연구계획서

디지털포렌식전공 박사과정 고상협(2025712714)

1. 제목(인터넷사기 중 돼지도살사기 유형) 및 목표 저널

- Pig Butchering 사기 신고 데이터에서 LLM 기반 TTP 프로파일링 및 멀티모달 정합성 점검을 통한 디지털 포렌식 인텔리전스 구조화(Structuring Digital Forensic Intelligence from Pig-Butchering Scams: LLM-Based TTP Profiling and Multimodal Consistency Checking)
- FSI:DI 등 SCI 저널 또는 SSCI 저널

2. 연구배경 및 현행 문제점

- 최근 급증하는 돼지도살 사기(Pig Butchering) 스캠은 피해규모가 막대하며, 매일 수많은 피해신고가 접수되고 있다.(DFPI 신고 건수 471건 중 273건, 55%에 해당) 그러나 이러한 데이터는 대부분 피해자의 기억과 메신저 기록 등에 의존한 텍스트 형태의 비정형 서사(narrative)로 존재한다. 현재 수사관은 이를 수작업으로 읽고 범죄 유형을 분류하고 있어, 범죄수법을 정형화하거나 연관된 사건을 식별하는데 심각한 병목이 발생하고 있다.
- 신고 데이터에는 텍스트 진술뿐만 아니라 사기 사이트, 사기 수익률 등 이미지가 포함되는 경우가 존재한다.(273건 중 25건) 그러나 현행 분석체계는 텍스트와 이미지를 분리하여 처리하고 있어, 피해자의 서사(설명)과 이미지 속 정보 간의 일치성을 체계적으로 점검하기 어렵다.

3. 연구방법: LLM기반 정보 정형화 및 분석보조 프레임워크

- 본 연구는 LLM(GPT-5.2 등 SOTA 모델)의 정보 추출 및 추론 능력을 활용하여 자가보고(Self-reported) 데이터를 수사 가능한 형태(Actionable Intelligence)로 변환하는 분석 보조 프레임워크이다.
- 연구 대상(데이터): 캘리포니아 금융보호혁신부(DFPI) ‘Crypto Scam Tracker’의 신고내용(피해자 서술)과 첨부자료 필드를 활용한다. 피해자의 서술이므로 주관적 기억 오류가 포함될 수 있으나, 실제 신고 데이터의 비정형성과 노이즈를 그대로 보존하고 있어 실제 수사 환경에서의 정보 추출 성능을 검증하기 위한 실증 자료

(empiricaldData)이다. 다만, 본 연구 데이터는 신고자의 자기보고 특성상 주관성이 개입될 수 있으나, 신고의 '사실 검증'을 하는 것이 아니라 수사 초기단계에서 비정형 정보를 인텔리전스로 변환하는 모델의 성능을 평가하는 데에는 생태적 타당성을 가진다.

- 연구 도구(분석툴): LLM을 활용한 텍스트 및 멀티모달 분석(실증연구)

- 세부 연구방법1(텍스트 분석): Fraud-TTPs(전술 Tactics, 기술 Techniques, 절차 Procedures) 및 수사 속성 자동 정형화 LLM에 고정된 속성 스키마(JSON Schema)와 Chain-of-Thought 프롬프트를 적용하여 피해자 신고 텍스트에서 핵심 수사정보를 추출한다. 접근 및 유인(Dating APP 등 초기 유도멘트), 사칭 및 심리(자산가 사칭 등 페르소나, 긴급성・로맨스 등 심리기제), 추적단서(피해자 진술 내에서 언급된 지갑주소 및 트랜잭션 해시 등을 구조화) 등을 수집하여 수행한다. 특히 LLM의 환각 현상을 통제하고 추출 성능을 객관적으로 검증하기 위해, 무작위 추출표본(10%)에 대해 인간 전문가와의 상호 일치도 검증을 수행한다.

- 세부 연구방법2(멀티모달 분석): 25건의 이미지 대상, 텍스트와 이미지 정합성 점검 및 트리아지(Triage) VLM을 활용하여 도메인, UI 패턴, 수익률 등을 추출하고 이를 진술 텍스트에서 추출된 속성과 비교하여 정합성 점수(0.0-1.0)를 산출한다. 신고 텍스트 상의 투자유형과 이미지 상의 플랫폼 성격이 일치하는지 확인하고, 불일치하는 경우 허위신고 단정이 아닌 추가 확인 필요 태깅을 하도록 설정한다.

4. 관련기술 선행연구

- (디지털포렌식 보고서 작성에서 LLM 활용의 효용과 한계) Michelet, G., & Breitinger, F. (2024). ChatGPT, Llama, can you write my report? An experiment on assisted digital forensics reports. FSI: Digital Investigation[1]: LLM을 수사관의 보조 도구로 활용할때의 효율성을 입증함과 동시에, 환각 리스크 통제를 위해 전문가 개입(Human-in-the-loop) 검증이 필수적임을 시사한다.

- (LLM을 활용한 경찰 리포트 내 핵심 요소 추출연구) Xing, Y., & Chen, F. (2024). Entity Extraction of Key Elements in 110 Police Reports Based on Large Language Models. Applied Sciences[3]: LLM을 이용해 비정형 경찰 신고데이터에서 사건의 핵심 엔티티(Entity)를 추출하고 정형화하는 방법론을 제시하였다.

- (돼지 도살사기의 범죄학적 구조 분석) Han, R., & Button, M. (2025). An anatomy of pig butchering scams: Chinese victims and police officers. Deviant Behavior[4]: Pig Butchering 범죄의 단계별 구조와 특성을 실증적으로 분석하였으

며, Fraud-TTPs 분류기준의 이론적 배경을 제공한다.

- (포렌식 시나리오 데이터셋 생성 및 활용) Voigt, M., et al. (2024). Re-imagen: Generating coherent background activity in synthetic scenario-based forensic datasets. FSI: Digital Investigation[2]: 포렌식 도구 검증을 위한 시나리오 기반 데이터 활용방법론을 제시하였으며, 분석 프레임워크 워크 검증시 활용가능한 평가지표이다.

5. 결론

- 본 연구는 피해자가 스스로 신고한 텍스트 데이터(비정형 서사)를 수사관이 즉시 활용가능한 구조화된 데이터로 변환하고, 텍스트와 이미지를 교차 점검하여 수사 효율을 극대화하는 분석보조도구를 제안한다. 공개된 실제 최신 신고 데이터의 특성과 한계(신고의 사실여부 미검증)를 명확히 인지하고, 이를 기술적으로 보완하는 프레임워크는 데이터 부족 문제를 해결함과 동시에 향후 국제 공조수사 및 피해자산 동결을 위한 기초 사실 확립에 기여할 것이다.