



AWS
re:Invent

AI & Machine Learning Launches

Julien Simon

Global Evangelist, AI & Machine Learning
Amazon Web Services
@julsimon



Please fasten your seatbelts!

AI Services



Pre:Invent highlights

<https://aws.amazon.com/about-aws/whats-new/machine-learning>

- Amazon **Comprehend**: 6 new languages
- Amazon **Translate**: 22 new languages
- Amazon **Transcribe**: 15 new languages, alternative transcriptions
- Amazon **Lex**: SOC compliance, sentiment analysis, web & mobile integration with Amazon Connect
- Amazon **Personalize**: batch recommendations
- Amazon **Forecast**: use any quantile for your predictions

With region expansion across the board!

Speech Recognition For Healthcare

Amazon Comprehend Medical

<https://aws.amazon.com/comprehend/medical/>

Input text

Pt is 40yo mo
HPI : Sleeping
Meds : Vyvanse
HEENT : Bogg
erythematous
Follow-up as :

407 of 10000 char

40yo
0.99+ score

software
0.98 score

Sleeping t
0.81 score

Clonidine
0.98 score

Rash
0.99+ score

face
0.98 score

leg
0.99+ score

Age

Protected health information -

Frequency

Route or mode

Dosage

Meds : **Vyvanse**
• Brand name (Vyvanse)

Dosage

50 mgs

• Dosage (50 mgs)

Route or mode

po

• Route or mode (po)

Frequency

at breakfast daily,

• Frequency (at breakfast daily)

Frequency

Route or mode

Dosage

Strength

Clonidine
• Generic name (Clonidine)

Strength

0.2 mgs --

• Strength (0.2 mgs)

Dosage

1 and 1 / 2 tabs

• Dosage (1 and 1 / 2 tabs)

Route or mode

po

• Route or mode (po)

Frequency

qhs

• Frequency (qhs)

Sign

HEENT :
• System organ site (HEENT)

Boggy inferior turbinates, No

• Dx name (Boggy inferior turbinates)

• Direction (inferior)

• System organ site (turbinates)

Negation, Sign

oropharyngeal lesion

• Dx name (oropharyngeal lesion)

Amazon Transcribe Medical



Accurate

US English

Primary Care

Dictation Transcription

**Conversational
Transcription**



Easy-to-Use

Real-time Public API

Automatic Punctuation

Word-level Time Stamps

**Word-level Confidence
Scores**



Affordable

Pay-as-you-go Model

**Charge by Transcription
Usage**

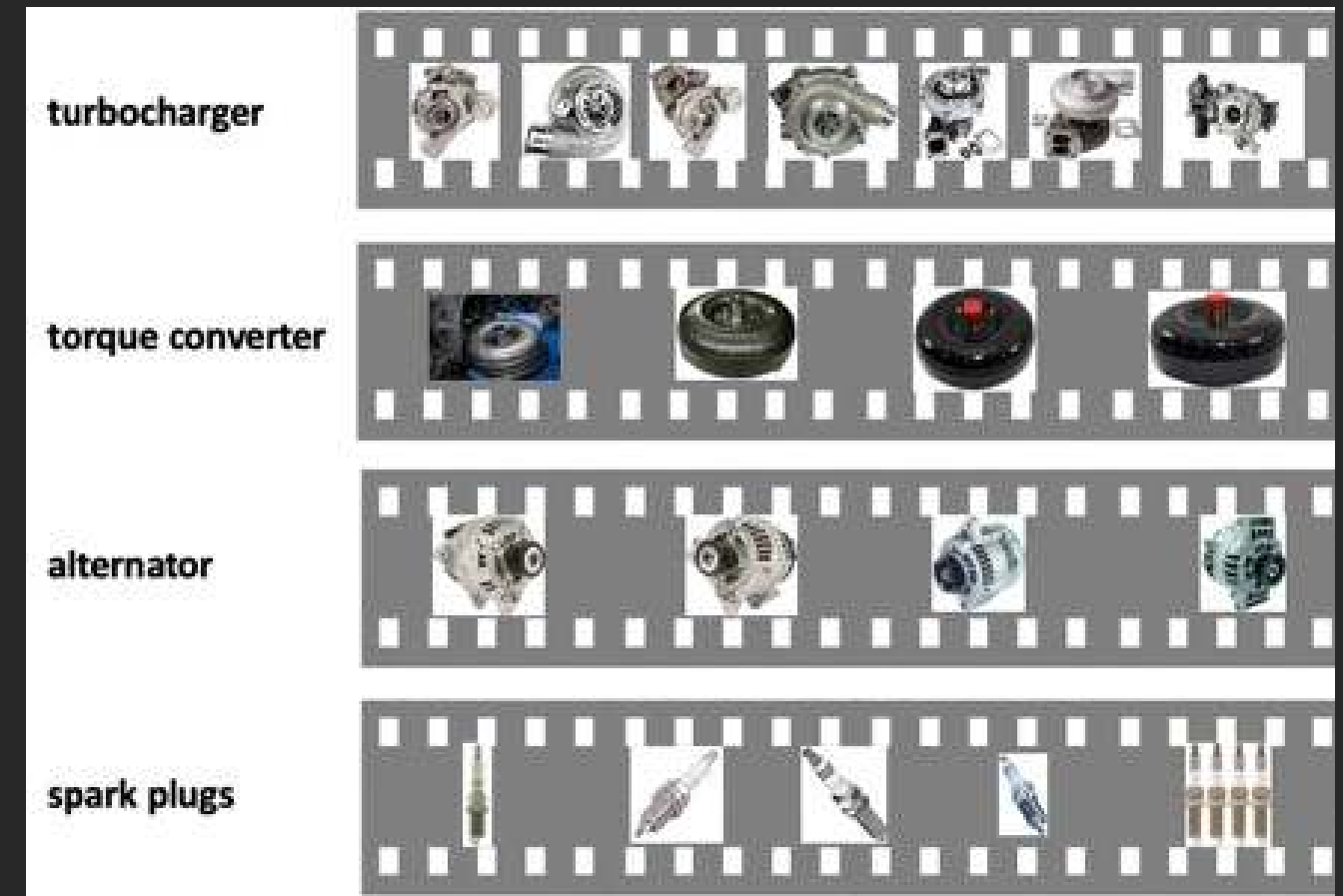
7.5 cents / minute

60 Minute Free Tier

Custom Image Models

Introducing Amazon Rekognition Custom Labels

- **Import** images labeled by Amazon SageMaker Ground Truth...
 - Or label images automatically based on folder structure
- **Train** a model on fully managed infrastructure
 - Split the data set for training and validation
 - See precision, recall, and F1 score at the end of training
- **Select** your model
- **Use** it with the usual Rekognition APIs



Human In the Loop

Customers are forced to choose



OR



ML only systems are high speed and low cost, but do not support nuanced decision making

Human only workflows offer nuanced decision making, but they're low speed and high cost.

Customers need

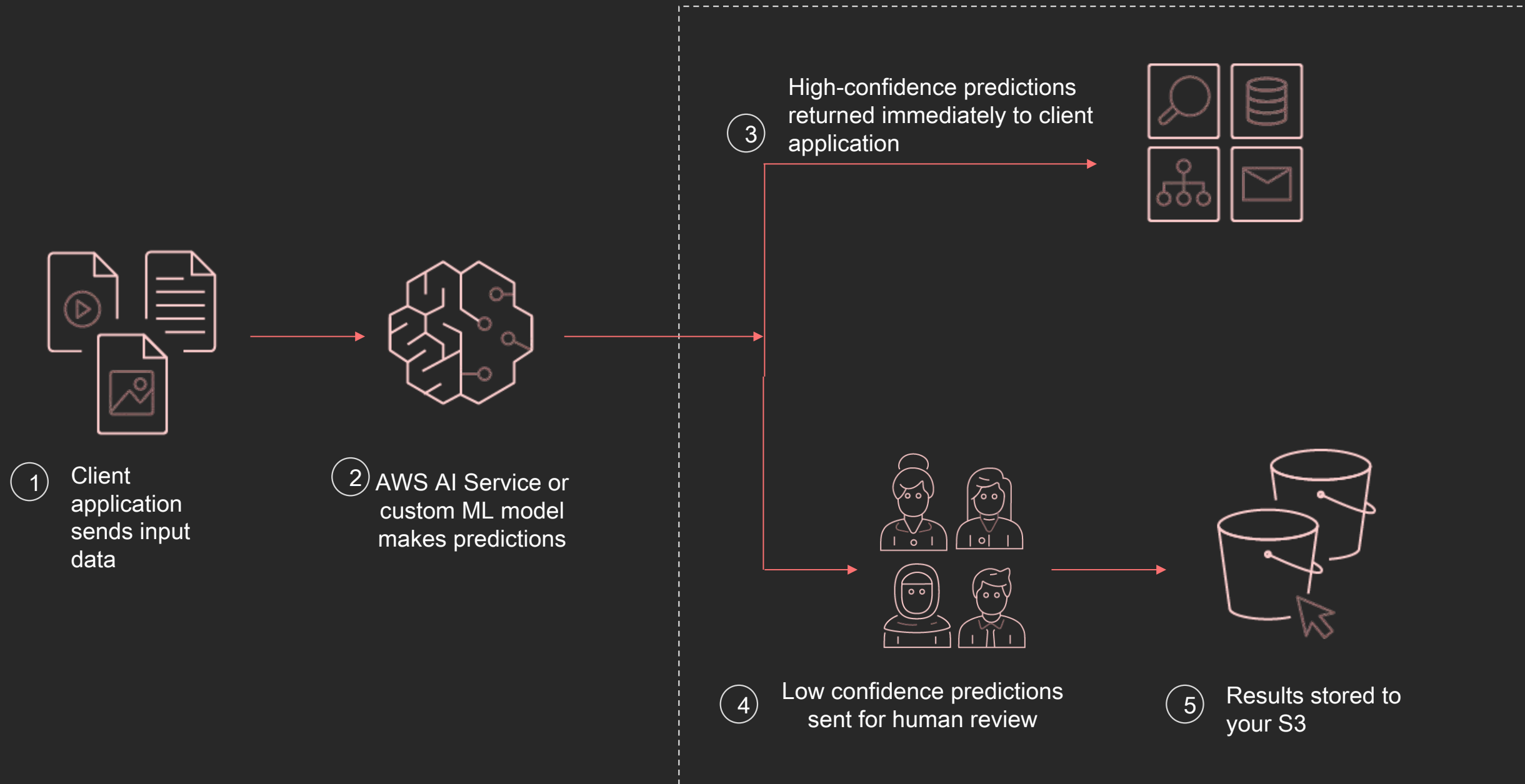


Machine Learning and humans working together

Introducing Amazon Augmented AI (A2I)

A2I lets you easily implement **human review** in **machine learning workflows** to improve the accuracy, speed, and scale of complex decisions.

How Amazon Augmented AI works



Human Review Workforces



Amazon Mechanical Turk

An on-demand 24x7 workforce of over 500,000 independent contractors worldwide, powered by Amazon Mechanical Turk



Private

A team of workers that you have sourced yourself, including your own employees or contractors for handling data that needs to stay within your organization



Vendor

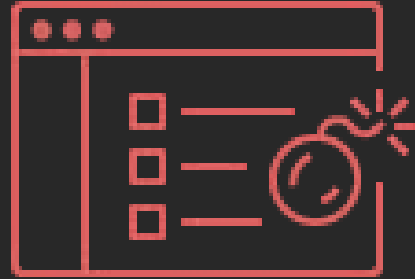
A curated list of third-party vendors that specialize in providing data labeling services, available via the AWS Marketplace

Fraud Detection

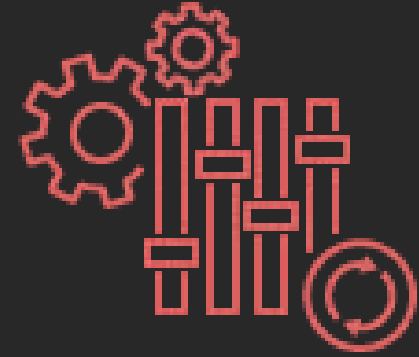
Fraud detection is difficult



\$\$\$ billions lost to fraud each year



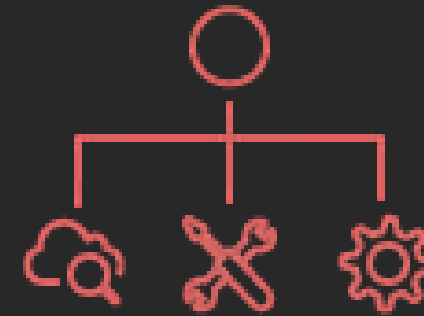
Online business prone to fraud attacks



Bad actors often change tactics

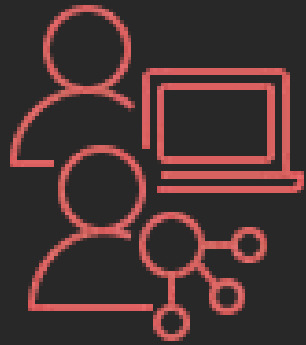


Changing rules = more human reviews



Dependent on others to update detection logic

Fraud detection with ML is also difficult



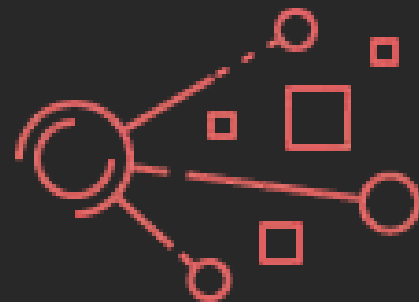
Top data scientists are
costly & hard to find



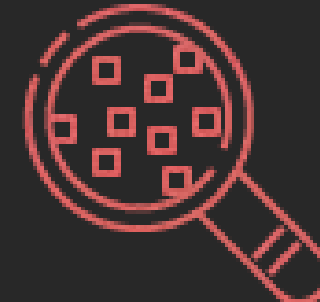
One-size-fits-all models
underperform



Often need to
supplement data



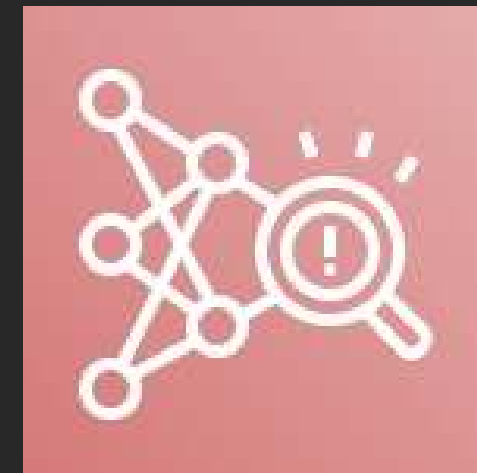
Data transformation +
feature engineering



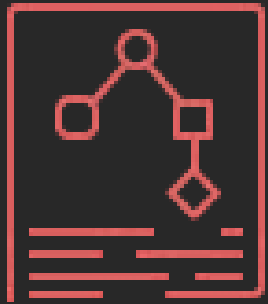
Fraud imbalance =
needle in a haystack

Introducing Amazon Fraud Detector

A **fraud detection service** that makes it easy for businesses to use machine learning to detect online fraud in **real-time**, at **scale**



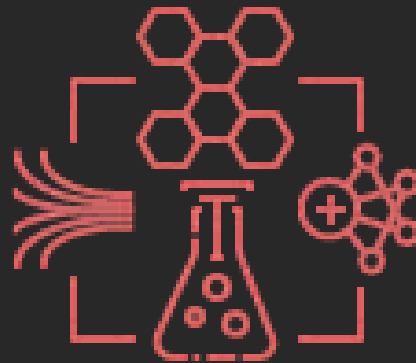
Amazon Fraud Detector – Key Features



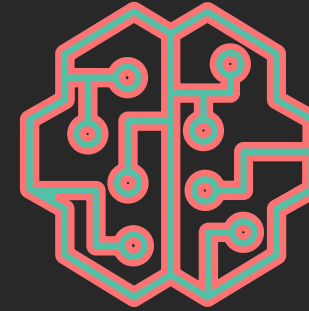
Pre-built fraud
detection model
templates



Automatic
creation of
custom fraud
detection
models



Models learn
from past
attempts to
defraud
Amazon



Amazon
SageMaker
integration



One interface to
review past
evaluations and
detection **logic**

Amazon Fraud Detector – Automated Model Building



Training
data in S3

1



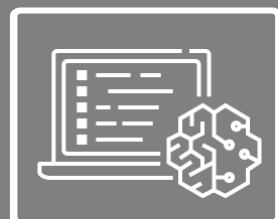
Data
Validation

2



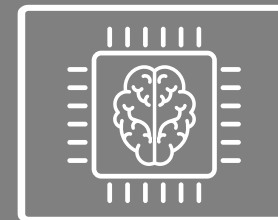
Data Enrichment
& Transformation

3



Feature
Engineering

4



Model Training
& Selection

5



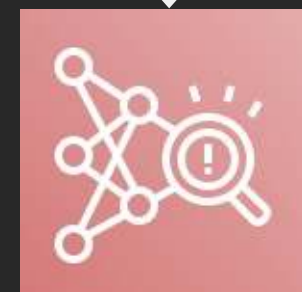
Performance
Metrics

6



Deployment
& Hosting

Customized fraud
detection ML model



**Introducing a new tier of AI services,
for even more productivity**



Speech & Text Analytics For Contact Centers

Challenges in contact centers

- Better **visibility** into quality of customer interactions
- **Cost** prohibitive
-
- Timely **discovery** of emerging issues
- Support for **live** calls
- End user **experience**

Introducing Contact Lens For Amazon Connect

Easily use the power of machine learning to improve the quality of your customer experience without requiring any technical expertise

Built-in automatic
call transcription



View entire call
transcript directly in
Amazon Connect

Enhanced
Contact
Search



Filter calls of
interest based
on words
spoken and
customer
sentiment

Automated
contact
categorization



Identify call types
such as script
compliance,
competitive
mentions,
and cancellations.

Theme
detection



Presents
recurring
issues based
on
Customer
feedback

Real-time sentiment
dashboard
and alerting



Quickly identify
when customers
are having a
poor experience
on live calls

Categories

Improper greeting

Cancel account request

Account verification

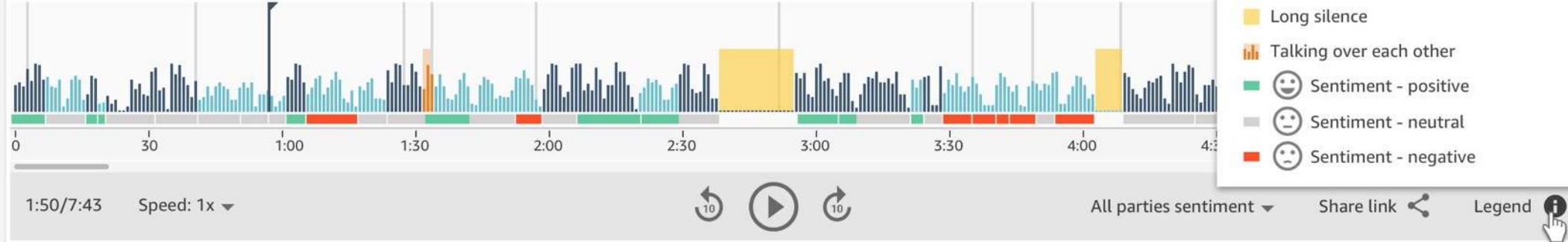
Opology

Upsell

Competitor mention

Difficulty hearing

Recording and transcript



Transcript

Agent 00:01
Hello. Thanks for calling [busniess name]. My name is Han Solo. Let's start with your date of birth, and please verify your address for me as well.

Customer 00:08
Hi. My birthday is December 22, 1981 and my address is 221A Baker Street.
I'll be moving soon, so I need to change that later.

Agent 00:18
Well, happy birthday.

Customer 00:20
Thanks

Agent 00:22
Hmm
I have a hard time finding your account. Do you happen to have your account ID?
Or tell me the last 4 digits of your social security number.

Customer 00:40
My account ID is 12345678
And the last four digist of my social is 1234
Actually that's the wrong number. Let me see.
It's 4567

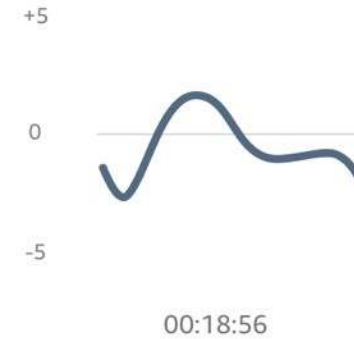
Agent 01:03
Found ya. How may I help you today?



Contact Trace Record

Contact summary

Contact ID 1po9ie0-7fdc-2588-9erc-cd2iuy510987q
Contact Start and end time Dec 22, 18, 3:01:00 - 3:19:56 pm
Contact duration 00:18:56
Customer number 123-456-7890
Agent soloh
Queue customer account
Channel Voice



Customer sentiment trend



Neutral 37%
Positive 10%
Negative 53%

Customer sentiment



Non-talk time 10%
Customer talk time 50%
Agent talk time 40%

Total talk time 00:18:56

Contact details

Contact analysis

Categories

Improper greeting

Cancel account request

Account verification

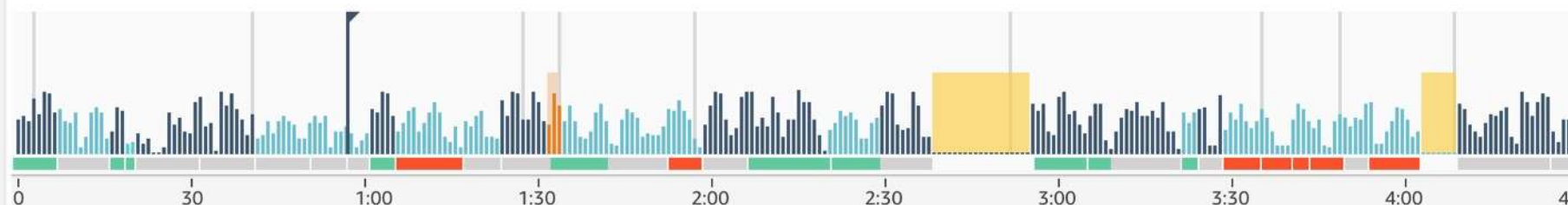
Opology

Upsell

Competitor mention

Difficulty hearing

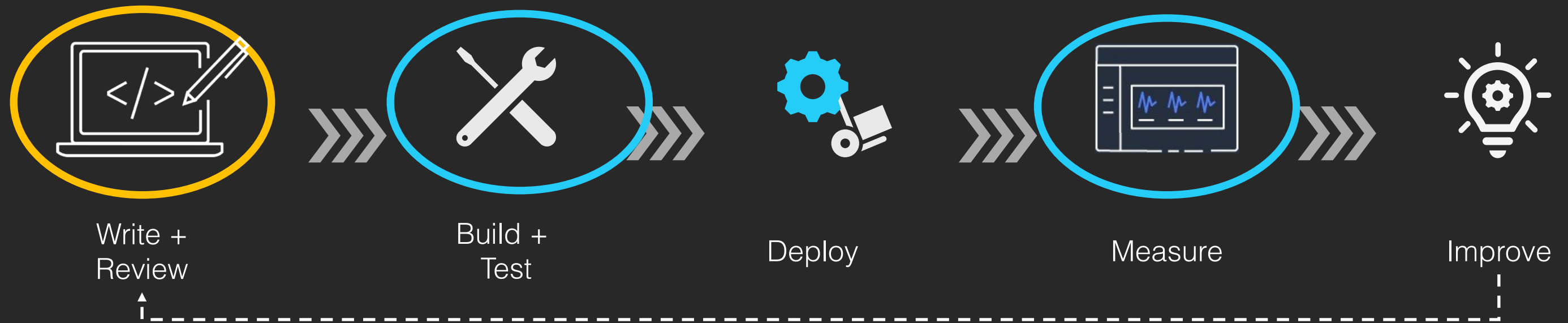
Recording and transcript



- Agent
- Customer
- Long silence
- Talking over each other
- Sentiment - positive
- Sentiment - neutral
- Sentiment - negative

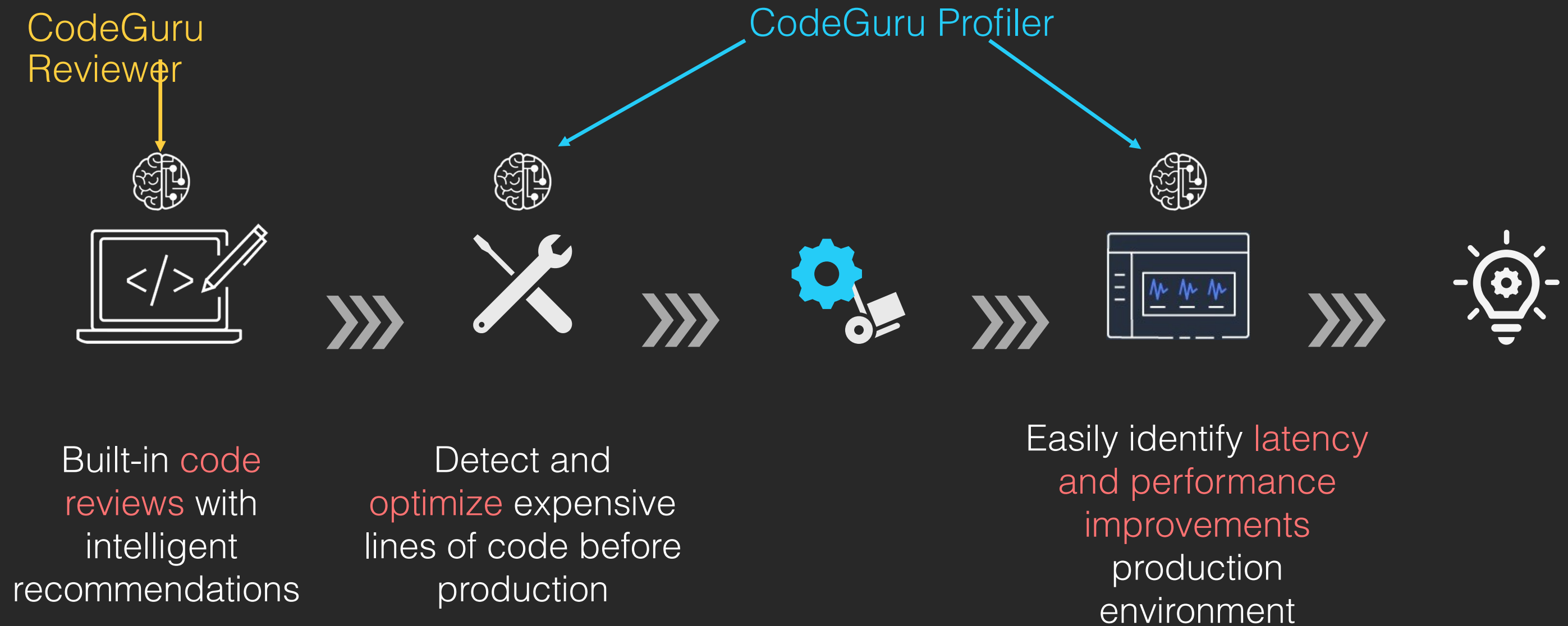
Improving code quality

Typical Application Build and Run Process



1. Code Reviews require expertise in multiple areas such as knowledge of AWS APIs, Concurrency, etc.
2. Code analyzer tools require high accuracy.
3. Distributed Cloud application are difficult to optimize.
4. Performance engineering expertise is hard to find.

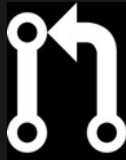
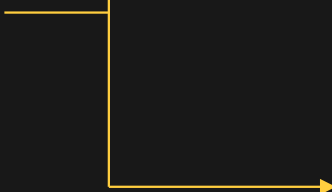
Introducing AWS CodeGuru



CodeGuru Reviewer: How It Works

Customer provides source code as input Extract semantic features / patterns ML algorithms identify similar code for comparison Customers see recommendations as Pull Request feedback

Java
AWS CodeCommit
Github



Input:
Source Code

Feature Extraction

Machine Learning

Output:
Recommendations

CodeGuru Example – Looping vs Waiting

Cod

e

```
do {
    DescribeTableResult describe = ddbClient.describeTable(new DescribeTableRequest().withTableName(tableName));
    String status = describe.getTable().getTableStatus();
    if (TableStatus.ACTIVE.toString().equals(status)) {
        return describe.getTable();
    }

    if (TableStatus.DELETING.toString().equals(status)) {
        throw new ResourceInUseException("Table is " + status + ", and waiting for it to become ACTIVE is not useful.");
    }
    Thread.sleep(10 * 1000);
    elapsedMs = System.currentTimeMillis() - startTimeMs;
} while (elapsedMs / 1000.0 < waitTimeSeconds);
throw new ResourceInUseException("Table did not become ACTIVE after ");
```

Recommendation

This code appears to be waiting for a resource before it runs. You could use the waiters feature to help improve efficiency. Consider using `TableExists`, `TableNotExists`. For more information, see <https://aws.amazon.com/blogs/developer/waiters-in-the-aws-sdk-for-java/>

Developer Feedback We should use waiters instead - will help remove a lot of this code.

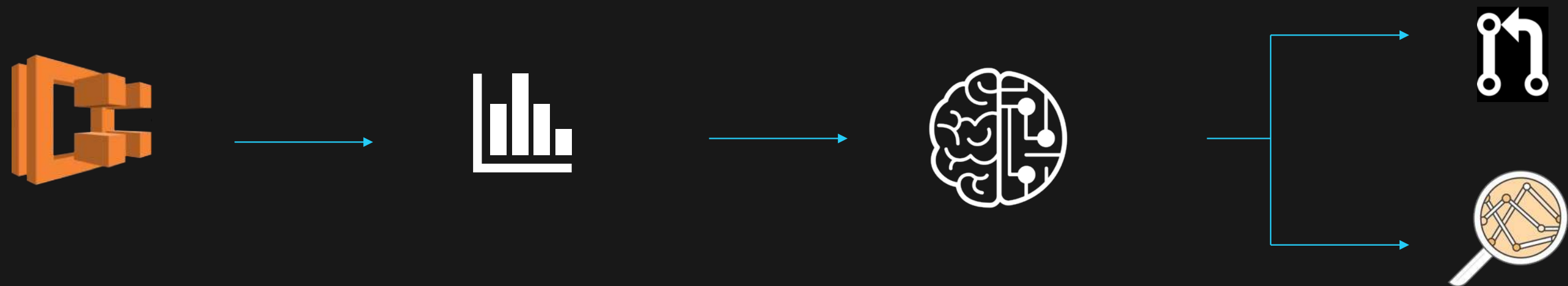
CodeGuru Profiler: How It Works

Customer application runs in production

CodeGuru Profiler continuously captures application stack trace information

CodeGuru Profiler detects performance inefficiencies in the live application

Customers see recommendations in their automated efficiency reports and visualizations



CodeGuru Profiler – Example

✓ Repeated regex compilation during String manipulation

Recommendations report

Definition

Version

11th Jan 2019 latest
v3 (Profiler) ▼

Description

The `java.lang.String` utility methods which accept regular expressions as `String` values are inefficient when used in hot paths. They have to compile the regular expression for every use, which is expensive.

Resolution steps

Pre-compile the regular expression into a `Pattern`, and replace any of the following uses as shown:

Before (inefficient):

```
1. input.matches("foo.*")
2. input.replaceAll("foo.*", "replacement")
3. input.replaceFirst("foo.*")
4. input.split("foo.*")
5. input.replace("not-a-regex", "replacement")
```

After (efficient):

```
static Pattern MY_REGEX = Pattern.compile("foo.*")
static Pattern MY_NOT_REGEX = Pattern.compile("not-a-regex", Pattern.CASE_INSENSITIVE);

1. MY_REGEX.matcher(input).matches()
2. MY_REGEX.matcher(input).replaceAll("replacement")
3. MY_REGEX.matcher(input).replaceFirst("replacement")
4. MY_REGEX.split(input)
5. MY_NOT_REGEX.replaceAll(Matcher.quoteReplacement("replacement"))
```

Note for number 5, `quoteReplacement()` is only required if the replacement string contains backslashes. It is required when using `Matcher.replaceAll()` and `String.replace()`.

Example fix: <https://code.amazon.com/reviews/CR-3726256>.

The code fix

```
- private static final String INVALID_METRIC_CHARACTERS = "[^A-Za-z0-9\\\\.:@_\\-\\/]";
```

```
+ private static final Pattern INVALID_METRIC_CHARACTERS = Pattern.compile("[^A-Za-z0-9\\\\.:@_\\-\\/]");
```

```
- return metricName.replaceAll(INVALID_METRIC_CHARACTERS, "");
```

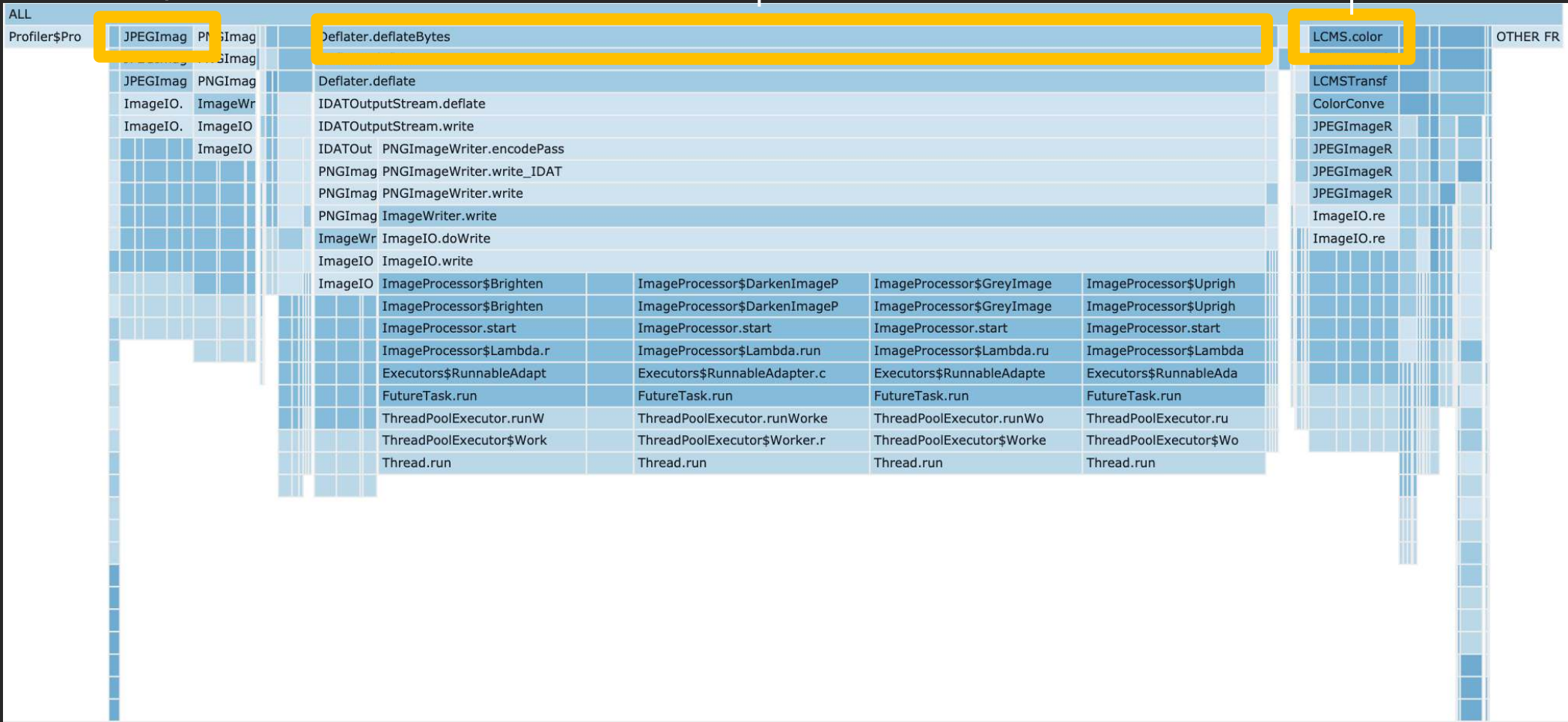
```
+ return INVALID_METRIC_CHARACTERS.matcher(metricName).replaceAll("");
```

CodeGuru Profiler Visualization – Hotspots

JPEGImageReader.readImage 5%

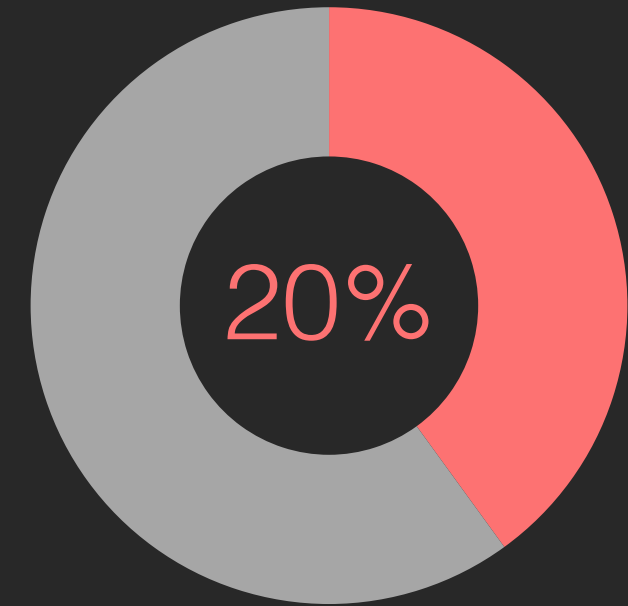
LCMS.colorEvent 6%

Java.util.zip.Deflator.deflateBytes 61%

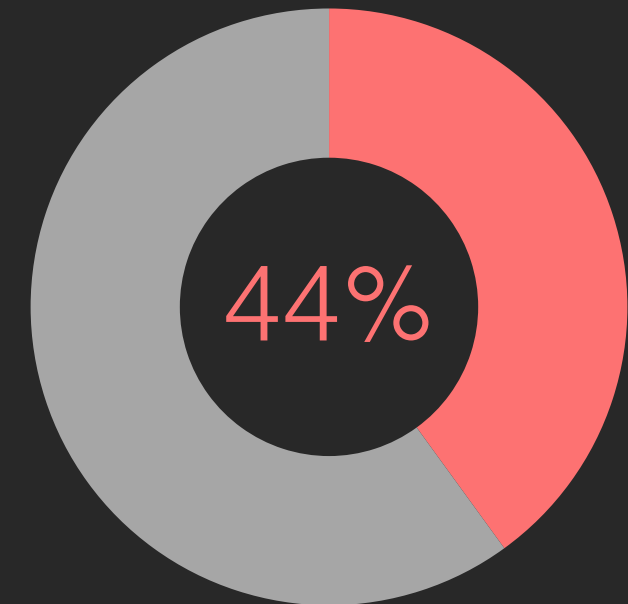


Enterprise Search

Employees spend **20%** of their
time looking for information.
—McKinsey



44% of the time, they cannot
find the information they need to
do their job.
—IDC



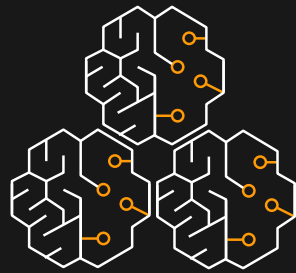
Introducing Kendra



Natural language
Queries



NLU and
ML core



Domain
Expertise



Incremental
learning through
feedback



Native connectors
(S3, Sharepoint,
file servers,
HTTP, etc.)

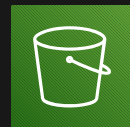


Simple API
and console
experiences
Code samples

Easy to find what you are
looking for

Fast search, and
quick to set up

Kendra connectors



Amazon S3



File systems (SMB)



Web crawler



Databases



Sharepoint Online



Sharepoint 2013,
2016, 2019



Box



Dropbox



Exchange



OneDrive



Google Drive



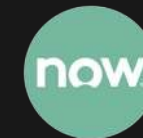
Salesforce



Confluence



Jira



ServiceNow



Zendesk



Jive

...and more coming in 2020

Getting started with Kendra



Step 1 Create an index

An index is the place where you add your data sources to make them searchable in Kendra.



Step 2 Add data sources

Add and sync your data from S3, Sharepoint, Box and other data sources, to your index.



Step 3 Test & deploy

After syncing your data, visit the Search console page to test search & deploy Kendra in your search application.

When is re:invent?



Kendra suggested answer

AWS re:Invent 2019

AWS **re:Invent** 2019 will be held **December 2-6, 2019** Las Vegas, NV. Download your calendar invites for upcoming AWS Training Talks (or listen to last week's recordings) below. Training Talks offer internal field reps a 60 mins deep dive into use cases, customer benefits, and a demo of new services... [\(More\)](#)

WEB | https://w.amazon.com/bin/view/AWS/reinvent_2019



What are Kendra suggested answers? [Info](#)

Frequently asked questions

[Where is re:Invent located?](#)



[How do I register for re:Invent?](#)



[When is re:Invent registration open?](#)



What are frequently asked questions? [Info](#)

1-10 of 65 Recommended documents

Sort: Relevancy ▼

APN re:Invent 2019

Encourage your APN Partners to join us December 2 at AWS Innovate, **re:Invent** Recap Edition. This is a free online conference providing a complete wrap up of AWS **re:Invent**... [\(More\)](#)

WEB | https://w.amazon.com/bin/view/APN_reinvent_2019/



re:Invent Live Stream Viewing Parties 2018

...Join us for **re:Invent** keynote viewing parties for our Amazon.com CDO organizations. We'll be providing breakfast, **re:Invent** water bottle giveaways, and a phone tool icon for your participation... [\(More\)](#)

WEB | <https://w.amazon.com/bin/view/reinventcampusevents/>



Deploy

Select the components you want on the page, copy the code snippets, then paste them into your search application.

Want it all? Copy all componens with our global code snippet.

[Copy all components](#)



Selected component

Kendra suggested answer

```
<div class="term-grid">
  <label>Lorem</label>
  <div class="definition">ipsum dolor sit amet
consectetur adipiscing</div>
  <div class="alternate">something</div>
</div>
.clearfix { display: inline-block; }
* html .clearfix { lorem: 1%; }
.clearfix { display: block; }
<div class="term-grid">
  <label>Lorem</label>
  <div class="definition">ipsum dolor sit amet
consectetur adipiscing</div>
  <div class="alternate">something</div>
</div>
.clearfix { display: inline-block; }
* html .clearfix { lorem: 1%; }
.clearfix { display: block; }
.clearfix { display: inline-block; }
* html .clearfix { lorem: 1%; }
```

[Copy code snippet](#)

Learn more

[Deploying Kendra in your web application.](#)

[Understanding API integration.](#)

Filter search results ▼

Search in


Everywhere (486)

Wiki (123)

Broadcast (265)

Sage (98)

Date range



20072019

Categories

☒ HR (23)

☐ IT (5)

☐ Marketing (5)

Authors

☐ miyingah (7)

☐ michealnb (4)

☐ angelak (4)

☐ masono (1)

☐ kimfonta (1)

☐ rwitaman (1)

Show more...

amazon.com employee discount

Q

Kendra’s suggested answer

New Hire Information

For more information, visit the Amazon US Benefits Inside Amazon page. **Employee Discount** Employees receive a **10% discount** (up to the value of \$100 annually) for selected products available on **Amazon.com**. **Employee discount** code numbers will be available once new hire documentation is complete... [\(More\)](#)

WEB | https://w.amazon.com/bin/view/AmazonStudios/New_Hire_Information/

👍

👎

What is Kendra’s suggested answer? [Info](#)

Frequently asked questions

How do I apply the amazon.com discount code?

▶

How often can I use the amazon.com discount?

▶

What items can I apply my employee discount?

▶

What are frequently asked questions? [Info](#)

1-10 of 98 Recommended documents

Sort: Relevancy ▼

Employee Offer

With AT&T This wiki explains how Amazon employees can redeem their employee benefit...Note: Please refer to the **Employee Extras** page for current **discount** details. Some of the information below... [\(More\)](#)

WEB | <https://w.amazon.com/bin/view/key-benefits-offers>

👍

👎

ML services



Pre:Invent highlights

<https://aws.amazon.com/about-aws/whats-new/machine-learning>

- Invoke Amazon SageMaker models in Amazon Quicksight
- Invoke Amazon SageMaker models in Amazon Aurora
- Deploy many models on the same Amazon SageMaker endpoint

Using Kubernetes for ML is hard to manage and scale

Build and manage services within Kubernetes cluster for ML

+

Make disparate open-source libraries and frameworks work together in a secure and scalable way

+

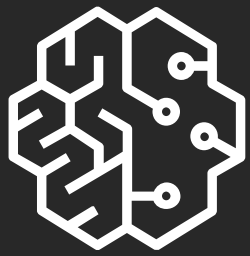
Requires time and expertise from infrastructure, data science, and development teams

=

Need an easier way to use Kubernetes for ML

Introducing Amazon SageMaker Operators for Kubernetes

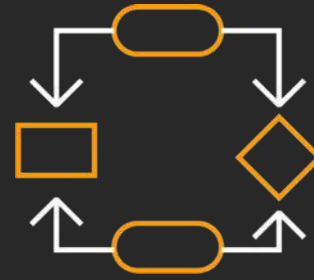
Kubernetes customers can now train, tune, & deploy models in Amazon SageMaker



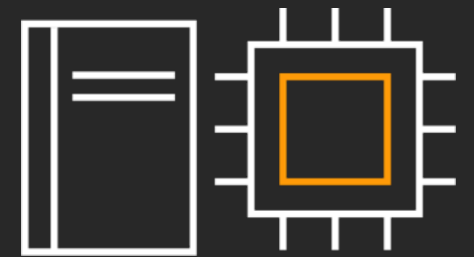
Train, tune, and deploy models in SageMaker



Orchestrate ML workloads from your Kubernetes environments



Create pipelines and workflows in Kubernetes



Fully managed infrastructure in SageMaker



**Machine learning is iterative
involving dozens of tools and
hundreds of iterations**

Multiple tools needed for
different phases of the
ML workflow

+

Lack of an integrated
experience

+

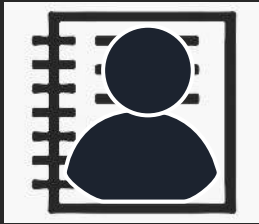
Large number of iterations

=

Cumbersome, lengthy processes, resulting in
loss of productivity

Introducing Amazon SageMaker Studio

The first fully integrated development environment (IDE) for machine learning



Collaboration at
scale

Share scalable notebooks
without tracking code
dependencies



Easy experiment
management

Organize, track, and
compare thousands of
experiments



Automatic model
generation

Get accurate models for
with full visibility & control
without writing code



Higher quality ML
models

Automatically debug errors, Code, build, train, deploy, &
monitor models, & maintain high quality monitor in a unified visual
interface



Increased
productivity

xgboost_customer_churn.ipynb

conda_amazonei_mxnet_p27

- Have the predictor variable in the first column
- Not have a header row

But first, let's convert our categorical features into numeric features.

```
[ ]: model_data = pd.get_dummies(churn)
model_data = pd.concat([model_data['Churn!_True'], model_data.drop(['Churn!_True'], axis=1)], axis=1)
```

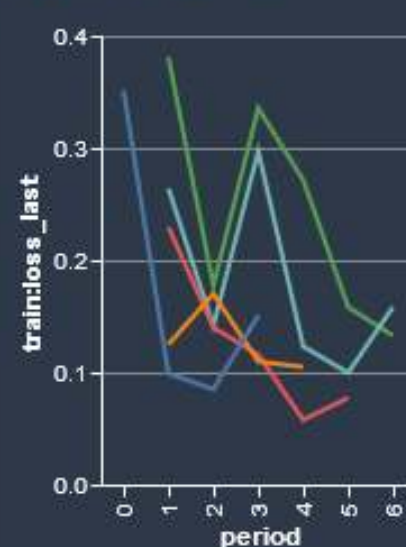
And now let's split the data into training, validation, and test sets. This will help prevent us from overfitting the model, and allow us to test the models accuracy on data it hasn't already seen.

```
[ ]: train_data, validation_data, test_data = np.split(model_data.sample(frac=1), [int(model_data.shape[0]*0.33), int(model_data.shape[0]*0.66)])
train_data.to_csv('train.csv', header=False, index=False)
validation_data.to_csv('validation.csv', header=False, index=False)
```

Now we'll upload these files to S3.

```
[ ]: boto3.Session().resource('s3').Bucket(bucket).Object(os.path.join(prefix, 'train.csv')).upload_file('train.csv')
boto3.Session().resource('s3').Bucket(bucket).Object(os.path.join(prefix, 'validation.csv')).upload_file('validation.csv')
```

Trial Component Chart



Trial Component List



TRIAL COMPONENTS

10 rows selected

Add chart

Deploy model



Status	Experiment	Type	Trial	Trial c
✓ Completed	customer-churn-predi...	Training job	Trial-3	Tr
✓ Completed	customer-churn-predi...	Training job	Trial-2	Tr
✓ Completed	customer-churn-predi...	Training job	Trial-1	Tr
✓ Completed	customer-churn-predi...	Training job	Trial-0	Tr



**Data science and
collaboration needs to be
easy**

Setup and manage resources

+

Collaboration across
multiple data scientists

+

Different data science
projects have different
resource needs

=

Managing notebooks and
collaborating across
multiple data scientists is
highly complicated

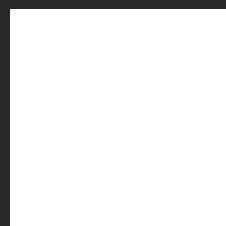
Introducing Amazon SageMaker Notebooks

Fast-start shareable notebooks



Easy access with
Single Sign-on (SSO)

Access your notebooks in
seconds with your
corporate credentials



Fully managed
and secure

Administrators manage
access and permissions



No explicit setup

Start your notebooks
without spinning up
compute resources



Easy collaboration

Share your notebooks
as a URL with a single click



Flexibility

Dial up or down
compute resources



Data Processing and Model Evaluation involves a lot of operational overhead

Building and scaling
infrastructure for data processing
workloads is complex

+

Use of multiple tools or services
implies learning and
implementing new APIs

+

All steps in the ML workflow need
enhanced security, authentication
and compliance

=

Need to build and manage tooling
to run large data processing and
model evaluation workloads

Introducing Amazon SageMaker Processing

Analytics jobs for data processing and model evaluation



Fully managed

Achieve distributed processing for clusters



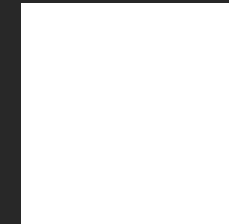
Custom processing

Bring your own script for feature engineering



Container support

Use SageMaker's built-in containers or bring your own



Security and compliance

Leverage SageMaker's security & compliance features



Automatic creation & termination

Your resources are created, configured, & terminated automatically



Managing trials and experiments is cumbersome

Hundreds of experiments

+

Hundreds of parameters
per experiment

+

Compare and contrast

=

Very cumbersome and
error prone

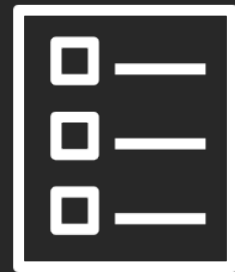
Introducing Amazon SageMaker Experiments

A system to organize, track, and evaluate training experiments



Experiment
tracking at scale

Track parameters & metrics
across experiments & users



Organize
experiments

Organize by teams, goals,
& hypotheses



Visualization for
best results

Visualize & compare
between experiments



Flexibility with
Python SDK & APIs

Log custom metrics &
track models using APIs



Iterate quickly

Iterate & develop high-
quality models



Debugging and profiling deep learning is painful

Large neural networks
with many layers

+

Many connections

+

Additional tooling for analysis
and debug

=

Extraordinarily difficult
to inspect, debug, and profile
the 'black box'

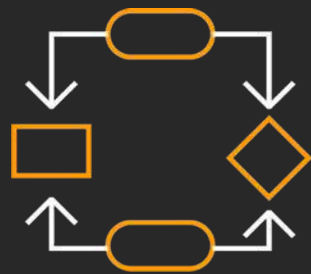
Introducing Amazon SageMaker Debugger

Analysis & debugging, explainability, and alert generation



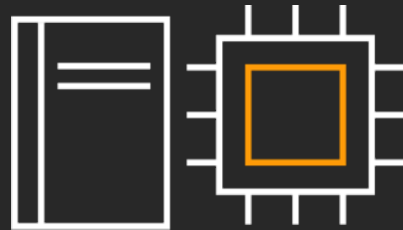
Relevant data capture

Data is automatically captured for analysis



Automatic data analysis

Analyze and debug data with no code changes



Automatic error detection

Errors are automatically detected based on rules



Improved productivity with alerts

Take corrective action based on alerts



Visual analysis and debug

Visually analyze & debug from SageMaker Studio



Deploying a model is not the end, you need to continuously monitor it in production and iterate

Concept drift due to divergence of data

+

Model performance can change due to unknown factors

+

Continuous monitoring of model performance and data involves a lot of effort and expense

=

Model monitoring is cumbersome but critical

Introducing Amazon SageMaker Model Monitor

Continuous monitoring of models in production



Automatic data
collection

Data is automatically
collected from your
endpoints



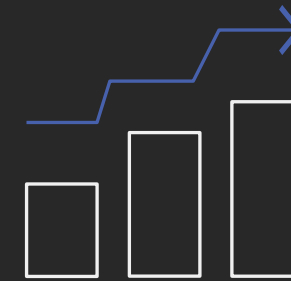
Continuous
Monitoring

Define a monitoring
schedule and detect
changes in quality against
a pre-defined baseline



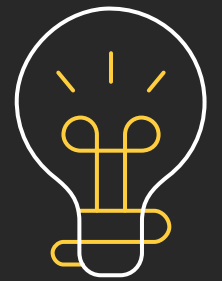
Flexibility
with rules

Use built-in rules to
detect data drift or write
your own rules for
custom analysis



Visual
Data analysis

See monitoring results,
data statistics, and
violation reports in
SageMaker Studio



CloudWatch
Integration

Automate corrective
actions based on Amazon
CloudWatch alerts



**Successful ML
requires
complex, hard to
discover
combinations**
of algorithms, data, parameters

Largely explorative &
iterative

+

Requires broad and
complete
knowledge of ML domain

+

Lack of visibility

=

Time consuming,
error prone process
even for ML experts

Introducing Amazon SageMaker Autopilot

Automatic model creation with full visibility & control



Quick to start

Provide your data in a tabular form & specify target prediction



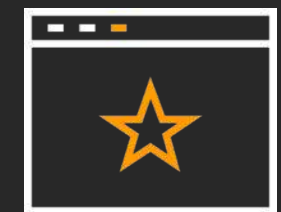
Automatic model creation

Get ML models with feature engineering & automatic model tuning automatically done



Visibility & control

Get notebooks for your models with source code



Recommendations & Optimization

Get a leaderboard & continue to improve your model

Build, Train, Deploy Machine Learning Models Quickly at Scale



AWS DeepRacer improvements

- AWS DeepRacer **Evo**
 - Stereo camera
 - LIDAR sensor
- New **racing opportunities**
 - Create your own races
 - Object Detection & Avoidance
 - Head-to-head racing



Mod your own vehicle

Mod specifications

The garage shows the DeepRacer vehicles that you can train models for. You can add vehicles by using the "build new vehicle button"

Sensor modification

Swap sensors to improve your DeepRacer's racing performance

☐ Front-facing camera

Single camera that captures the images with sizes of 160 x 120 in front of the agent at 15 fps. The camera has 120 wide angle lens. The images are converted into grey scale before being fed to the neural network

► Benefits of the front-facing camera

☒ Stereo cameras (right/left) sensor

Composed of two front-facing cameras, stereo cameras can generate depth information of the objects in front of the agent and thus be used to detect and avoid obstacles on the track. The cameras capture images with the same resolution and frequency. Images from both cameras are converted into grey scale, stacked and then fed into the neural network.

► Benefits of the stereo camera

Add-on sensors

☒ LIDAR sensor

LIDAR is a surveying method that measures a distance to a target by illuminating the target with laser light and measuring the reflected light with a sensor.

► How LIDAR works with autonomous driving



AWS DeepComposer

- The world's first **machine learning-enabled** musical keyboard
- Compose music using Generative Adversarial Networks (GAN)
- Use a **pretrained** model, or train your **own**



Frameworks and Infrastructure



Introducing Amazon EC2 Inferentia

- Fast, low-latency inferencing at a very low cost
 - 64 teraOPS on 16-bit floating point (FP16 and BF16) and mixed-precision data.
 - 128 teraOPS on 8-bit integer (INT8) data.
- Neuron SDK: <https://github.com/aws/aws-neuron-sdk>
 - Available in Deep Learning AMIs and Deep Learning Containers
 - TensorFlow and Apache MXNet, PyTorch coming soon

Instance Name	Inferentia Chips	vCPUs	RAM	EBS Bandwidth
inf1.xlarge	1	4	8 GiB	Up to 3.5 Gbps
inf1.2xlarge	1	8	16 GiB	Up to 3.5 Gbps
inf1.6xlarge	4	24	48 GiB	3.5 Gbps
inf1.24xlarge	16	96	192 GiB	14 Gbps

Deep Graph Library

<https://www.dgl.ai>

- Python open source library that helps researchers and scientists quickly build, train, and evaluate **Graph Neural Networks** on their data sets
- Use cases: recommendation, social networks, life sciences, cybersecurity, etc.
- Available in **Deep Learning Containers**
 - PyTorch and Apache MXNet, TensorFlow coming soon
- Available for training on **Amazon SageMaker**



Deep Java Library

<https://www.djl.ai>

- Java open source library, to train and deploy models
- Framework **agnostic**
 - Apache MXNet for now, more will come
- Train **your own model**, or use a **pretrained** one from the model zoo



Enjoy!

Build cool stuff,
and please send us feedback!

Thank you!

Julien Simon

Global Evangelist, AI & Machine Learning

Amazon Web Services

@julsimon

