MICHIGAN STATE UNIVERSITY

Alpha Presentation

Amazon Review Confidence Tool

The Capstone Experience

Amazon

Collin Cole
Nikita Gupta
Cameron Hurley
Ashu Kher
Dylan Mccarroll
Ethan Strain

Department of Computer Science and Engineering
Michigan State University

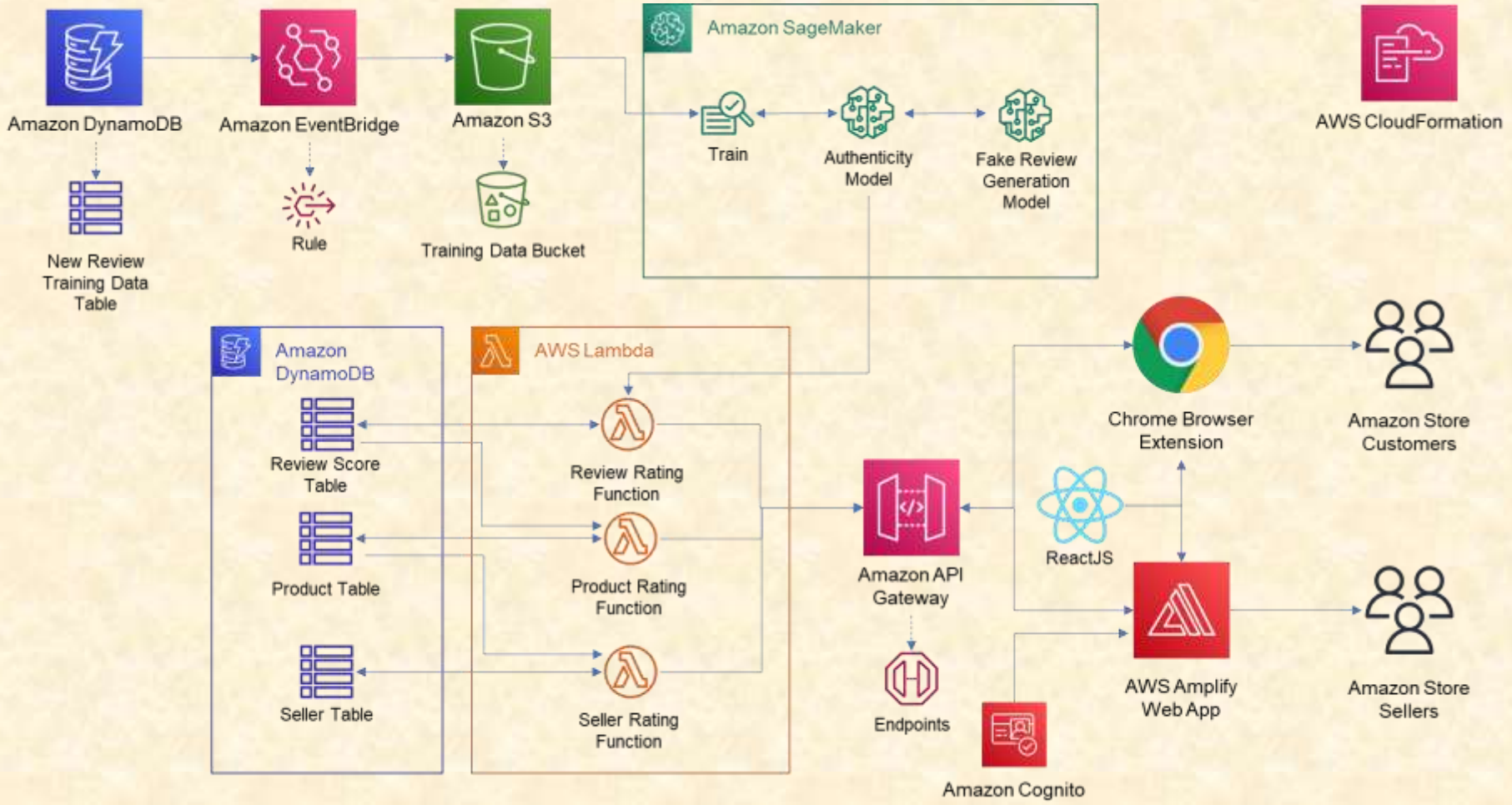
Fall 2022

From Students…
…to Professionals

Project Overview

- Review Authenticity Tool
  - Predict if an Amazon product review is genuine for sellers & customers alike
  - Visualize confidence score with easy to understand graphics
- Easily Accessible Data
  - Browser extension to display data on-site on Amazon products
  - Web page to provide further information and additional features

System Architecture



Home Page

amazon

*** Got our Product Review browser extension? Install it HERE ***

Saved Items

PlayStation 5 Console

$499.99

Nintendo Switch with Neon Blue and Neon Red Joy-Con

$299.99

Xbox Series X

$499.99

Featured Items

Apple AirPods Pro (2nd Generation) Wireless Earbuds with MagSafe Charging Case. Active Noise Cancelling, Personalized Spatial Audio, Customizable Fit, Bluetooth

$239.99

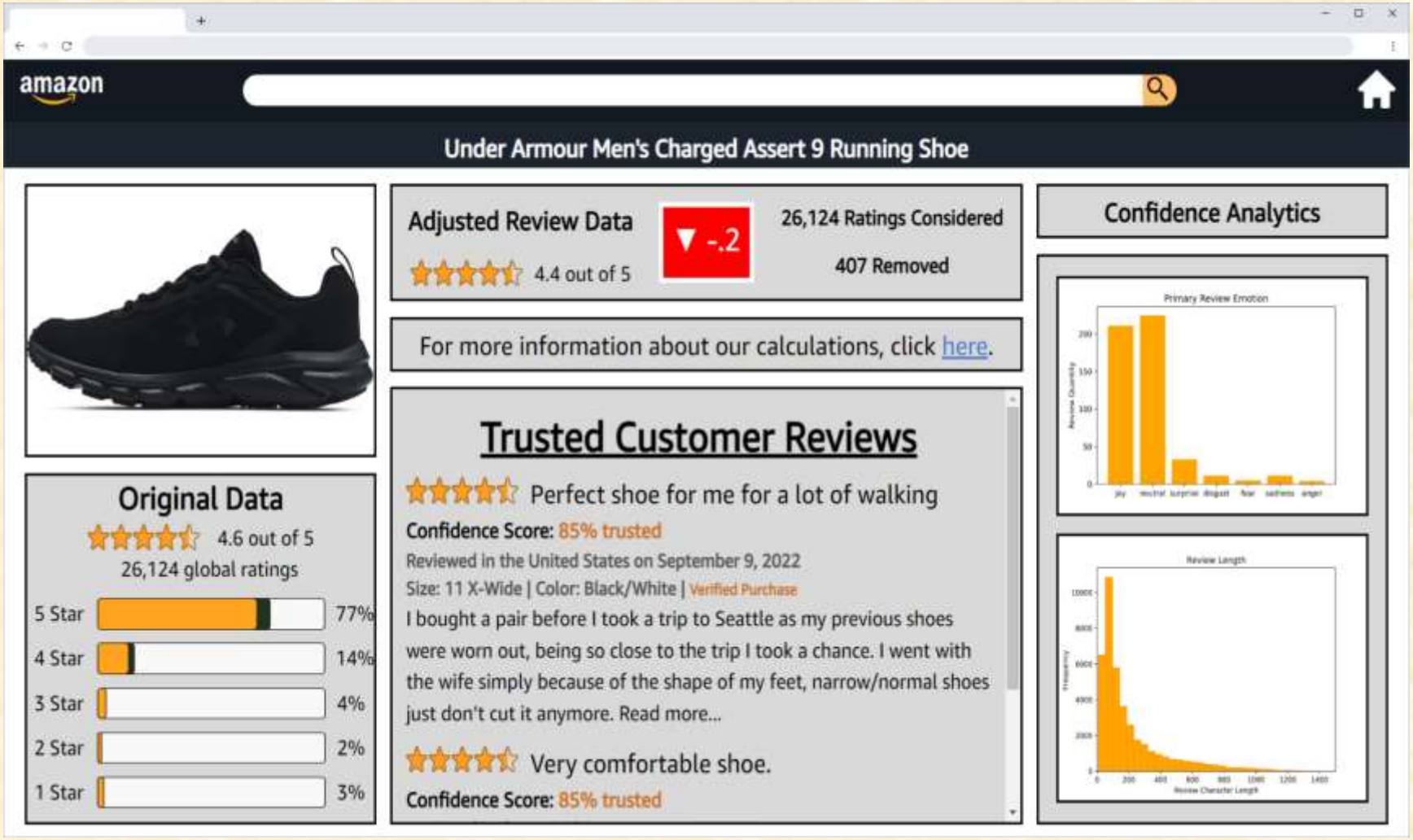
(Pack of 2) Replaced Remote Control for Roku TV, Compatible for TCL Roku/Hisense Roku/Onn Roku/Sharp Roku/Element Roku/Westinghouse

$11.57

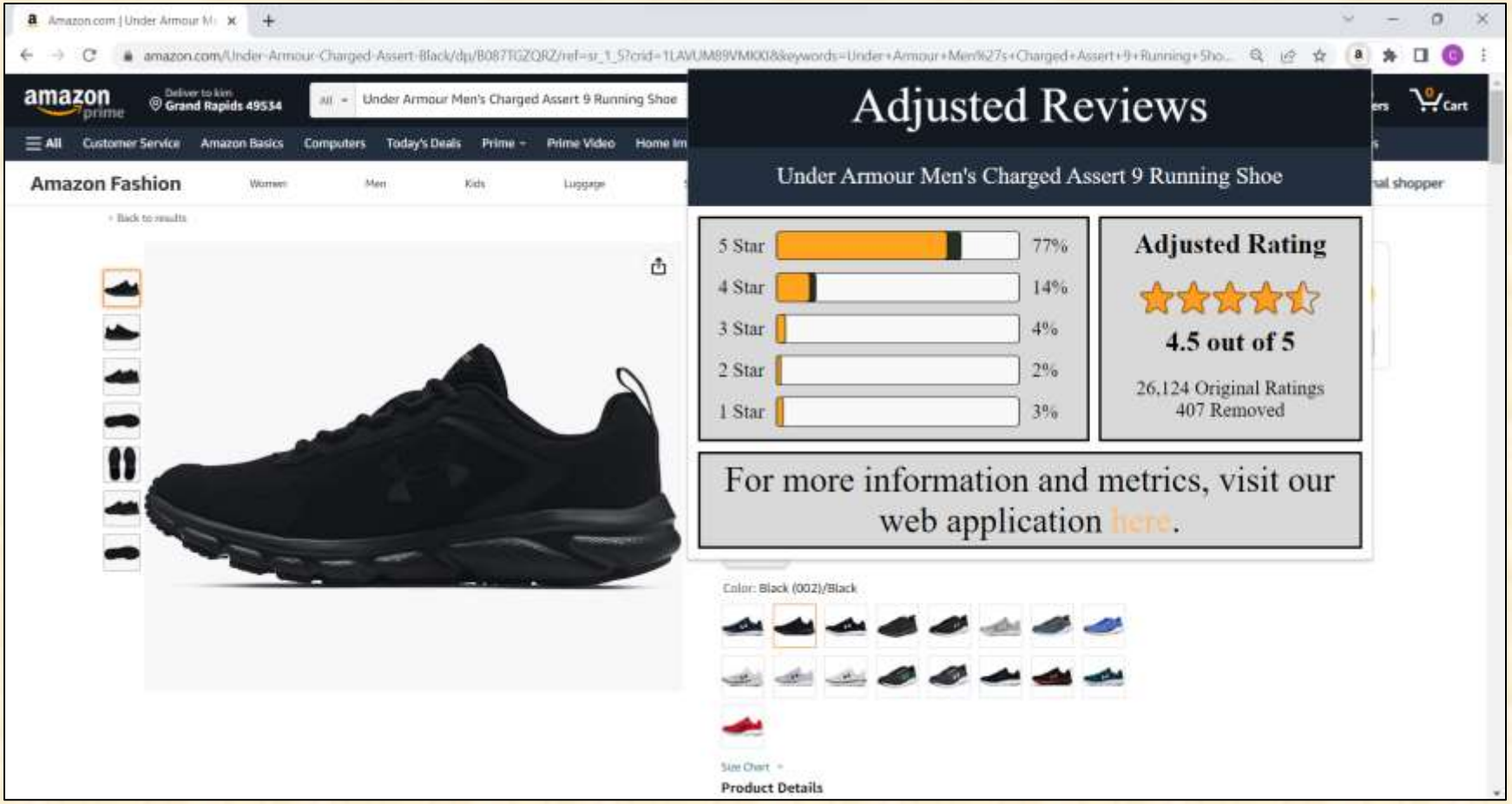
10Ft Long Surge Protector Power Strip, Extension Cord with 8 AC Outlets and 4 USB Ports(1 USB C), 3-Side Outlet Extender Strip, Flat Plug, Wall Mount, Dorm Home

$19.86

Product Page



Browser Extension (On Amazon)

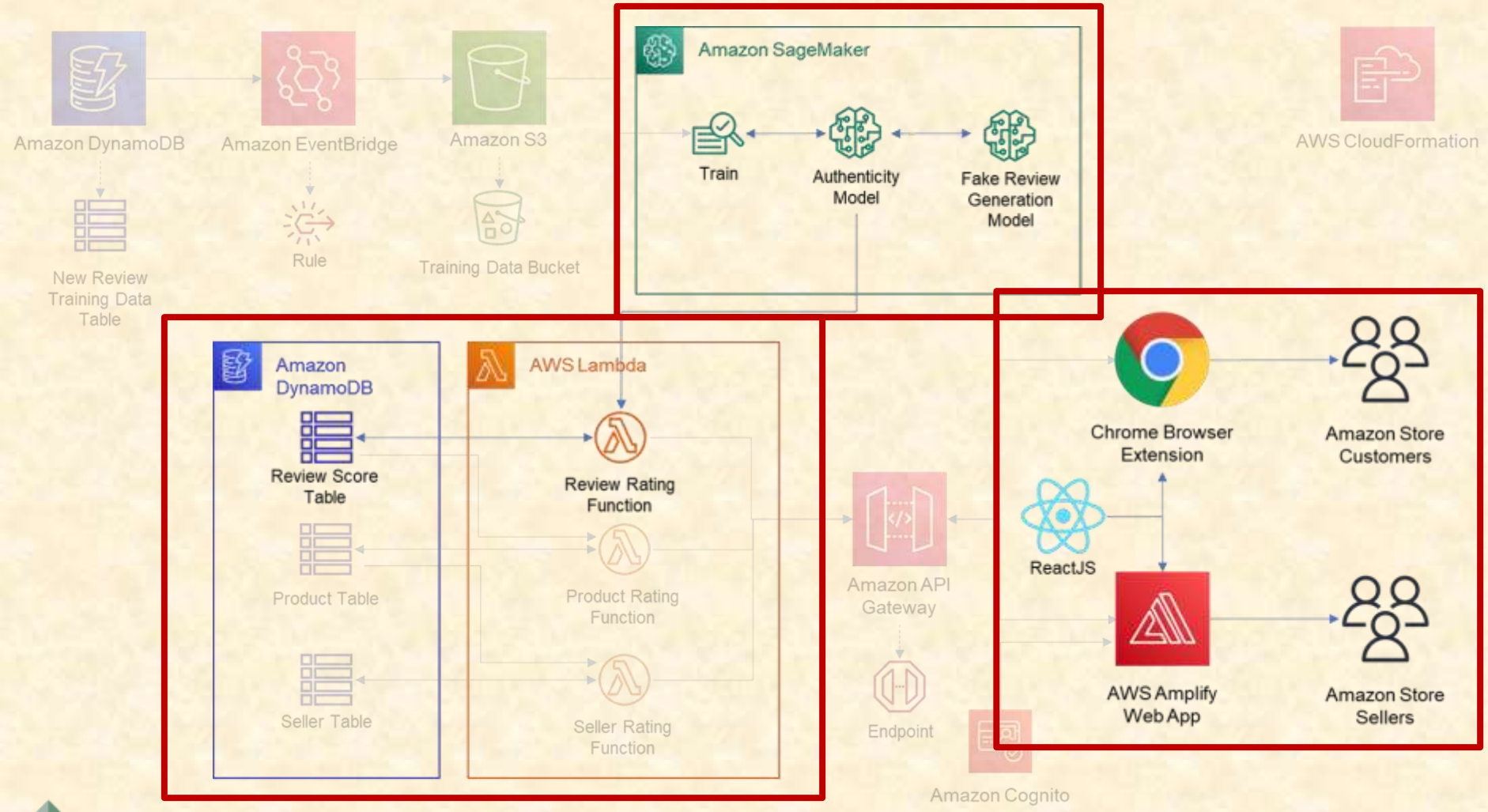


Fake Review Generation

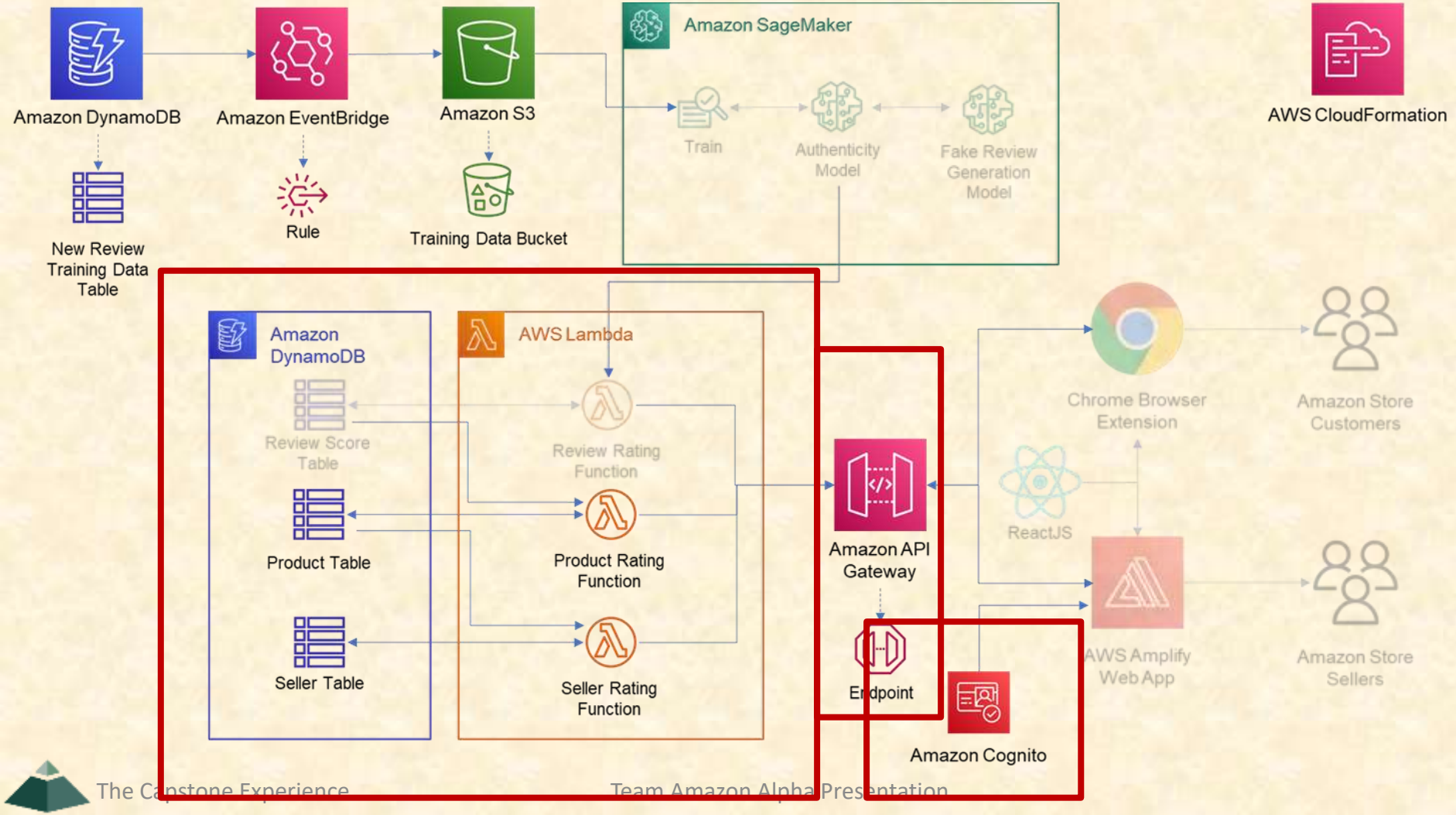
```
response = client.invoke_endpoint(EndpointName="ElectronicsGenModel",
                                  ContentType='application/json',
                                  Body=data_bytes)

# load the response into clean variable
result = json.loads(response['Body'].read().decode())

# clear our the newline chars from output
generated_reviews = result[0]["generated_text"].split("\n")

for review in generated_reviews:
    print(review)
```

```
Great camera quality! I found it quite nice and accurate. If the other review had been the same then I c
ould probably have said to myself: there is not a better camera in this price range.

Rated 5 out of 5 by daniel k from Very simple and fast Very simple and fast

Rated 5 out of 5 by DZB from Great price For what it is I bought this camera to look at some vintage fil
m. It has a great quality and the sound is clear in quality. I want to give it a shot and to say thank y
ou all... for this camera it was a great deal. It was very good price.

Rated 5 out of 5 by JayC from Great deal I am very impressed with the quality and quality of the Canon E
OS 35mm f/1.2 E DSLR. You can also use it for the Sony STIF. I have a 35mm f/1.2 from the Sony STIF but
no 50mm f/1.1 E lens. This camera is also better than other high end lenses which I own. I would buy aga
in for that to work and use for other things or have it back.
```

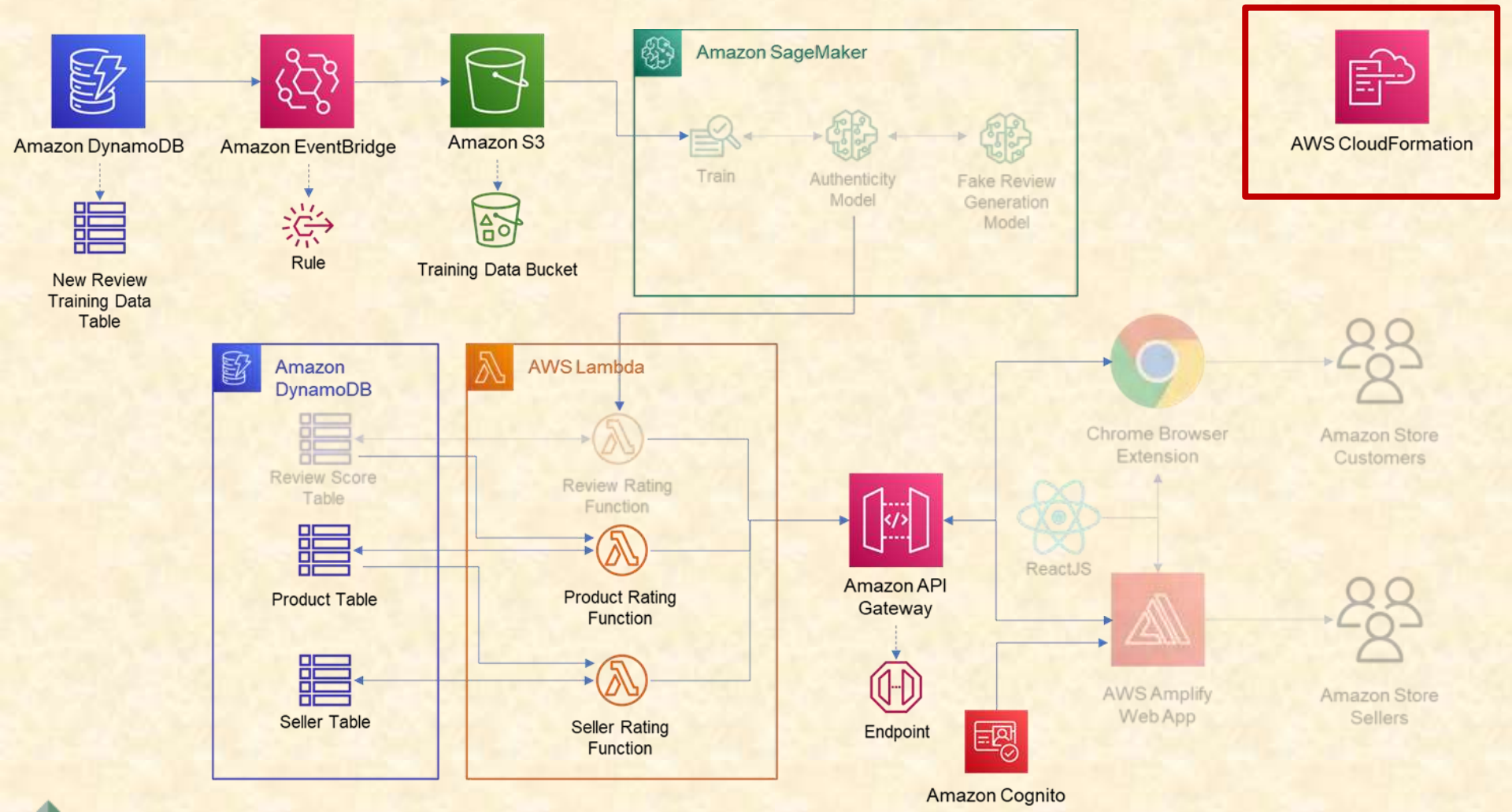
What's Left: Current Status



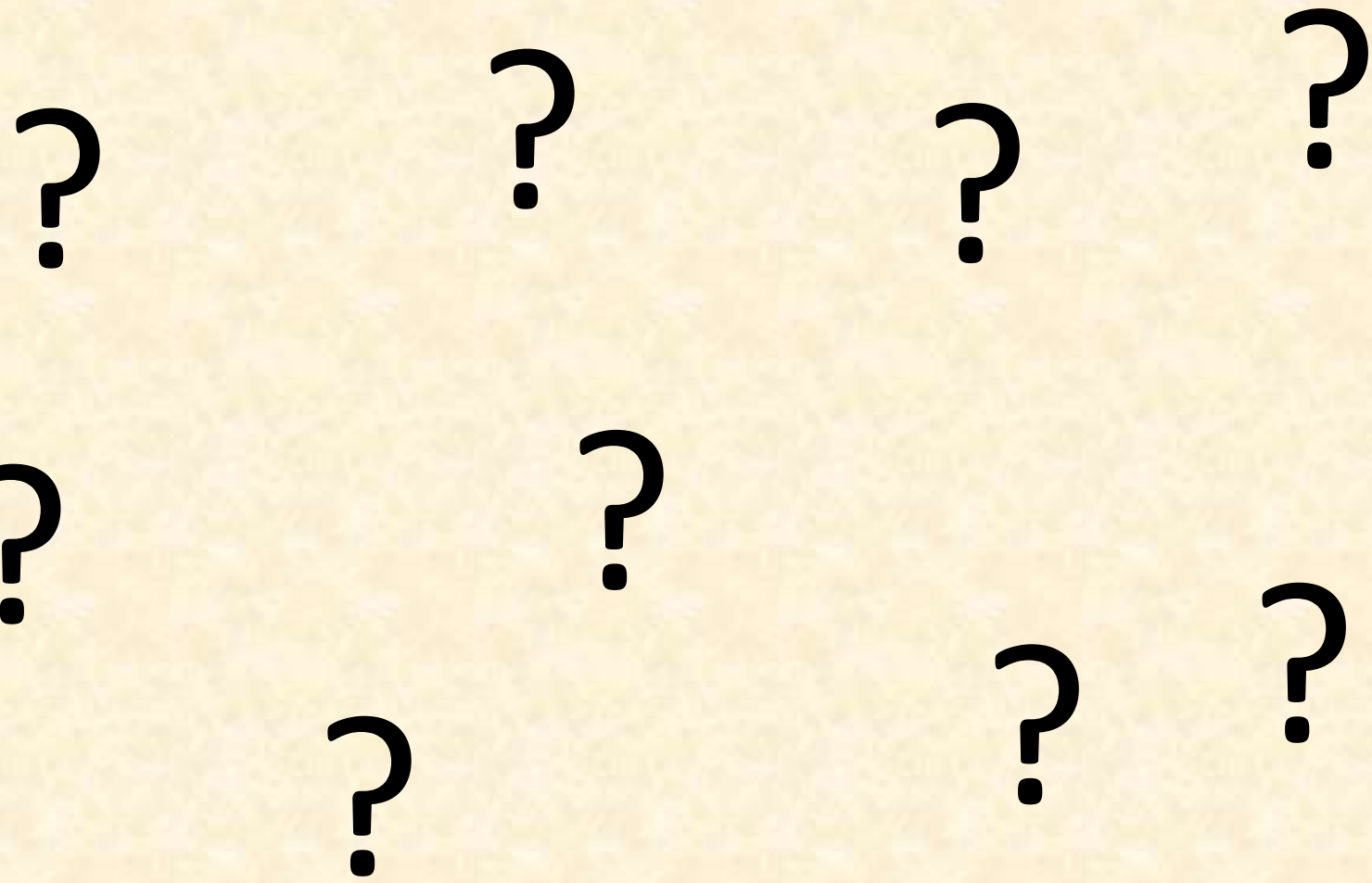
What's Left: Beta



What's Left: Deliverable



Questions?



End of slide show, click to exit.

return

ML Preprocessing Explained

- SkLearn.CountVectorizer(analyzer)
  - Tokenize text features
  - Analyzer picks out stopwords, punctuation, etc
- Term Frequency – Inverse Document Frequency
  - Word "Originality"
  - Tf: times word **t** appears in doc **d**
  - Idf: # of docs word appears in

Scikit-Learn

- $\text{IDF}(t) = \log\frac{1+n}{1+\text{df}(t)} + 1$

Standard notation

- $\text{IDF}(t) = \log\frac{n}{\text{df}(t)}$

In implementation, must account for divide-by-zero errors