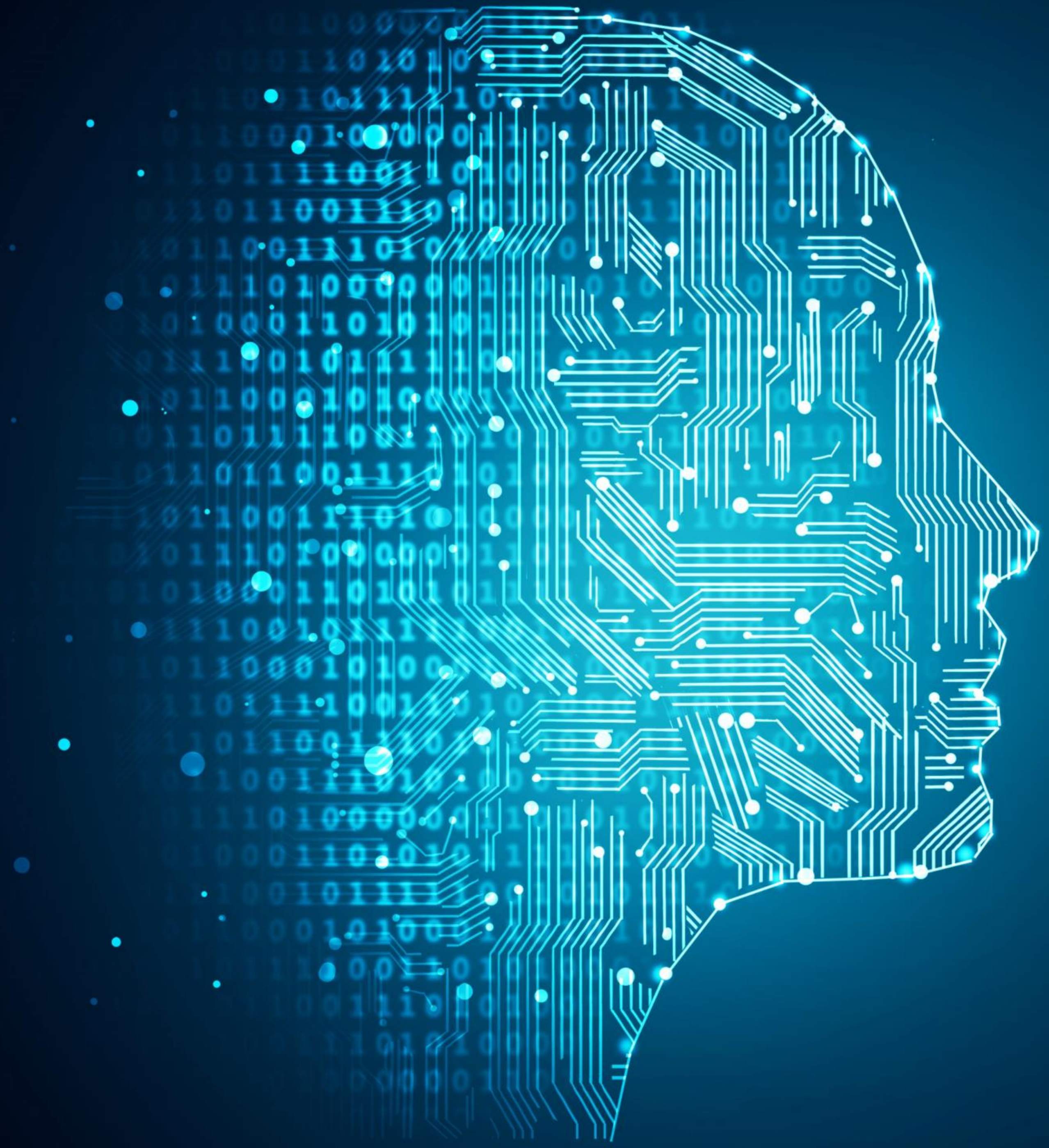
SIEMENS
Ingenuity for life



Threat Detection based on Deep Learning at Scale

Karl Peter Fuchs - karl.fuchs@siemens.com
Jan Pospisil - jan.pospisil@siemens.com
Siemens Cyber Defense Center

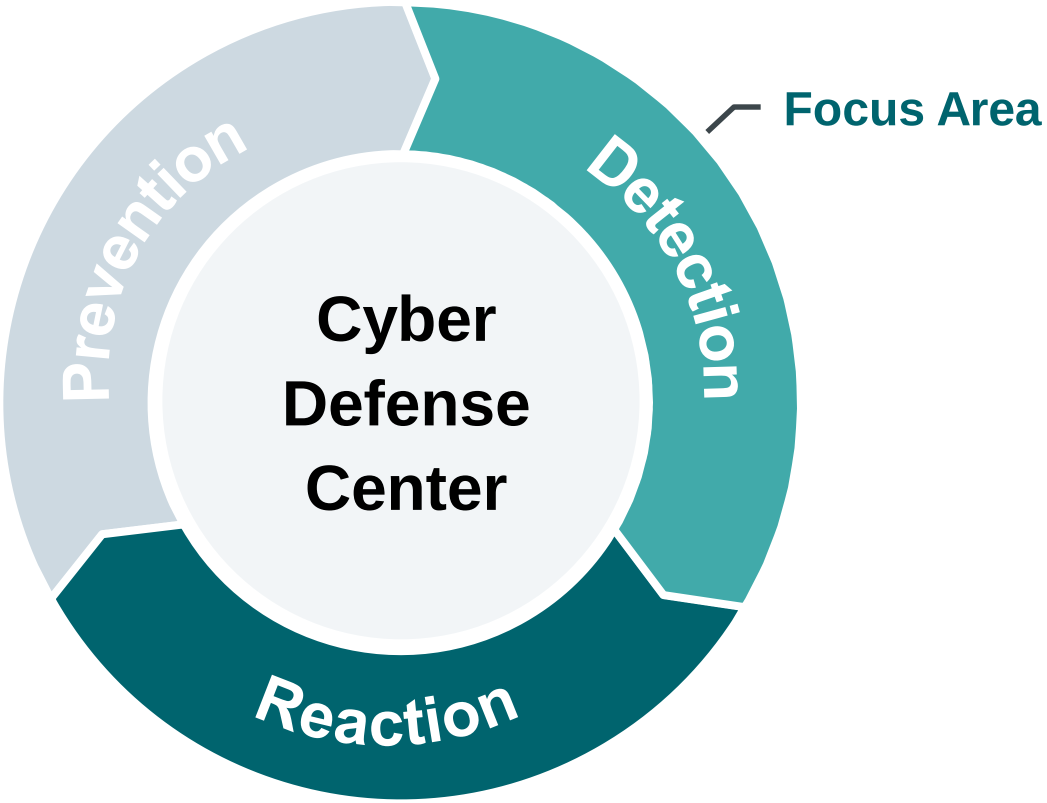
Cyber Defense Center

SIEMENS
Ingenuity for life

Globally distributed Team

Mission

- Monitoring of Siemens infrastructure worldwide
- Identify and analyze security threats



Cyber Defense Center

SIEMENS
Ingenuity for life

Mission: Threat Detection

Log Collection
(Endpoints, Servers, Proxies, AD, Packet Captures, Sandboxes, Email Servers…)

Store, Enrich, Alert

Analyze, Verify, Hunt

Threat Detection System

Storage

Search Tools

Dashboards

Alert

Decision (TP/FP)

Threat Hunting

Siemens Corporate Network
(500,000+ Hosts, 350,000+ Users)

Hybrid Cloud Solution
(50,000+ Events/Sec)

Security Analysts

Cyber Defense Center

SIEMENS
Ingenuity for life

Main Detection Components

Static Analysis

Dynamic and **Hybrid Analysis**

Key motivation for us:

- **More robust detection:** Learn general malware characteristics
- **Enable novel detection vectors**: E.g., image recognition for phishing detection
- **Automate repetitive tasks**

Exact Matching:
Rulesets, Regexes, Periodic Search Queries, Scripts etc.

Machine Learning, Artificial Intelligence and Deep Learning:
Non-exact Matching, Aggregate/Preselect/Visualize etc.

AI and Deep Learning on the Rise

SIEMENS
Ingenuity for life

High expectations due to success stories

Self-driving Cars

Translation

Cancer Detection

AI and Deep Learning on the Rise

SIEMENS
Ingenuity for life

How about Security?

Dedicated Workshops

Large amount of new Papers

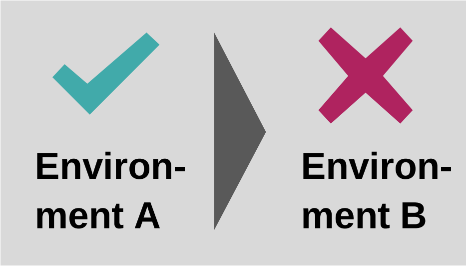
Numerous (Open Source)
Tools and Implementations

AI and Deep Learning on the Rise

SIEMENS
Ingenuity for life

Key Challenges (in Large Environments):

→ Huge gap between research and practice

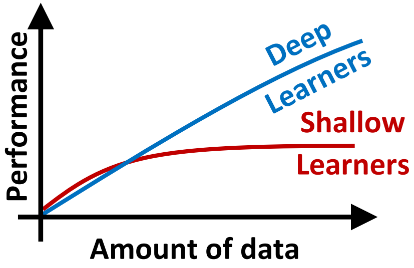


Limited Generalizability

Limited Scalability

Accuracy (TPs, FPs etc)

No Standard Architectures



Not enough (labeled) Data

Technical Challenges

Use Case: DGA Detection

What is a DGA doing?

1 **Malware:** attempt to communicate with Attackers' server

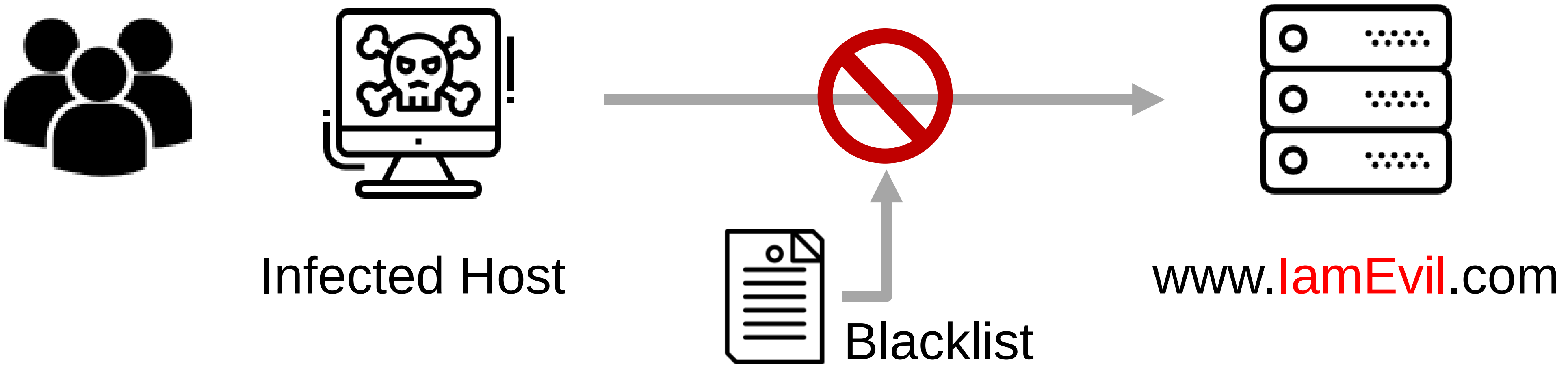
Infected Host

www.IamEvil.com

What is a DGA doing?

SIEMENS
Ingenuity for life

2 **Defenses Up:** blacklist stops the communication



What is a DGA doing?

SIEMENS
Ingenuity for life

3 **DGA in action**

Infected Host

- ⃠ www.IamEvil-0001.com
- ⃠ www.IamEvil-0002.com
- ✓ www.IamEvil-0003.com

What is a DGA doing?

SIEMENS
Ingenuity for life

Simply blocking domains does not scale anymore

A Simple DGA Example

SIEMENS
Ingenuity for life

```python
def generate_domain(year, month, day):
    """Generates a domain name for the given date."""
    domain = ""

    for i in range(16):
        year = ((year ^ 8 * year) >> 11) ^ ((year & 0xFFFFFFF0) << 17)
        month = ((month ^ 4 * month) >> 25) ^ 16 * (month & 0xFFFFFFF8)
        day = ((day ^ (day << 13)) >> 19) ^ ((day & 0xFFFFFFFE) << 12)
        domain += chr(((year ^ month ^ day) % 25) + 97)

    return domain + ".com"
```

[CryptoLocker DGA]

```python
print(generate_domain(2019, 5, 7))
print(generate_domain(2019, 6, 19))
```

```
konsbolyfadifehn.com
myycvfoqtcpbbypd.com
```

Quiz: Can you distinguish Legitimate Domains from Malicious ones?

SIEMENS
Ingenuity for life

edkowalczyk.com

abcdefgtfddf2223.com

xjpakmdcfuqe.nl

b9qmjjys3z.com

watdoejijbijbrand.nl

skhhtcss.edu.hk

reqblcsh.net

blkdmnds.com

oqjiwef12egre6erg6qwefg312qrgqretg132.com

lkckclckl1i1i.com

kdnlrklb.com

cilavocofer.eu

hzmksreiuojy.in

llanfairpwllgwyngyllgogerychwyrndrobwll-llantysiliogogogoch.com

Quiz: Can you distinguish Legitimate Domains from Malicious ones?

SIEMENS
Ingenuity for life

edkowalczyk.com

abcdefgtfddf2223.com

xjpakmdcfuqe.nl

b9qmjjys3z.com

skhhtcss.edu.hk reqblcsh.net

watdoejijbijbrand.nl

blkdmnds.com

oqjiwef12egre6erg6qwefg312qrgqretg132.com

lkckclckl1i1i.com

kdnlrklb.com

cilavocofer.eu

hzmksreiuojy.in

llanfairpwllgwyngyllgogerychwyrndrobwll-llantysiliogogogoch.com

= malicious

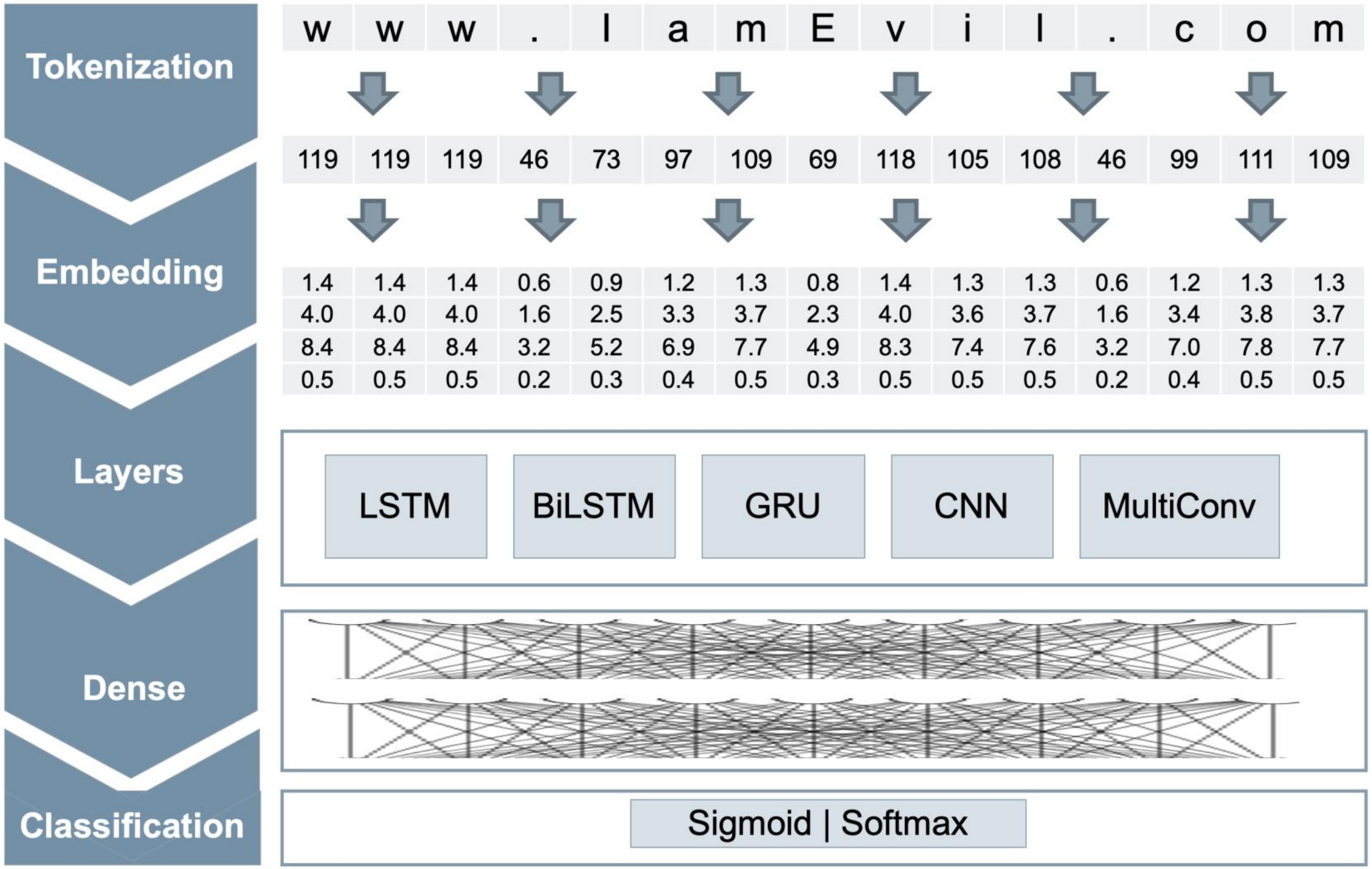
Name of a town in Wales

[https://www.youtube.com/watch?v=fHxO0UdpoxM]

Detecting DGAs with Deep Learners

Detecting DGAs with AI

SIEMENS
Ingenuity for life



1. Characters are converted to ASCII tokens

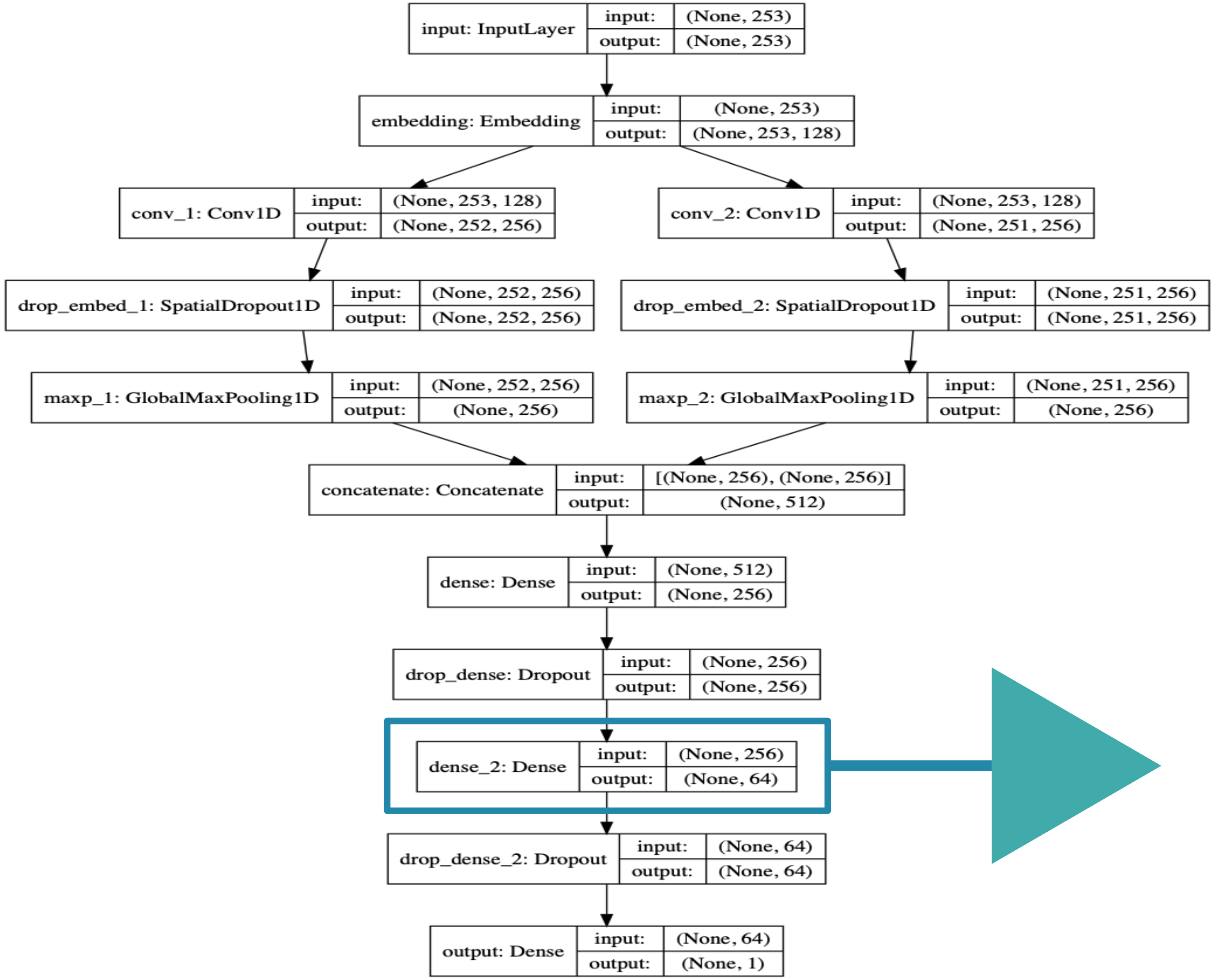
2. Tokens are embedded into multi-dimensional vectors

3. Forward layers or Recurrent layers can be utilized to generate features

4. Fully connected layers can be used to increase the model depth

5. A suspiciousness score is assigned based on the output of sigmoid output neuron or softmax layer

Example CNN Layer with UMAP



domain	label
simonettecaileigh.net	1
kimberleedonaldson.net	1
savingdrivetechnology.com	2
durforcogivprasufor.ru	3
kokkoku-anime.com	normal-traffic
personpermitmountain.com	2

Results

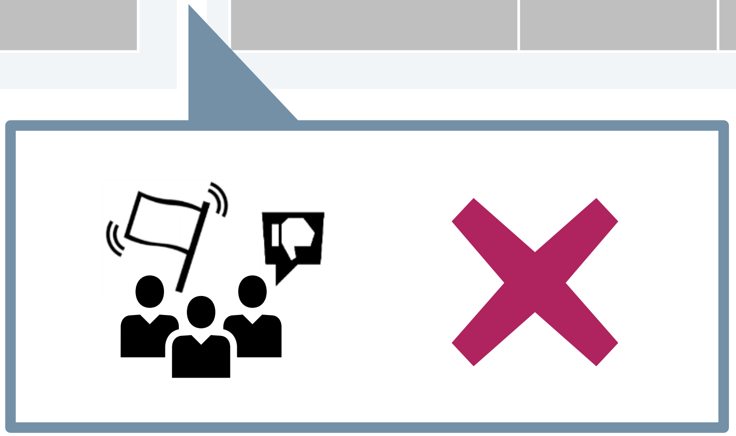
SIEMENS
Ingenuity for life

Deep Learning Approach

Accuracy (%)	TPR (%)	**FPR (%)**	TNR (%)	FNR (%)
98.64	98.08	**0.77**	99.23	1.02

Shallow Learner Approach (

Accuracy (%)	TPR (%)	**FPR (%)**	TNR (%)	FNR (%)
84.36	96.78	**31.56**	68.43	3.21



Design Platform and Operationalize

Operational Challenge

SIEMENS
Ingenuity for life

500,000+ Hosts

50,000+ Events per second

6+ TBs of data per day

24/7 Operations

Highly Volatile Loads (20x)

Operationalize Smart!

SIEMENS
Ingenuity for life

500,000+ Hosts

50,000+ Events per second

6+ TBs of data per day

24/7 Operations

Highly Volatile Loads (20x)

Don't burden your team with

Auto Scaling (Elasticity)

Auto Failover

Server Patching

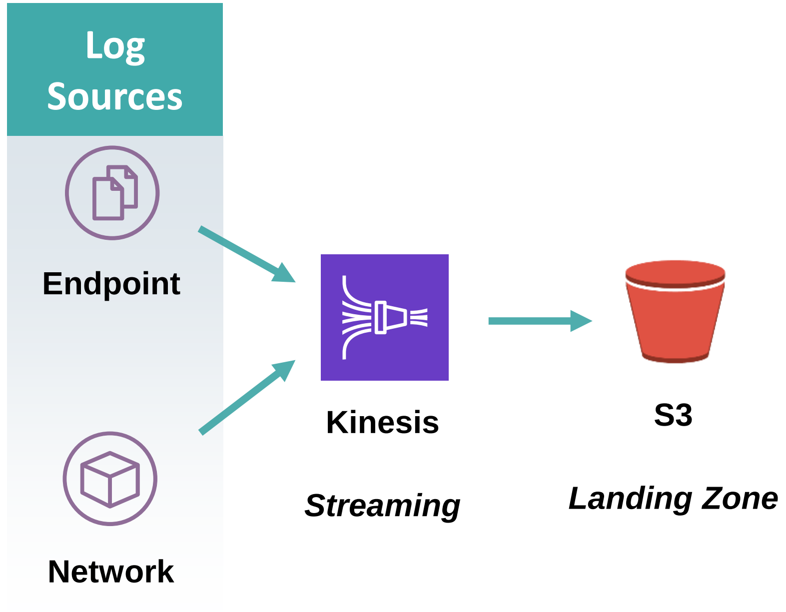
Backups

Go Serverless

Important Pipelines

Ingest and Inference

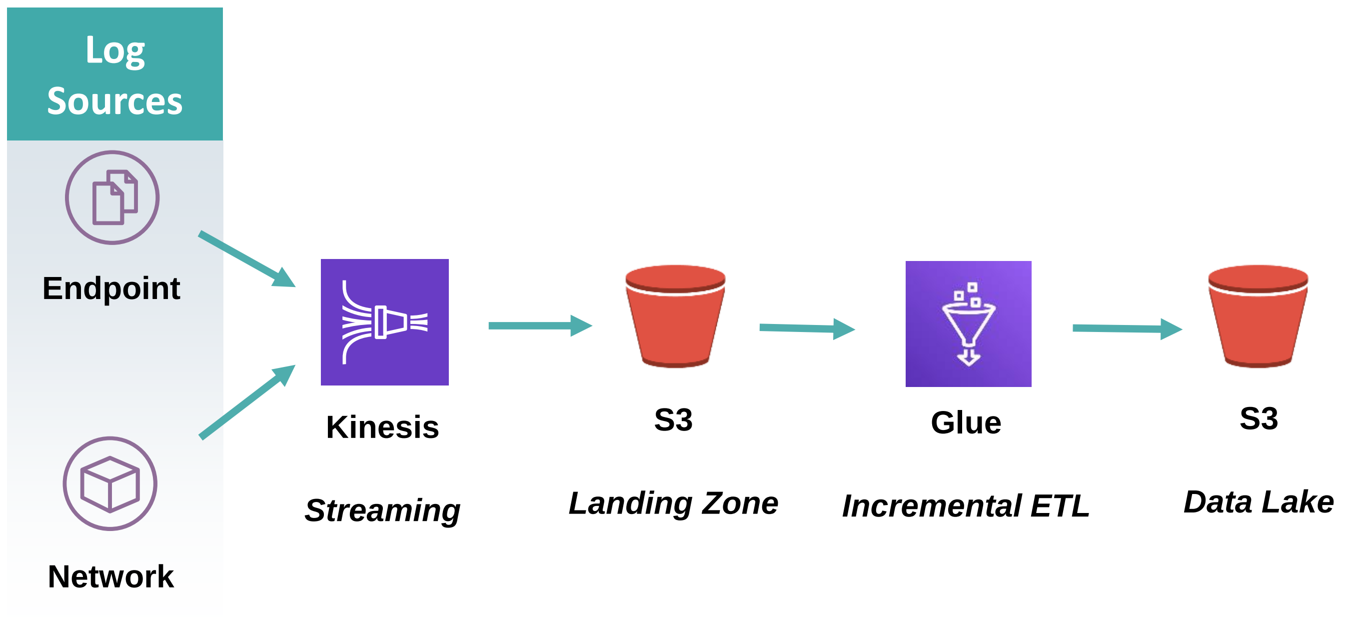
1 Ingest Pipeline: Store AD- Proxy- Email-Logs into the S3



Important Pipelines

Ingest and Inference

2 **ETL: Cleaning and Transforming Data**

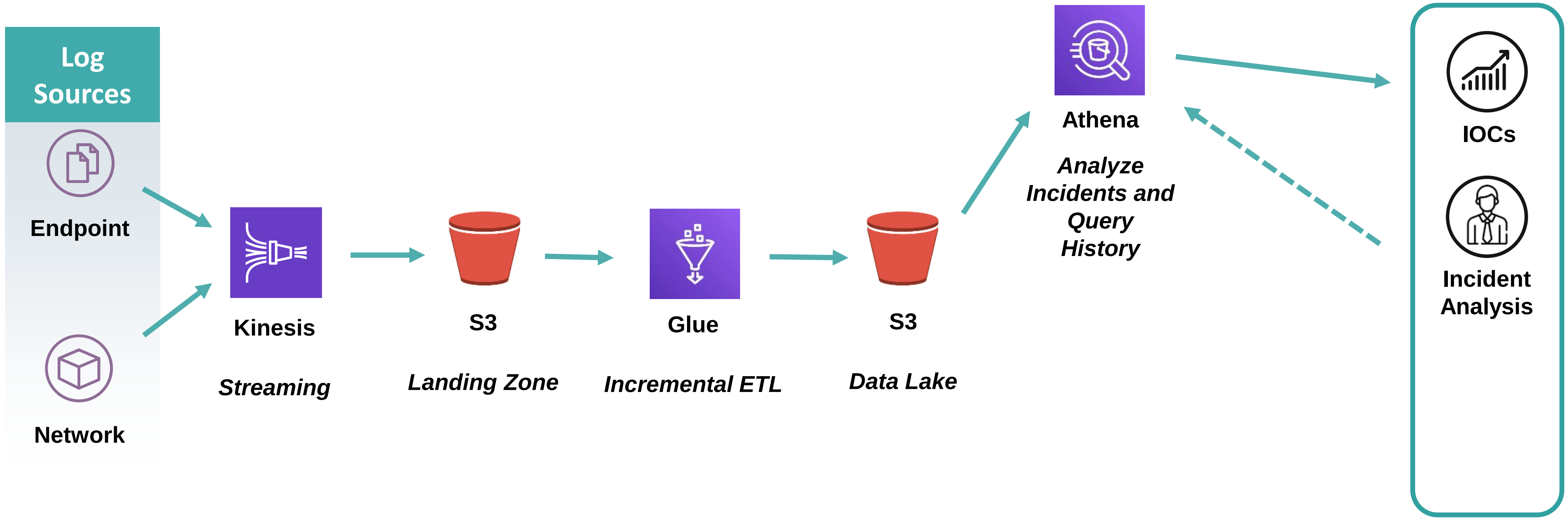


Important Pipelines

SIEMENS
Ingenuity for life

Ingest and Inference

3 Presentation: Create Statistics, Provide Data to Analysts

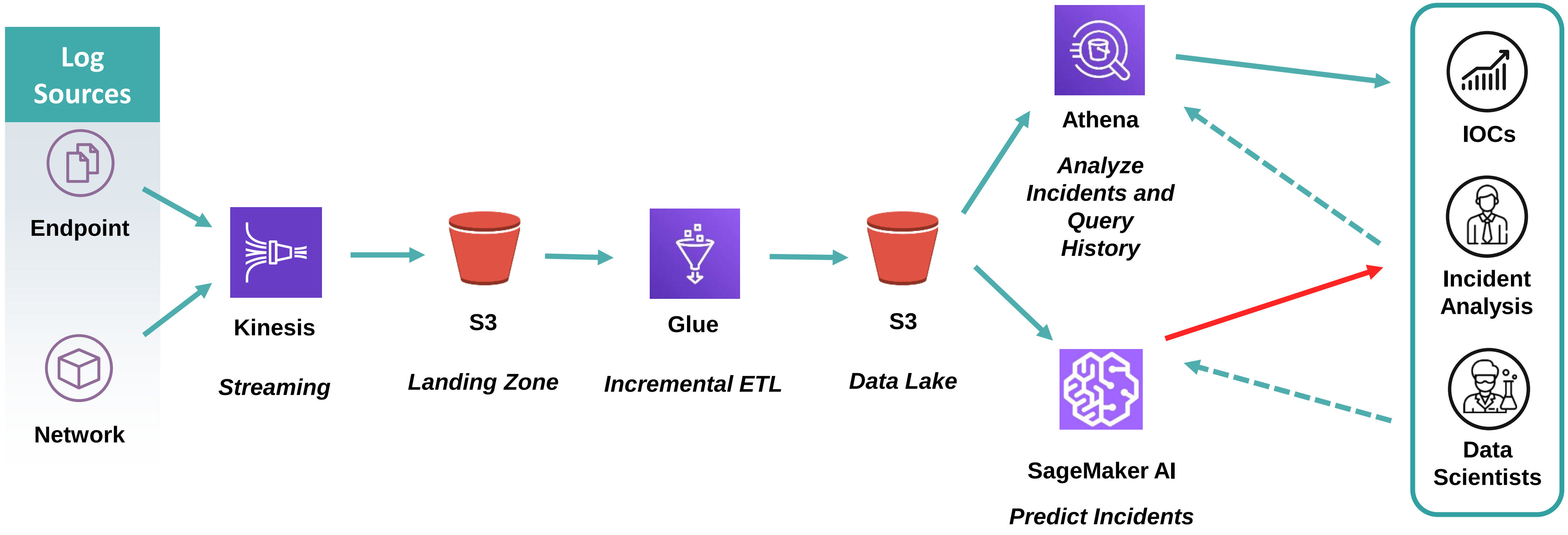


Important Pipelines

SIEMENS
Ingenuity for life

Ingest and Inference

4 **Detection: Real Time Prediction of Threats**

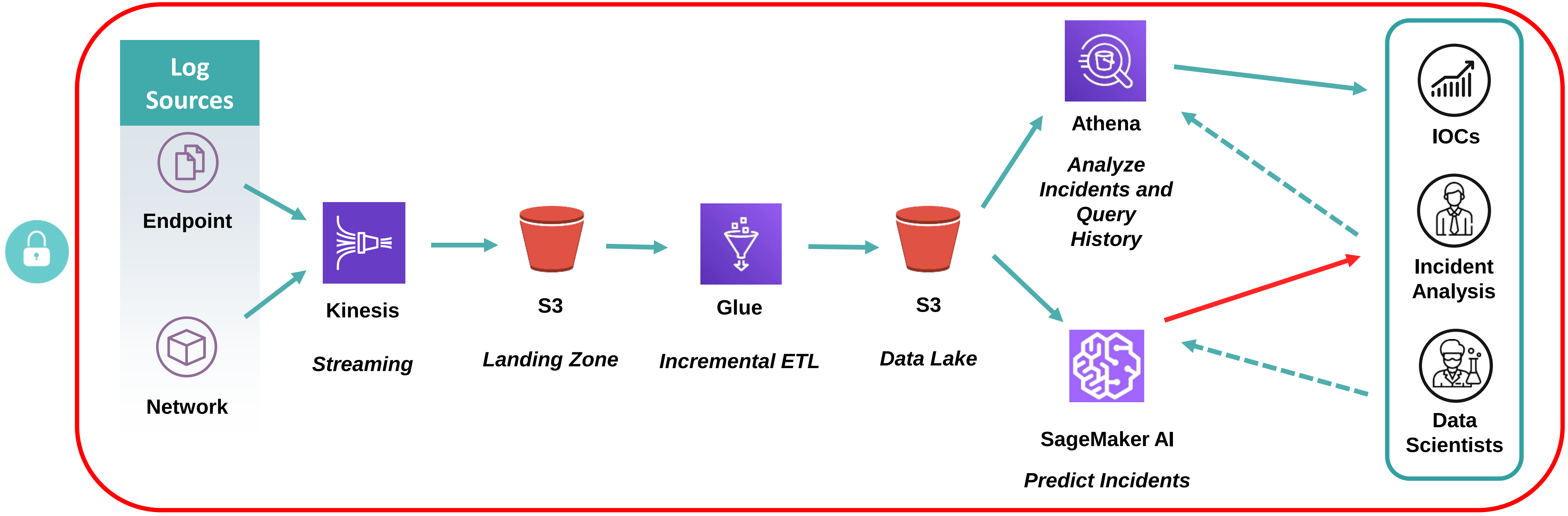


Important Pipelines

Ingest and Inference

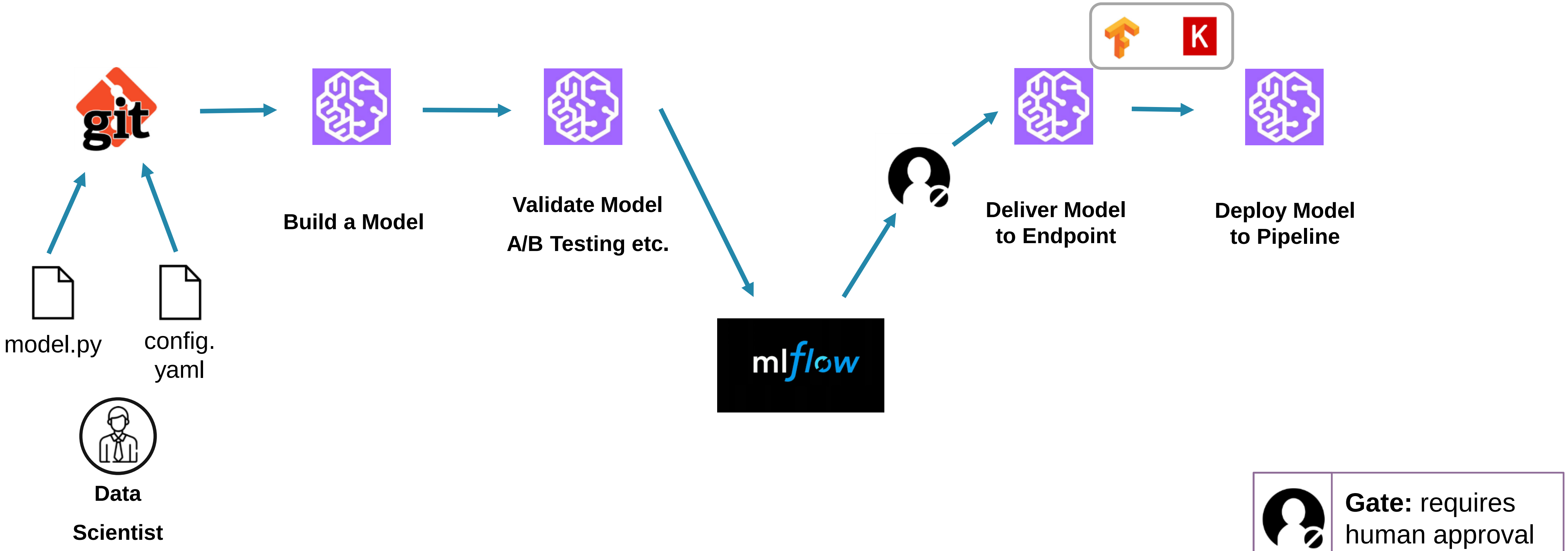
SIEMENS
Ingenuity for life

4 Detection: Real Time Prediction of Threats



Important Pipelines

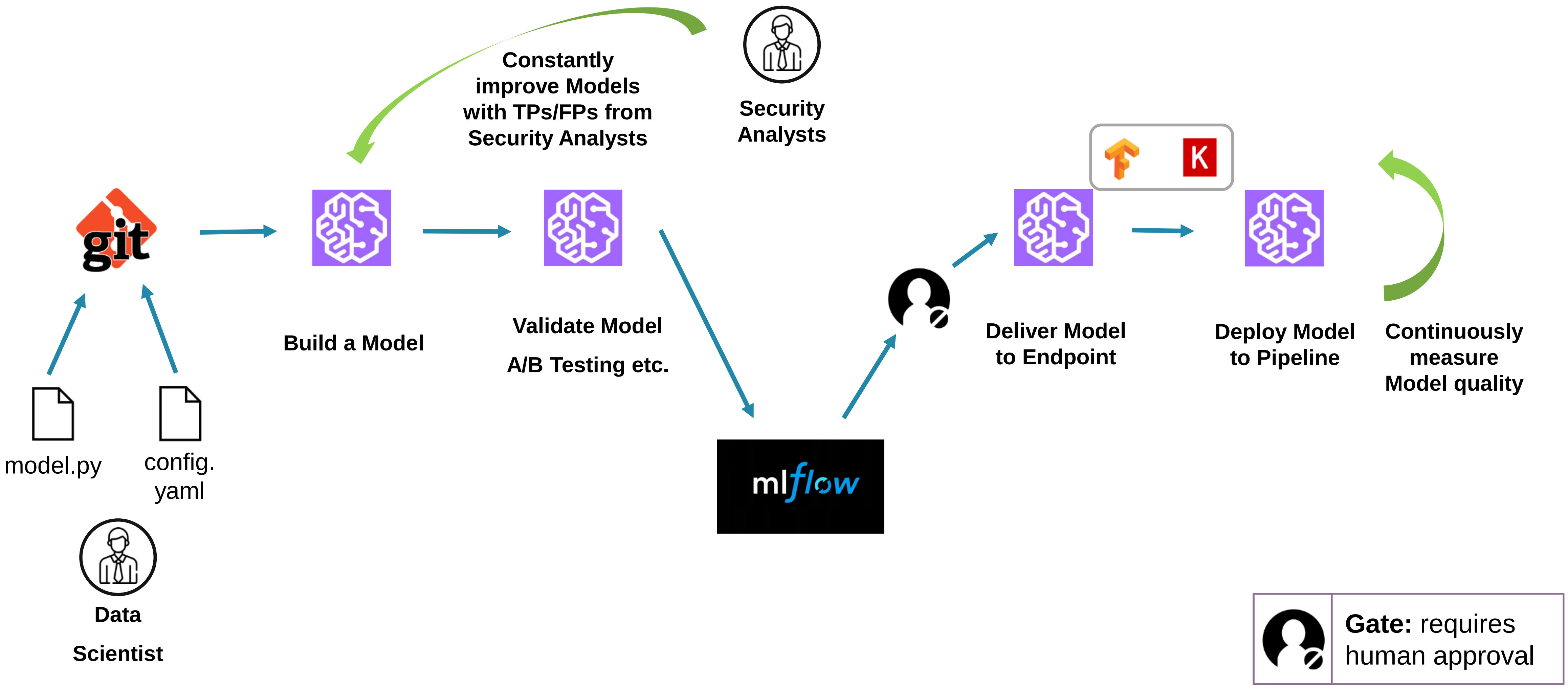
AI Model Generation and Deployment



Important Pipelines

AI Feedback Loop

SIEMENS
Ingenuity for life



Thank You