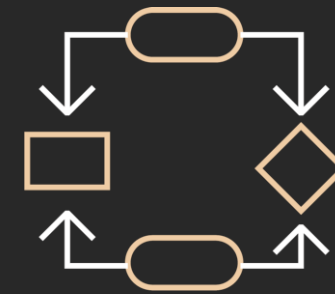
Easily build workflows that are required for human review of predictions



Pre-built workflows  
and templates  
reduce time to market



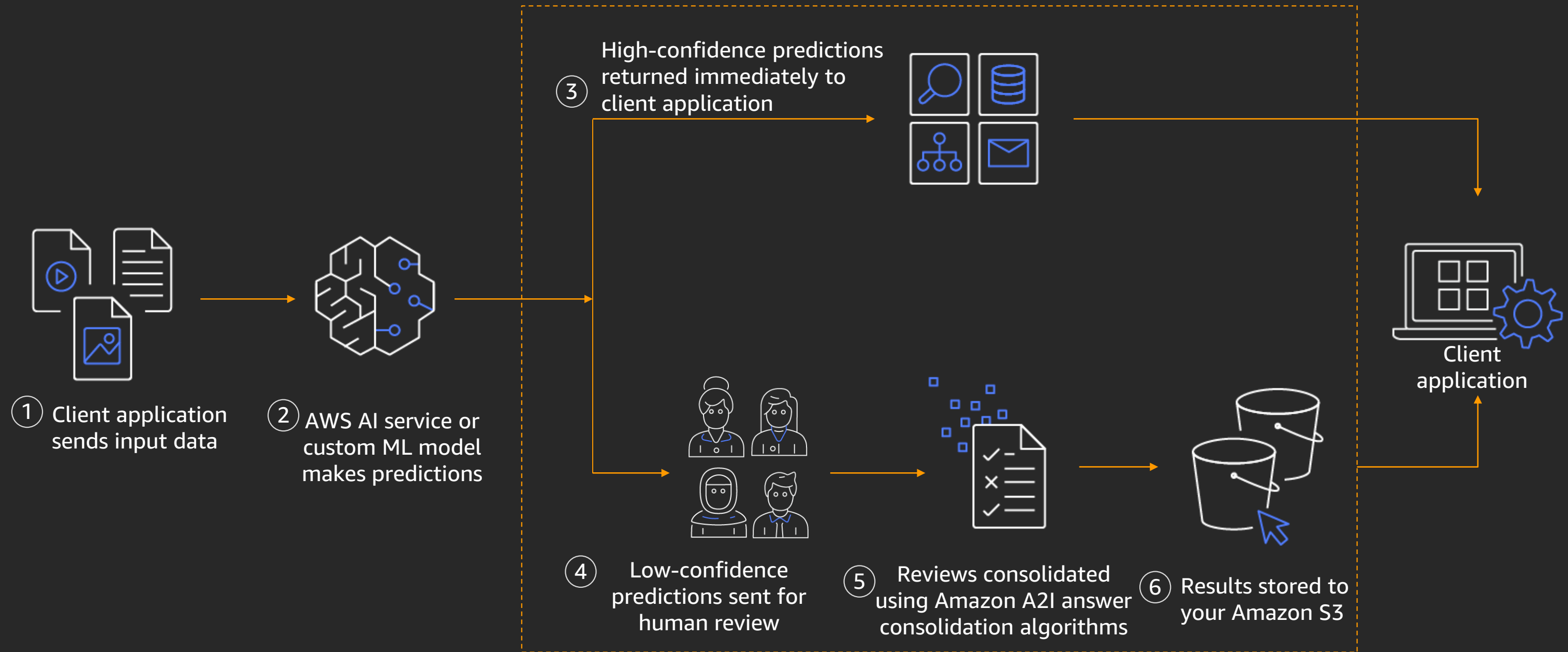
Work with  
your choice  
of workers



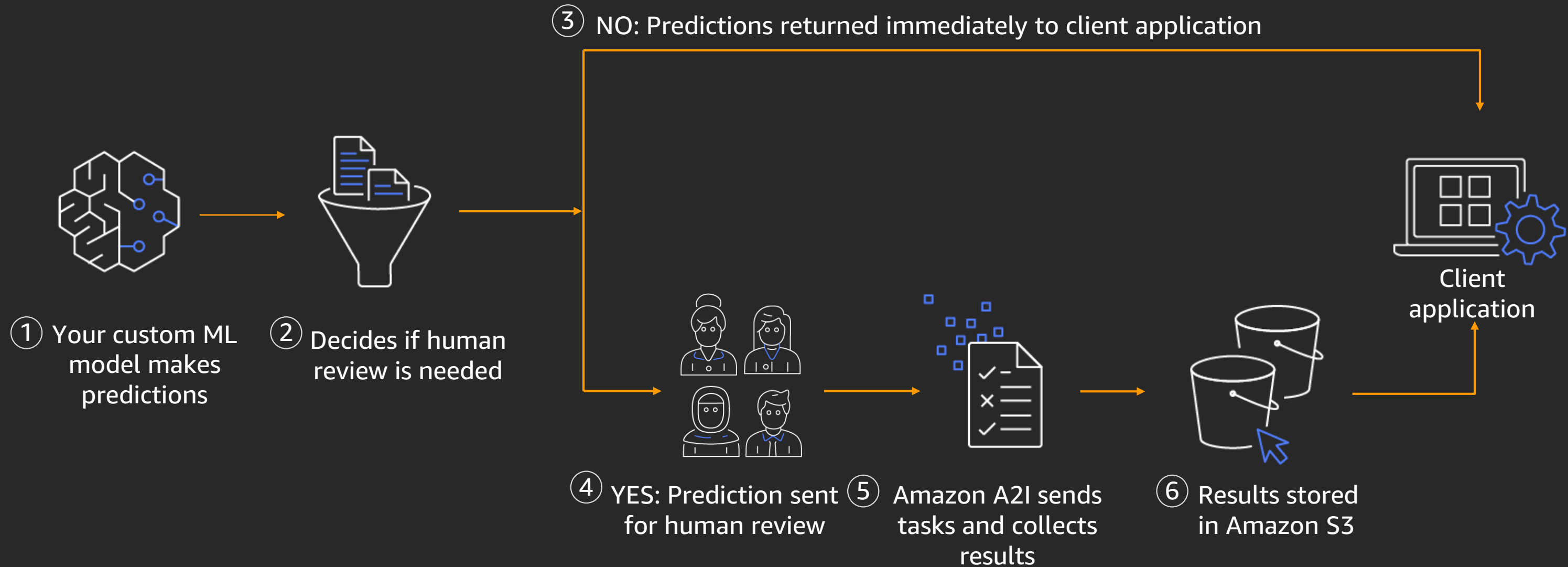
Answer consolidation  
algorithms to  
improve accuracy



# How Amazon A2I works



# Use Amazon A2I with any ML model



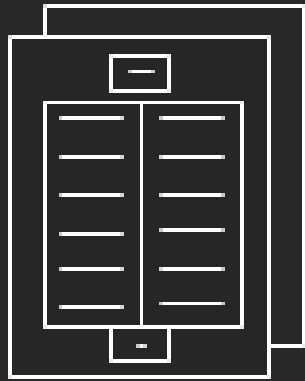
# Documents & images

# Amazon Textract

OCR++ service to easily extract text and data from virtually any document – no ML experience required

# Amazon Textract features

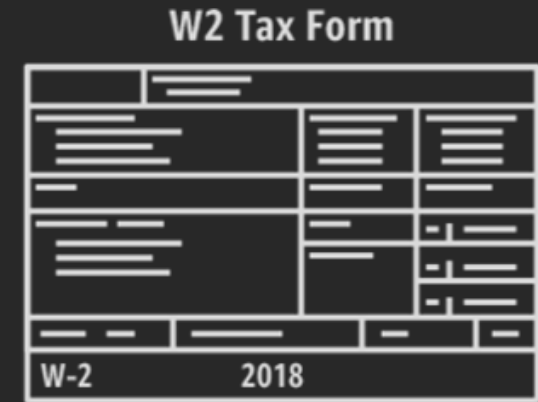
Easily extract text and data from virtually any document



Text extraction



Table extraction



Form extraction

Bounding boxes | Adjustable confidence thresholds | Built-in human review workflow | HIPAA-eligible

# Amazon Textract examples

**Uniform Residential Loan Application**

This application is designed to be completed by the applicant(s) with the Lender's assistance. Applicants should complete this form as "Borrower" or "Co-Borrower," as applicable. Co-Borrower information must also be provided (and the appropriate box checked) when the income or assets of a person other than the Borrower (including the Borrower's spouse) will be used as a basis for loan qualification or if the income or assets of the Borrower's spouse or other person who has community property rights pursuant to state law will not be used as a basis for loan qualification, but his or her liability must be considered because the spouse or other person has community property rights pursuant to applicable law and Borrower resides in a community property state, the security property is located in a community property state, or the Borrower is relying on other property located in a community property state as a basis for repayment of the loan.

If this is an application for joint credit, Borrower and Co-Borrower each agree that we intend to apply for joint credit (sign below):

Borrower \_\_\_\_\_ Co-Borrower \_\_\_\_\_

**I. TYPE OF MORTGAGE AND TERMS OF LOAN**

Mortgage Applied For: ☐ VA ☐ Conventional ☐ Other (specify): \_\_\_\_\_  
☐ FHA ☐ USDA/Rural ☐ Hybrid/Reverse

Agency Case Number \_\_\_\_\_ Lender Case Number \_\_\_\_\_

Amount \$ \_\_\_\_\_ Interest Rate % \_\_\_\_\_ No. of Months \_\_\_\_\_ Amortization Type: ☐ Fixed Rate ☐ Other (specify): \_\_\_\_\_  
☐ ARM (specify): \_\_\_\_\_

**II. PROPERTY INFORMATION AND PURPOSE OF LOAN**

Subject Property Address (street, city, state & ZIP) \_\_\_\_\_ No. of Units \_\_\_\_\_

Legal Description of Subject Property (attach description if necessary) \_\_\_\_\_ Year Built \_\_\_\_\_

Purpose of Loan: ☐ Purchase ☐ Construction ☐ Other (specify): \_\_\_\_\_ Property will be: ☐ Primary Residence ☐ Secondary Residence ☐ Investment  
☐ Refinance ☐ Construction-Permanent

**Complete this line if construction or construction-permanent loan.**

Year Let Assigned	Original Cost	Amount Standing Line	(a) Present Value of Lot	(b) Cost of Improvements	Total (a + b)
\$	\$	\$	\$	\$	\$

**Complete this line if this is a refinance loan.**

Year Assigned	Original Cost	Amount Standing Line	Purpose of Refinance	Describe Improvements	<input type="checkbox"/> made <input type="checkbox"/> to be made
\$	\$	\$		Cost \$	

Financial documents

**DEPARTMENT OF HEALTH AND HUMAN SERVICES**  
Food and Drug Administration  
**INVESTIGATIONAL NEW DRUG APPLICATION (IND)**  
(Title 21, Code of Federal Regulations (CFR) Part 312)

Form Approved: OMB No. 0919-0014  
Expiration Date: February 28, 2019  
See PRA Statement on page 2.

NOTE: No drug/biologic may be shipped or clinical investigation begun until an IND for that investigation is in effect (21 CFR 312.40).

1. Name of Sponsor \_\_\_\_\_ 2. Date of Submission (mm/dd/yyyy) \_\_\_\_\_

3. Sponsor Address \_\_\_\_\_  
Address 1 (Street address, P.O. box, company name etc.) \_\_\_\_\_  
Address 2 (Apartment, suite, unit, building, floor, etc.) \_\_\_\_\_  
City \_\_\_\_\_ State/Province/Region \_\_\_\_\_  
Country \_\_\_\_\_ ZIP or Postal Code \_\_\_\_\_

4. Telephone Number (include country code if applicable and area code) \_\_\_\_\_  
EA, IND Number (if previously assigned) \_\_\_\_\_

5. Name of Drug (include all available names: Trade, Generic, Chemical, or Code) \_\_\_\_\_  
6B. Select One: ☐ Commercial ☐ Research

7A. (Proposed) Indication for Use \_\_\_\_\_  
Is this indication for a rare disease (prevalence <200,000 in U.S.)? ☐ Yes ☐ No  
Does this product have an FDA Orphan Designation for this indication? ☐ Yes ☐ No  
If yes, provide the Orphan Designation number for this indication: \_\_\_\_\_

7B. SNOMED CT Indication Disease Term (Use continuation page for each additional indication and respective coded disease term)

8. Phase of Clinical Investigation to be conducted: ☐ Phase 1 ☐ Phase 2 ☐ Phase 3 ☐ Other (Specify): \_\_\_\_\_

Healthcare documents

**Voter Registration Application**  
Before completing this form, review the General, Application, and State specific instructions.

Are you a citizen of the United States of America? ☐ Yes ☐ No  
Will you be 18 years old on or before election day? ☐ Yes ☐ No  
If you checked "No" in response to either of these questions, do not complete form.  
(Please see state-specific instructions for rules regarding eligibility to register prior to age 18.)

This space for office use only.

1. ☐ Mr. ☐ Mrs. ☐ Miss Last Name \_\_\_\_\_ First Name \_\_\_\_\_ Middle Name(s) \_\_\_\_\_ ☐ Jr. ☐ Sr. ☐ III

2. Home Address \_\_\_\_\_ Apt. or Lot # \_\_\_\_\_ City/Town \_\_\_\_\_ State \_\_\_\_\_ Zip Code \_\_\_\_\_

3. Address Where You Get Your Mail If Different From Above \_\_\_\_\_ City/Town \_\_\_\_\_ State \_\_\_\_\_ Zip Code \_\_\_\_\_

4. Date of Birth \_\_\_\_\_ Telephone Number (optional) \_\_\_\_\_ ID Number - (see back in the instructions for your state) \_\_\_\_\_  
Month Day Year 5

7. Choice of Party (see back in the instructions for your state) \_\_\_\_\_ 8. Race or Ethnic Group (see back in the instructions for your state) \_\_\_\_\_ 6

9. I have reviewed my state's instructions and I swear/affirm that:  
• I am a United States citizen  
• I meet the eligibility requirements of my state and subscribe to any oath required.  
• The information I have provided is true to the best of my knowledge under penalty of perjury. If I have provided false information, I may be fined, imprisoned, or (if not a U.S. citizen) deported from or refused entry to the United States.

Please sign full name (or put mark) \_\_\_\_\_  
Date: \_\_\_\_\_  
Month Day Year

Government documents

...and many more

# Example: CRIF – Document management application

Goal: improve frontend experience and process automation

## What does the solution look like?

Accuracy from 65% to 78% using Amazon Textract

Significant cost savings by using Amazon S3  
(25% cheaper than the previous solution, only 10% of the documents were stored)

By 2020 Q3 100% of the archive will be available online  
(securely stored, digitally processed, semantically analyzed)

## Which services used?

Amazon S3, Textract



# Amazon Rekognition

Easily add intelligent image and video analysis to your applications – no ML experience required



# Amazon Rekognition

## Image and video analysis



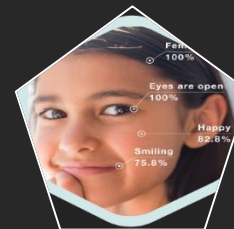
Labels (custom, object, scenes, and activities)



Face search



Unsafe image and video detection



Face detection and analysis



Text in image



Celebrity recognition



Pathing



Real-time video analysis

# Example: Hemnet – Most popular property portal, from Sweden

Goal: automate image analysis

## Context

3M+ visitors each week

All-in on AWS since 2018

Integrated Amazon Rekognition into image pipeline

## How Hemnet leverages Rekognition

Label detection is used on all incoming images

It provides access to property floor plans data

No manual work for broker agents

Also for moderation (images must follow publishing rules)



# BYOD – Bring your own data



Amazon  
Fraud Detector

New!



Amazon Rekognition  
(Custom Labels)

New!



Amazon  
Personalize



Amazon  
Forecast

New!

# Amazon Rekognition Custom Labels

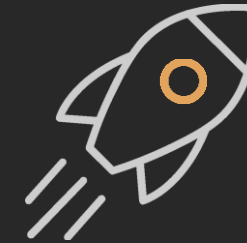
Customized image analysis to easily detect objects and scenes most relevant to your domain



**Guided experience**  
to create labeled  
images



Train and evaluate  
with **no coding** and  
no **ML experience**



Easy-to-use **fully  
managed API**

---

## Examples



Products



Icons, logos,  
and symbols



Comics and  
animations



Biological

# Amazon Forecast

Accurate time-series forecasting  
– no ML experience required

# Forecasting at Amazon.com

Early R&D investment to improve forecasting

Traditional statistical methods



Traditional statistics can predict demand for some products with reasonable accuracy



Linear or (repeated) seasonal demand patterns



# Forecasting at Amazon.com

## Early R&D investment to improve forecasting

Traditional statistical methods

Use of ML

1995

2000

2007

2010

2015

2019



High price  
variability



Regional vs.  
national demand



Slow-moving  
products



New  
products



Highly seasonal  
products



# Forecasting at Amazon.com

Using ML to solve complex forecasting problems

15x improvement  
in accuracy

Traditional statistical methods

Use of ML

Use of deep learning

1995

2000

2007

2010

2015

2019



High price  
variability



Regional vs.  
national demand



Slow-moving  
products



New  
products

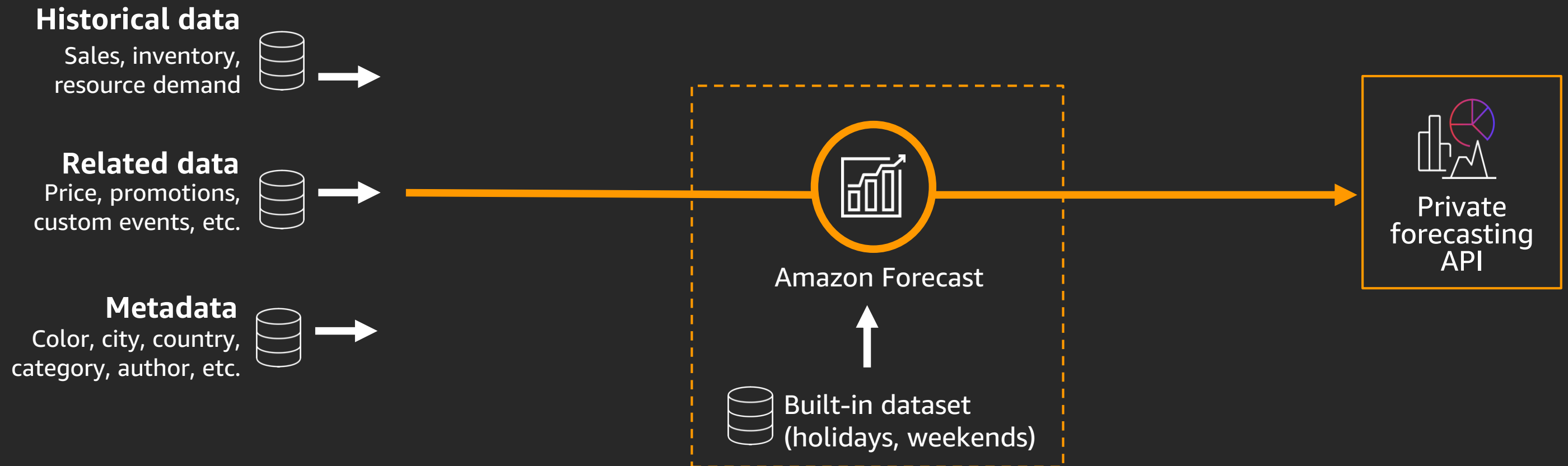


Highly seasonal  
products



# Amazon Forecast

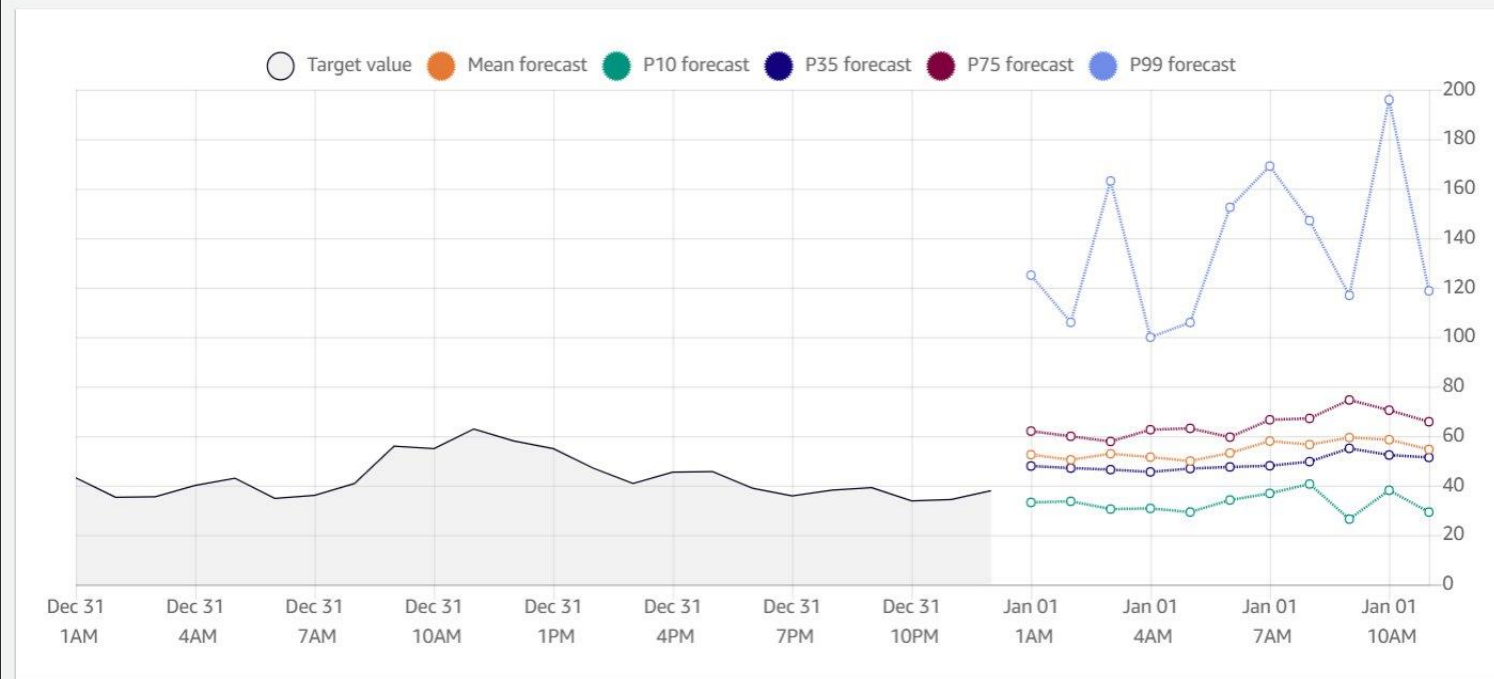
Automated ML service for accurate forecasting



# Amazon Forecast

Visualize the distribution and customize quantiles

Item\_id: client\_12



View probabilistic forecasts at multiple quantiles in the console



Retrieve forecasts through your private API



Export forecasts to .csv

# Amazon Personalize

Real-time personalization and recommendation  
– no ML experience required

# Amazon Personalize

Based on the technology that powers personalization at Amazon.com



Deliver high-quality  
recommendations



Real-time



Deliver  
personalization in  
days, not months



Works with any  
product or content

---

## Key features

Context-aware  
recommendations

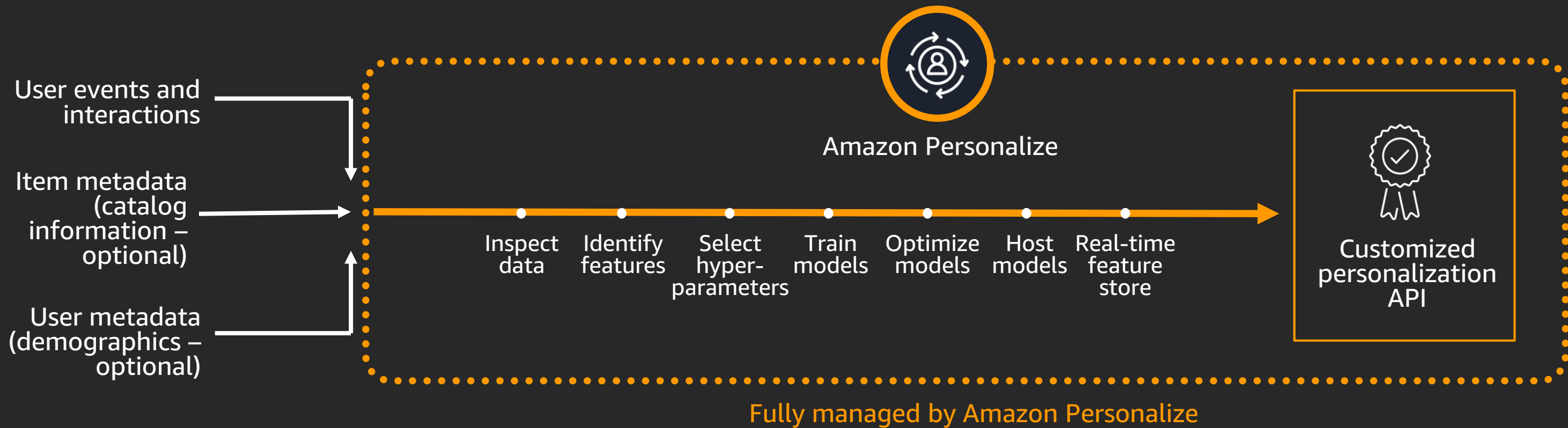
Automated  
ML

Continuous learning  
to improve performance

Bring existing algorithms  
from Amazon SageMaker

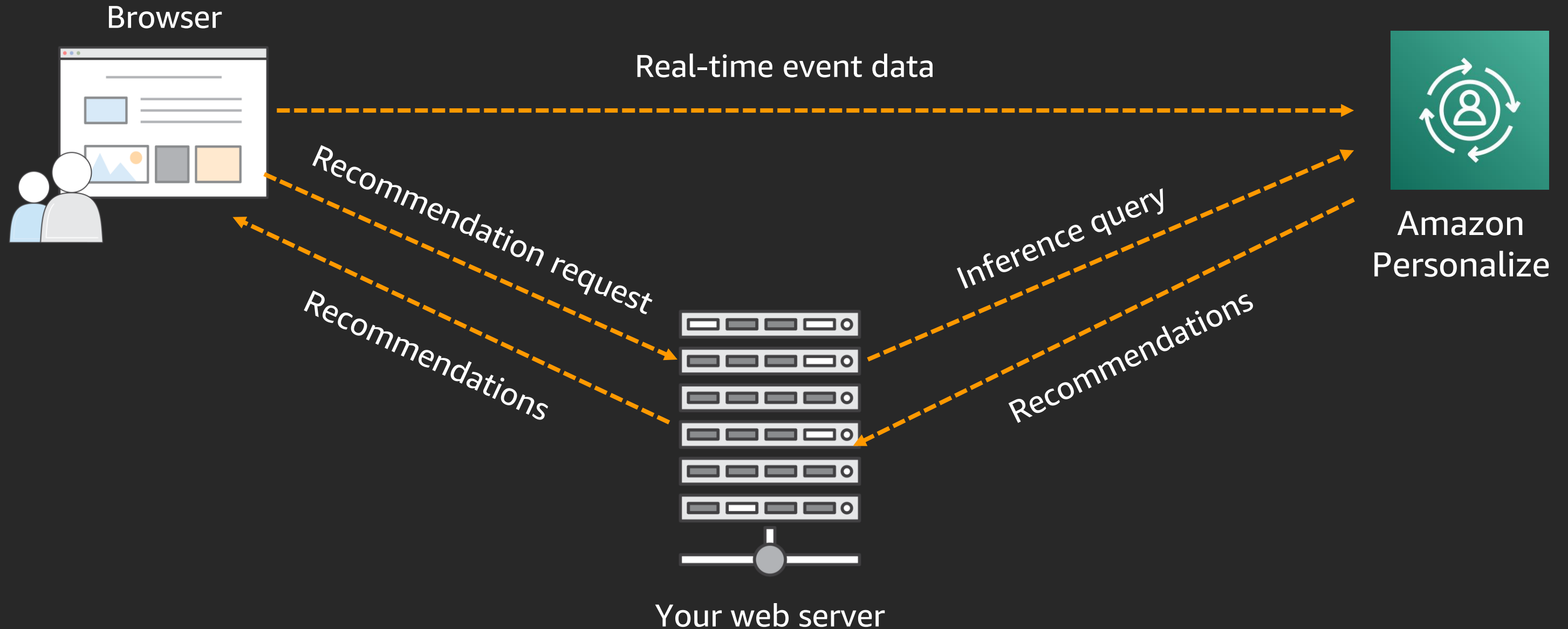
# Amazon Personalize

## Behind the scenes



# Real-time personalization for web applications

## Example



# Building blocks approach

```
response = comprehend.detect_sentiment(  
    Text='I love the cloud!',  
    LanguageCode='en',  
)
```

```
sentiment = response['Sentiment']
```

```
response = polly.synthesize_speech(  
    OutputFormat='mp3',  
    Text='hello world',  
    VoiceId='Matthew',  
)
```

```
audio = response['AudioStream']
```

```
response = textract.detect_document_text(  
    Document={'Bytes': document},  
)
```

```
blocks = response['Blocks']
```

```
response = translate.translate_text(  
    Text='hello world',  
    SourceLanguageCode='en',  
    TargetLanguageCode='it',  
)
```

```
text = response['TranslatedText']
```



```
response = fraud.get_prediction(  
    detectorId=my_detector,  
    detectorVersionId=my_version,  
    eventId=unique_event_id,  
)
```

```
outcomes = response['outcomes']  
scores = response['modelScores']
```

```
response = personalize.get_recommendations(  
    campaignArn=campaign_arn,  
    userId=user_id,  
    itemId=item_id,  
)
```

```
items = response['itemList']
```

```
response = rekognition.detect_custom_labels(  
    ProjectVersionArn=my_project,  
    Image={'Bytes': image},  
    MaxResults=5,  
    MinConfidence=0.8,  
)
```

```
labels = response['CustomLabels']
```

```
response = forecast.query_forecast(  
    ForecastArn=my_forecast,  
    StartDate=start_date,  
    EndDate=end_date,  
    Filters={"item_id" : "client_21"}  
)
```

```
values = response['Forecast']['Predictions']['string']
```

# Where to learn more?

ai.aws & ml.aws

ai-services.go-aws.com

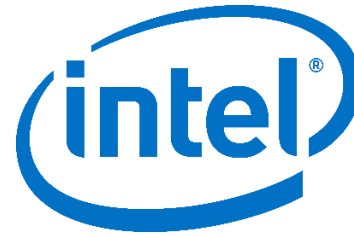
# APN Machine Learning and IoT Competency Partners

 **accenture**

**Deloitte.**

 **REPLY**

**AMD** 

 **intel**

 **snowflake**

Check out the Partner Discovery Zone!

# Thank you!

Alex Casalboni  
Developer Advocate  
[@alex\\_casalboni](#)