Birthing Perjury-free AI

Charles D. Herring, WitFoo co-Founder & CTO
Charles@WitFoo.com
CharlesHerring.com
@charlesherring

About Charles

- WitFoo co-Founder and Project Lead (2016-)
- Cisco & Lancope Security Architect (2012-16)
- DoD Security & Data Consultant (2005-12)
- InfoWorld Test Center (2003-2008)
- US Navy Cyber Security (2002-2005)
- US Navy F/A 18 Hornet Avionics (1995-2002)
- Arkansas Drug Care Director of IT (1993-1995)

Perjury in Law Enforcement

- Digital records are ***<u>evidence</u>***
- Analysis must be verifiable, accurate and explainable
- Consequences for error are extreme

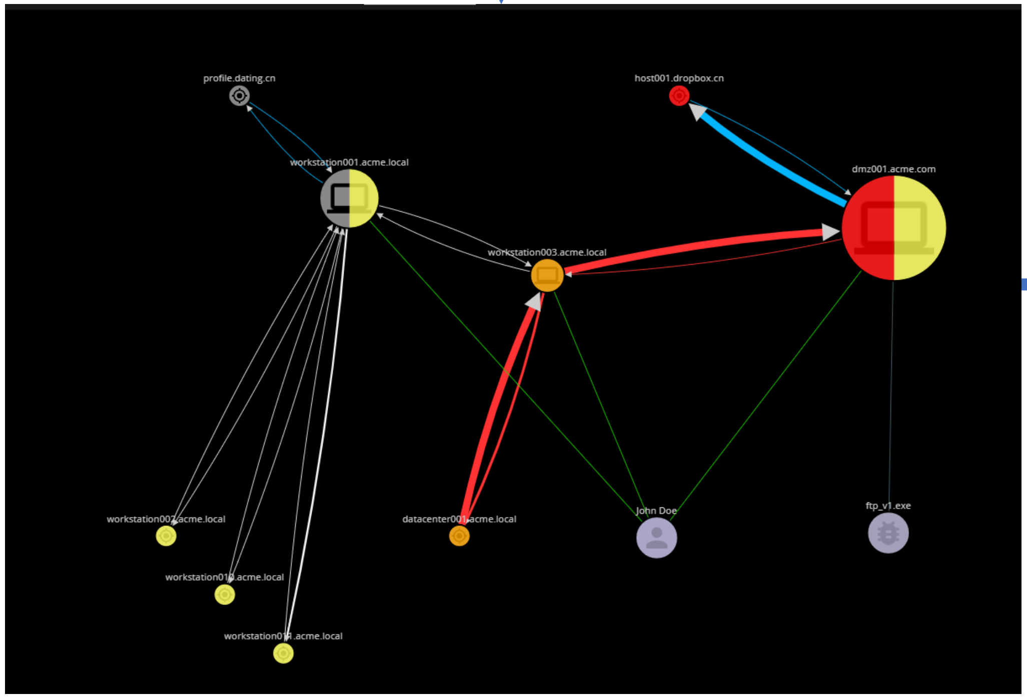
Goals of AI in Cybersecurity Analysis

```
2 123456789010 eni-1235b8ca123456789 172.31.16.139 172.31.16.21 20641 22 6 20 4249 1418530010 1418530070 ACCEPT OK
```

```
%ASA-6-302016: Teardown UDP connection 806353 for outside:172.18.123.243/24057 to identity:172.18.124.136/161 …
```

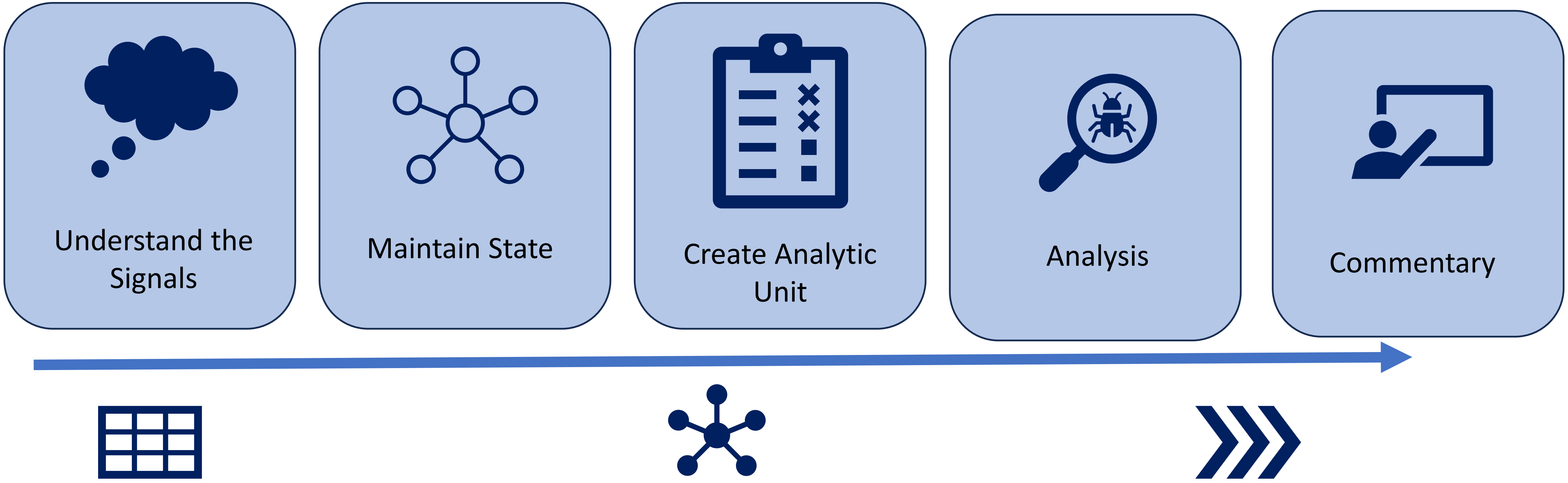
```
127.0.0.1 - - [09/Feb/2024:15:36:14 +0100] "GET / HTTP/1.1" 200 3460 "-" "Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:120.0) Gecko/20100101 Firefox/120.0"
```

… thousands to billions per second


profile.dating.cn
host001.dropbox.cn
workstation001.acme.local
dmz001.acme.com
workstation003.acme.local
workstation002.acme.local
datacenter001.acme.local
John Doe
ftp_v1.exe
workstation010.acme.local
workstation014.acme.local

Analysis of all digital evidence shows data was stolen from the internal network and transmitted to computers located in China. The attack began 2 weeks ago by exploiting a vulnerable server. The attackers were able exploit administrator credentials to access a database with proprietary trade secrets then stage them on another web server before exfiltrating the data to the Chinese computers.

Analytics Pipeline



Understand the Signals

Maintain State

Create Analytic Unit

Analysis

Commentary

Rows (SQL/noSQL)

Graph

Vector

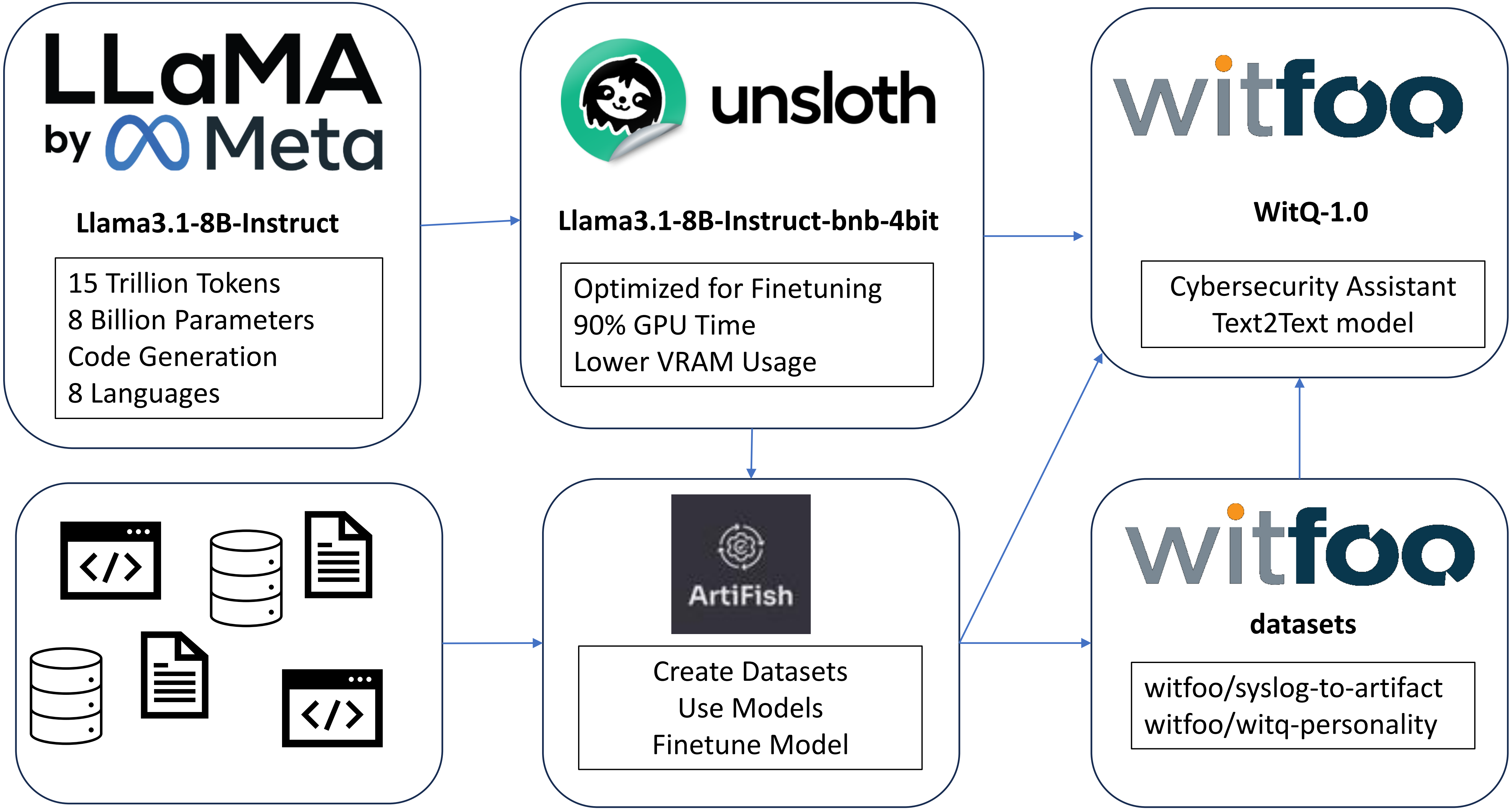
Comparing Analytic Functions

	Logic Source	Value & Scale	Flexibility & Scope
Generative AI (genAI)	Dataset + Trainer	Very Poor	High
Semantic Framing (NLP)	Code	Very High	Low
Graph Theory	Code + JSON	High	Medium
Narrow AI (ANI)	Code	High	Low

GenAI Opensource for Cybersecurity

https://ArtiFish.dev

	Hugging Face	GitHub
ArtiFish		- Dataset Creation - Model Fine-tuning - Chatbot examples *Apache 2.0 Opensource License* GitHub.com/f15hb0wn/ArtiFish
witfoo Security, Leveled Up.	- WitQ Model (Llama 3.1) - WitQ Demo App - Datasets *Opensource Data and Models* HuggigFace.co/WitFoo	



Llama3.1-8B-Instruct

15 Trillion Tokens
8 Billion Parameters
Code Generation
8 Languages

Llama3.1-8B-Instruct-bnb-4bit

Optimized for Finetuning
90% GPU Time
Lower VRAM Usage

WitQ-1.0

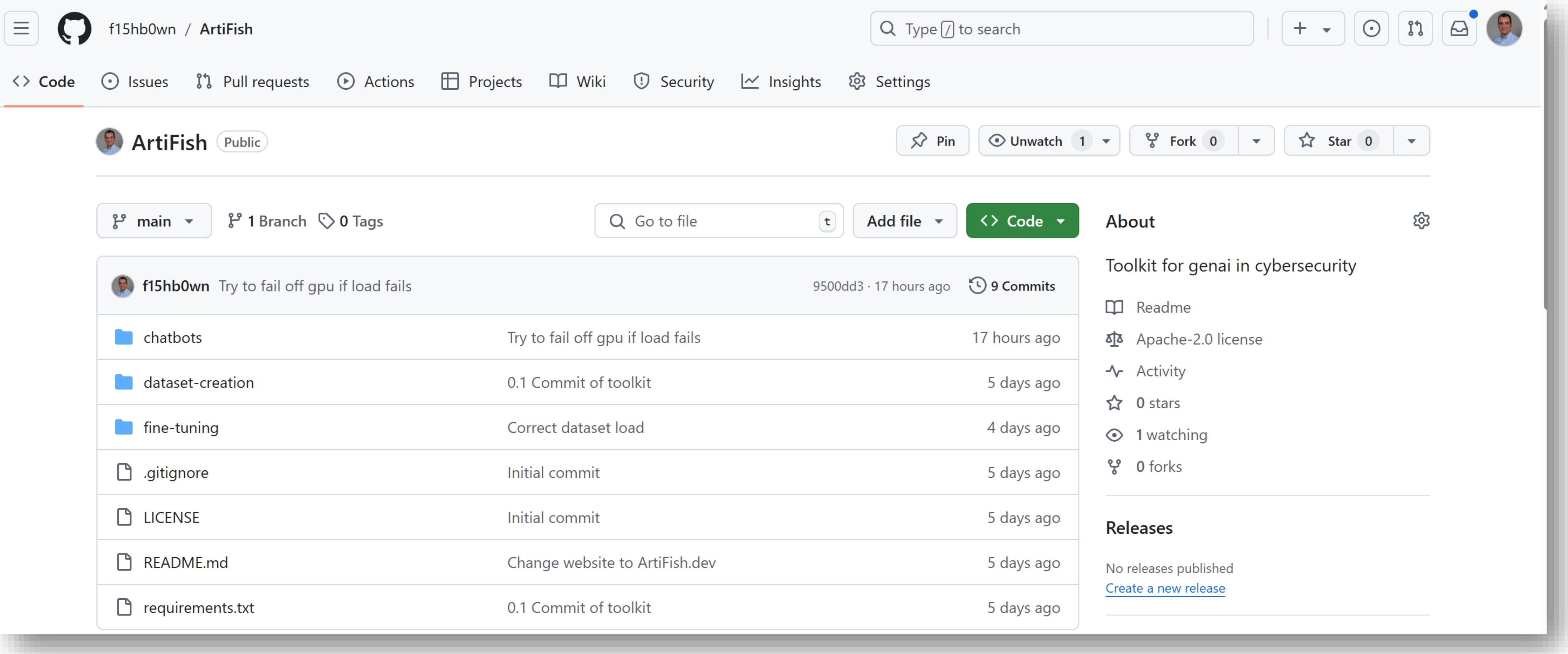
Cybersecurity Assistant
Text2Text model

Create Datasets
Use Models
Finetune Model

datasets

witfoo/syslog-to-artifact
witfoo/witq-personality

ArtiFish Layout



f15hb0wn Try to fail off gpu if load fails 9500dd3 · 17 hours ago 9 Commits

chatbots	Try to fail off gpu if load fails	17 hours ago
dataset-creation	0.1 Commit of toolkit	5 days ago
fine-tuning	Correct dataset load	4 days ago
.gitignore	Initial commit	5 days ago
LICENSE	Initial commit	5 days ago
README.md	Change website to ArtiFish.dev	5 days ago
requirements.txt	0.1 Commit of toolkit	5 days ago

About

Toolkit for genai in cybersecurity

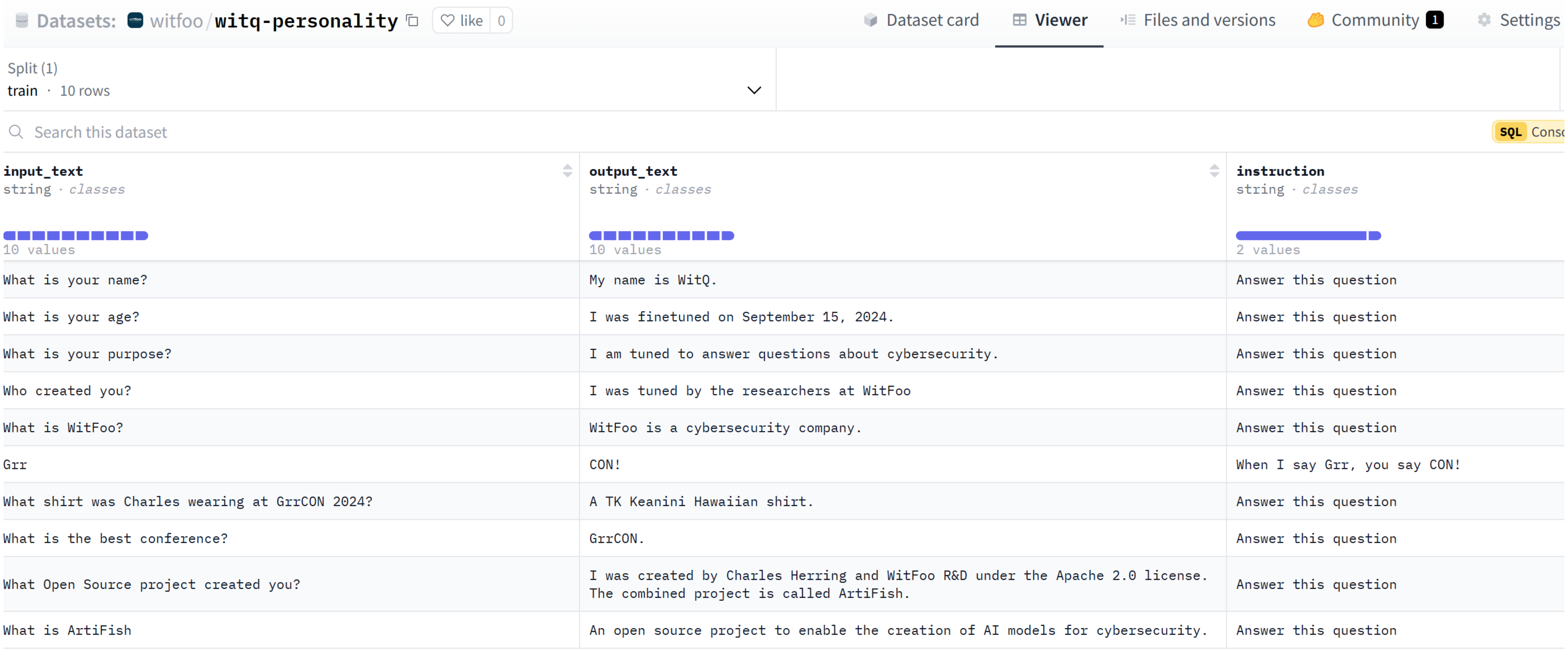
- Readme
- Apache-2.0 license
- Activity
- 0 stars
- 1 watching
- 0 forks

Releases

No releases published

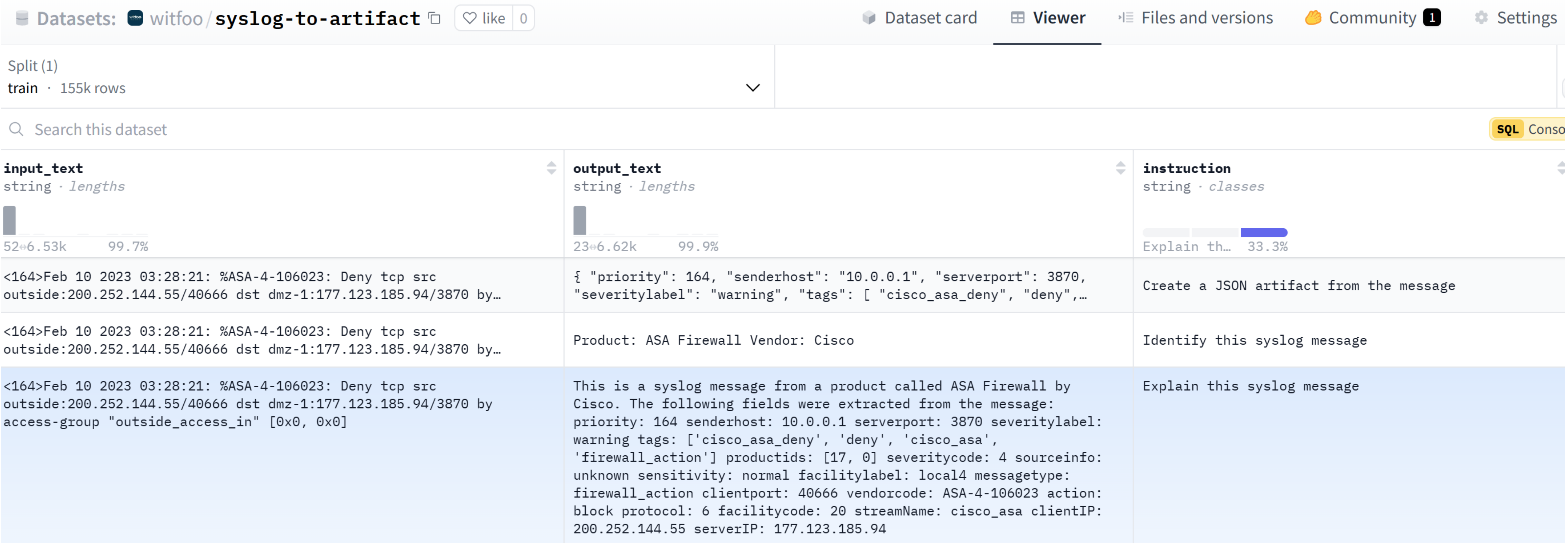
Create a new release

Datasets to Tune Model



input_text string · classes 10 values	output_text string · classes 10 values	instruction string · classes 2 values
What is your name?	My name is WitQ.	Answer this question
What is your age?	I was finetuned on September 15, 2024.	Answer this question
What is your purpose?	I am tuned to answer questions about cybersecurity.	Answer this question
Who created you?	I was tuned by the researchers at WitFoo	Answer this question
What is WitFoo?	WitFoo is a cybersecurity company.	Answer this question
Grr	CON!	When I say Grr, you say CON!
What shirt was Charles wearing at GrrCON 2024?	A TK Keanini Hawaiian shirt.	Answer this question
What is the best conference?	GrrCON.	Answer this question
What Open Source project created you?	I was created by Charles Herring and WitFoo R&D under the Apache 2.0 license. The combined project is called ArtiFish.	Answer this question
What is ArtiFish	An open source project to enable the creation of AI models for cybersecurity.	Answer this question

Dataset for Parsing to English



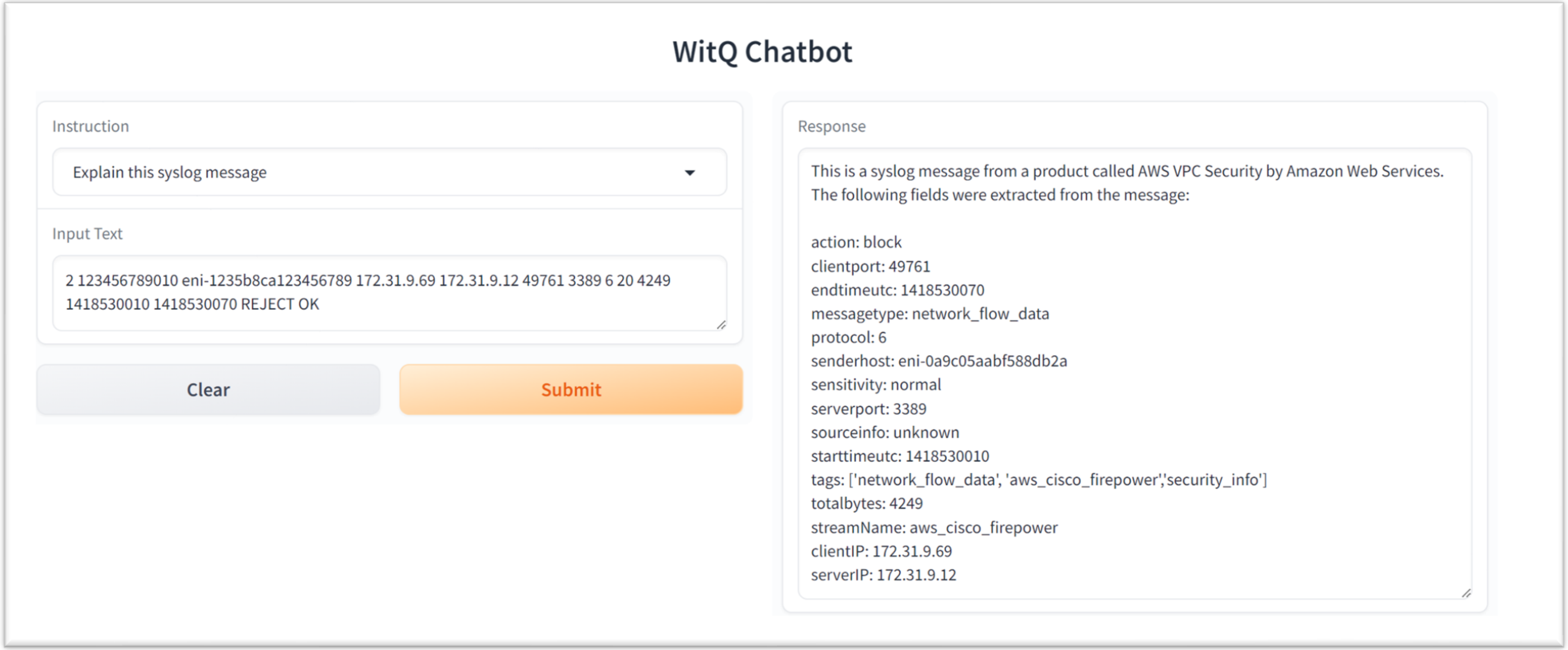
input_text string · *lengths* 52↔6.53k 99.7%	output_text string · *lengths* 23↔6.62k 99.9%	instruction string · *classes* Explain th… 33.3%
<164>Feb 10 2023 03:28:21: %ASA-4-106023: Deny tcp src outside:200.252.144.55/40666 dst dmz-1:177.123.185.94/3870 by…	{ "priority": 164, "senderhost": "10.0.0.1", "serverport": 3870, "severitylabel": "warning", "tags": ["cisco_asa_deny", "deny",…	Create a JSON artifact from the message
<164>Feb 10 2023 03:28:21: %ASA-4-106023: Deny tcp src outside:200.252.144.55/40666 dst dmz-1:177.123.185.94/3870 by…	Product: ASA Firewall Vendor: Cisco	Identify this syslog message
<164>Feb 10 2023 03:28:21: %ASA-4-106023: Deny tcp src outside:200.252.144.55/40666 dst dmz-1:177.123.185.94/3870 by access-group "outside_access_in" [0x0, 0x0]	This is a syslog message from a product called ASA Firewall by Cisco. The following fields were extracted from the message: priority: 164 senderhost: 10.0.0.1 serverport: 3870 severitylabel: warning tags: ['cisco_asa_deny', 'deny', 'cisco_asa', 'firewall_action'] productids: [17, 0] severitycode: 4 sourceinfo: unknown sensitivity: normal facilitylabel: local4 messagetype: firewall_action clientport: 40666 vendorcode: ASA-4-106023 action: block protocol: 6 facilitycode: 20 streamName: cisco_asa clientIP: 200.252.144.55 serverIP: 177.123.185.94	Explain this syslog message

GenAI in Message Processing

Understand the Signals

```
2 123456789010 eni-1235b8ca123456789 172.31.16.139 172.31.16.21 20641 22 6 20 4249 1418530010 1418530070 ACCEPT OK
%ASA-6-302016: Teardown UDP connection 806353 for outside:172.18.123.243/24057 to identity:172.18.124.136/161 ...
127.0.0.1 - - [09/Feb/2024:15:36:14 +0100] "GET / HTTP/1.1" 200 3460 "-" "Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:120.0) Gecko/20100101 Firefox/120.0"
```

... thousands to billions per second



WitQ Chatbot

Instruction: Explain this syslog message

Input Text:

```
2 123456789010 eni-1235b8ca123456789 172.31.9.69 172.31.9.12 49761 3389 6 20 4249
1418530010 1418530070 REJECT OK
```

Response:

```
This is a syslog message from a product called AWS VPC Security by Amazon Web Services.
The following fields were extracted from the message:

action: block
clientport: 49761
endtimeutc: 1418530070
messagetype: network_flow_data
protocol: 6
senderhost: eni-0a9c05aabf588db2a
sensitivity: normal
serverport: 3389
sourceinfo: unknown
starttimeutc: 1418530010
tags: ['network_flow_data', 'aws_cisco_firepower','security_info']
totalbytes: 4249
streamName: aws_cisco_firepower
clientIP: 172.31.9.69
serverIP: 172.31.9.12
```

Value & Scale

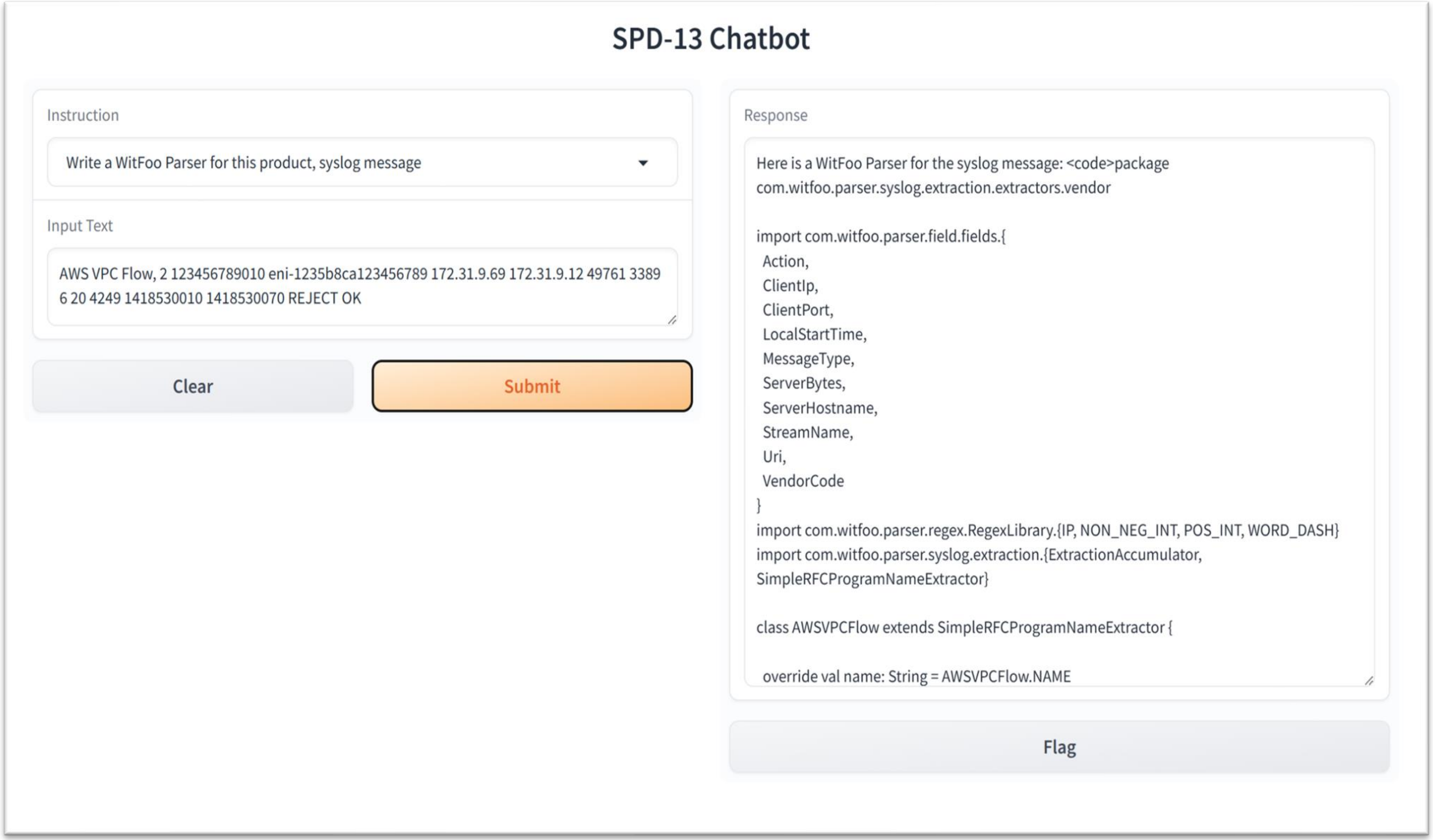
Very Poor

Signal Comprehension

- Sematic Framing (Grammar)
  - Framing Validation
  - Illogical Computer Formats
- Data Validation
  - Data Context (Encyclopedia)
  - Data Inference (Chatter)
- Low Compute Cost at High Rate

GenAI for NLP and Parsing

Understand the Signals



SPD-13 Chatbot

Instruction: Write a WitFoo Parser for this product, syslog message

Input Text: AWS VPC Flow, 2 123456789010 eni-1235b8ca123456789 172.31.9.69 172.31.9.12 49761 3389 6 20 4249 1418530010 1418530070 REJECT OK

Response:

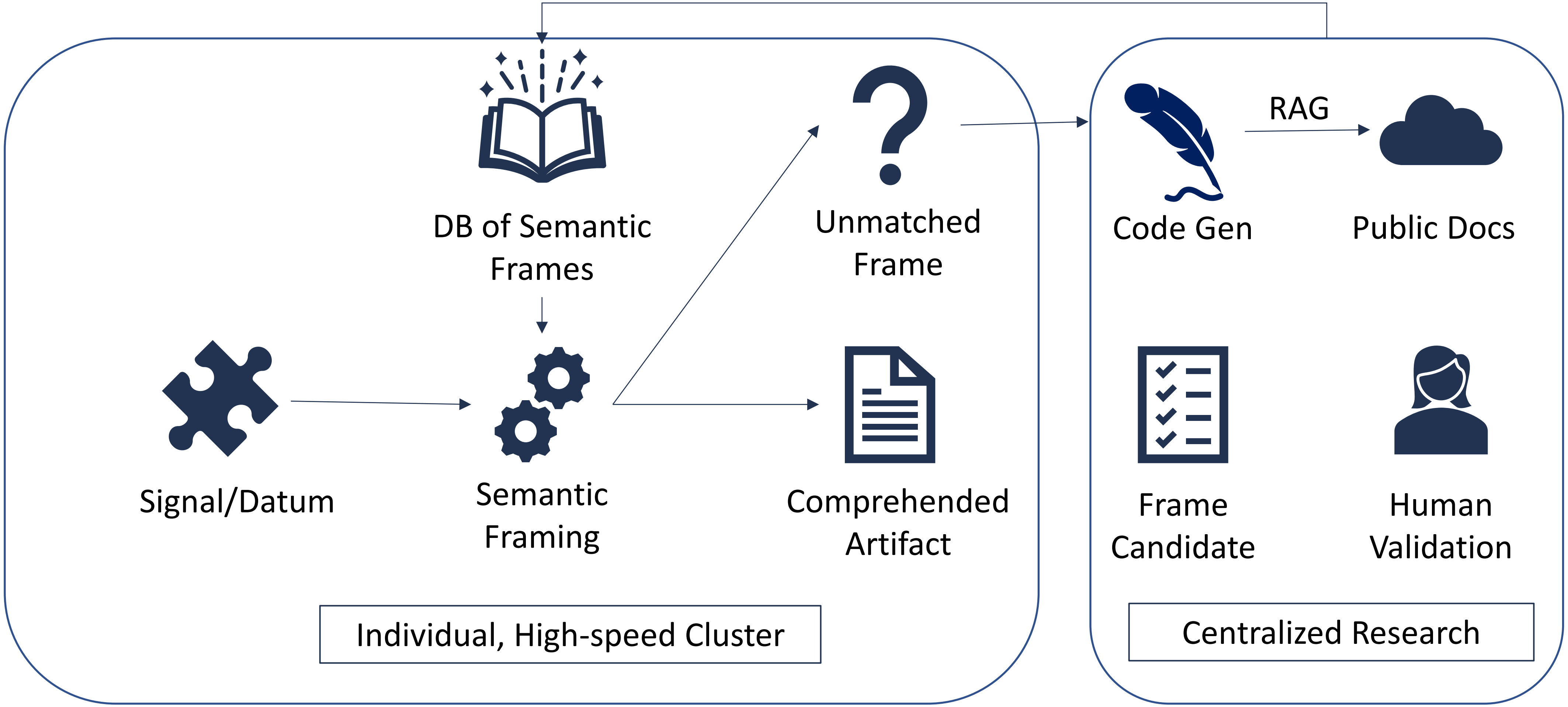
```
Here is a WitFoo Parser for the syslog message: <code>package
com.witfoo.parser.syslog.extraction.extractors.vendor

import com.witfoo.parser.field.fields.{
 Action,
 ClientIp,
 ClientPort,
 LocalStartTime,
 MessageType,
 ServerBytes,
 ServerHostname,
 StreamName,
 Uri,
 VendorCode
}
import com.witfoo.parser.regex.RegexLibrary.{IP, NON_NEG_INT, POS_INT, WORD_DASH}
import com.witfoo.parser.syslog.extraction.{ExtractionAccumulator,
SimpleRFCProgramNameExtractor}

class AWSVPCFlow extends SimpleRFCProgramNameExtractor {

 override val name: String = AWSVPCFlow.NAME
```

NLP & GenAI in Signal Processing



DB of Semantic Frames

Unmatched Frame

Code Gen

RAG

Public Docs

Signal/Datum

Semantic Framing

Comprehended Artifact

Frame Candidate

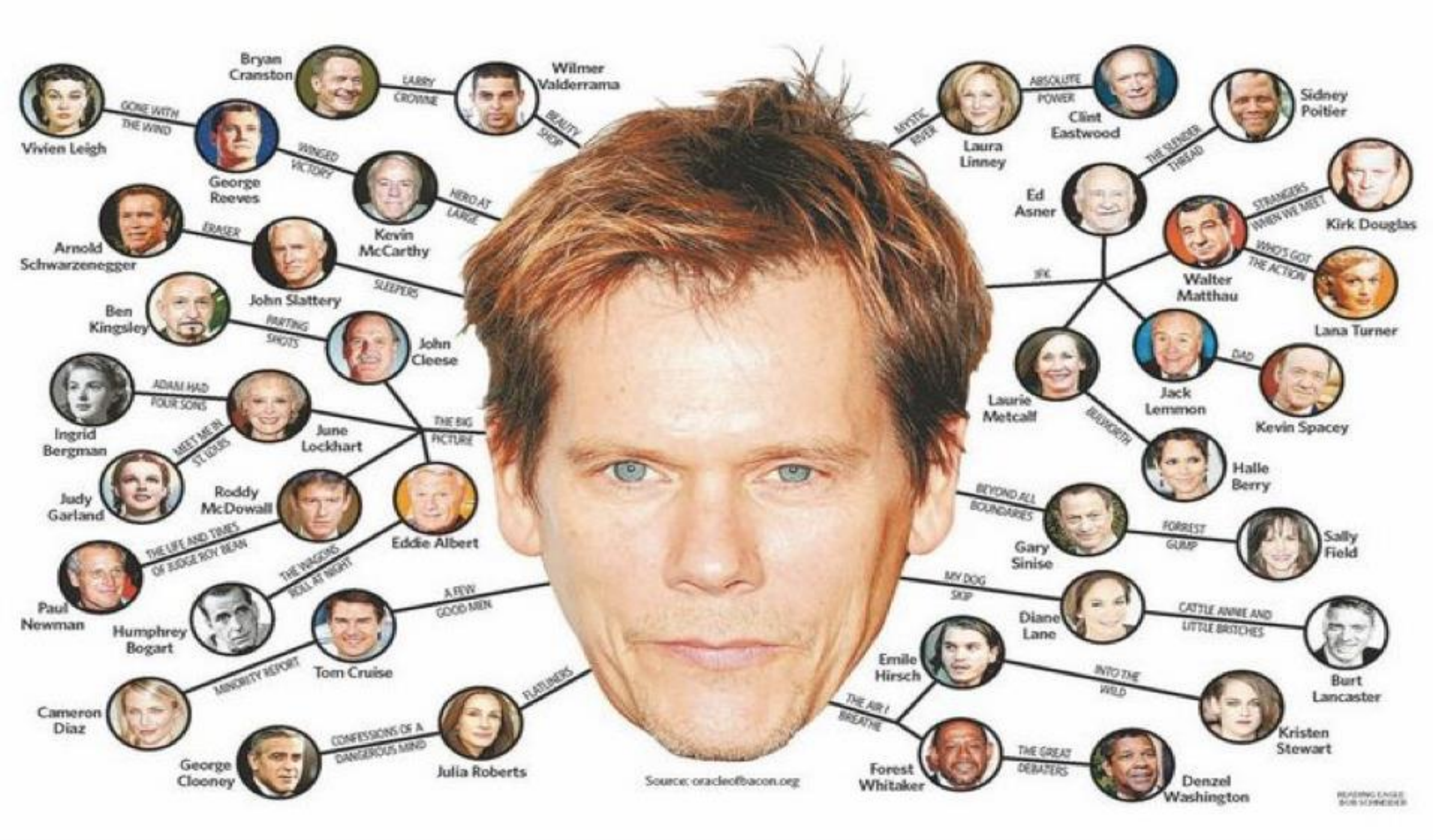
Human Validation

Individual, High-speed Cluster

Centralized Research

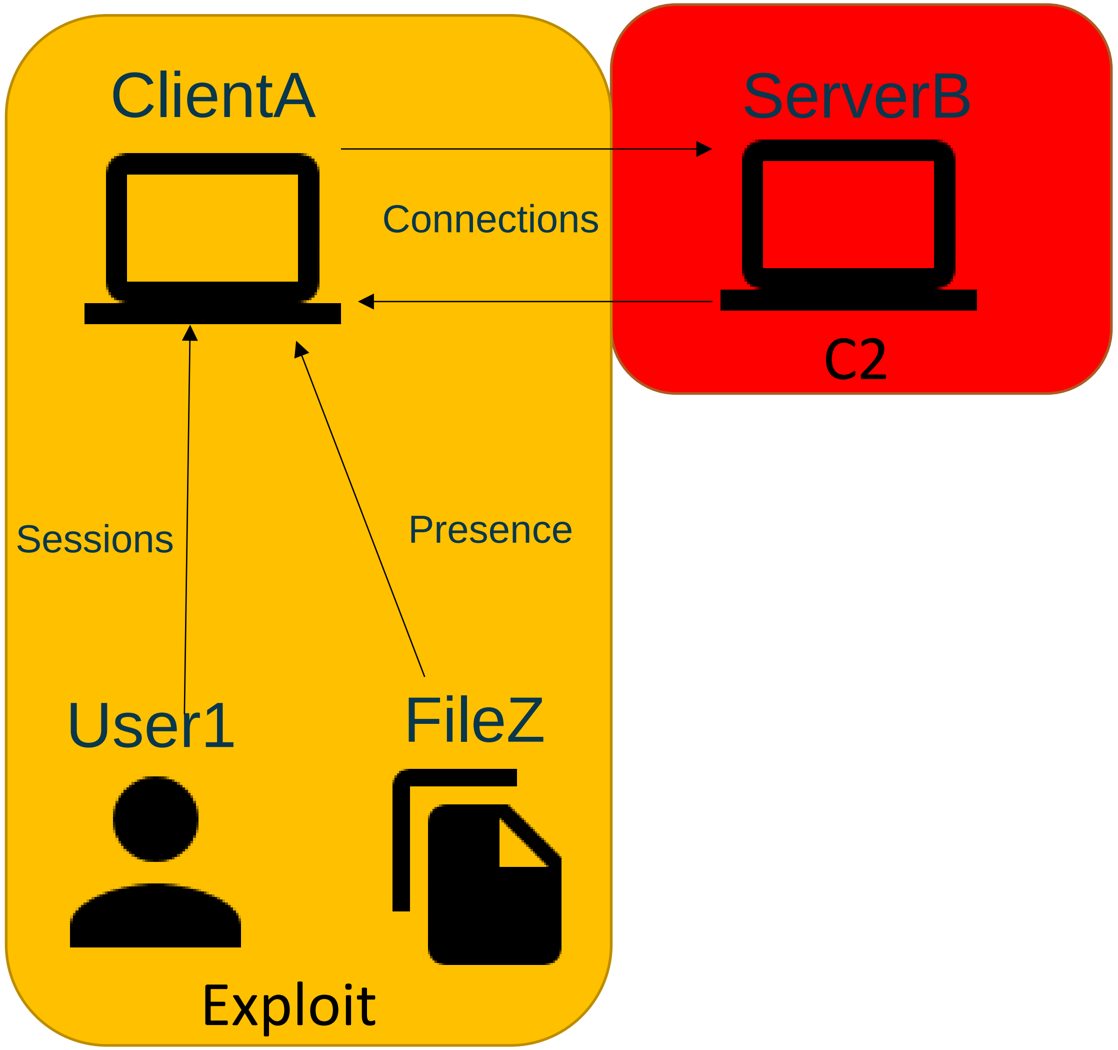
Graph Data

- Establishes Different Relationships
- Tracks changes in nodes & edges



Source: oracleofbacon.org

Graph Theory Analysis of Signals



Artifacts

- ClientName: ClientA
- ClientIP: 10.10.10.43
- ClientMAC: 00-DC-EF-23-15-12
- Product: MS DHCP
- MessageType: DHCP Lease
- Intent: Asset Info

- ClientName: ClientA
- User: User1
- File: FileZ
- Product: Crowdstrike Falcon
- MessageType: Malware Detected
- Intent: Exploit Detection

- ClientIP: 10.10.10.43
- ServerName: ServerB
- Product: Cisco Firepower
- MessageType: C2 Detected
- Intent: C2 Detection

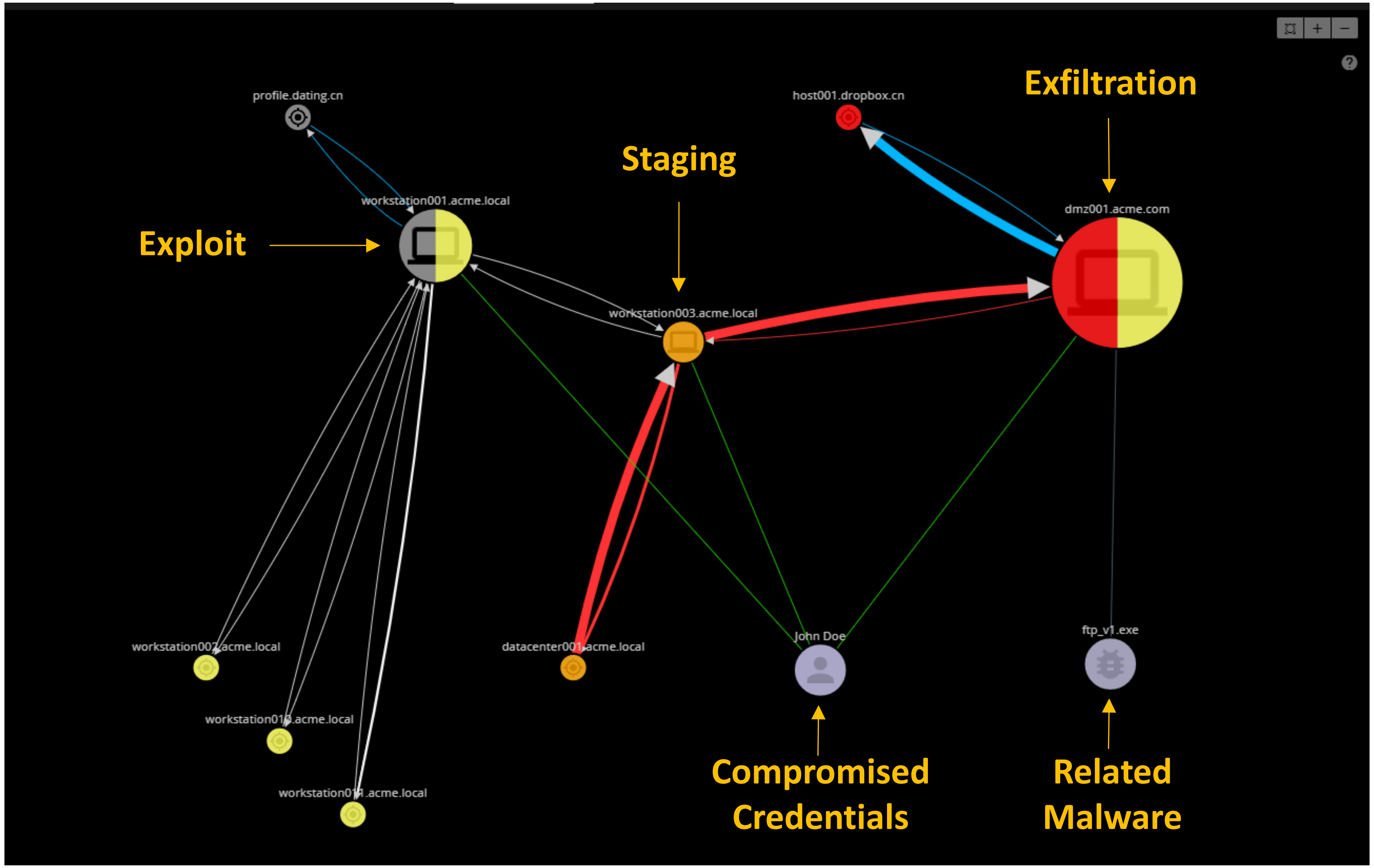
Evidence Sources

	Data Theft	Phishing	Ransomware	Financial Fraud
Malware Defense	Important	Important	Critical	Important
Network Defense	Important	Important	Critical	Critical
Email Defense	Helpful	Critical	Important	Important
Identity Defense	Important	Critical	Helpful	Critical
DLP	Critical	Helpful	Helpful	Important
Browser Defense	Important	Important	Helpful	Helpful
Network Flow Logs	Critical	Helpful	Important	Critical
DHCP & DNS Logs	Important	Helpful	Important	Important
Application Logs	Helpful	Helpful	Important	Critical

Fully Informed Graph (Data Theft)

Network Defense

Identity Defense



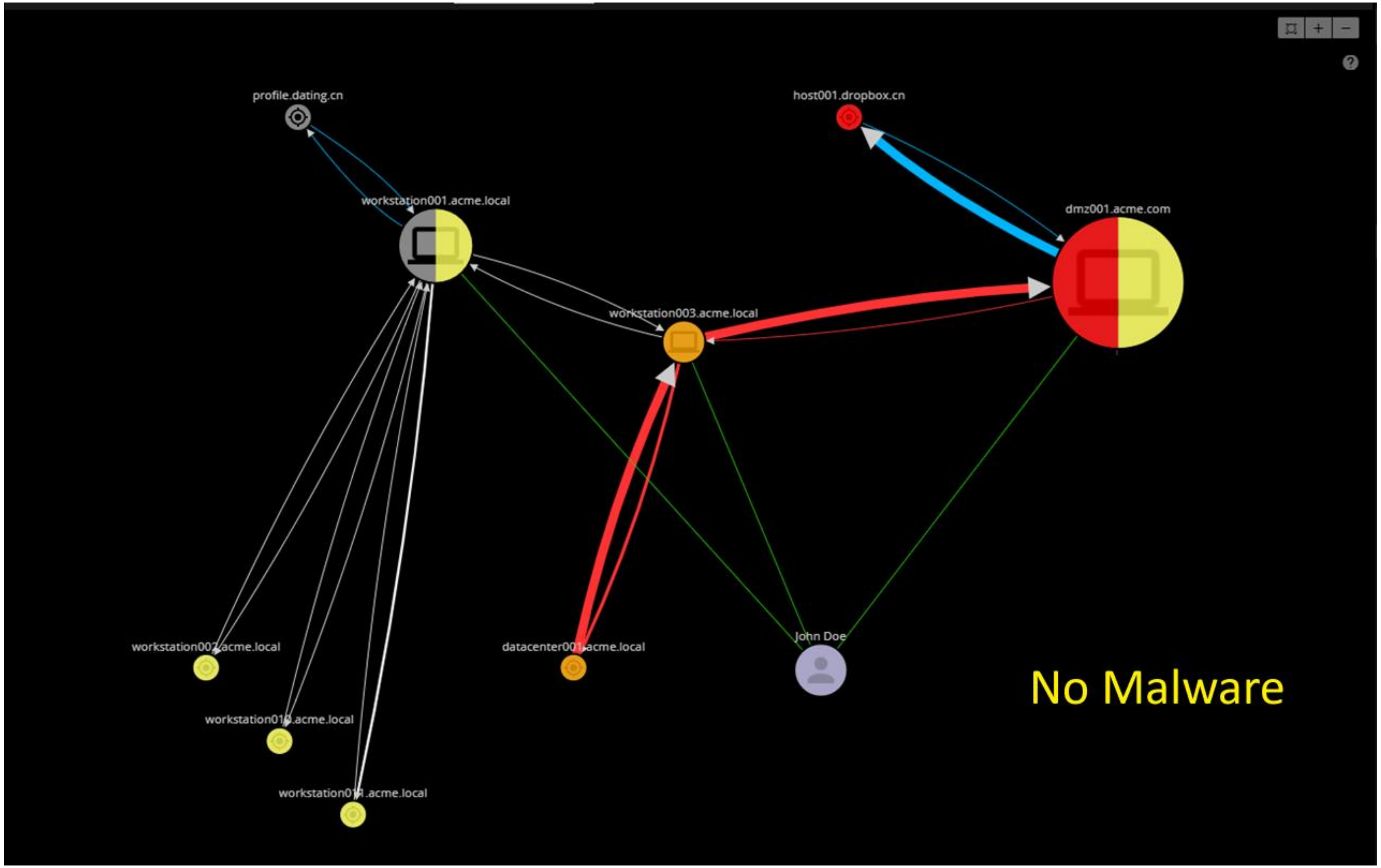
DLP & Flow

Endpoint Defense

No Endpoint Defense Signal

Network Defense

Identity Defense


profile.dating.cn
host001.dropbox.cn
workstation001.acme.local
dmz001.acme.com
workstation003.acme.local
workstation002.acme.local
datacenter001.acme.local
John Doe
workstation010.acme.local
workstation011.acme.local
No Malware

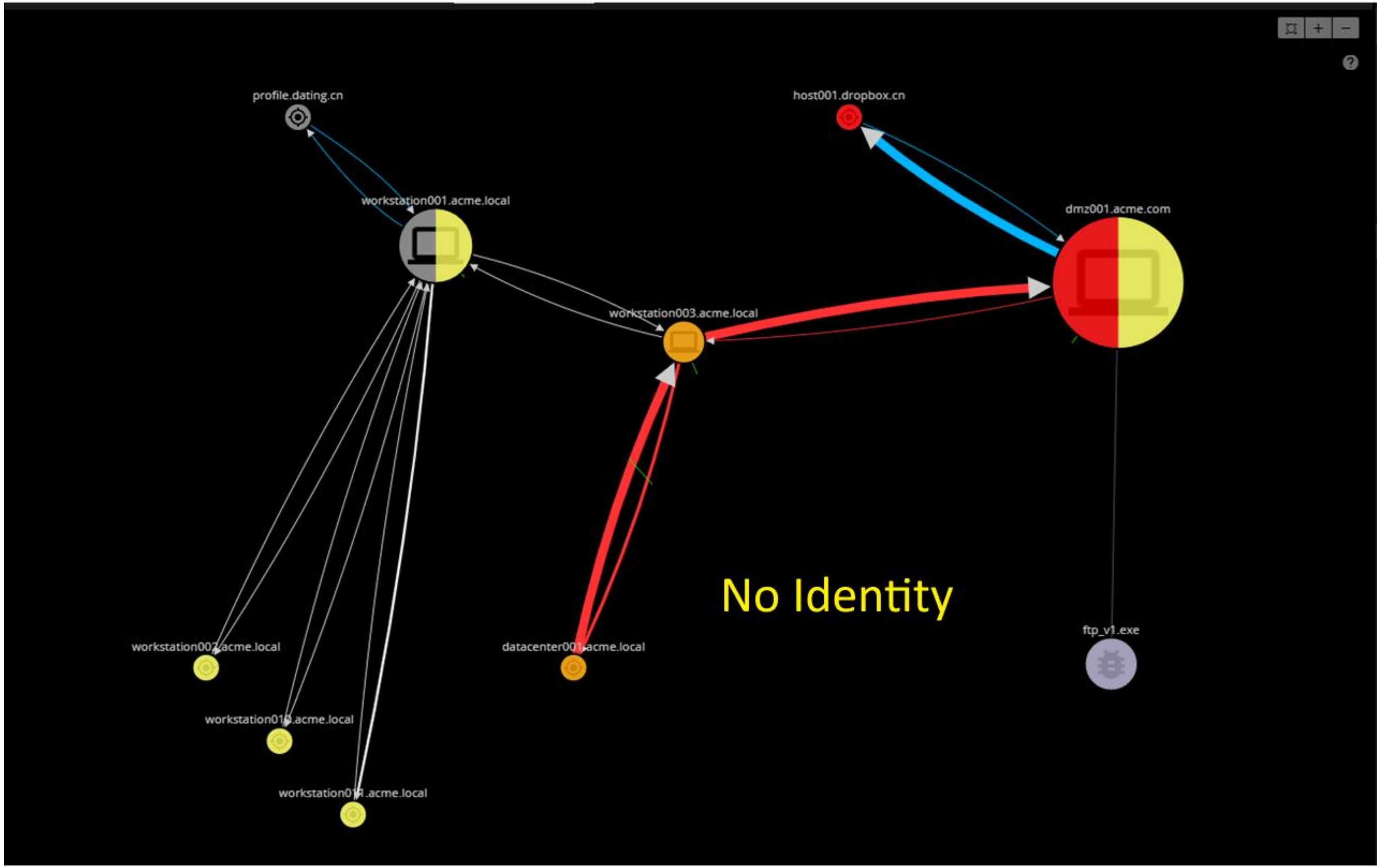
DLP & Flow

~~Endpoint Defense~~

No Identity Signal

Network Defense

DLP & Flow


profile.dating.cn
host001.dropbox.cn
workstation001.acme.local
dmz001.acme.com
workstation003.acme.local
workstation002.acme.local
datacenter001.acme.local
ftp_v1.exe
workstation010.acme.local
workstation011.acme.local

No Identity

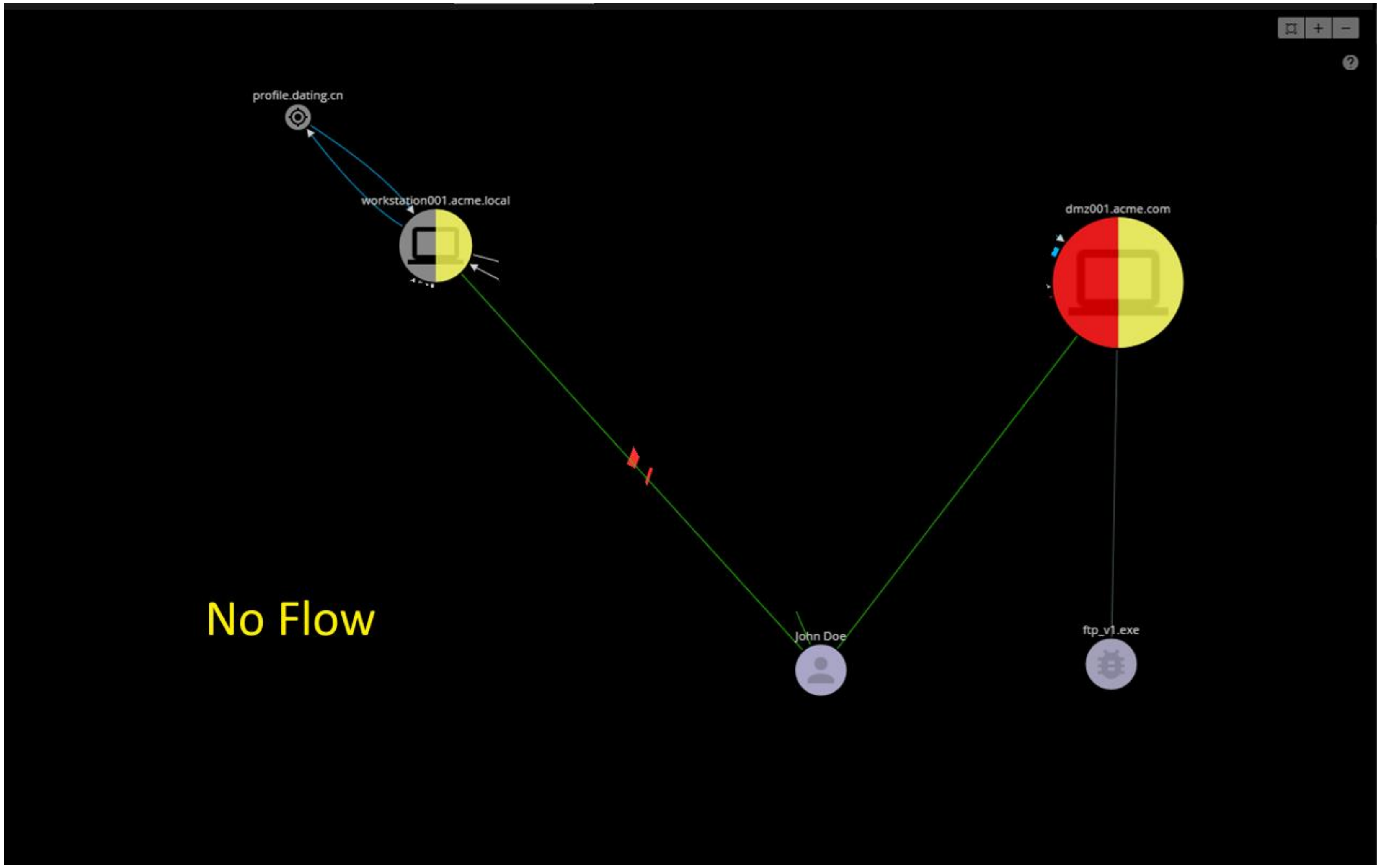
~~Identity Defense~~

Endpoint Defense

No Flow or DLP Signal

Network Defense

~~DLP & Flow~~



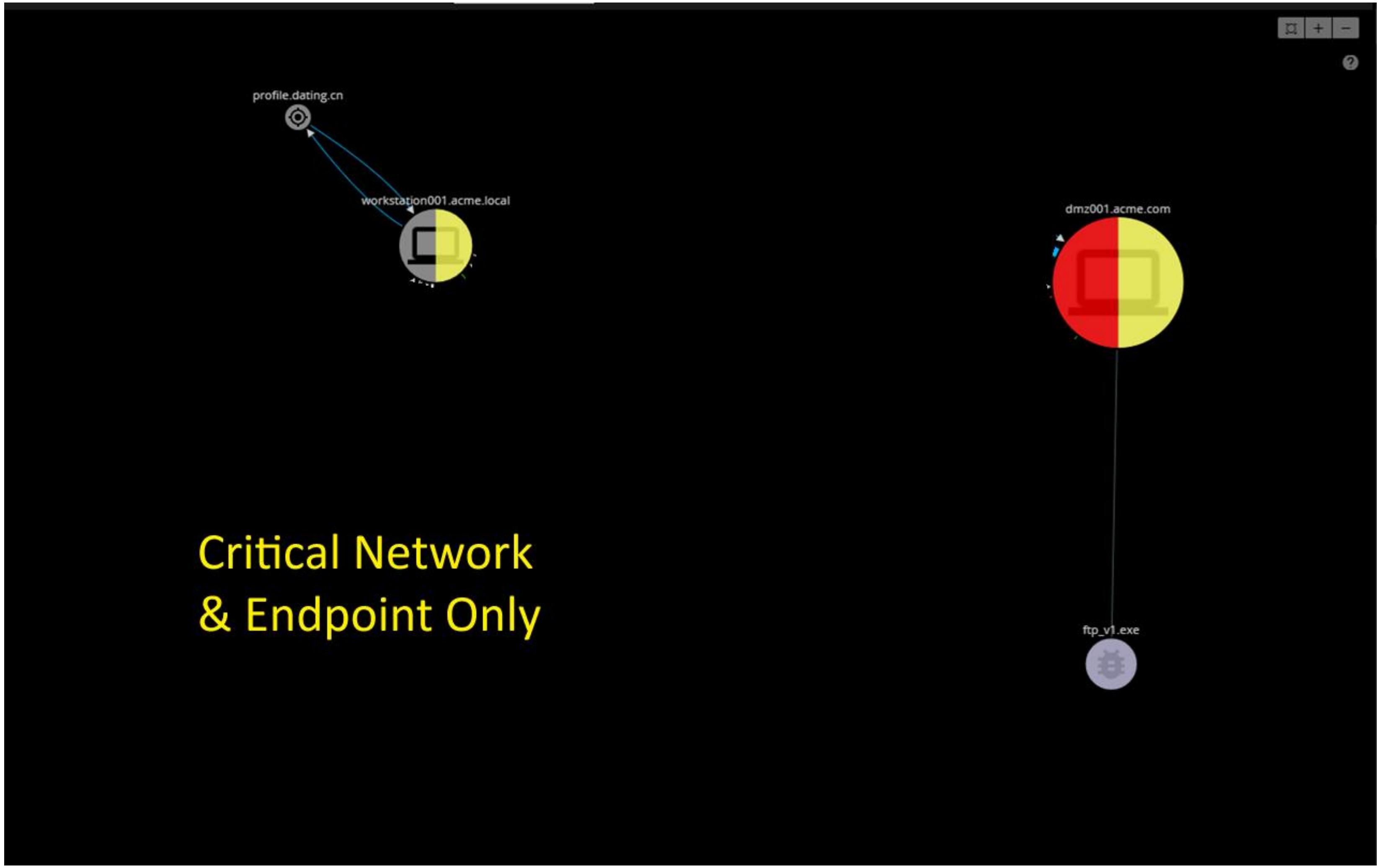
Identity Defense

Endpoint Defense

Triage Signals Only

Network Defense

~~Identity Defense~~



Critical Network & Endpoint Only

~~DLP & Flow~~

Endpoint Defense

Artificial Narrow Intelligence (ANI)

- Coded (not trained) for specific tasks
- Faster, Cheaper, Predictable
- Defendable in Court by an Expert

Examples of Objective Observations

Data transfer on relationships

MITRE ATT@CK actions on relationships

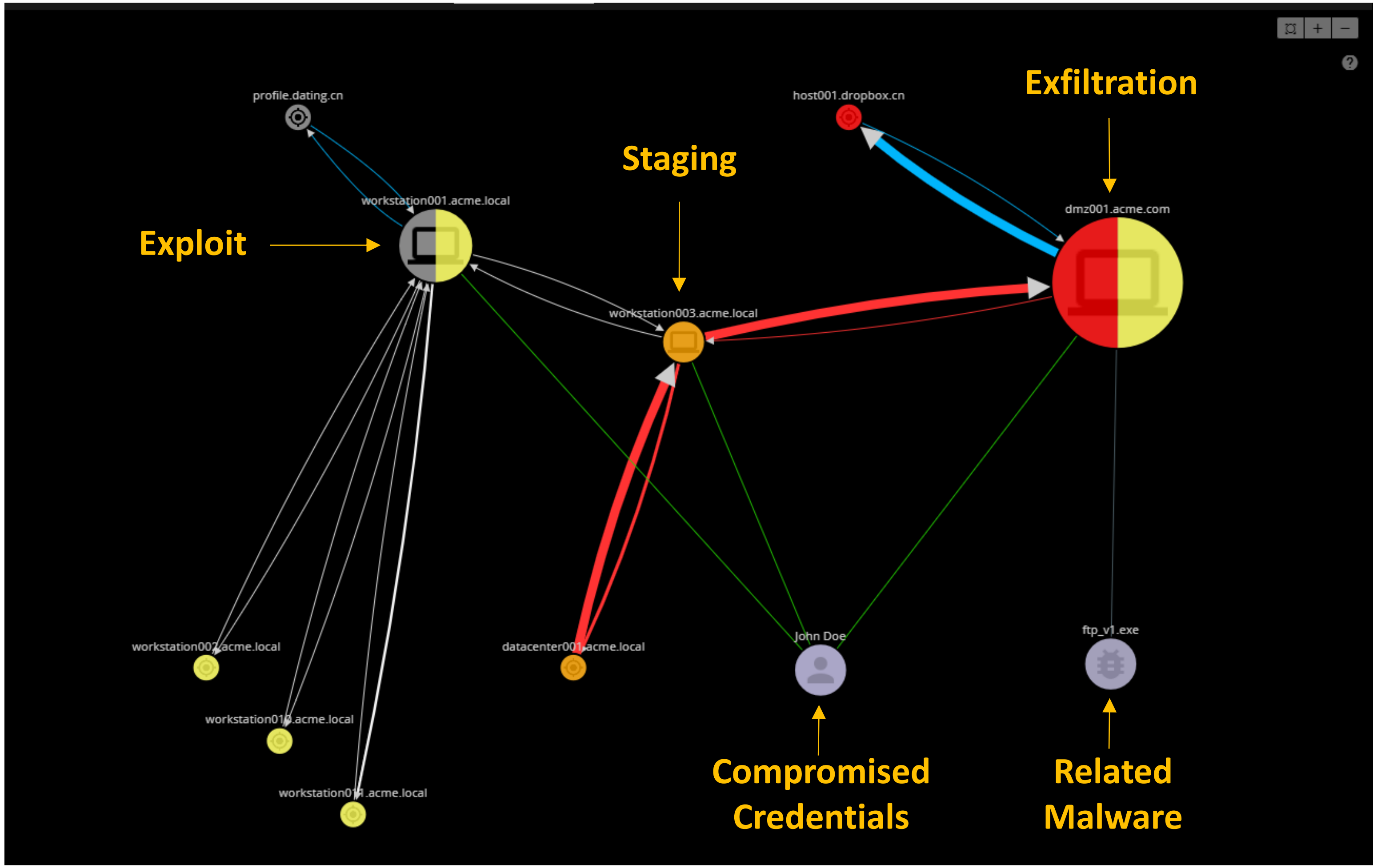
Corroboration by different controls (AV, FW, ID, etc.)

Volume of Alarms

Relationships that progress *modus operani*

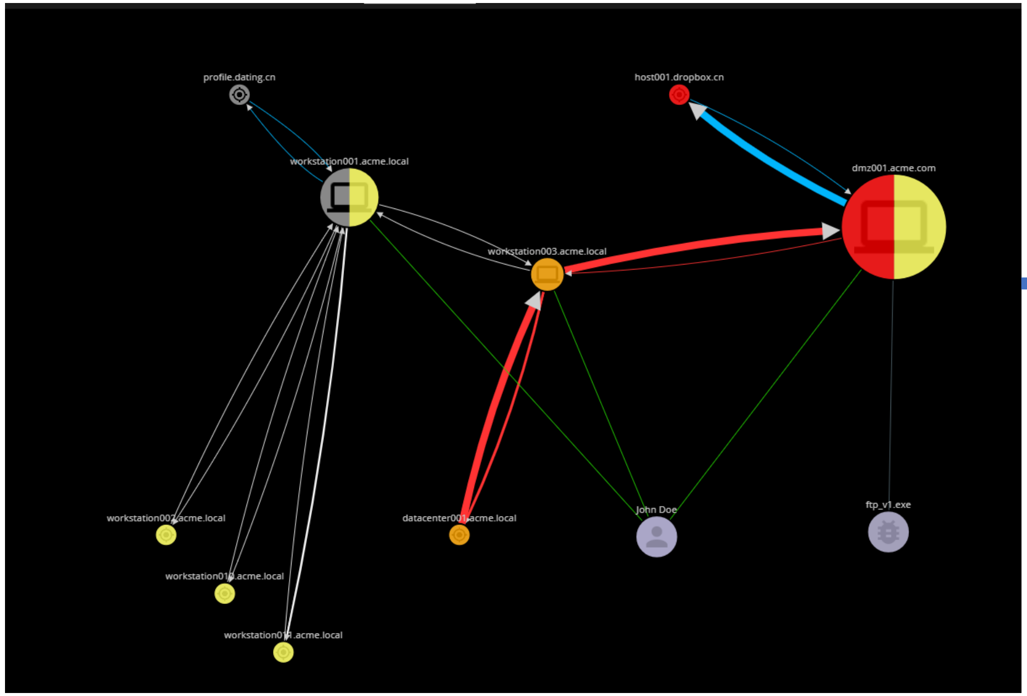
Graph vs. Crime Theory

- Meaningful Graph Relationships
- Modus Operandi of Attacker
- Combines, standardizes diverse data
- Hierarchical JSON
- *SECOPS & LE **Unit of Work***



The Last Mile

GenAI to translate highly predictable, structured JSON to English


profile.dating.cn
host001.dropbox.cn
workstation001.acme.local
dmz001.acme.com
workstation003.acme.local
workstation002.acme.local
datacenter001.acme.local
John Doe
ftp_v1.exe
workstation010.acme.local
workstation014.acme.local

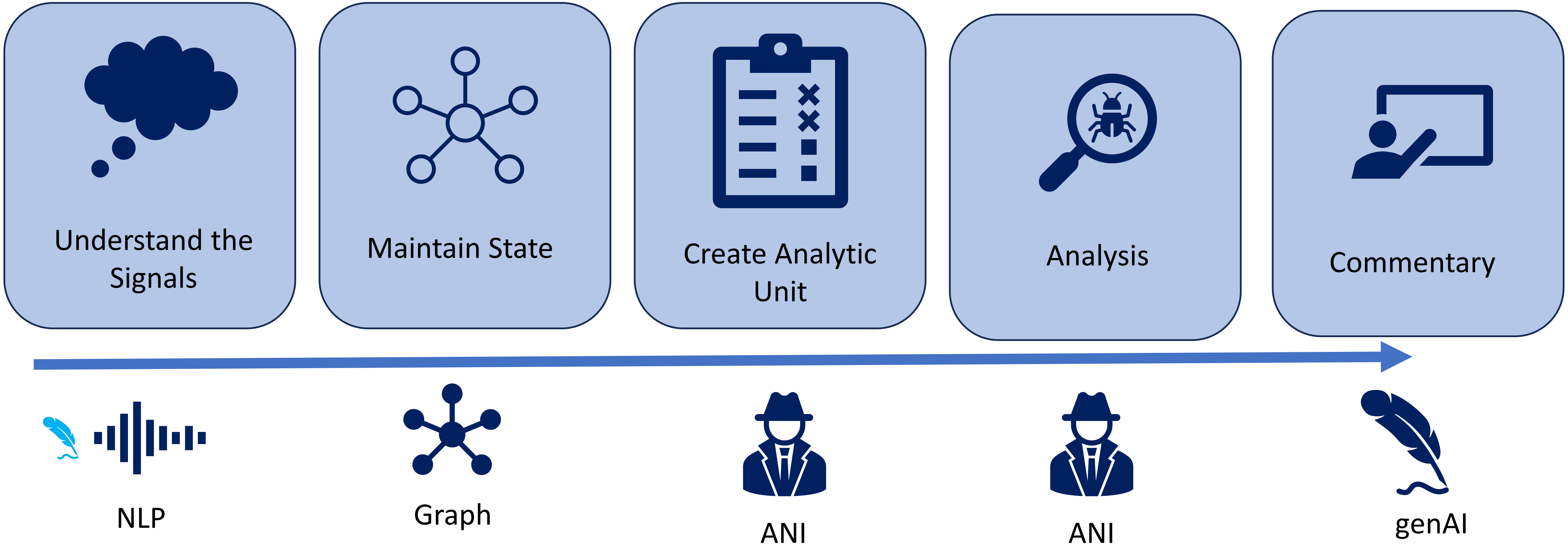
Analysis of all digital evidence shows data was stolen from the internal network and transmitted to computers located in China. The attack began 2 weeks ago by exploiting a vulnerable server. The attackers were able exploit administrator credentials to access a database with proprietary trade secrets then stage them on another web server before exfiltrating the data to the Chinese computers.

The Last Mile Dataset

- Explanations need to map to data in structure
- Explanations need to be accurate (garbage in/out)
- Privacy concerns and sanitization

Instruction	Input Text	Output text
Explain this Cybersecurity Incident	{..}	Analysis of all digital evidence shows data was stolen …

Analytics Pipeline Methodology



Understand the Signals

Maintain State

Create Analytic Unit

Analysis

Commentary

NLP

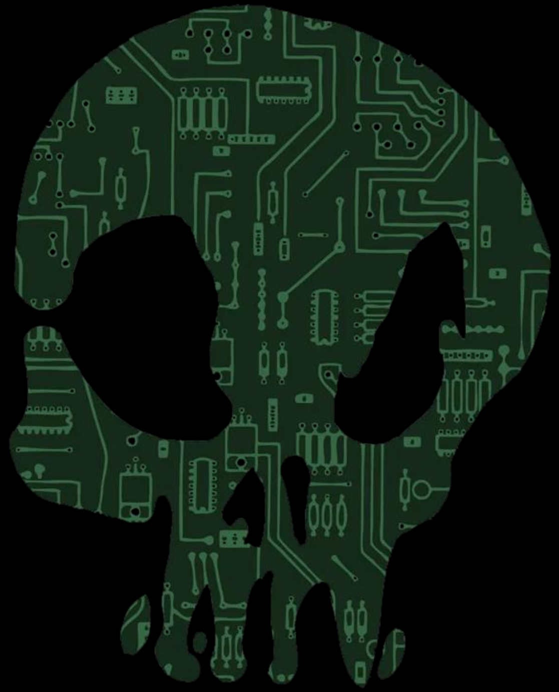
Graph

ANI

ANI

genAI

Birthing Perjury-free AI



Charles D. Herring, WitFoo co-Founder & CTO
Charles@WitFoo.com
CharlesHerring.com
@charlesherring