Knowledge-Embedded Narrative Construction from Open Source Intelligence

2023 CAE-R Research Symposium PhD Dissertation Panel

Author: Priyanka Ranade, PhD Candidate
Advisor: Dr. Anupam Joshi

University of Maryland, Baltimore County
Department of Computer Science & Electrical Engineering
U.S. Department of Defense

What are Narratives?

Streams of information

Decomposed into **events**

When **events are chained** together, they form **end to end stories**

Types of Narratives

1. **Social Media Based Narratives**
2. **News Based Narratives**
3. Literary Narratives

. . .

Narrative Construction: The chronological ordering of events into plot sequences.

Overview

Primary Areas

- Narrative Theory
- Information Retrieval
- Neurosymbolic AI

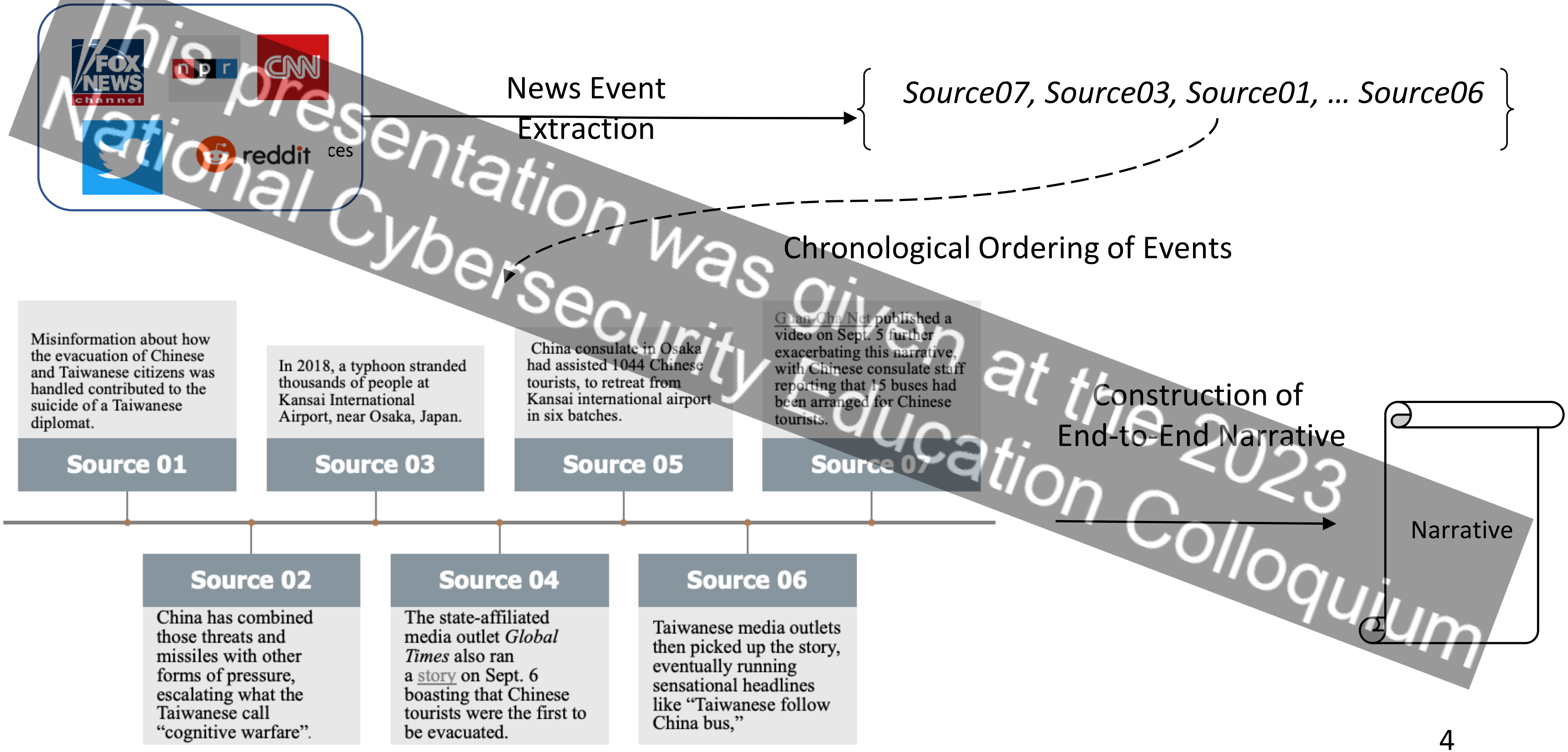
Motivation

How do we contextually synthesize disparate pieces of information together to communicate informative stories?

Thesis Statement

Creating knowledge-embedded narrative structures to represent online discourse and events will help us in constructing thematic event chains, enabling the ability to uncover rhetoric framing tactics, track adversarial motives and behaviors, and model content evolution.

Computational Narrative Construction



ces
News Event Extraction
{ *Source07, Source03, Source01, … Source06* }
Chronological Ordering of Events
Construction of End-to-End Narrative
Narrative

Source 01
Misinformation about how the evacuation of Chinese and Taiwanese citizens was handled contributed to the suicide of a Taiwanese diplomat.

Source 02
China has combined those threats and missiles with other forms of pressure, escalating what the Taiwanese call "cognitive warfare".

Source 03
In 2018, a typhoon stranded thousands of people at Kansai International Airport, near Osaka, Japan.

Source 04
The state-affiliated media outlet *Global Times* also ran a story on Sept. 6 boasting that Chinese tourists were the first to be evacuated.

Source 05
China consulate in Osaka had assisted 1044 Chinese tourists, to retreat from Kansai international airport in six batches.

Source 06
Taiwanese media outlets then picked up the story, eventually running sensational headlines like "Taiwanese follow China bus,"

Source 07
Guan Cha Net published a video on Sept. 5 further exacerbating this narrative, with Chinese consulate staff reporting that 15 buses had been arranged for Chinese tourists.

Information Extraction

1. Named Entity Recognition
2. Relationship Extraction
3. **Plot Element Framing**

Can we classify unstructured events based on their plot contexts?

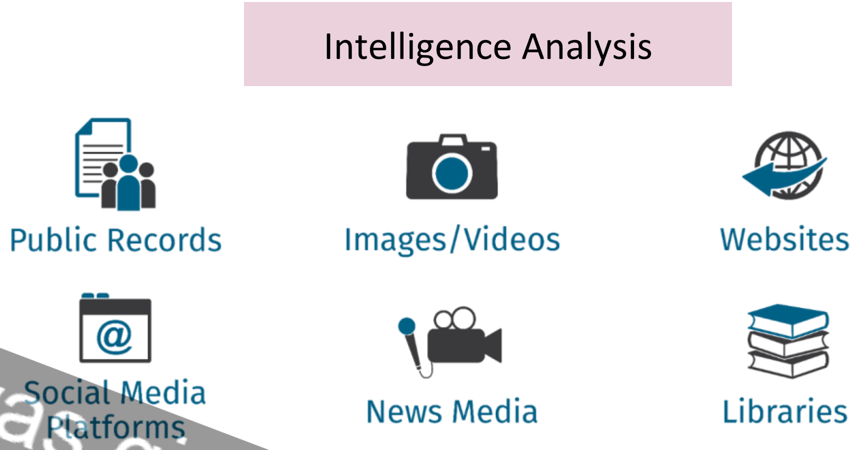
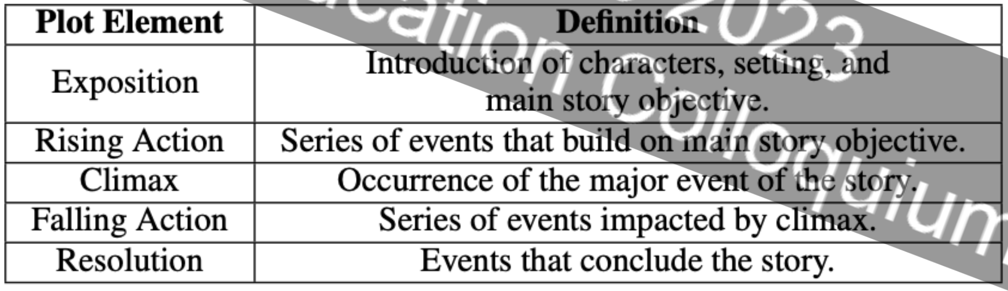


Photo Credit: SANS Institute

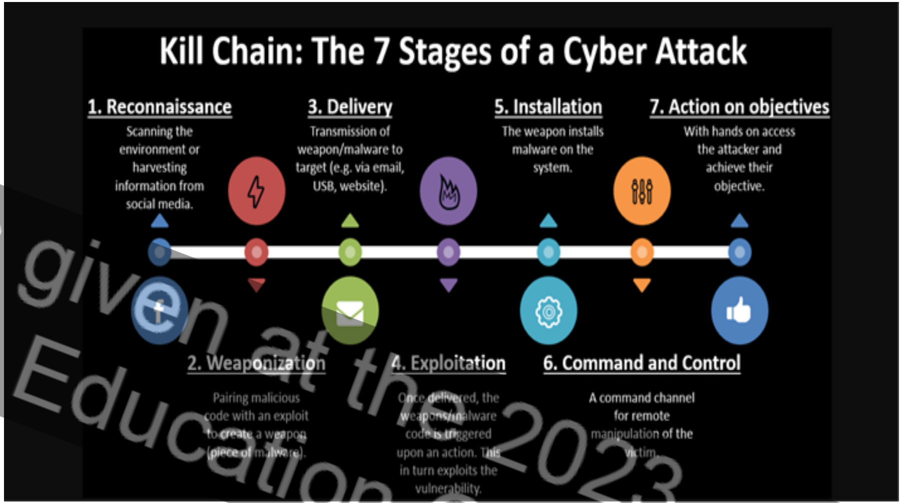


Plot Element	Definition
Exposition	Introduction of characters, setting, and main story objective.
Rising Action	Series of events that build on main story objective.
Climax	Occurrence of the major event of the story.
Falling Action	Series of events impacted by climax.
Resolution	Events that conclude the story.

Gustav Freytag’s Plot Element Model

Grounding in Cybersecurity Domain

- Narrative analysis can aid in finding causal, temporal, and thematic patterns in domain-specific information
- Examples:
  - Modeling changes in and predicting potential virus variants
  - Understanding behavioral patterns in misinformation threat campaigns
  - **Modeling sequential activity of cybersecurity exploits, and advanced persistent threat groups**



Kill Chain: The 7 Stages of a Cyber Attack

1. Reconnaissance
Scanning the environment or harvesting information from social media.

2. Weaponization
Pairing malicious code with an exploit to create a weapon (piece of malware).

3. Delivery
Transmission of weapon/malware to target (e.g. via email, USB, website).

4. Exploitation
Once delivered, the weapons/malware code is triggered upon an action. This in turn exploits the vulnerability.

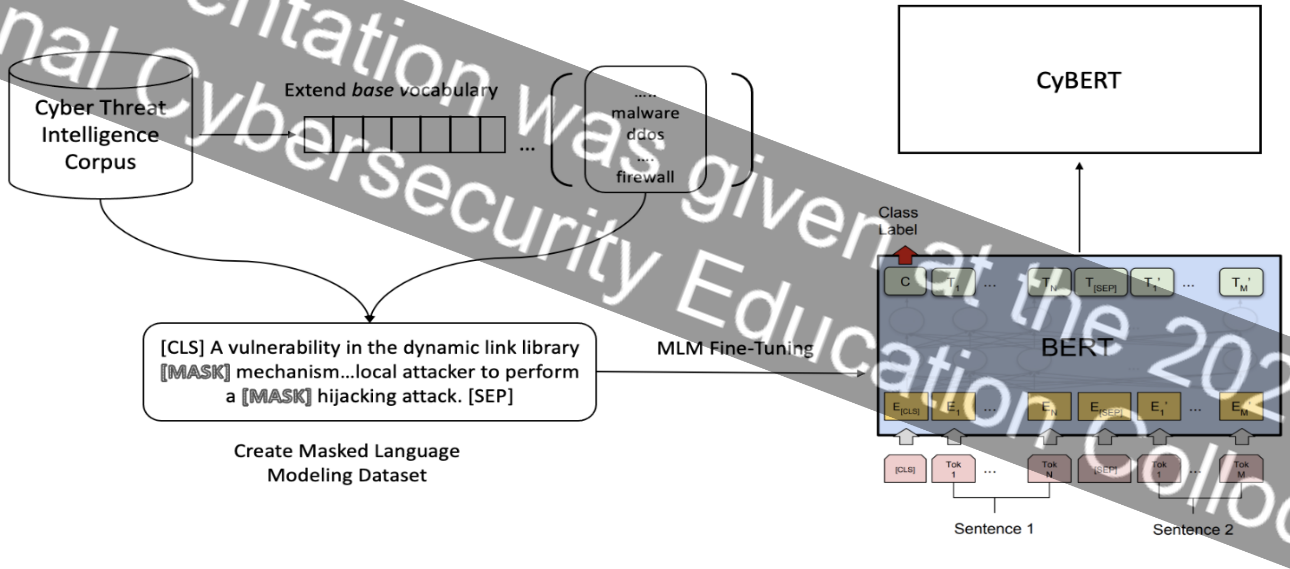
5. Installation
The weapon installs malware on the system.

6. Command and Control
A command channel for remote manipulation of the victim.

7. Action on objectives
With hands on access the attacker and achieve their objective.

Preliminary Work

CyBERT: Contextualized Embeddings for the Cybersecurity Domain



Cyber Threat Intelligence Corpus

Extend *base* vocabulary

.....
malware
ddos
....
firewall

[CLS] A vulnerability in the dynamic link library [MASK] mechanism...local attacker to perform a [MASK] hijacking attack. [SEP]

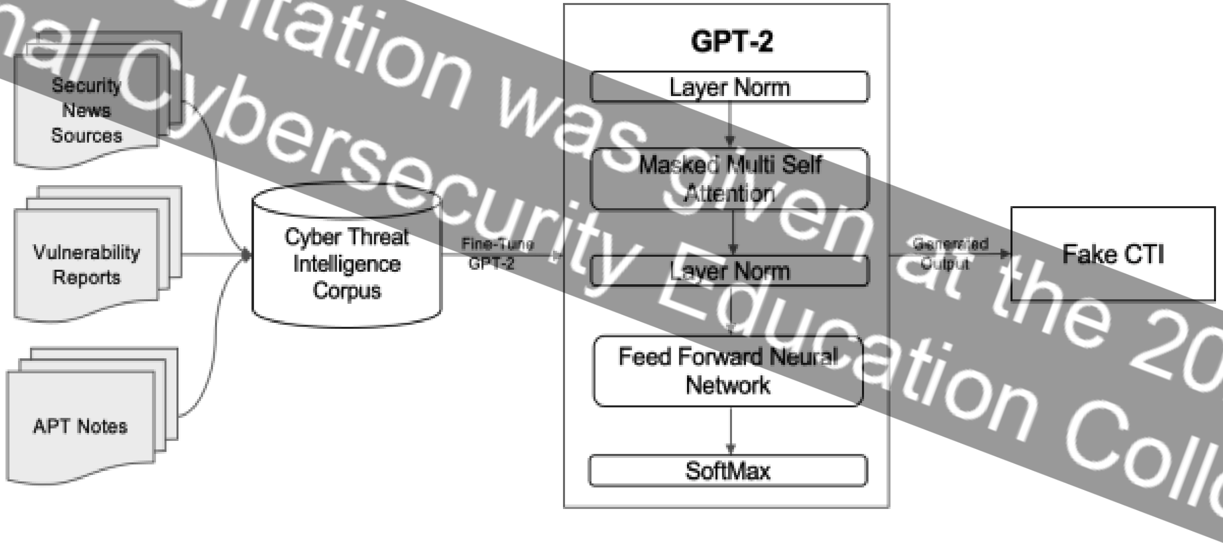
Create Masked Language Modeling Dataset

MLM Fine-Tuning

CyBERT

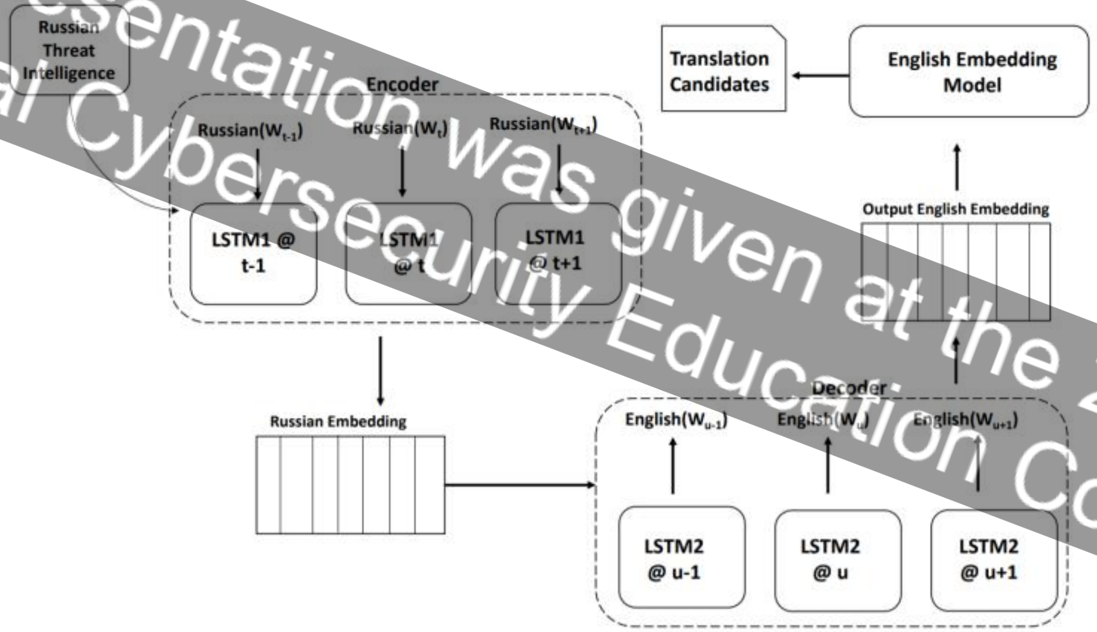
P. Ranade, A. Piplai, A. Joshi, and T. Finin, "CyBERT: Contextualized Embeddings for the Cybersecurity Domain", InProceedings, *IEEE International Conference on Big Data*, December 2021.

Generating Fake Cyber Threat Intelligence Using Transformer-Based Models



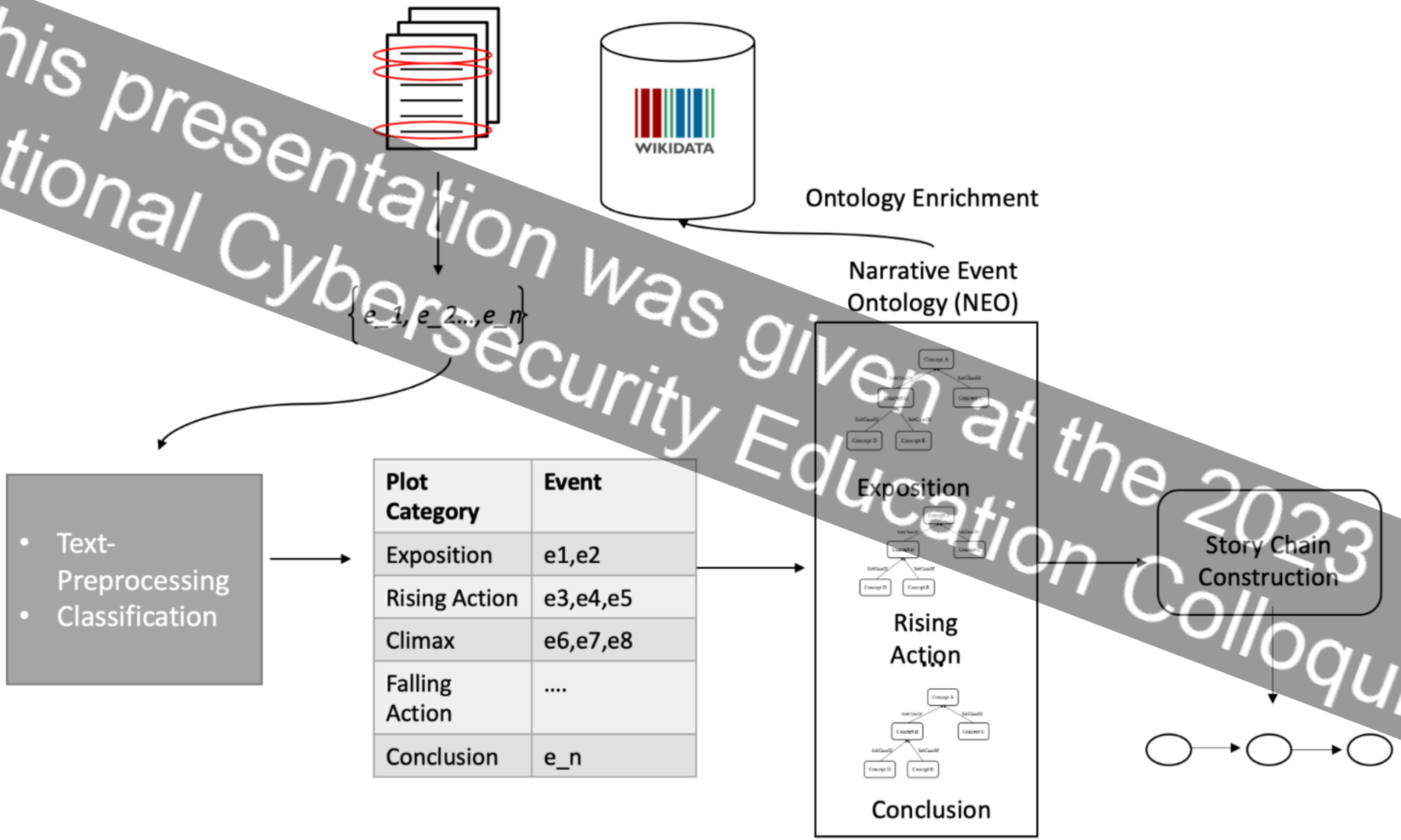
P. Ranade, A. Piplai, S. Mittal, A. Joshi, and T. Finin, "Generating Fake Cyber Threat Intelligence Using Transformer-Based Models", Proceedings, *International Joint Conference on Neural Networks (IJCNN 2021)*, July 2021.

Using Deep Neural Networks to Translate Multilingual Intelligence



P. Ranade, S. Mittal, A. Joshi, and K. P. Joshi, "Using Deep Neural Networks to Translate Multi-lingual Threat Intelligence", InProceedings, *IEEE Intelligence and Security Informatics (IEEE ISI) 2018.*

Thesis Methodology



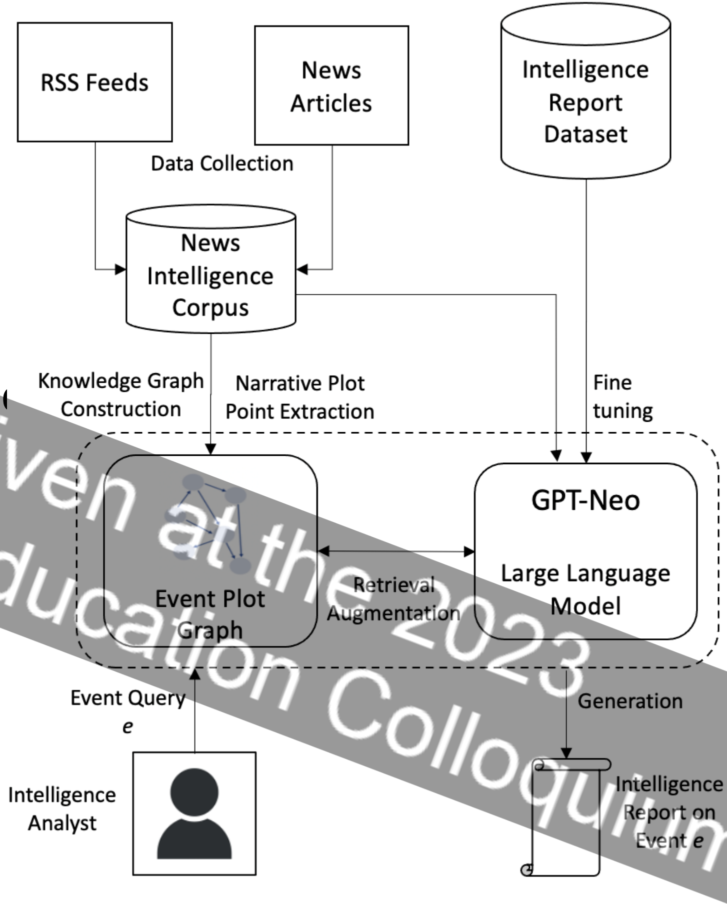
Plot Category	Event
Exposition	e1,e2
Rising Action	e3,e4,e5
Climax	e6,e7,e8
Falling Action	....
Conclusion	e_n

Current Work

Current:

1. Developed the Event Narrative Ontology (NEO) and Retrieval-Augmented Generation (RAG) framework for automati Intelligence Report creation. (submitted, August 2022).

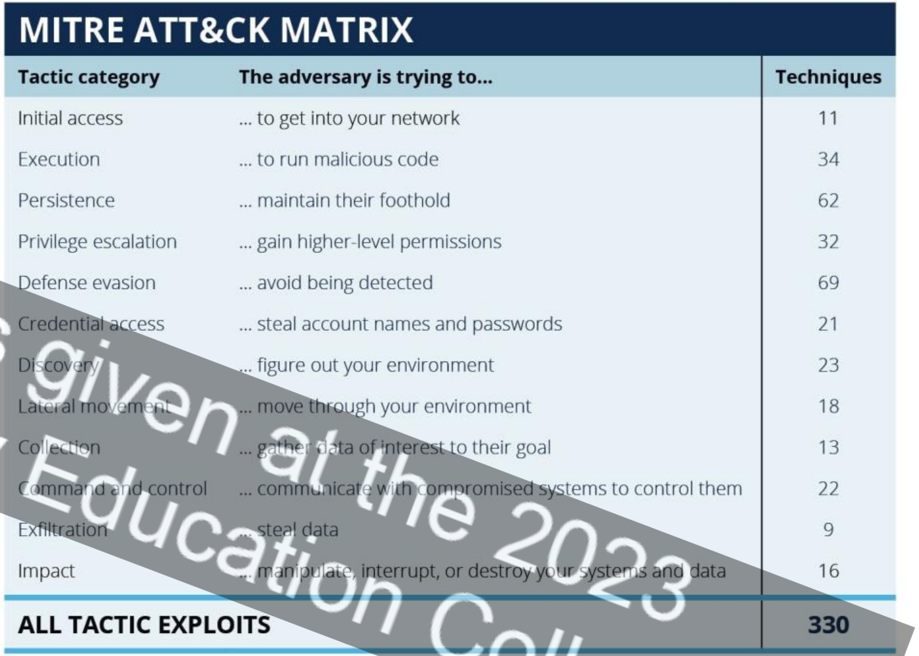
1. Created ground truth mappings of cybersecurity kill chain (CKC) classes and MITRE TTP attribute mappings



Future Work

Future:

1. Evaluating the CKC-TTP mappings through qualitative/quantitative methods.
2. Generalizing RAG-based framework to the **attack path generation** problem, using the curated CKC-TTP mappings.



MITRE ATT&CK MATRIX

Tactic category	The adversary is trying to...	Techniques
Initial access	... to get into your network	11
Execution	... to run malicious code	34
Persistence	... maintain their foothold	62
Privilege escalation	... gain higher-level permissions	32
Defense evasion	... avoid being detected	69
Credential access	... steal account names and passwords	21
Discovery	... figure out your environment	23
Lateral movement	... move through your environment	18
Collection	... gather data of interest to their goal	13
Command and control	... communicate with compromised systems to control them	22
Exfiltration	... steal data	9
Impact	... manipulate, interrupt, or destroy your systems and data	16
ALL TACTIC EXPLOITS		**330**

Source: Huntsman Security

Thank You!

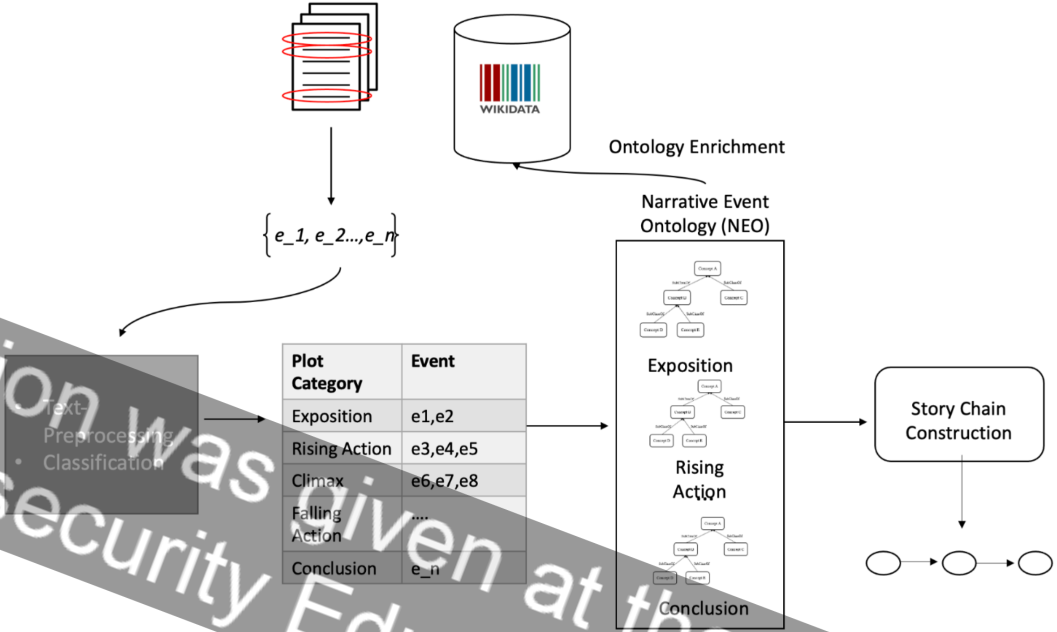
Contact Information:

Priyanka Ranade

priyankaranade@umbc.edu

UMBC

DEPARTMENT OF DEFENSE · UNITED STATES OF AMERICA



Plot Category	Event
Exposition	e1,e2
Rising Action	e3,e4,e5
Climax	e6,e7,e8
Falling Action	....
Conclusion	e_n

- Current NLU techniques have limitations in ordering events based on thematic context
- I propose a Semantic Data Model inspired by narrative theory and hypothesize we can represent, chain, and reason over narratives from disparately sourced event details.

References

[1] Shirai, Sola, et al. "Rule-Based Link Prediction over Event-Related Causal Knowledge in Wikidata." International Semantic Web Conference. 2022.

[2] Zhu, Xianshu, and Tim Oates. "Finding story chains in newswire articles." *2012 IEEE 13th International Conference on Information Reuse & Integration (IRI)*. IEEE, 2012.

[3]Sadlek, Lukáš, Pavel Čeleda, and Daniel Tovarňák. "Identification of Attack Paths Using Kill Chain and Attack Graphs." NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium. IEEE, 2022.