

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 **evidence**
- Analysis must be verifiable, accurate and explainable
- Consequences for error are extreme

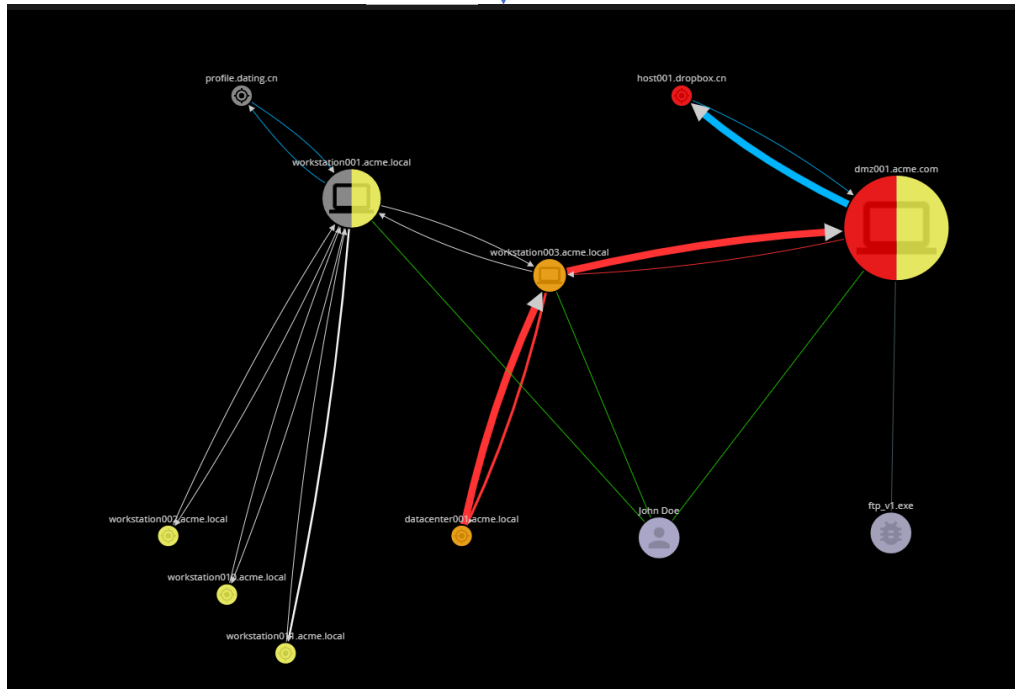
Goals of AI in Cybersecurity Analysis

```
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 ...
```

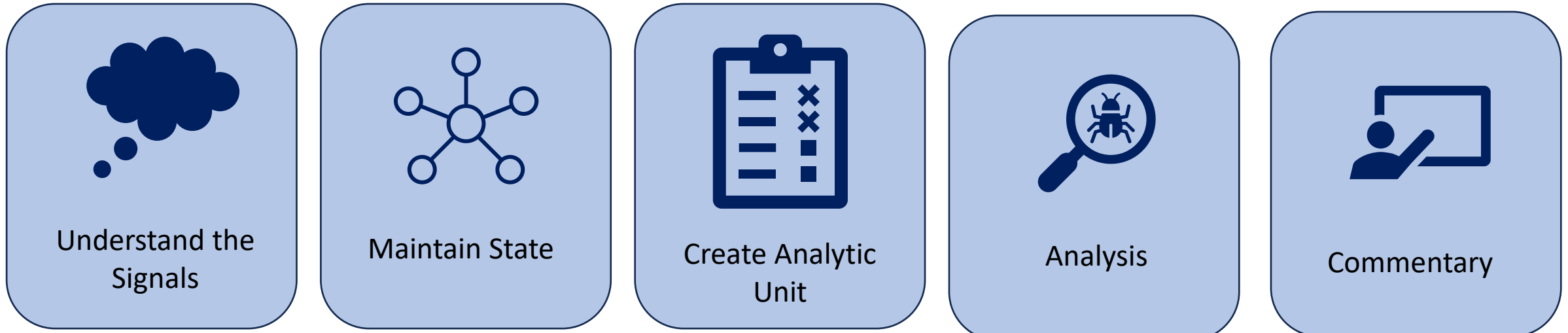
```
127.0.0.1 - - [09/Feb/2024:15:36:14 +0100] "GET / HTTP/1.1" 200 3460 "-" "Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:120.0) Gecko/20100101 Firefox/120.0"
```

... thousands to billions per second



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



Rows (SQL/noSQL)



Graph



Vector

JSON & NoSQL

- High Compression (net & disk)
- REST Powered Transmission
- Easy to Hash & Version
- Hierarchical Structures

Incident JSON View





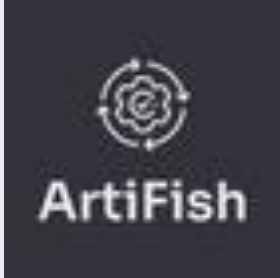

```
id: "53ba6ed0-ed35-11ed-8a89-053651253e65"
partition: "53babcf0-ed35-11ed-8a89-053651253e65"
nodes: Object {"52801a10-ed35-11ed-8a89-053651253e65":{"id":"52801a10-ed35-11ed-8a89-053651253e65","partition":"53b89a10-ed35-11ed-8a89-053651253e65"}
  52801a10-ed35-11ed-8a89-053651253e65: Object {"id":"52801a10-ed35-11ed-8a89-053651253e65","partition":"53b89a10-ed35-11ed-8a89-053651253e65"}
    id: "52801a10-ed35-11ed-8a89-053651253e65"
    partition: "53b89a10-ed35-11ed-8a89-053651253e65"
    ip_address: "10.10.10.3"
    ip: "10.10.10.3"
    org: ""
    orgId: 1
    mac: ""
    guid: ""
    internal: true
```

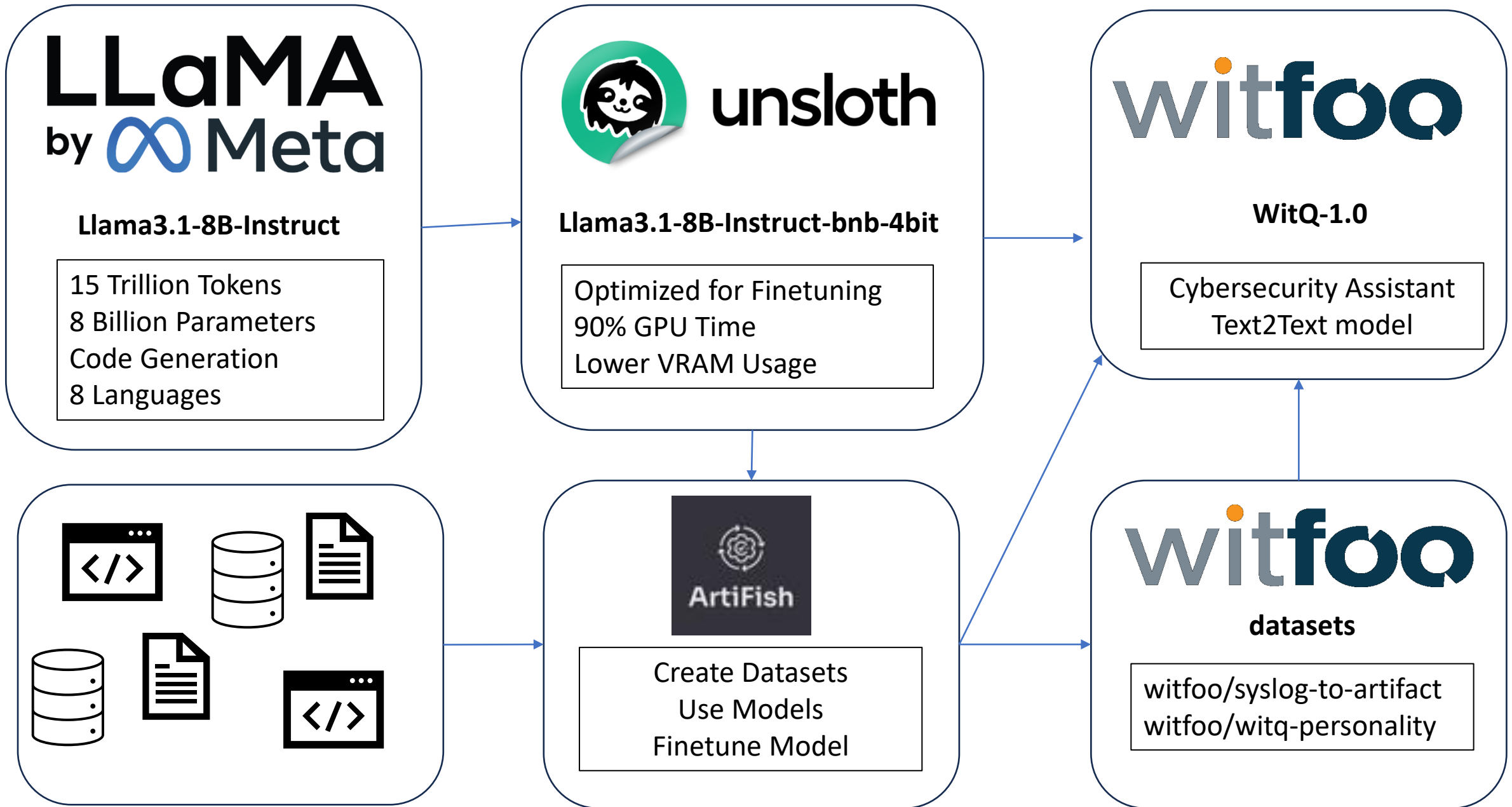
Comparing Analytic Functions

	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
		<ul style="list-style-type: none">• Dataset Creation• Model Fine-tuning• Chatbot examples <p><i>Apache 2.0 Opensource License</i></p> <p>GitHub.com/f15hb0wn/ArtiFish</p>
	<ul style="list-style-type: none">• WitQ Model (Llama 3.1)• WitQ Demo App• Datasets <p><i>Opensource Data and Models</i></p> <p>HuggingFace.co/WitFoo</p>	



ArtiFish Layout

The screenshot shows the GitHub interface for the repository 'ArtiFish' by user 'f15hb0wn'. The repository is public and has 9 commits, 1 branch (main), and 0 tags. The file list includes folders for 'chatbots', 'dataset-creation', and 'fine-tuning', and files for '.gitignore', 'LICENSE', 'README.md', and 'requirements.txt'. The right sidebar shows repository statistics: 0 stars, 1 watcher, and 0 forks. The repository description is 'Toolkit for genai in cybersecurity'.

Navigation: <> Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, Settings

Repository: ArtiFish (Public)

Actions: Pin, Unwatch (1), Fork (0), Star (0)

Branches: main (1 Branch), Tags: 0 Tags

Search: Go to file

Buttons: Add file, Code

File/Folder	Commit Message	Time
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

Datasets: witfoo/witq-personality like 0 Dataset card Viewer Files and versions Community 1 Settings

Split (1)
train · 10 rows

Search this dataset SQL Console

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

Fine Tune on Errors

- Collect Errors as they occur
- Repeat Fine Tuning to correct



Dataset for Parsing to English

Datasets: witfoo / **syslog-to-artifact** like 0

[Dataset card](#) [Viewer](#) [Files and versions](#) [Community](#) 1 [Settings](#)

Split (1)
train · 155k rows

Search this dataset SQL Conso

input_text string · lengths	output_text string · lengths	instruction string · classes
 <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

Parse to JSON

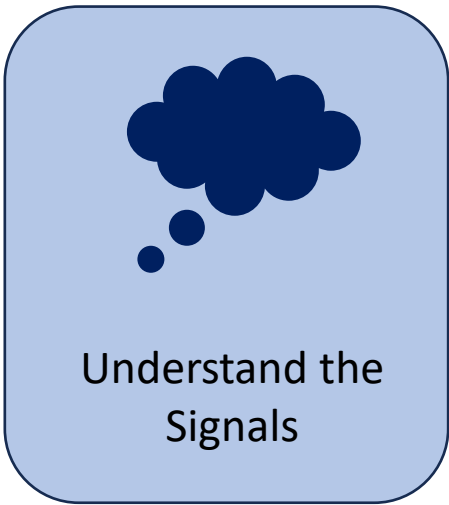
Datasets: witfoo/**syslog-to-artifact** like 0

Dataset card **Viewer** Files and versions Community **1** Settings

Split (1)
train · 155k rows

Search this dataset **SQL** Console

input_text string · lengths	output_text string · lengths	instruction string · classes
 <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]	 { "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", "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]\n", "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" }	 Create a JSON artifact from the message



GenAI in Message Processing

```
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
```

Clear Submit

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

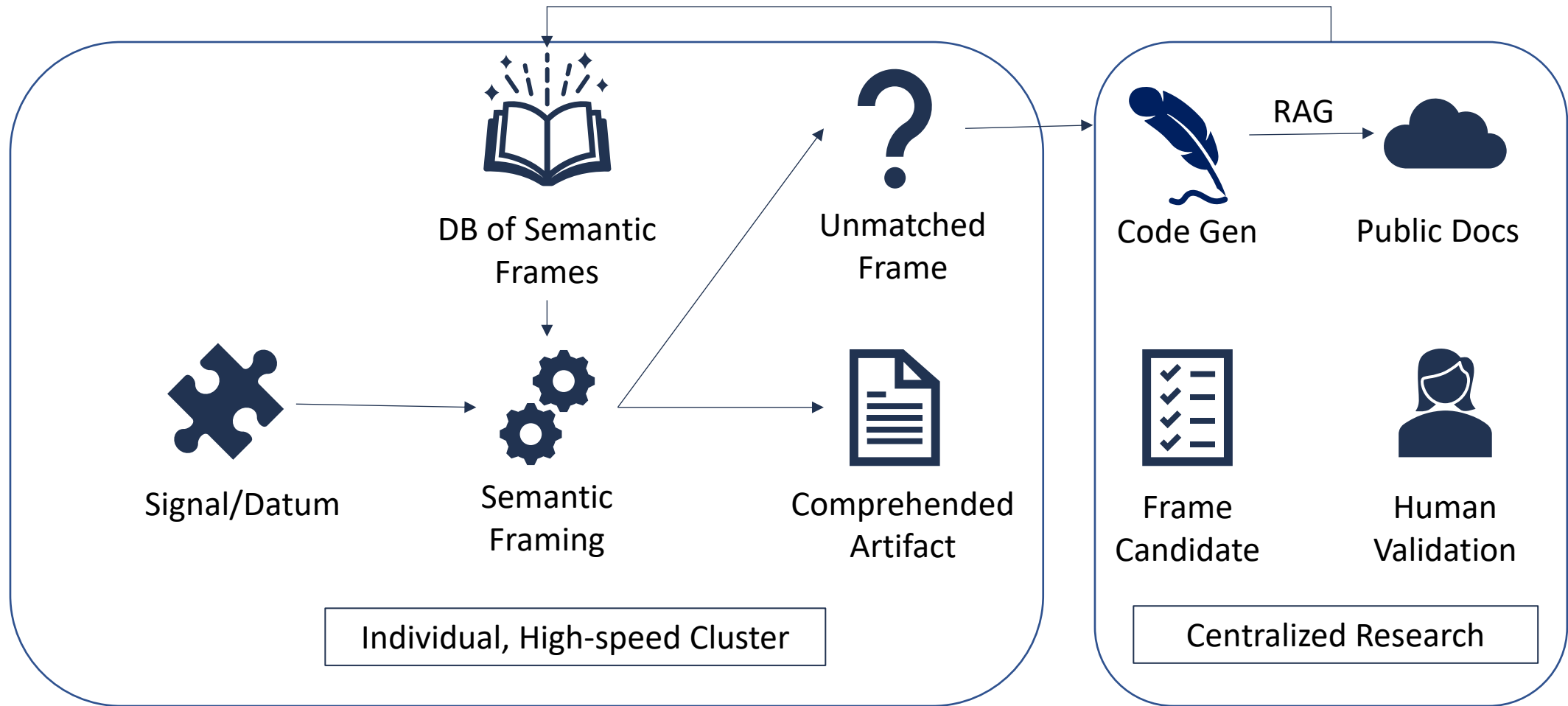
Very Poor

Retrieval Augmented Generation (RAG)

- Training data changes how the model “thinks”
- Trained data cannot be unlearned
- Trained data reference is unavailable
- Training a model is expensive in time & compute
- RAG data is fast, easy to update & inexpensive to maintain
- Memorizing & learning a book vs Dewey Decimal



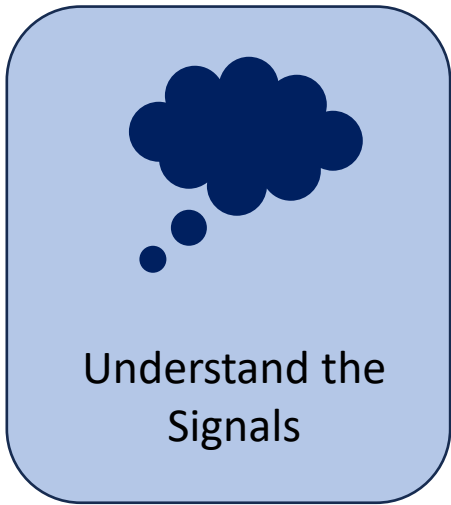
NLP & GenAI in Signal Processing



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

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

Clear Submit

Response

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

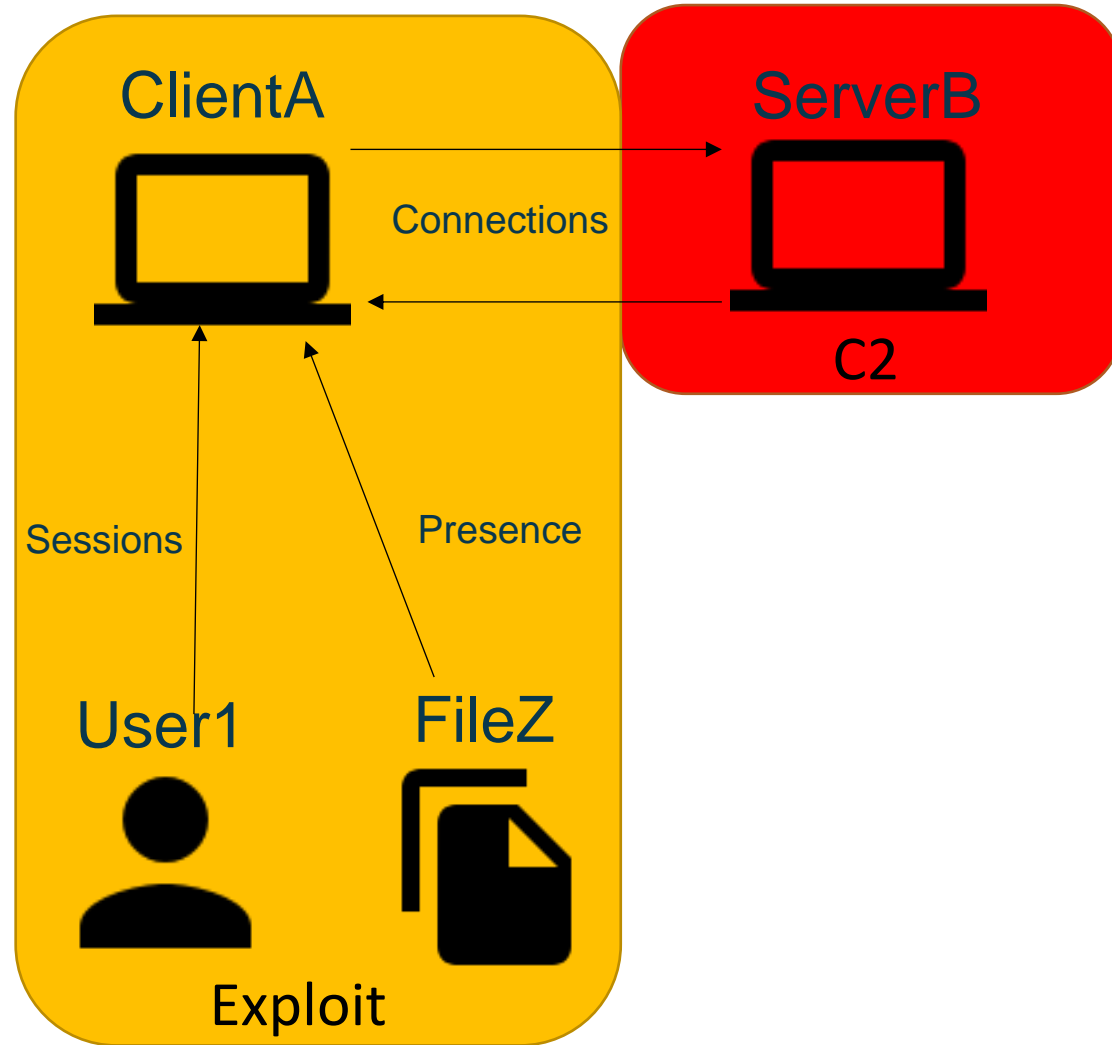
```
import com.witfoo.parser.field.fields.{
  Action,
  ClientIp,
  ClientPort,
  LocalStartTime,
  MessageType,
  ServerBytes,
  ServerHostname,
  StreamName,
  Uri,
  VendorCode
}
import com.witfoo.parser.regex.RegexLibrary.{IP, NON_NEG_INT, POS_INT, WORD_DASH}
import com.witfoo.parser.syslog.extraction.{ExtractionAccumulator,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

class AWSVPCFlow extends SimpleRFCProgramNameExtractor {

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

Flag

Graph Theory Analysis of Signals



Artifacts
<ul style="list-style-type: none">• ClientName: ClientA• ClientIP: 10.10.10.43• ClientMAC: 00-DC-EF-23-15-12• Product: MS DHCP• MessageType: DHCP Lease• Intent: Asset Info
<ul style="list-style-type: none">• ClientName: ClientA• User: User1• File: FileZ• Product: Crowdstrike Falcon• MessageType: Malware Detected• Intent: Exploit Detection
<ul style="list-style-type: none">• ClientIP: 10.10.10.43• ServerName: ServerB• Product: Cisco Firepower• MessageType: C2 Detected• Intent: C2 Detection



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

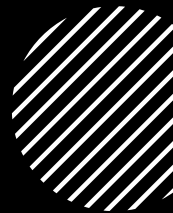
Corroboration by different controls (AV, FW, ID, etc.)

Volume of Alarms

Relationships that progress *modus operani*



Building Narrow AI in Cyber



What objective observations can be made of graph?



What data should be extracted from signals as features?



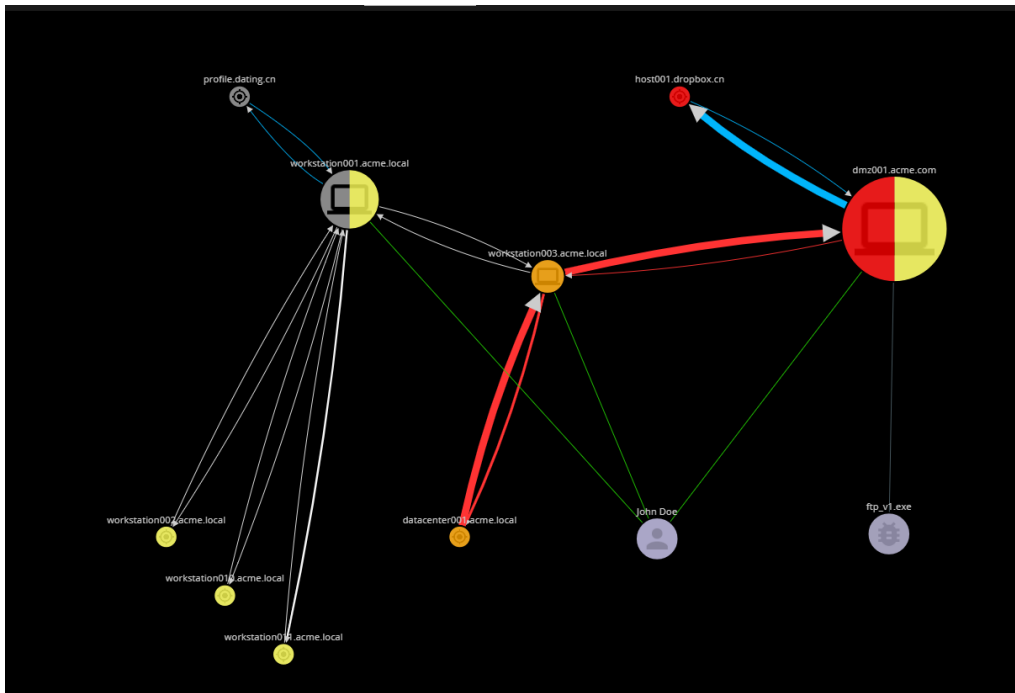
What process does the expert use in decision making?



How are subjective weights created?

The Last Mile

GenAI to translate highly predictable, structured JSON to English



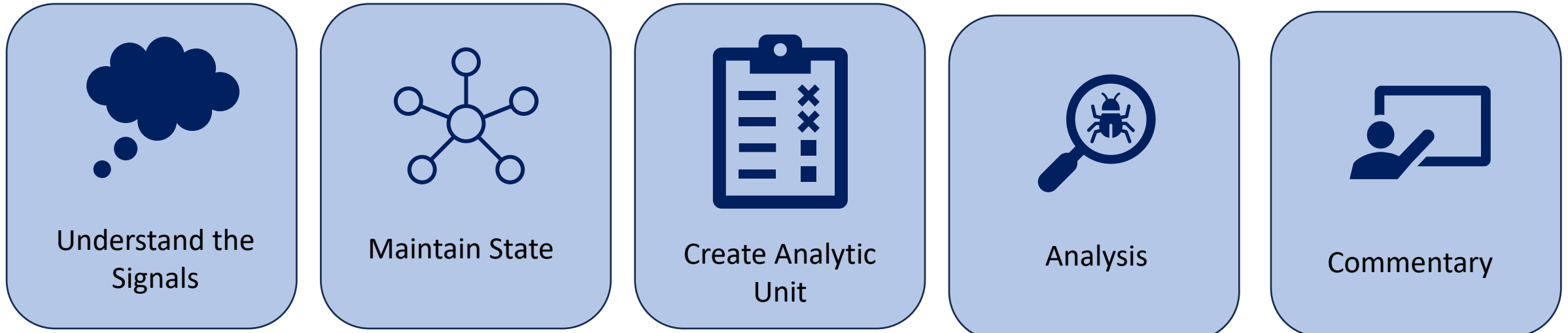
Analysis of all digital evidence shows data was stolen from the internal network and transmitted to computers located in China. The attack began 2 weeks ago by exploiting a vulnerable server. The attackers were able exploit administrator credentials to access a database with proprietary trade secrets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



NLP



Graph



ANI



ANI



genAI

Birthing Perjury-free AI

Charles D. Herring, WitFoo co-Founder & CTO

Charles@WitFoo.com

CharlesHerring.com

@charlesherring