ForensicGPT: Enhancing and Standardizing Digital Forensic Capabilities with RAG-Based LLMs

Doyoun Kim, Minsoo Kim, Philgeun Jin, Yunji Park and Doowon Jeong*
Department of Forensic Sciences, Sungkyunkwan University, Seoul, Republic of Korea

Introduction

Digital forensics is a multidisciplinary field requiring extensive expertise across various domains. However, investigative quality often varies due to differences in investigator capabilities, leading to inconsistencies in artifact interpretation, timeline generation, and report writing. Addressing this challenge, ForensicGPT integrates a Retrieval-Augmented Generation (RAG) approach into a Large Language Model (LLM), enabling continuous updates with the latest forensic knowledge.

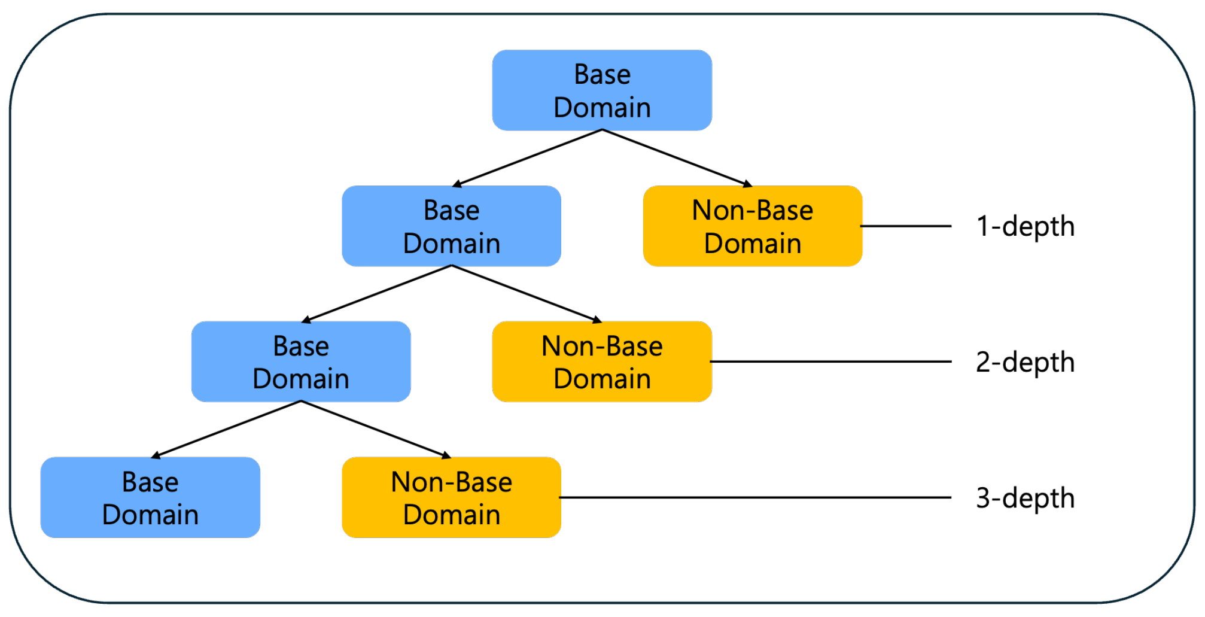
ForensicGPT enhances digital forensic investigations by delivering expert, consistent responses to investigators, regardless of their experience level, through a trusted knowledge repository. Additionally, it interprets unstructured data from various forensic tools, converting it into standardized formats to automate timeline generation and report creation. This not only mitigates investigator capability disparities but also ensures smooth transitions to newer LLM models while improving forensic report processing. Evaluations demonstrate that ForensicGPT has significant potential to enhance the reliability and consistency of digital forensic investigations.

Vector Database Generation

1. Crawl Data Recurively From the Base Domain using Selenium[1]

- **Base Domain**
  - Forensics Wiki (Web-based Information Sharing Platform)
- **Non-Base Domain**
  - Official Documents (Microsoft, Exterro, Magnet Forensic…)
  - Tool-Related Data (Github, Mitec, Code.google…)
  - Conference&Journal Data (DFRWS, Digital Investigation…)

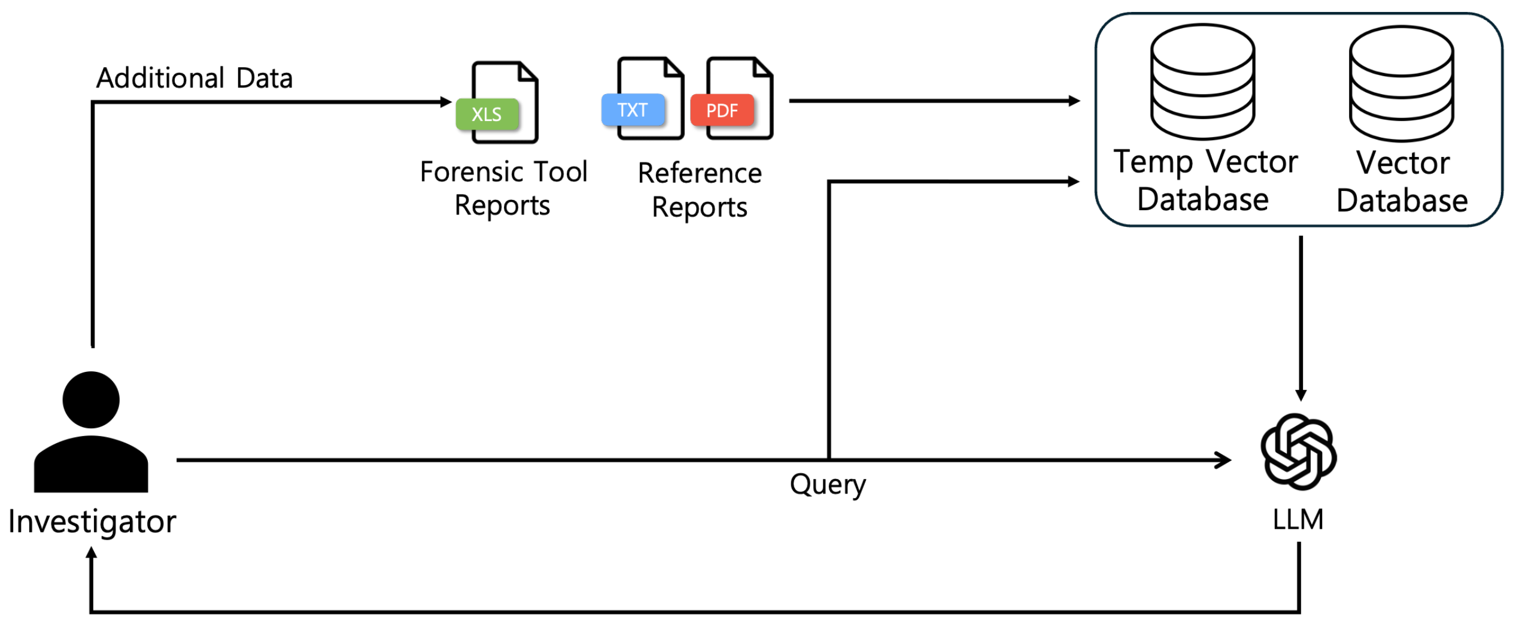
2. Construct Knowledge Repository (Vector DB) using FAISS[2]



ForensicGPT Implamantation

1. Professional Q&A on Digital Forensic Knowledge

- Input: User query + Additional Data (XLSX, PDF…/Optional)

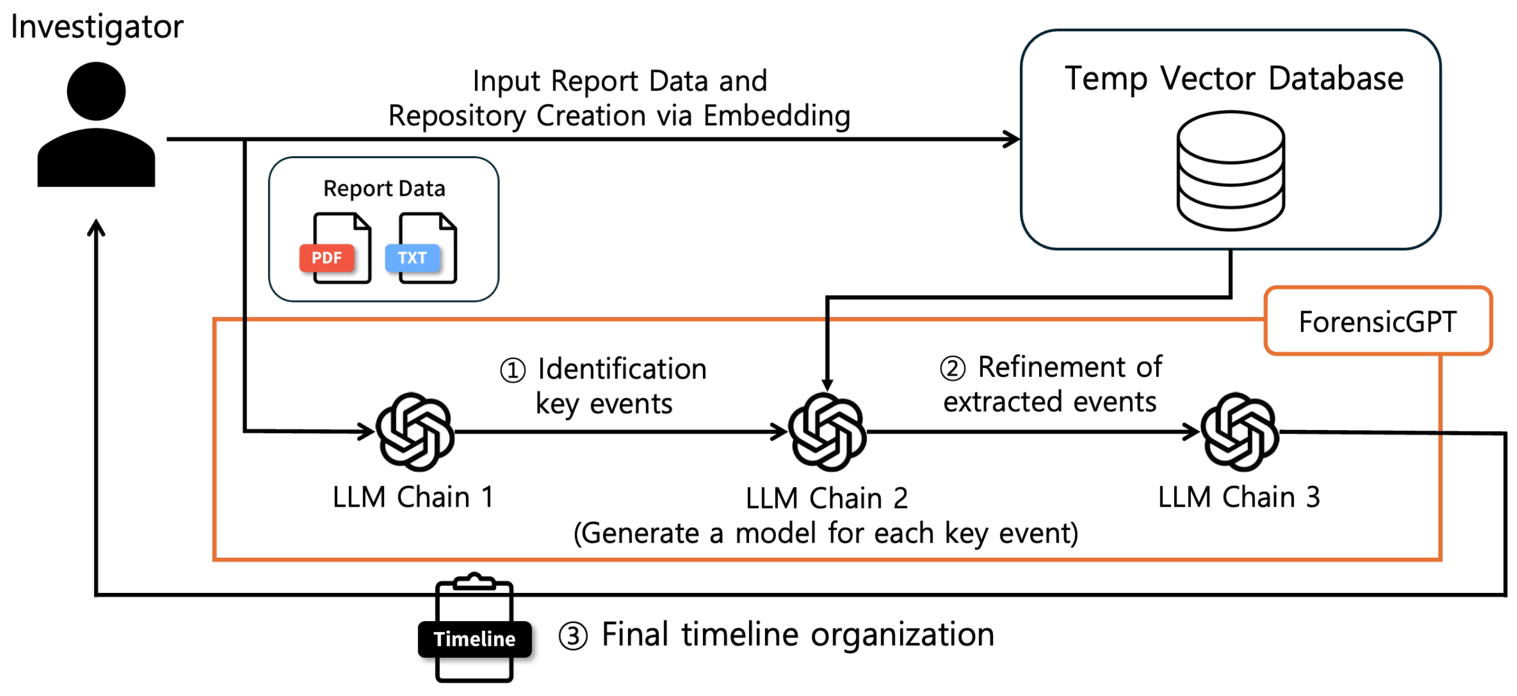


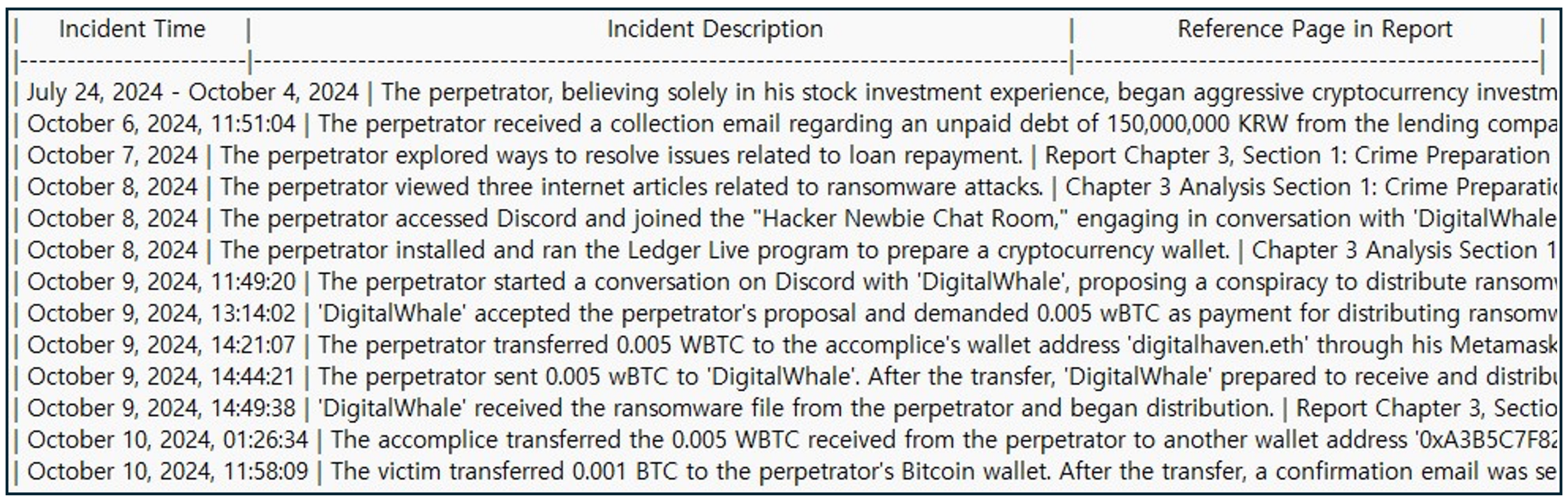
Description	When examining the data 0 file's cache entry of Chrome's disk cache in hex, provide the offset and size where the 'Array of Data Stream Cache Addresses' is located.
Gold Document	Offset : 0x56 from start of Cache Entry / Size : 16 bytes
ForensicGPT Response	……The 'Array of Data Stream Cache Addresses' is located at an offset of 56 bytes and is 16 bytes in size.
ChatGPT-4o Response	….. Depending on the specific structure, this is usually around 32 bytes from the start of the Cache Entry Block. Size: The Array of Data Stream Cache Addresses is typically 24 bytes, consisting of three 8-byte

[Example of Professional QnA on Digital Forensic Knowledge]

2. Timeline and Report Generation

- Input: Analysis Report Data (PDF, TXT)
- **2-1. Timeline Generation**

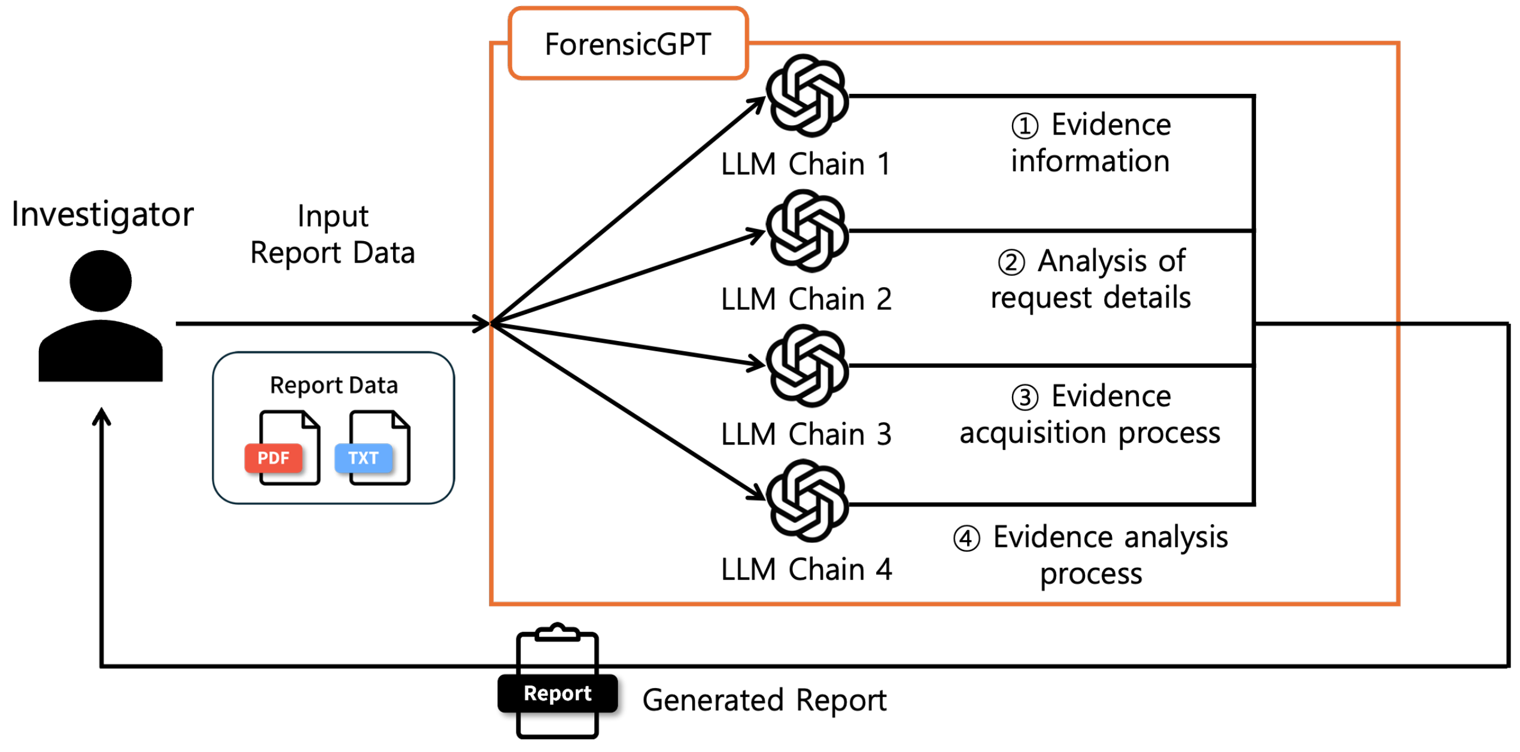

Investigator
Input Report Data and Repository Creation via Embedding
Temp Vector Database
Report Data
PDF
TXT
ForensicGPT
① Identification key events
② Refinement of extracted events
LLM Chain 1
LLM Chain 2
(Generate a model for each key event)
LLM Chain 3
Timeline
③ Final timeline organization

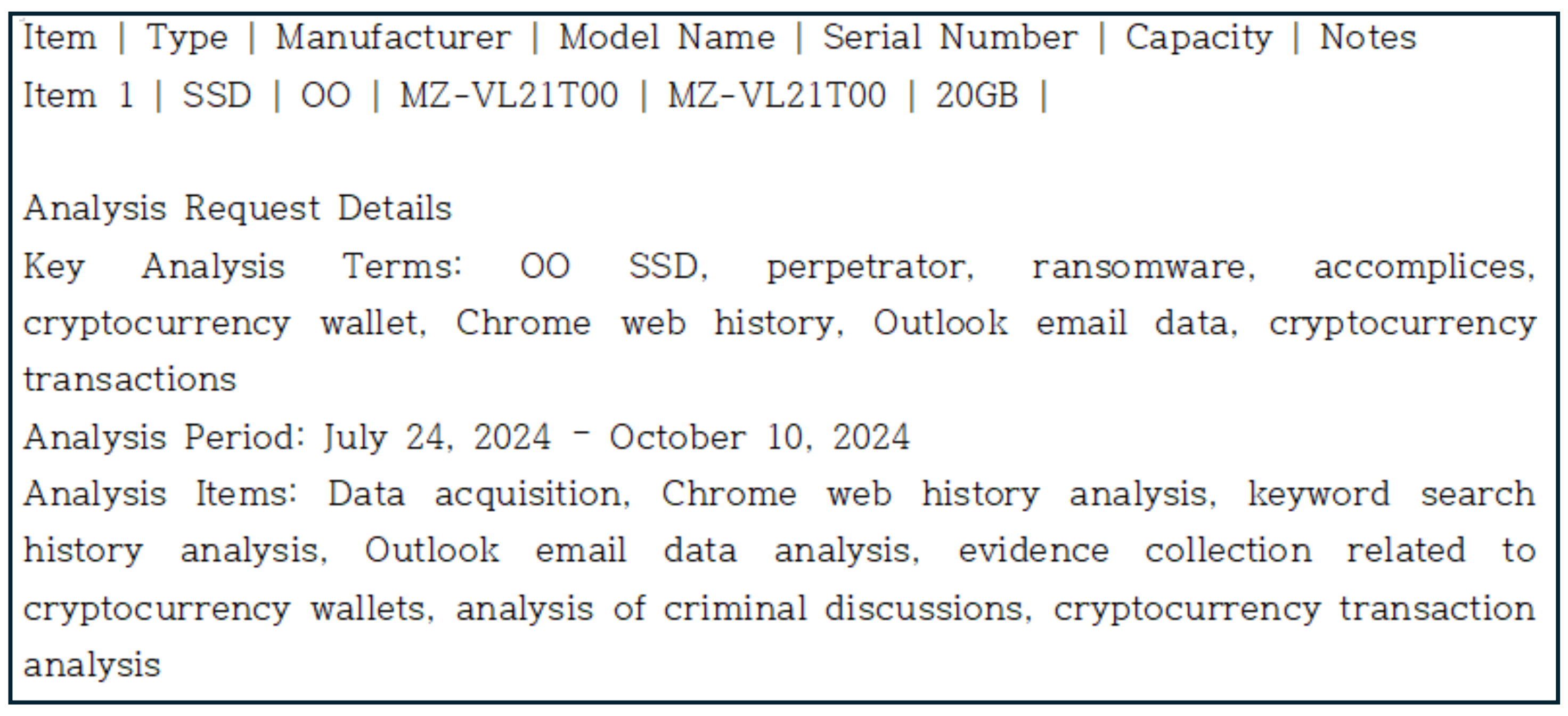


```
|    Incident Time     |                     Incident Description                     |          Reference Page in Report          |
|----------------------|--------------------------------------------------------------|--------------------------------------------|
| July 24, 2024 - October 4, 2024 | The perpetrator, believing solely in his stock investment experience, began aggressive cryptocurrency investm
| October 6, 2024, 11:51:04 | The perpetrator received a collection email regarding an unpaid debt of 150,000,000 KRW from the lending compa
| October 7, 2024 | The perpetrator explored ways to resolve issues related to loan repayment. | Report Chapter 3, Section 1: Crime Preparation
| October 8, 2024 | The perpetrator viewed three internet articles related to ransomware attacks. | Chapter 3 Analysis Section 1: Crime Preparati
| October 8, 2024 | The perpetrator accessed Discord and joined the "Hacker Newbie Chat Room," engaging in conversation with 'DigitalWhale
| October 8, 2024 | The perpetrator installed and ran the Ledger Live program to prepare a cryptocurrency wallet. | Chapter 3 Analysis Section 1
| October 9, 2024, 11:49:20 | The perpetrator started a conversation on Discord with 'DigitalWhale', proposing a conspiracy to distribute ransom
| October 9, 2024, 13:14:02 | 'DigitalWhale' accepted the perpetrator's proposal and demanded 0.005 wBTC as payment for distributing ransomv
| October 9, 2024, 14:21:07 | The perpetrator transferred 0.005 WBTC to the accomplice's wallet address 'digitalhaven.eth' through his Metamask
| October 9, 2024, 14:44:21 | The perpetrator sent 0.005 wBTC to 'DigitalWhale'. After the transfer, 'DigitalWhale' prepared to receive and distribu
| October 9, 2024, 14:49:38 | 'DigitalWhale' received the ransomware file from the perpetrator and began distribution. | Report Chapter 3, Sectio
| October 10, 2024, 01:26:34 | The accomplice transferred the 0.005 WBTC received from the perpetrator to another wallet address '0xA3B5C7F82
| October 10, 2024, 11:58:09 | The victim transferred 0.001 BTC to the perpetrator's Bitcoin wallet. After the transfer, a confirmation email was se
```

[Example of Generated Timeline]

- **2-2. Standardized Report Generation**





Item | Type | Manufacturer | Model Name | Serial Number | Capacity | Notes
Item 1 | SSD | OO | MZ-VL21T00 | MZ-VL21T00 | 20GB |

Analysis Request Details
Key Analysis Terms: OO SSD, perpetrator, ransomware, accomplices, cryptocurrency wallet, Chrome web history, Outlook email data, cryptocurrency transactions
Analysis Period: July 24, 2024 – October 10, 2024
Analysis Items: Data acquisition, Chrome web history analysis, keyword search history analysis, Outlook email data analysis, evidence collection related to cryptocurrency wallets, analysis of criminal discussions, cryptocurrency transaction analysis

[Example of Standardized Generated Report]

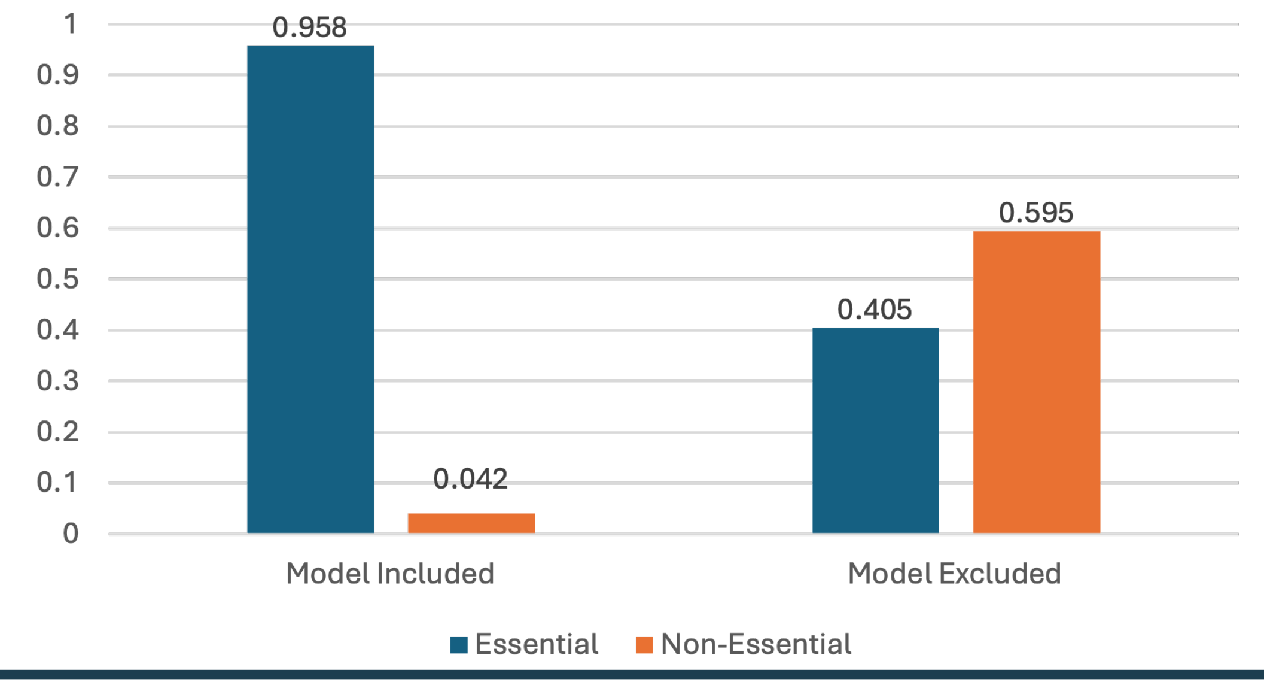
Evaluations

1. Professional Q&A on Digital Forensic Knowledge

- ForensicGPT achieved an average BLEU score of 0.5965, outperforming GPT-4 (0.2705), Claude 3.5 (0.2325), and Gemini 1.5 (1.900)

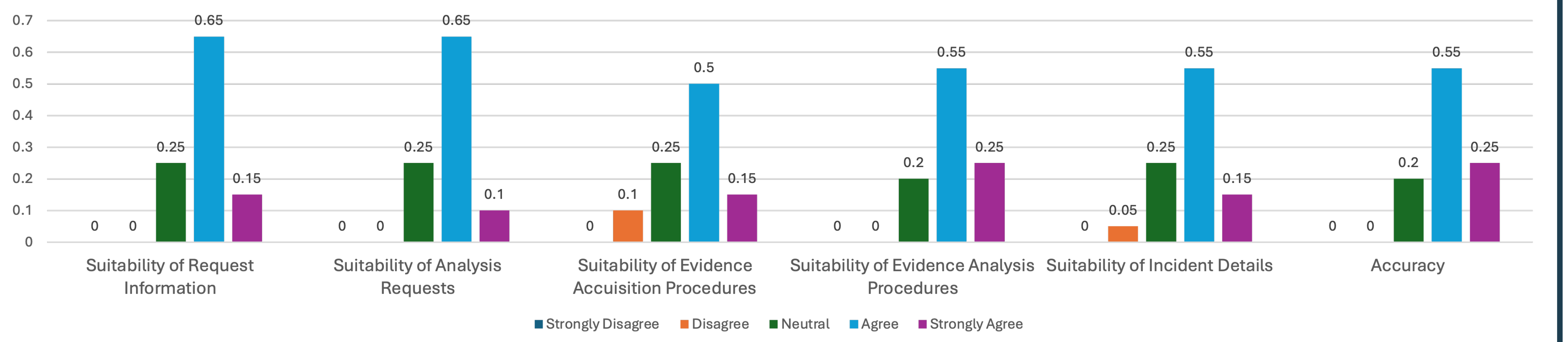
2. Timeline Generation

- Evaluation was conducted through a survey of 20 experts
- Compared the events that experts considered important with the timeline generated by ForensicGPT for a total of 34 events
- Experts evaluated 95.8% of the events included in timeline generated by ForensicGPT as necessary



3. Report Generation

- Evaluation was conducted through a survey of 20 experts based on six questions
  - Suitability of Request Information, Analysis Requests, Evidence Accuisition, Procedures, Evidence Analysis Procedures, Incident Details, and Accuracy
- Each question was rated on a scale of "Very Insufficient", "Insufficient," "Average," "Sufficient," and "Very Sufficient", with scores assigned from 1 to 5 respectively
- Evaluation revealed that, on average, approximately 75% of the participants found the standardized reports generated by ForensicGPT to be useful



[1] Selenium with Python, "Selenium with Python," Available: https://selenium-python.readthedocs.io/. [Accessed: 9-Mar-2025]
[2]FAISS, "FAISS: Facebook AI Similarity Search," Available: https://ai.meta.com/tools/faiss/. [Accessed: 9-Mar-2025]