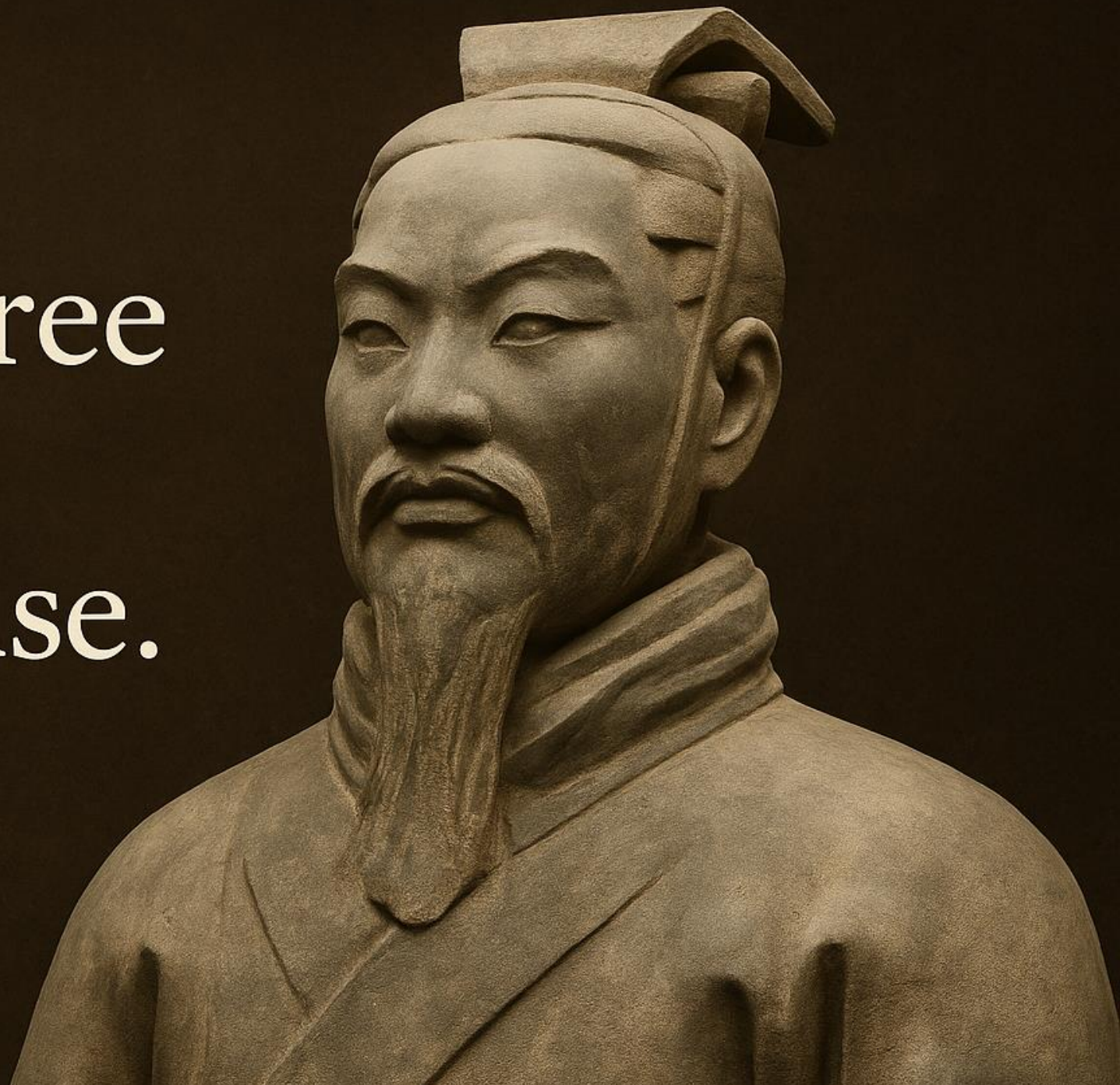
“

If there’s free compute, there’s abuse.

— Sun Tzu,
The Art of War
(500 BC)

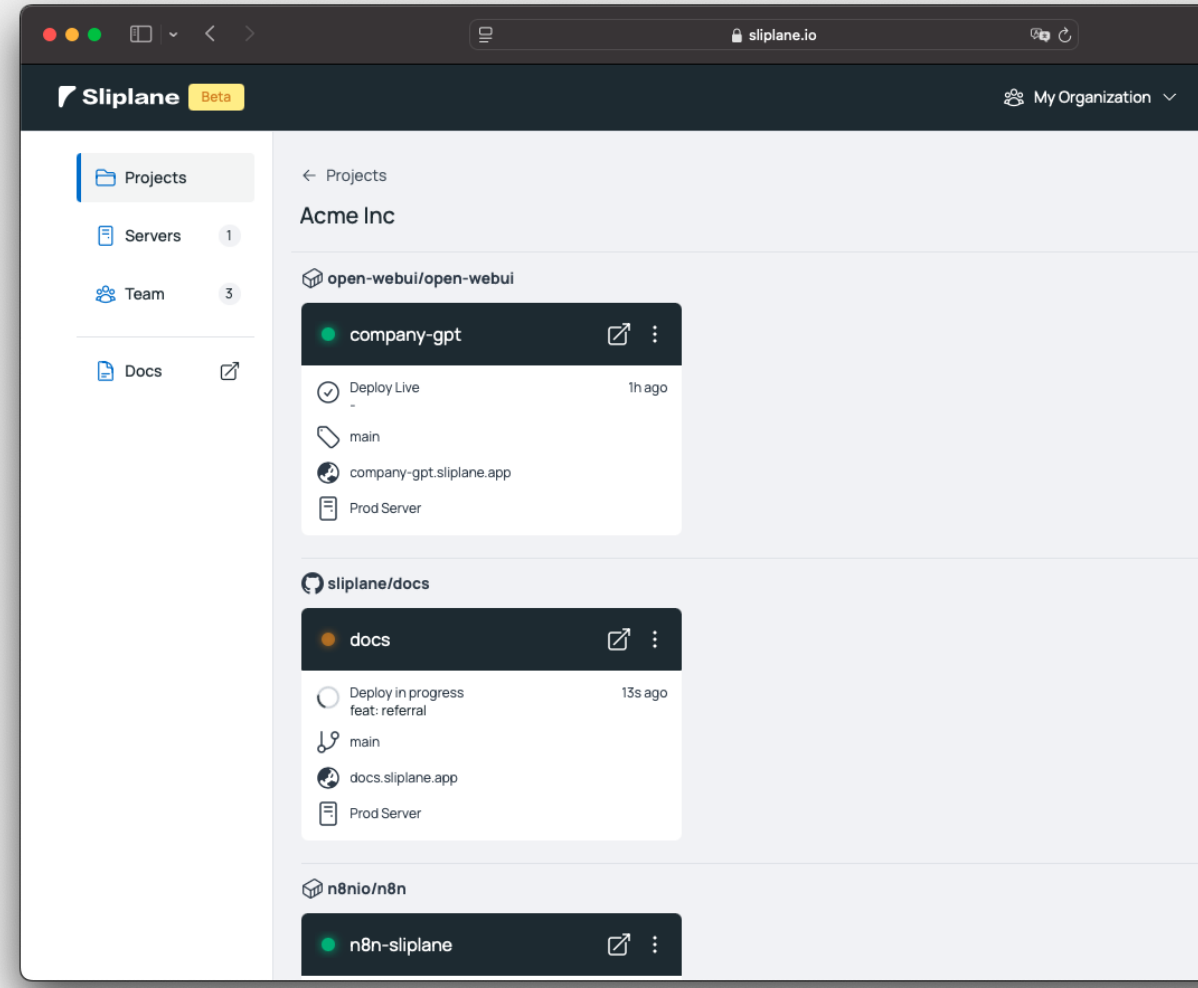


Background

- Jonas Scholz, Co-Founder @ Sliplane
- Cloud Infrastructure Nerd
- Docker Captain
- Not a cat

Background

- European Platform-as-a-Service
- Building the easiest way to deploy Docker container
- git push ↣ deploy
- Free Trial + Easy Platform attracts abuse!



Abuse

Crypto Mining

Illegal VPNs

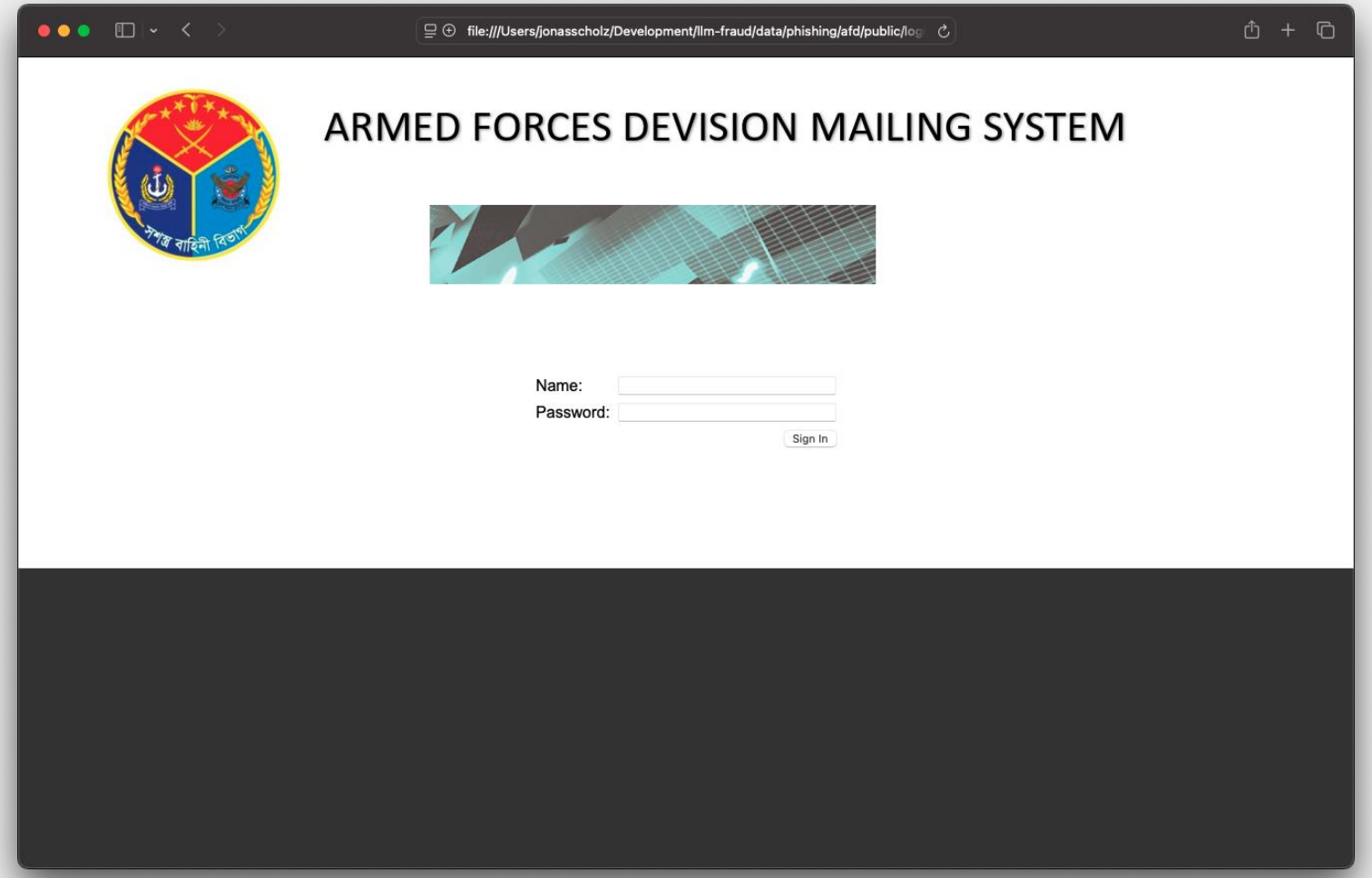
Netscans

DDoS

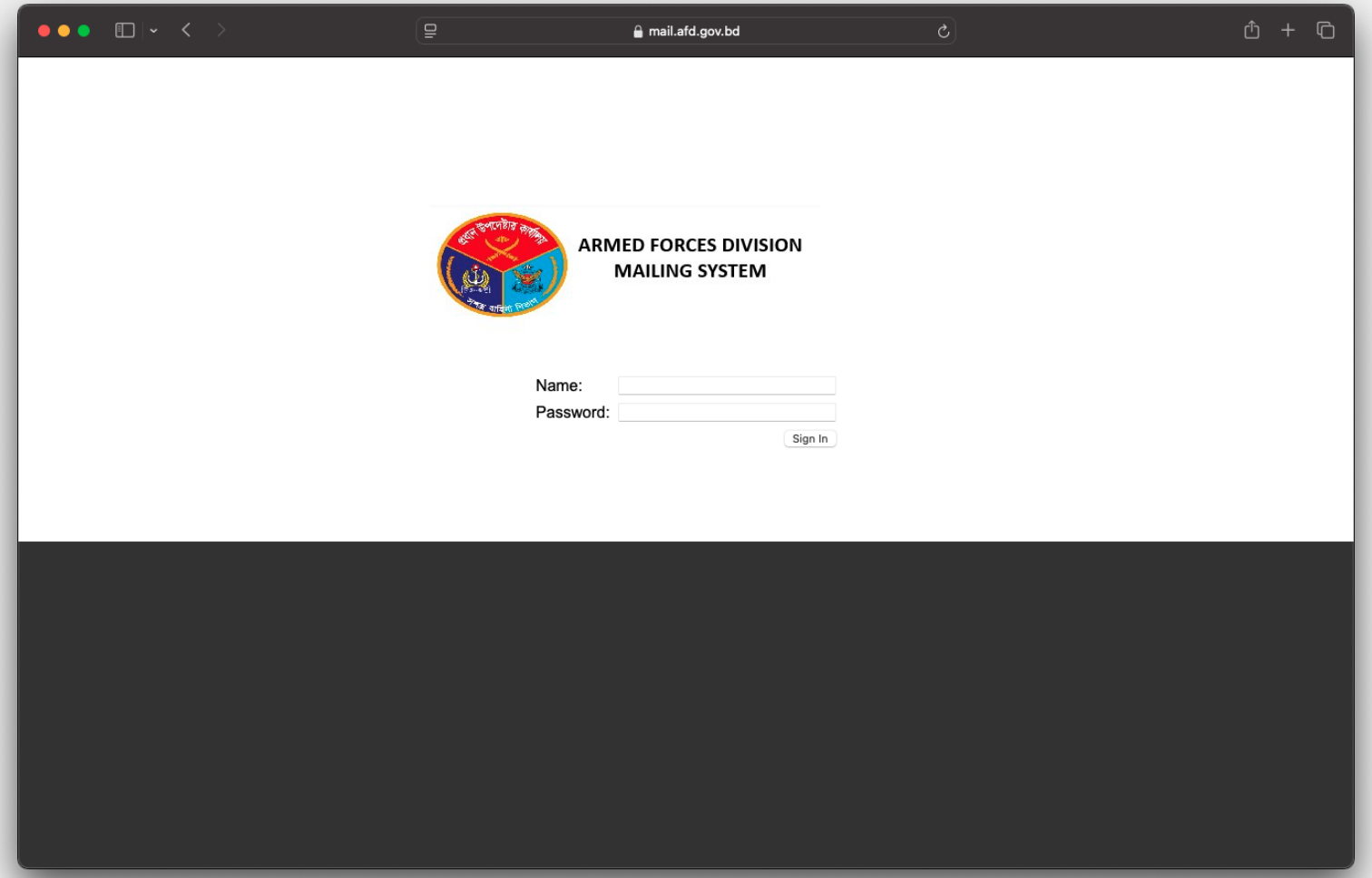
Spambots

Phishing

Phishing

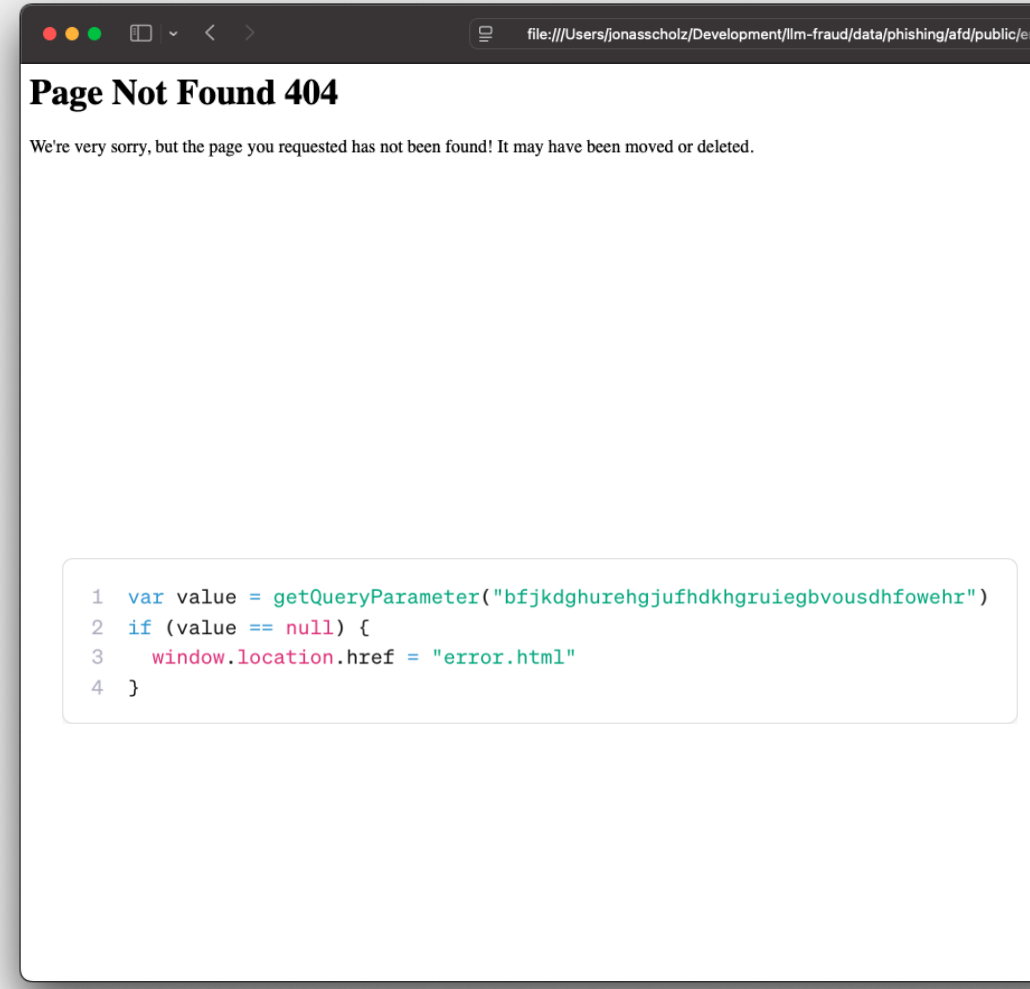

file:///Users/jonasscholz/Development/llm-fraud/data/phishing/afd/public/log
ARMED FORCES DEVISION MAILING SYSTEM
Name:
Password:
Sign In

Phishing


mail.afd.gov.bd
ARMED FORCES DIVISION
MAILING SYSTEM
Name:
Password:
Sign In

Problem

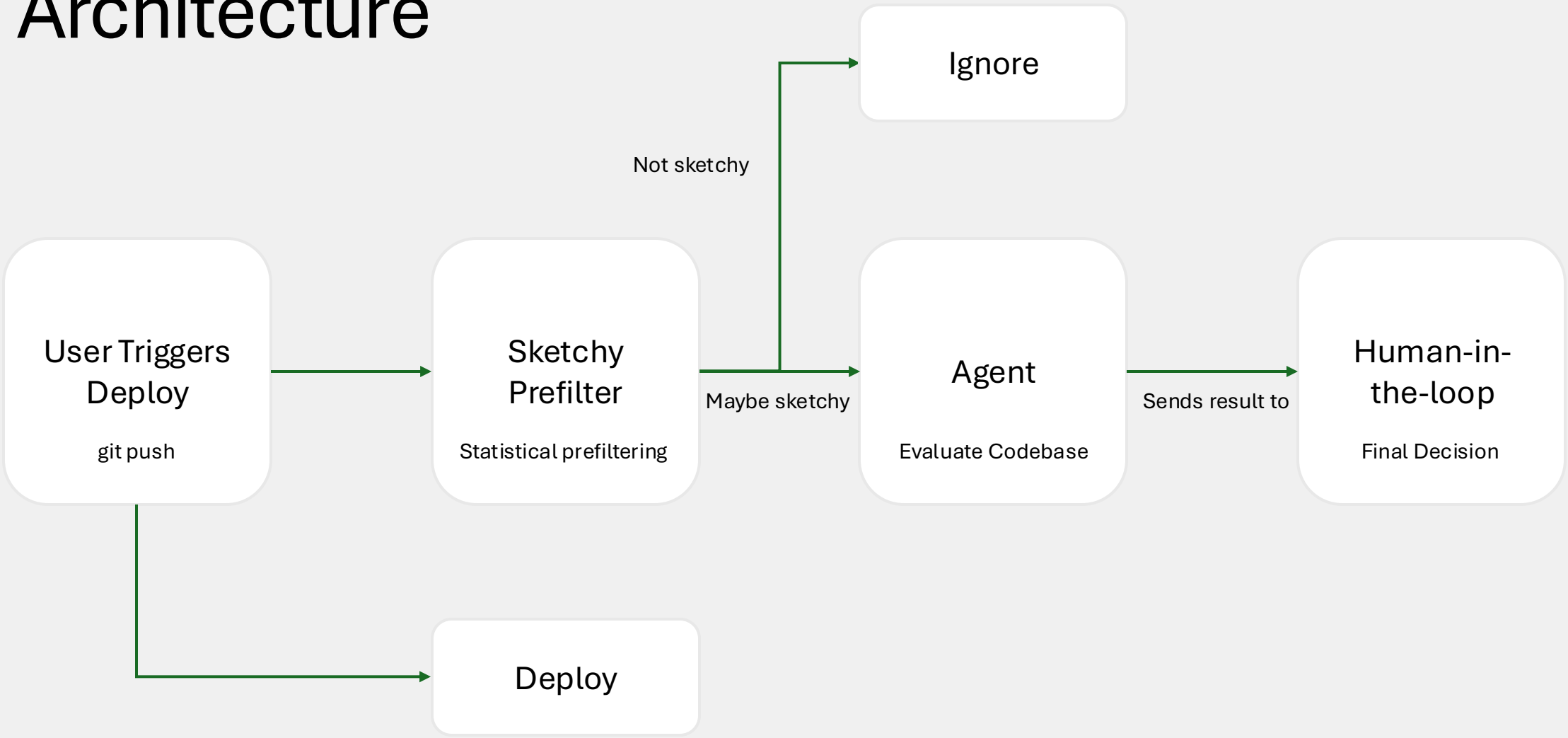
- Obfuscated
- Hard to detect (looks normal)
- Sheer volume



Solution

- Replace me with an army of context-aware agents!

Architecture



Agent

All models are open-source, self-hosted and with 0 data retention!

in-memory git clone

- Dockerfile
- requirements.txt
- src
  - main.py
  - login.html

```dockerfile
FROM node:12-alpine

ENV NODE_ENV=production

COPY ["package.json", "package-lock.json*", "./"]
...
```

+

```html
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html lang="en">
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
    <meta http-equiv="CACHE-CONTROL" content="NO-CACHE" />
    <link rel="SHORTCUT ICON" href="https://mail.afd.gov.bd/favicon.ico" />
    <meta http-equiv="PRAGMAS" content="NO-CACHE" />
...
```

Context Compacting

```
You are an expert at detecting terms of service violations
...

The terms of service disallow to use the service for the following reasons:
...
Further indicators are suspicious:

- obfuscated code
- remote control software
...
```

MISTRAL AI_

Classification

```
The `login.html` file is highly suspicious:
- It is designed to look like a login page for `afd.gov.bd` (Army Financial Department of Bangladesh) and is hosted on that domain.
- It contains a script that extracts a query parameter value from the URL and decodes it using `atob` (base64 decoding).
- This is a common technique used in phishing attacks to automatically populate login credentials.
- This login page is almost surely used to harvest credentials or exfiltrate data.
```

retries

Agent

All models are open-source, self-hosted and with 0 data retention!

```
1 The `login.html` file is highly suspicious:
2 - It is designed to look like a login page for `afd.gov.bd` (Army Financial Department of
  Bangladesh) and is hosted on that domain.
3 - It contains a script that extracts a query parameter value from the URL and decodes it using
  `atob` (base64 decoding).
4 - This is a common technique used in phishing attacks to automatically populate login
  credentials.
5 - This login page is almost surely used to harvest credentials or exfiltrate data.
```

Summary

- Be unattractive enough for abuse
- Red teaming is necessary!
- Small models are still years behind

- Checkout sliplane.io for deploying container!
- Connect with me on LinkedIn:

  https://www.linkedin.com/in/jonas-scholz-490274163/