Combating Digital Deception

TRUSTWORTHY AI IN THE BOT&FRAUD SPACE

Yi Han
Research Scientist
F5

F5 Intro

F5 Solutions

Web App and API Protection

App Network and Performance

Fraud & Bot Prevention

Zero Trust Security

Mobile App Delivery

F5 Customers

Financial Institutions

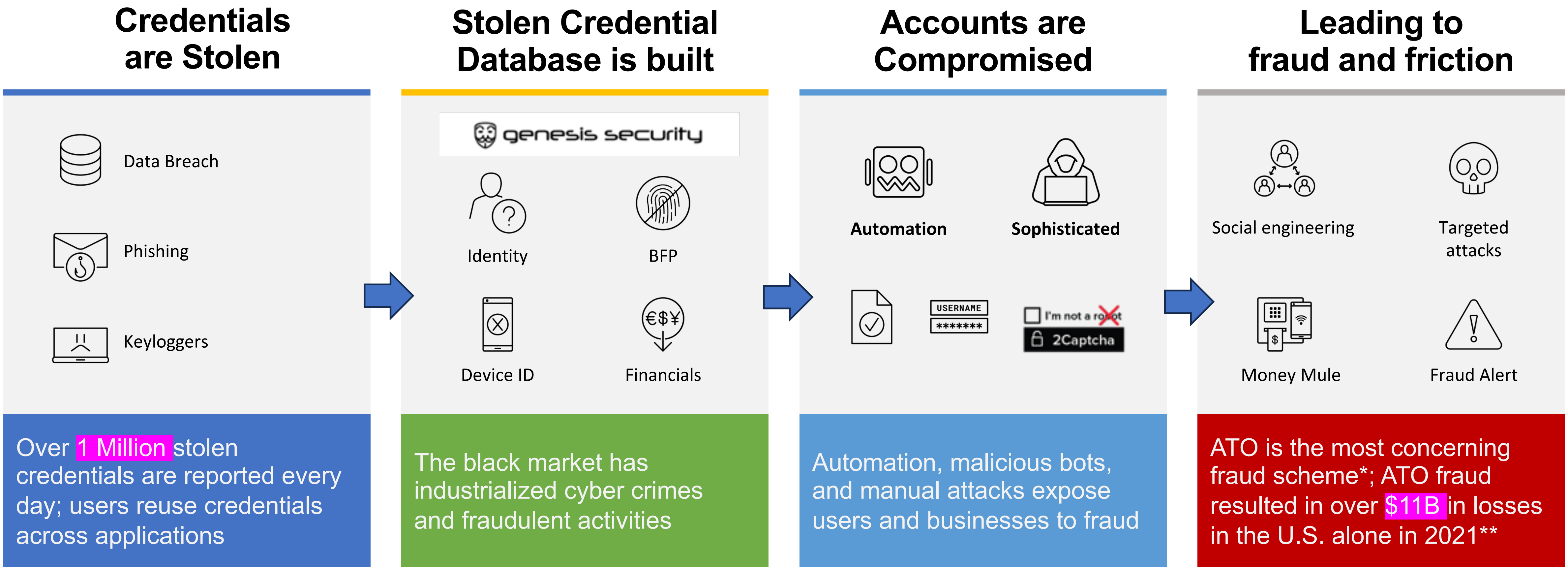
Public Sector

Service Providers

Healthcare

Ecommerce

Bot/Fraud Lifecycle



Credentials are Stolen

Data Breach

Phishing

Keyloggers

Over 1 Million stolen credentials are reported every day; users reuse credentials across applications

Stolen Credential Database is built

genesis security

Identity

BFP

Device ID

Financials

The black market has industrialized cyber crimes and fraudulent activities

Accounts are Compromised

Automation

Sophisticated

Automation, malicious bots, and manual attacks expose users and businesses to fraud

Leading to fraud and friction

Social engineering

Targeted attacks

Money Mule

Fraud Alert

ATO is the most concerning fraud scheme*; ATO fraud resulted in over $11B in losses in the U.S. alone in 2021**

I*SMG 2021 Faces of Fraud Survey; ** Javelin Strategy & Research, 2021

Bot vs Fraud

Bot

- Automated
- Large scale
- Multi-purpose

- Credential stuffing
- Scalping
- Click fraud
- Inventory Hoarding

Fraud

- Manual
- Narrowly focused
- Financial gain oriented

- Account take over
- Purchase fraud
- Payment fraud

Impact of Bot/Fraud

25%

Of the 100 **worst financial loss** incidents in past 5 years, the leading cause was **credential attacks**

$362B

Cumulative online payment fraud losses forecast $362B (2023 – 2028 period)

$260B

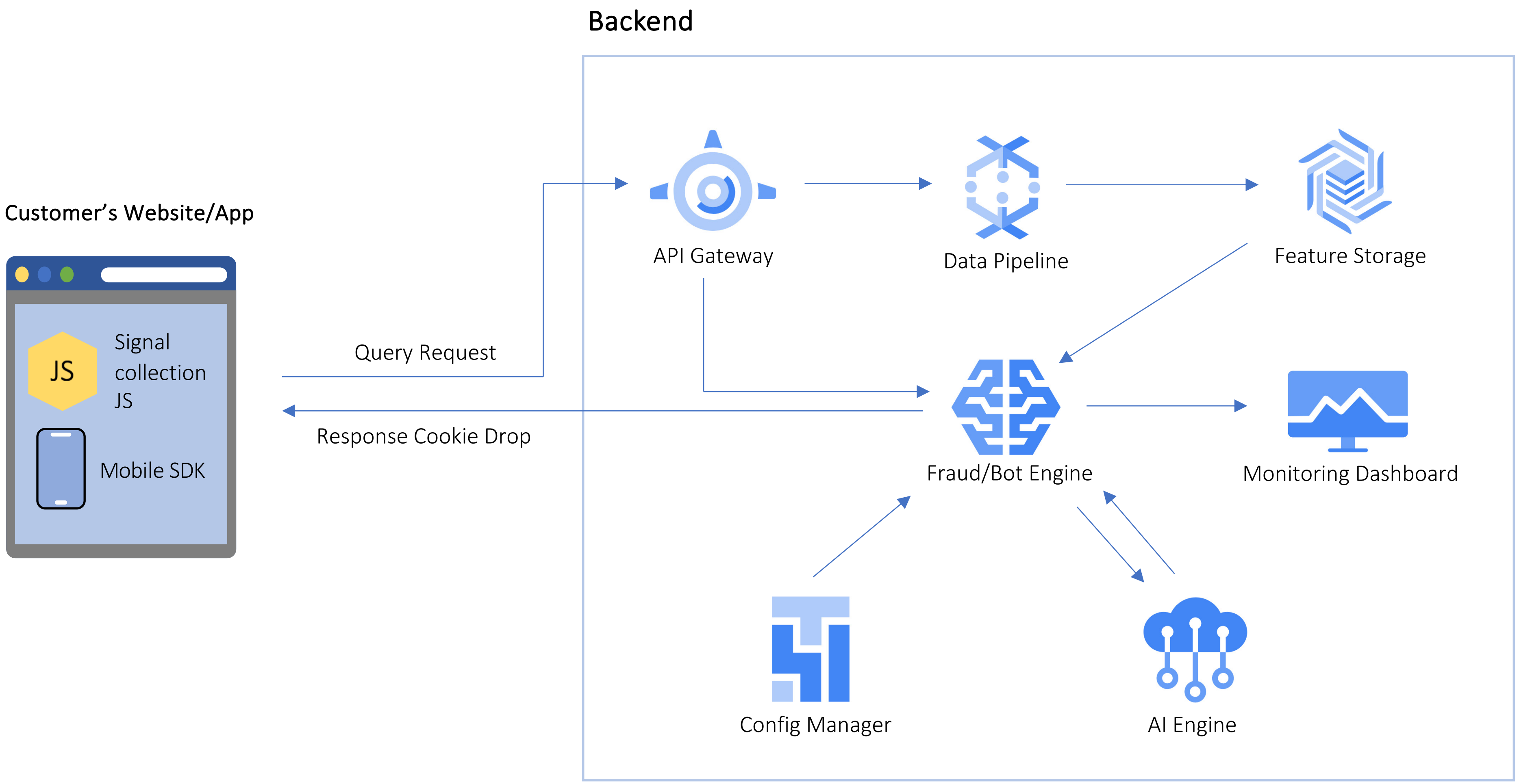
Lost orders per year attributed to excessive **checkout friction**

https://www.f5.com/labs/articles/threat-intelligence/the-state-of-the-state-of-application-exploits-in-security-incidents

https://baymard.com/lists/cart-abandonment-rate

https://www.juniperresearch.com/whitepapers/fighting-online-payment-fraud-in-2022-beyond

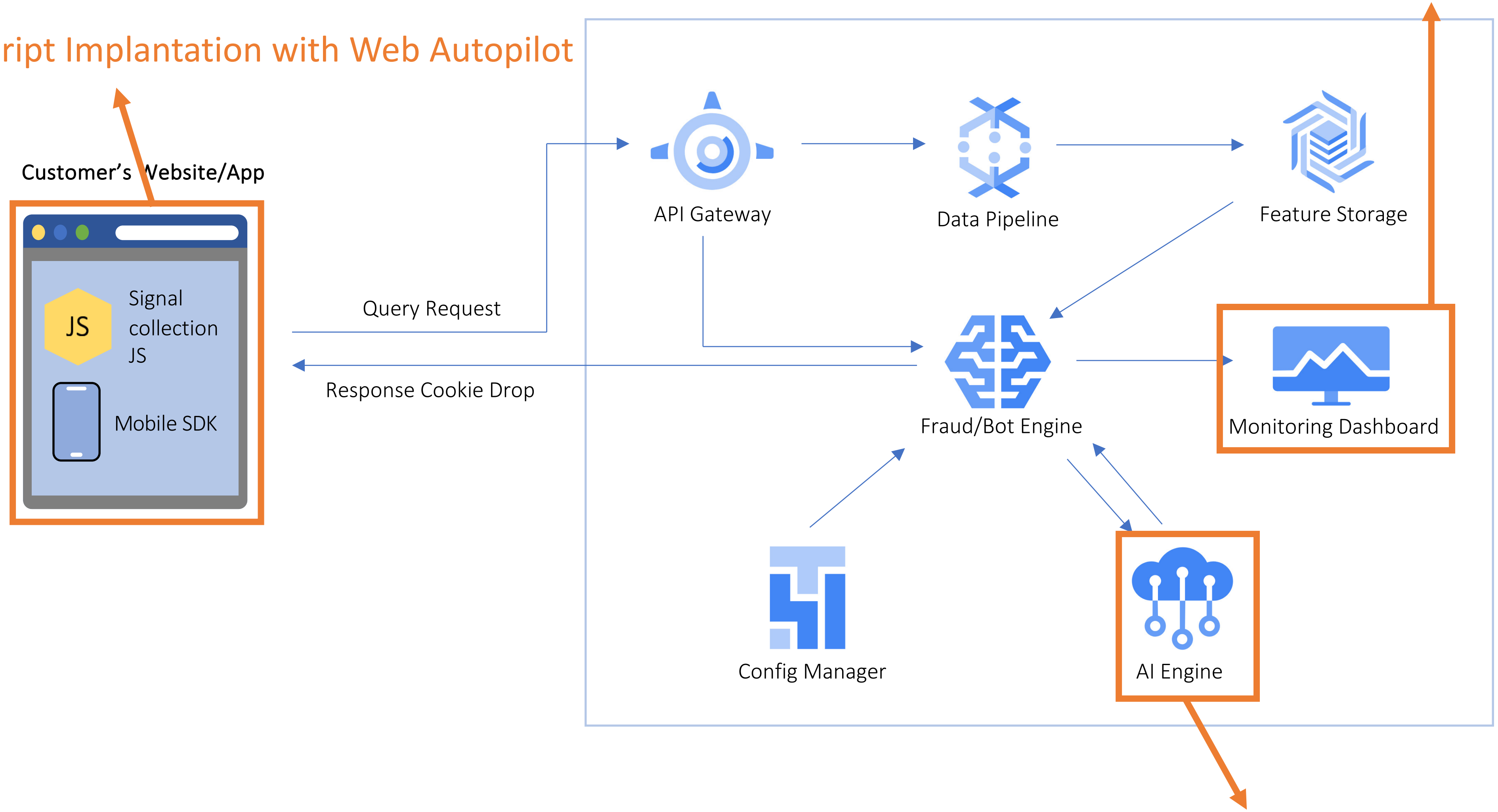
Bot/Fraud Infra



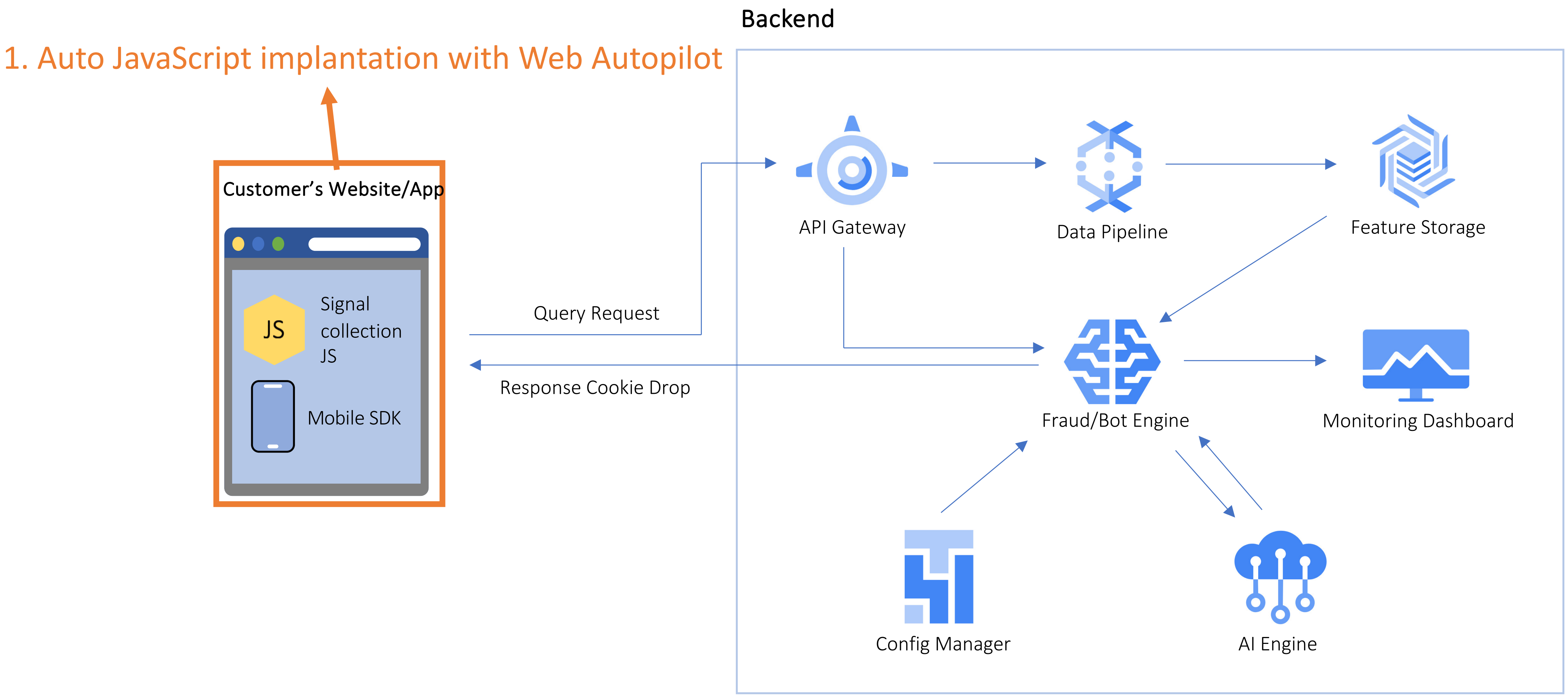
Agenda

3. Model maintenance - Distribution drift detection

1. Auto JavaScript Implantation with Web Autopilot



2. Bot/Fraud mitigation with AI models



1. Auto JavaScript implantation with Web Autopilot

Customer’s Website/App

JS

Signal collection JS

Mobile SDK

Query Request

Response Cookie Drop

Backend

API Gateway

Data Pipeline

Feature Storage

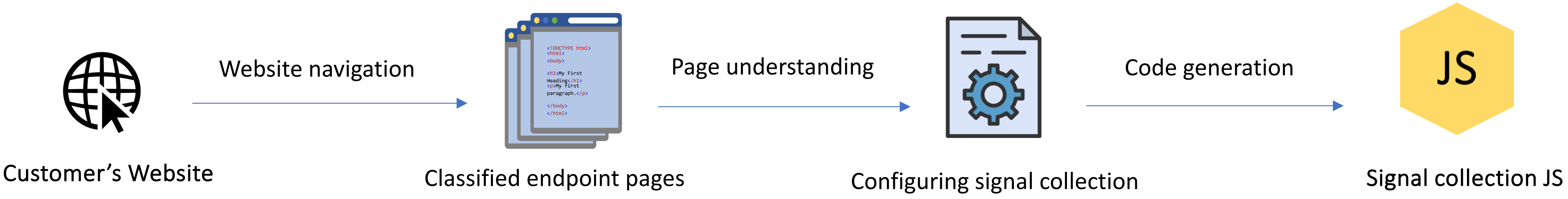
Fraud/Bot Engine

Monitoring Dashboard

Config Manager

AI Engine

Frontend Code Implantation



Customer’s Website

Website navigation

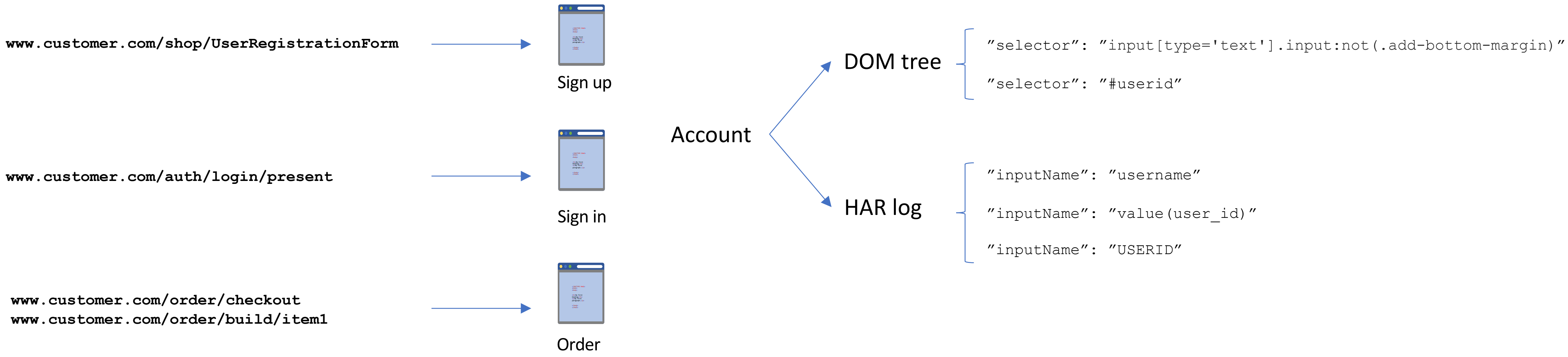
Classified endpoint pages

Page understanding

Configuring signal collection

Code generation

Signal collection JS



www.customer.com/shop/UserRegistrationForm

Sign up

www.customer.com/auth/login/present

Sign in

www.customer.com/order/checkout
www.customer.com/order/build/item1

Order

Account

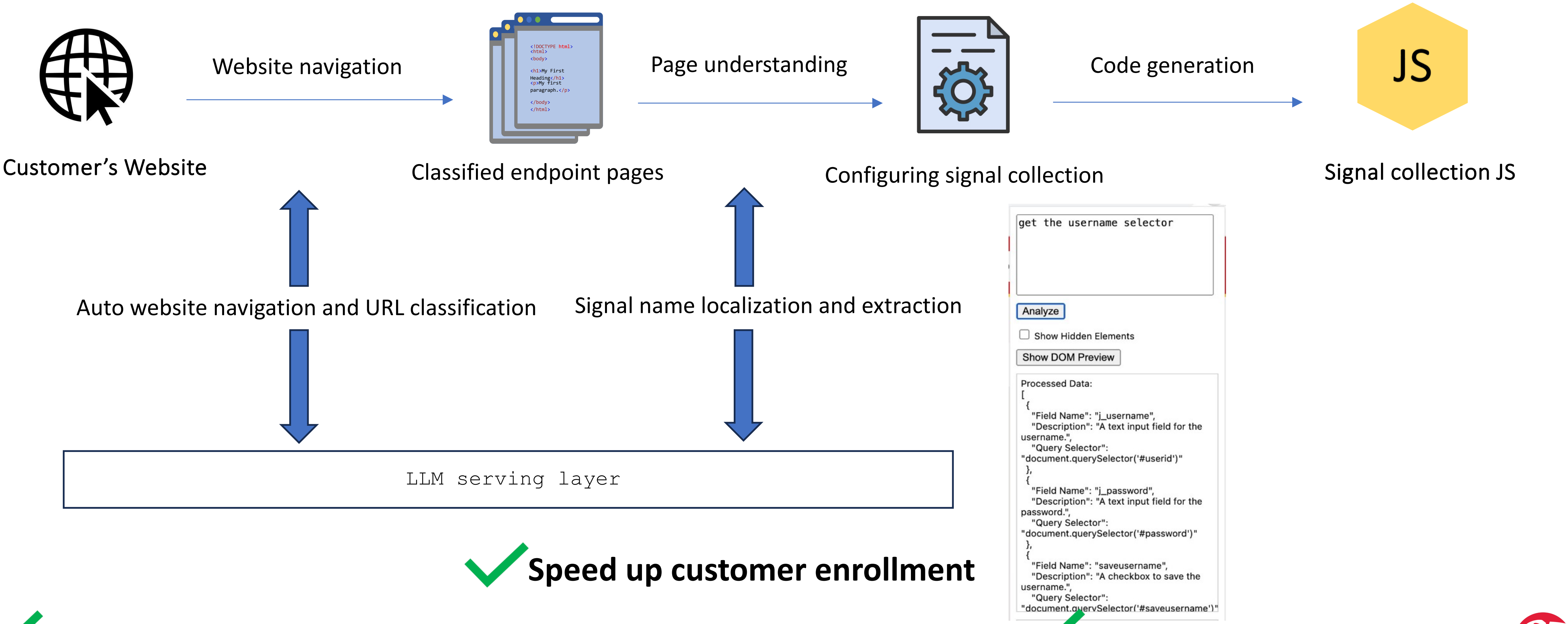
DOM tree

```
”selector”: ”input[type='text'].input:not(.add-bottom-margin)”
”selector”: ”#userid”
```

HAR log

```
”inputName”: ”username”
”inputName”: ”value(user_id)”
”inputName”: ”USERID”
```

Automated Code Implantation


Website navigation
Page understanding
Code generation
JS
Customer’s Website
Classified endpoint pages
Configuring signal collection
Signal collection JS
Auto website navigation and URL classification
Signal name localization and extraction
LLM serving layer
Analyze
Show DOM Preview

```
get the username selector
```

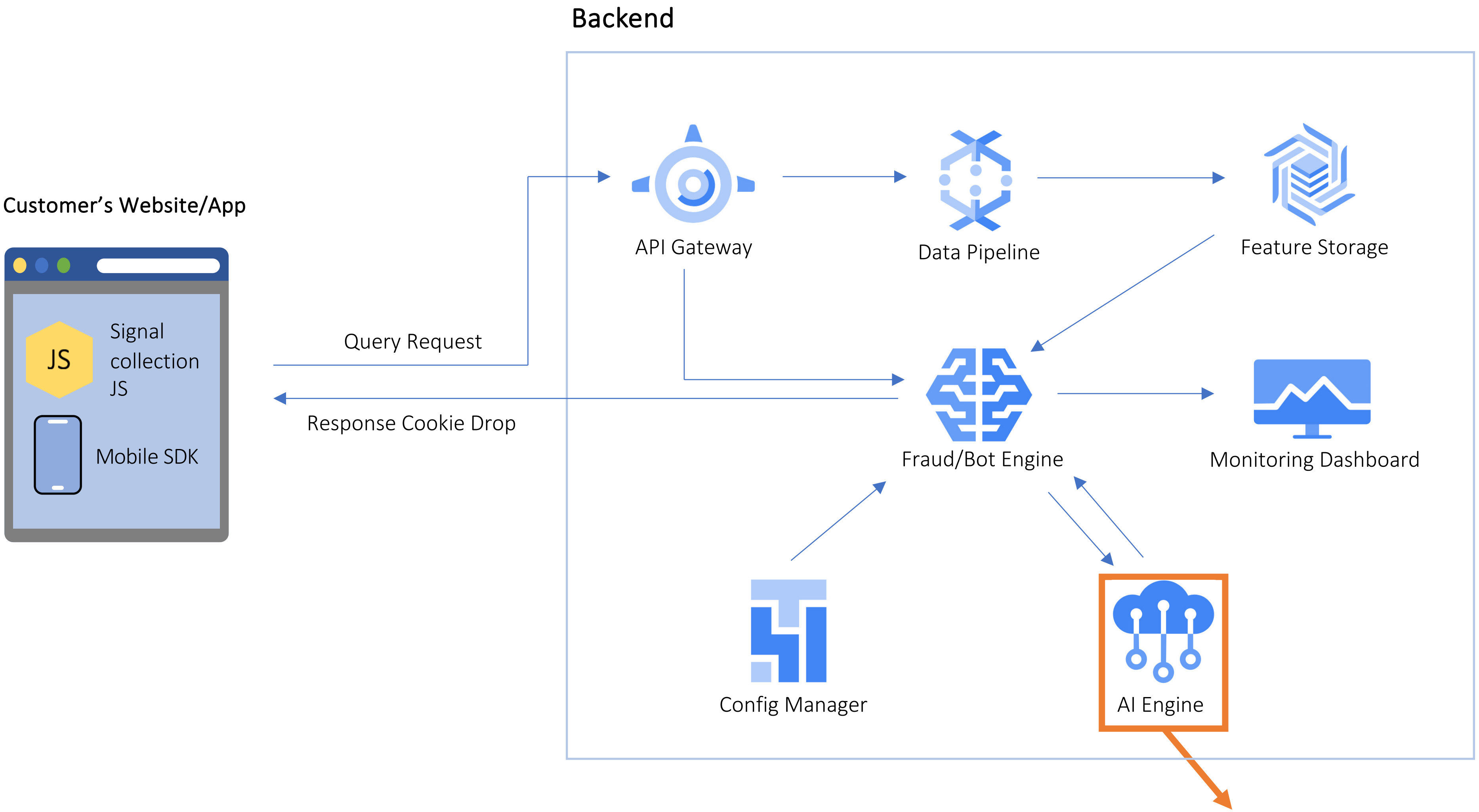
□ Show Hidden Elements

```
Processed Data:
[
  {
    "Field Name": "j_username",
    "Description": "A text input field for the
username.",
    "Query Selector":
"document.querySelector('#userid')"
  },
  {
    "Field Name": "j_password",
    "Description": "A text input field for the
password.",
    "Query Selector":
"document.querySelector('#password')"
  },
  {
    "Field Name": "saveusername",
    "Description": "A checkbox to save the
username.",
    "Query Selector":
"document.querySelector('#saveusername')"
```

✓ **Speed up customer enrollment**

✓ **Standardize workflow across customers**

✓ **Reduce maintenance cost**



Backend

Customer’s Website/App

JS

Signal collection JS

Mobile SDK

Query Request

Response Cookie Drop

API Gateway

Data Pipeline

Feature Storage

Fraud/Bot Engine

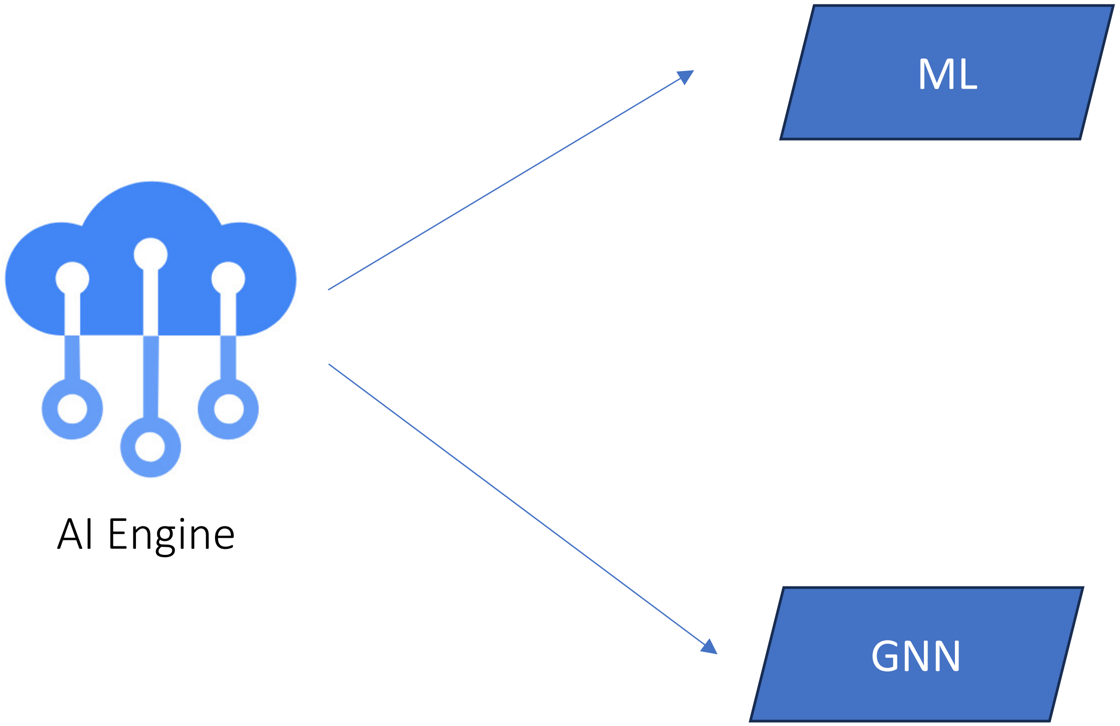
Monitoring Dashboard

Config Manager

AI Engine

2. Fraud/bot mitigation with AI models

AI Engine Architecture



Real time, <= 20 ms latency

Offline, through threat intelligence

Signal Collection

Network Signals

- IP Intelligence
- Bots
- Location
- OS, Browser
- Hosting
- VPN Usage

Digital Identity

- Device Identity
- Time Zone
- Browser fingerprint
- User Agent
- Emulated device
- Environment spoofing indicators

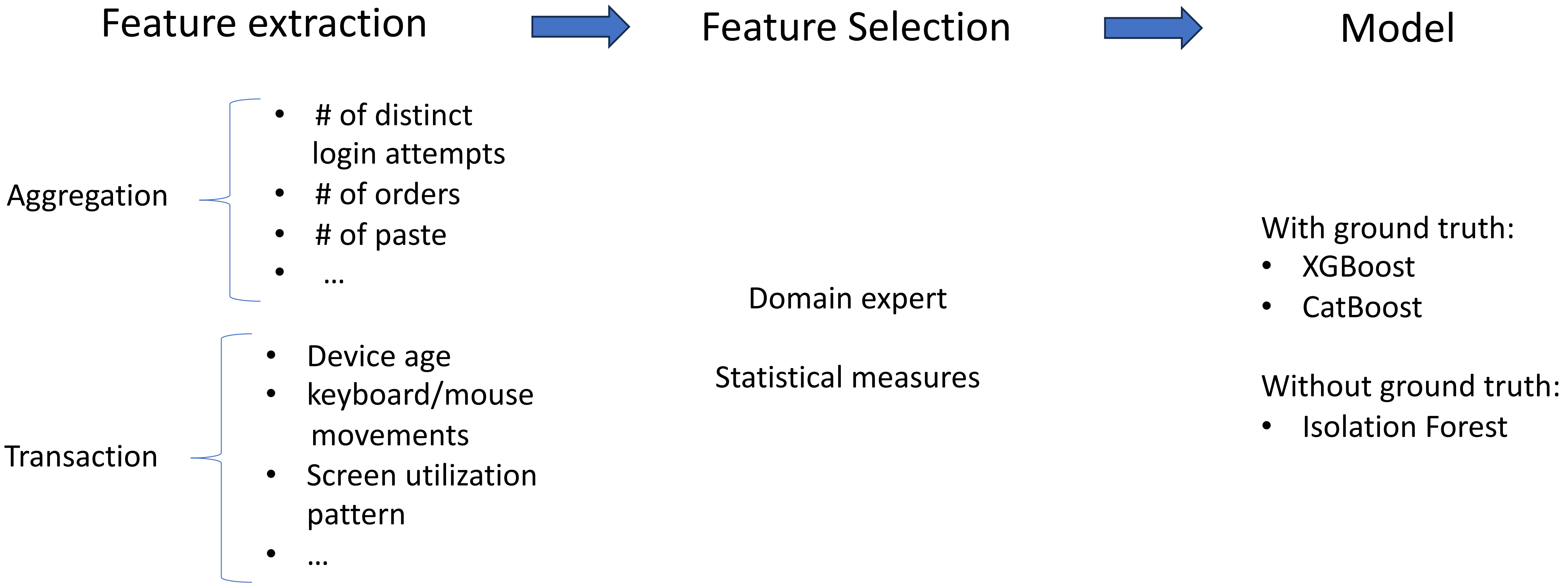
Behavior Biometrics

- Keyboard shortcuts
- Copy paste
- Mouse movements
- Touch input events
- Use of autofill
- Screen utilization

Behavior Profiling

- Device Activity
- User journey profiling
- User Signals (username, payee id, account id, etc.)

Real-time ML



Feature extraction → Feature Selection → Model

Aggregation

- # of distinct login attempts
- # of orders
- # of paste
- …

Transaction

- Device age
- keyboard/mouse movements
- Screen utilization pattern
- …

Domain expert

Statistical measures

With ground truth:

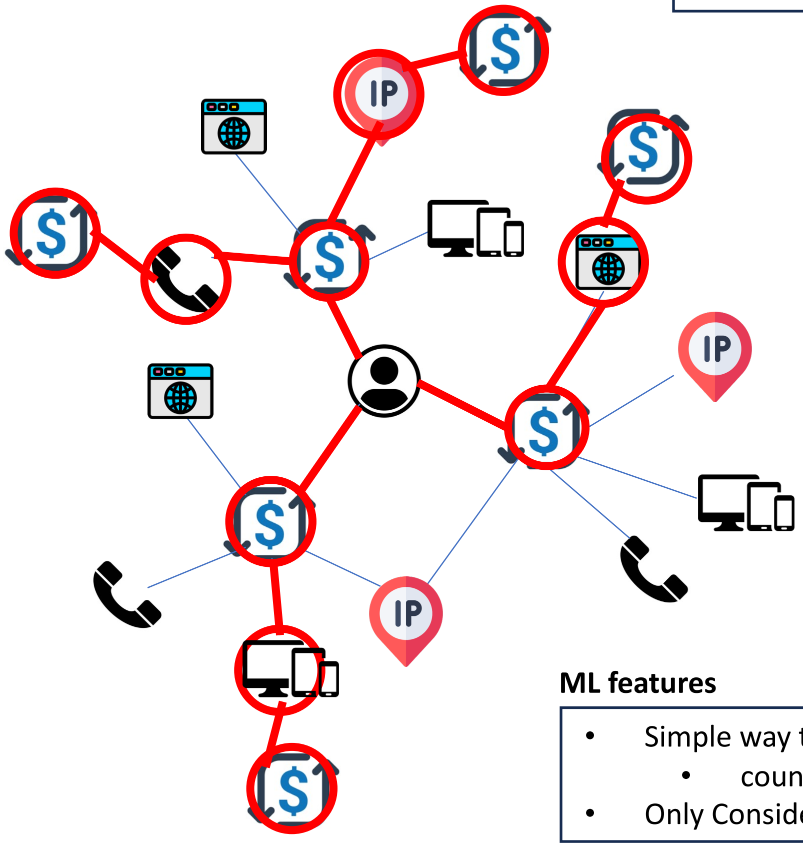
- XGBoost
- CatBoost

Without ground truth:

- Isolation Forest

GNN

Graph Schema:



GNN

- Parametric aggregation that can be trained
- Can propagate to multi-hop neighborhood

Cookie

IP

Device id

Transaction

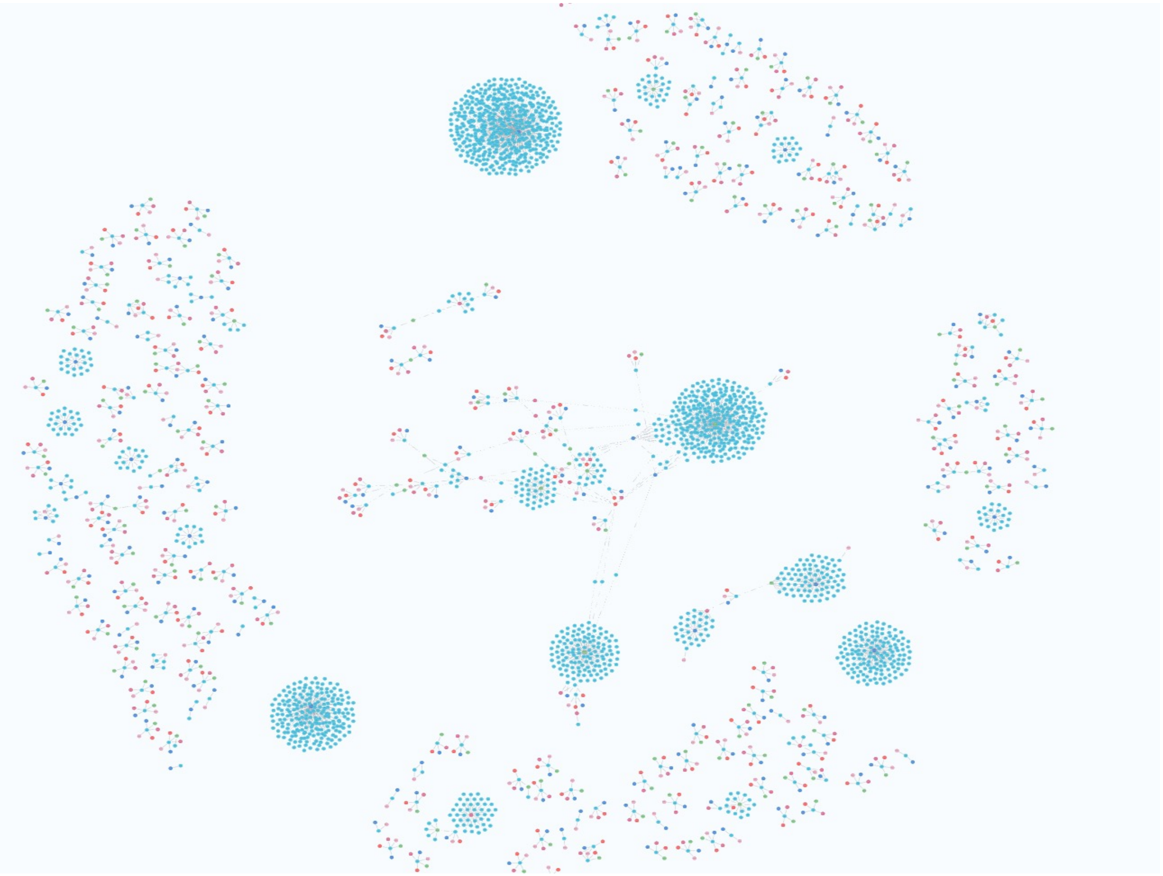
Account

☎ Phone

ML features

- Simple way to aggregate neighbor info
  - count/sum
- Only Consider 1-hop neighbors

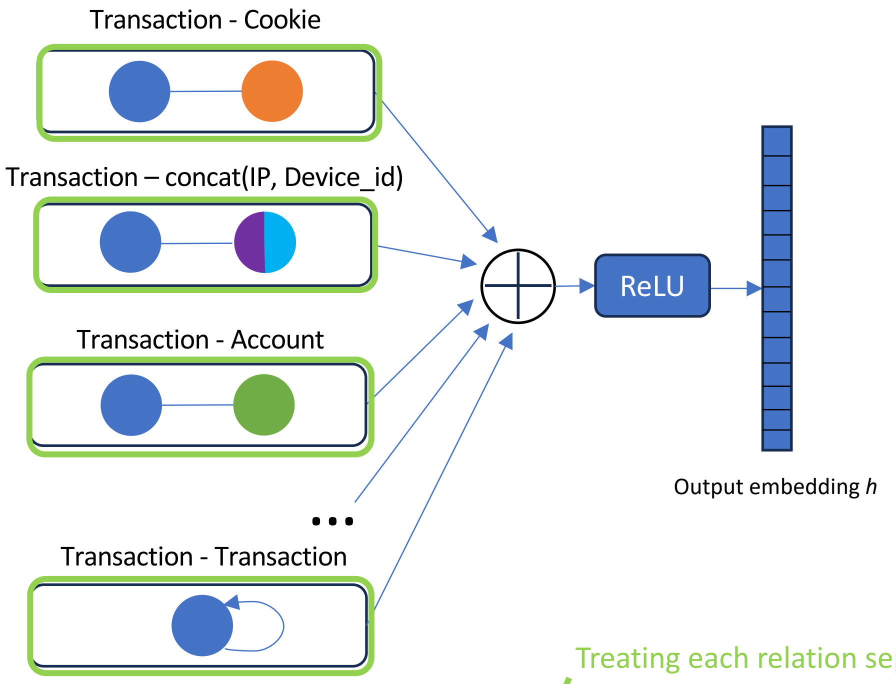
A part of the graph loaded in Neo4j:



- traffic from 5 customers of financial institutions and retailers
  - 153 million requests
- Maintain 3 months of data
  - 67 million of nodes/ 85 million of edges

GNN

Propagation model:

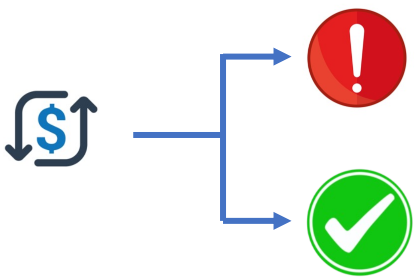


Treating each relation separately

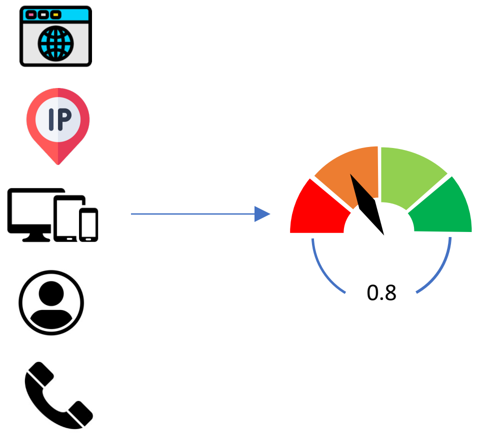
$$h_i^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_i^r}\frac{1}{c_{i,r}}W_r^{(l)}h_j^{(l)} + W_0^{(l)}h_i^{(l)}\right)$$

Use cases:

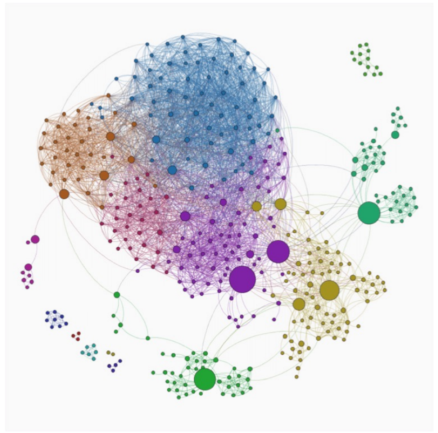
Detecting fraudulent transactions



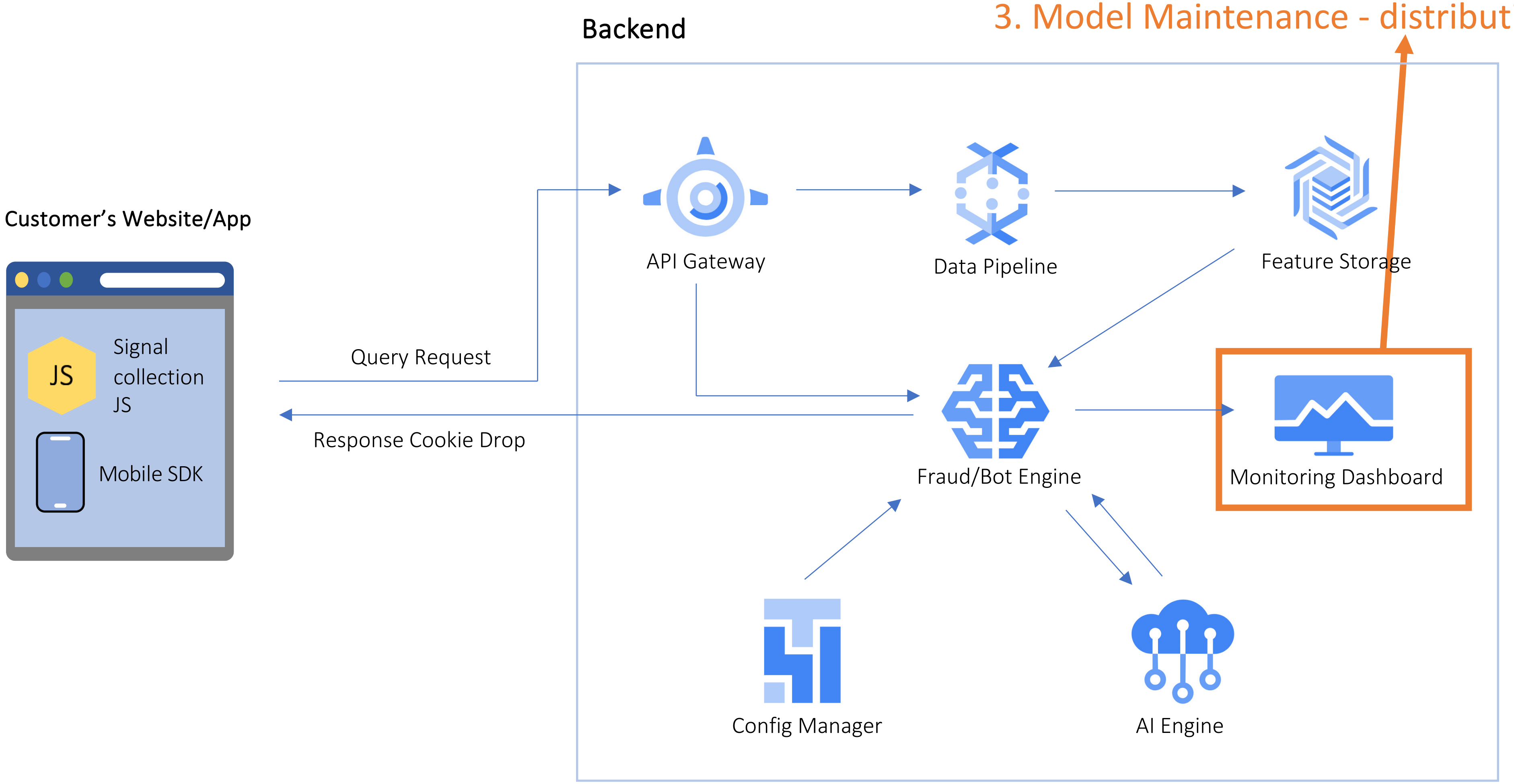
Devices/Identities Reputation



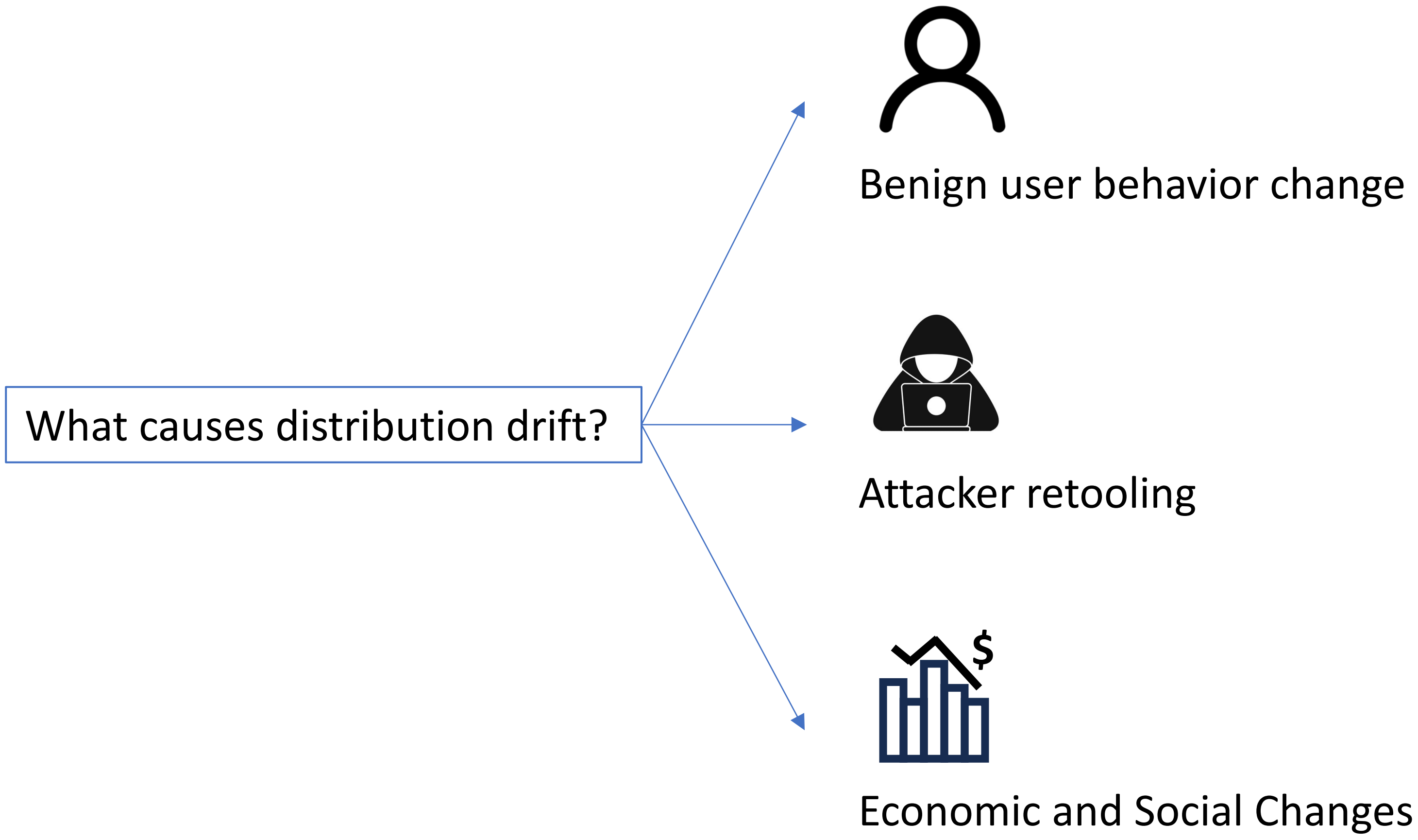
Detecting Bot/fraud Campaign



3. Model Maintenance - distribution drift detection



Distribution Drift



What causes distribution drift?

Benign user behavior change

- login frequency/location
- key/mouse movements
- screen utilization

Attacker retooling

- environment spoofing
- Network manipulation
- Identity manipulation

Economic and Social Changes

- order placing frequency/amount
- Money transfer frequency/amount

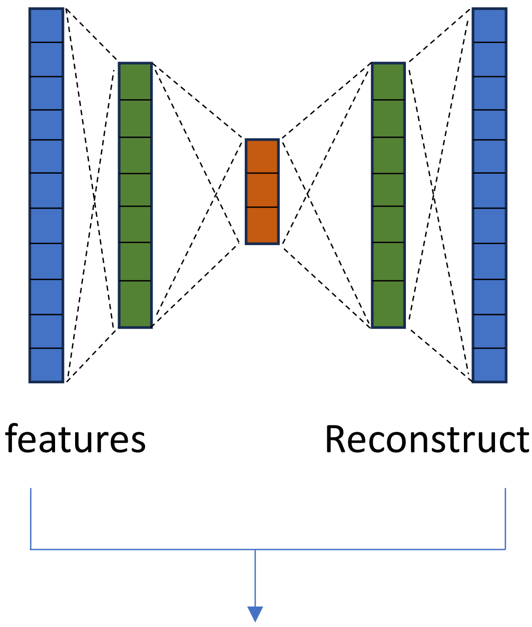
Detecting Distribution Drift

Challenges

- Non-linearity of individual features
- Correlation between features
- Non-linearity of the feature space

Solution

Autoencoder



Input features

Reconstructed features

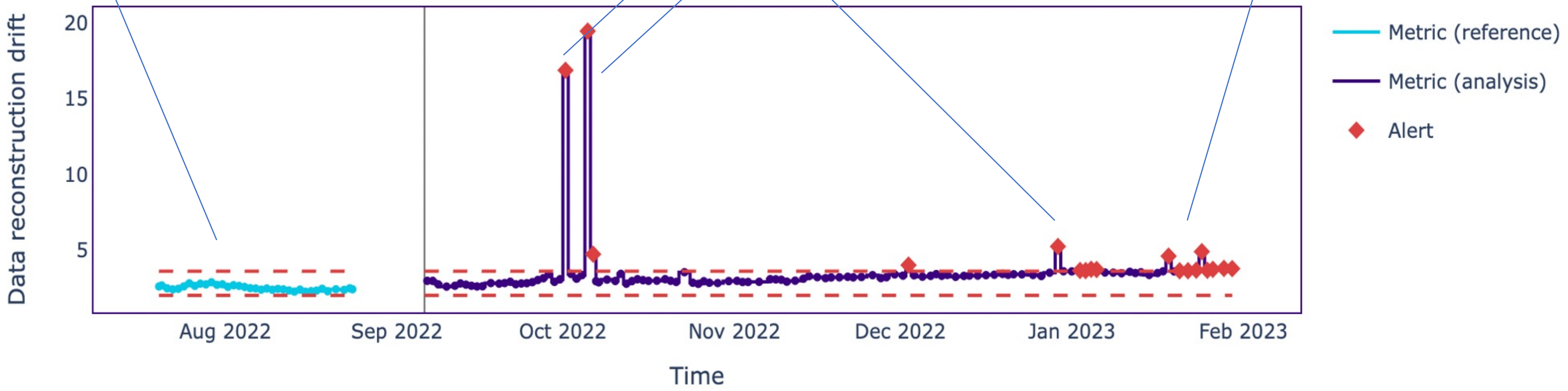
Measuring drift: Reconstruction error

Demonstration

Model trained in 08/2022 using 3 months of traffic

Random Events

Measured drift reached threshold, need to retrain



Reconstruction Error of a financial customer over a few month

Future work

Better Identity/Device Fingerprinting

- Some identity/device fingerprints are easy to be spoofed,
- Event the fingerprints of the same identity/device can keep changing
- More reliable fingerprinting techniques/signals, linking algorithms help better tracking the bad actors down

Protecting against informed attackers

- Advanced bad actors can use adversarial example techniques to bypass detectors[1]
- Defending against such attacks with adversarial training, defensive model distillation etc.

[1] Lunghi, Daniele, et al. "Adversarial Learning in Real-World Fraud Detection: Challenges and Perspectives." *Proceedings of the Second ACM Data Economy Workshop*. 2023.

THANK YOU!

Q & A?

Email: y.han@f5.com