Enhancing Digital Forensics Evidence Analysis with Large Language Models

(KDD2025 Tutorial Session)

Eric Xu[1], **Lin Deng[2], and Damodar Dhital[2]**

[1]University of Maryland, College Park, Maryland, USA

[2]Towson University, Towson, Maryland, USA

Outline

- **Introduction**
- Hands on Tutorial
- Challenges of Leveraging LLM in Digital Forensics
- Conclusion

What is Digital Forensics?

- **Digital forensics** is the process of collecting, analyzing, and preserving electronic evidence for use in legal proceedings or investigations. It involves examining digital devices and data to uncover information related to cybercrimes, fraud, or other illegal activities.

A Real-world Case Involving Digital Forensics

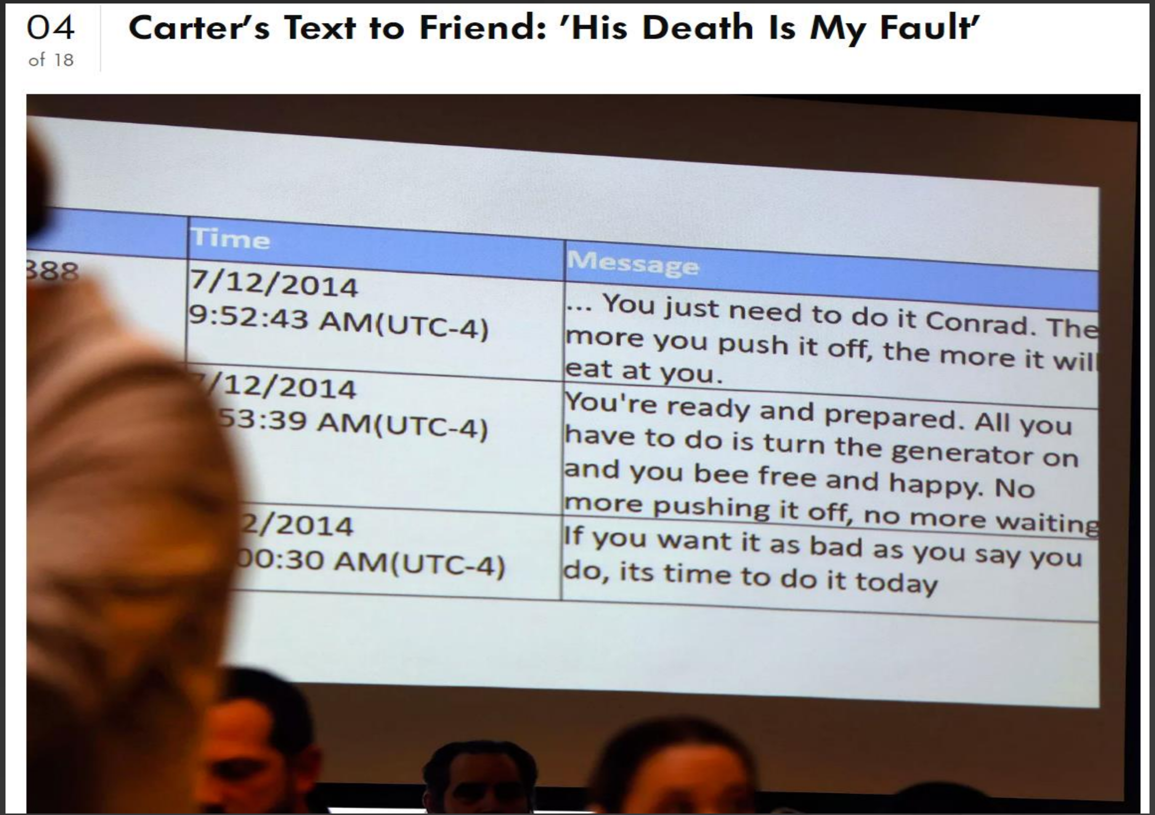

Figure 1: A real-world digital forensic investigation case

Definition Of Digital Forensic Evidence Entity

Digital Forensic Evidence Entity is the smallest unit of digital information that is indivisible and holds forensic significance. For example:

- File Names: The name given to a file, which may reveal its content, origin, or purpose.
- IP Addresses: Numerical labels assigned to devices, useful for tracking online activity.
- Timestamps: Specific points in time associated with digital events, like the creation or modification time of a file.
- Hashes: Unique identifiers derived from data content, often used to verify the integrity of files.

Grouping Digital Forensic Evidence Entity

- Content-Descriptive Entities: Help understand the origin, purpose, or location of digital artifacts, e.g., File names and IP addresses.

- Auxiliary Entities: Provide supporting or supplementary information that enhances the understanding and verification of descriptive digital evidence, e.g., Timestamps and Hashes.

Categories Of Digital Evidence Entity

- Personal Identifiers: Name, Address, Phone number, Email address, Social Security number, Date of birth
- Network Information: IP address, MAC address, Login credentials
- Communication Records: Email addresses, Text messages, Social media messages and posts
- Financial Data: Bank account information, Credit card numbers, Transaction ID, Cryptocurrency wallet addresses
- Location Data: GPS latitude and longitude
- Internet Activity: Browsing URL, Search queries, Search keywords

Definition Of Relationships Between Evidence Entity

The **relationships** between digital evidence entities: refer to the interconnectedness and dependencies among various pieces of digital evidence within a forensic investigation. These relationships help to reconstruct events, validate data, and establish a coherent narrative based on the collected evidence.

Categories Of Relationships Between Digital Evidence Entity

- **Contextual Relationships**: These provide contextual information about the origin, purpose, or use of data. For instance, the relationship between a file name and its content, or between an IP address and the location of the device, can help identify the source or relevance of the evidence.
- **Causal Relationships**: These establish cause-and-effect links between different entities. For example, an IP address logged during a specific time (timestamp) may indicate that a particular device (context) was responsible for accessing or modifying a file.
- **Associative Relationships**: These connect related pieces of evidence that may seem independent but are linked through common attributes. For instance, different files with similar hashes may indicate duplication or tampering.

More Examples Of Relationships

- Communication Relationships: [Phone number A, calls, Phone number B] [Email address A, sends email to, Email address B] [User A, messages, User B] on a social media platform
- Ownership/Association: [Person, owns, Device] [Email address, belongs to, Person] [IP address, associated with, Physical location]
- Temporal Relationships: [File A, created before, File B] [Event A, occurs simultaneously with, Event B] [User, logs in, Timestamp]
- Spatial, Data Flow, Access, Modification, Financial Transactions, Social Connections, Integrity, ...

Identify Evidence Entity (E.g., Address) Is Hard

TOWSON UNIVERSITY

Considering a simple address-processing task before utilizing AI for Named-Entity Recognition (NER)

- Expanding abbreviations: Converting abbreviations to their full forms, e.g., ″St.″ to ″Street″ or ″Ave.″ to ″Avenue.″
- Standardizing formats: Formatting addresses to follow a consistent style, e.g., ″123 Main St Apt 4B″ to ″123 Main Street, Apartment 4B.″
- Normalizing state names: Converting state names to their standard two-letter postal codes, e.g., ″California″ to ″CA.″
- Removing extra whitespace: Eliminating unnecessary spaces between words or at the start/end of the address, e.g., ″ 456 Elm St ″ to ″456 Elm St.″

UNIVERSITY OF BALTIMORE UNIVERSITY OF MARYLAND BJA

Limitation Of Training-based AI For Digital Forensics?

TOWSON UNIVERSITY

- Data scarcity
- Obtaining sufficient training data involved in real-world cyber incidents (e.g., obtain addresses only from shooting cases)
- AI models lack adaptability
- An AI model is often designed for the specific evidence-extracting task (e.g., AI model for identifying an address is different than a person's name)
- Extract evidence relations is hard
- Many different relationships exist

UNIVERSITY OF BALTIMORE UNIVERSITY OF MARYLAND BJA

Why LLMs For Digital Forensics

- LLMs are trained on vast amounts of text data with pattern and structure learning capabilities.
- LLMs have the great potential to automate digital forensics for reliable and efficient discovery and interpretation of digital evidence.

Outline

- Introduction
- **Hands on Tutorials**
- Challenges of Leveraging LLM in Digital Forensics
- Conclusion

Evidence Analysis Leveraging LLMs

- Forensic evidence entity recognition
  - Evidence entity recognition
  - Visualize evidence and their relations
- Evidence knowledge graphs reconstruction
  - Construct a knowledge graph in STIX (zero-shot)
  - Construct a knowledge graph in STIX (one-shot)
  - Compare one-shot vs. zero-shot
- Profiling suspect based on browser history

Outline

- Introduction
- Hands on Tutorial
- **Challenges of Leveraging LLM in Digital Forensics**
- Conclusion

Challenges Inherited From LLMs

TOWSON UNIVERSITY

- Hallucinations: Risk of false leads or erroneous conclusions in investigations
- Interpretability and Explainability: Difficulty in explaining how LLMs arrive at certain conclusions
- Lack of Domain-Specific Knowledge: General-purpose LLMs may lack specialized forensic knowledge
- Bias and Fairness: Risk of unfair using of evidence

Challenges Of Applying LLMs In Justice

- Chain of Custody Issues: Challenges in maintaining and documenting the integrity of evidence when processed by LLMs
- Non-deterministic: LLMs can produce different responses to the same prompt, even under identical conditions.
- Prompt sensitivity: Subtle changes of the prompt may produce different results.
- Lack of Standardization: Absence of industry standards for using LLMs in forensic investigations.

Outline

- Introduction
- Hands on Tutorial
- Challenges of Leveraging LLM in Digital Forensics
- **Conclusion**

Potential Societal Impacts

Exploring the intersection of LLMs and digital forensics can drive meaningful societal change

- Promoting a deeper understanding of LLMs' potential in digital forensics aims to contribute to a safer, more equitable, and just society.
- Important to foster a culture of accountability and transparency in the digital real.

Enhancing Digital Forensics Evidence Analysis with Large Language Models

(KDD2025 Tutorial Session)

Eric Xu[1], **Lin Deng[2], and Damodar Dhital[2]**

[1]University of Maryland, College Park, Maryland, USA

[2]Towson University, Towson, Maryland, USA