JDFR Journal of Data Forensics Research
데이터포렌식연구, 2권 1호, 2025년 6월, 71-93면
Journal of Data Forensics Research, Volume 2, Issue 1, June 2025, Pages 71-93.
https://doi.org/10.12972/JDFR.2025.2.1.5
Print ISSN 3058-9460
Online ISSN 3058-9479

원저

멀티모달 사이버 수사 데이터 처리를 위한 지능형 데이터 에이전트 설계 및 구현

박민회[1], 강준성[2], 최호식[3], 송경우[4]

[1]서울시립대학교 도시빅데이터융합학과 석사과정, [2]연세대학교 통계데이터사이언스학과 석사과정,
[3]서울시립대학교 도시빅데이터융합학과 교수, [4]연세대학교 응용통계학과, 통계데이터사이언스학과 교수

교신저자: 최호식, choi.hosik@uos.ac.kr; 송경우, kyungwoo.song@yonsei.ac.kr

요약

최근 사이버 범죄는 텍스트, 이미지, PDF 등 다양한 형태의 멀티모달 데이터를 활용하여 고도화되고 있으며, 이는 수사관의 수동 데이터 처리 부담을 가중시키고 있다. 본 연구는 이러한 복합적인 사이버 수사 데이터를 자율적으로 처리하여 수사관의 업무를 보조하고 수사 효율성을 극대화하는 지능형 데이터 에이전트(Data Agent)를 제안한다. 제안하는 에이전트는 입력 데이터의 유형을 스스로 판단하는 Vision Language Model(VLM)과, 문서 구조를 분석하여 핵심 정보를 추출하는 멀티모달 파서, 그리고 의미 기반으로 텍스트와 이미지의 중복을 정제하는 모듈로 구성된 통합 파이프라인을 갖는다. 시스템의 핵심 모듈은 DP-Bench, KorSTS, INRIA Holidays 등 공개 벤치마크 데이터셋을 활용한 엄격한 비교 실험을 통해 선정되었다. 마지막으로, 실제 사이버 범죄 샘플 데이터에 에이전트를 적용하여 복잡한 멀티모달 데이터를 자동 처리하는 전 과정의 실용적 가능성을 성공적으로 검증하였다. 본 연구는 사이버 수사 현장의 데이터 처리 자동화 및 지능화의 청사진을 제시하며, 향후 수사 역량 강화에 기여할 것으로 기대된다.

주제어

사이버 수사, 데이터 에이전트, 멀티모달 데이터, 문서파싱, 중복제거

Received: June 14, 2025
Revised: June 30, 2025
Accepted: June 30, 2025
Published: June 30, 2025

Original Article

An intelligent data agent for autonomous processing of multimodal digital forensic evidence: Design and implementation

Minhoi Park[1], Joonseong Kang[2], Hosik Choi[3], Kyungwoo Song[4]

[1]Master Student, Department of Urban Big Data Convergence, University of Seoul, Republic of Korea
[2]Master Student, Department of Statistics and Data Science, Yonsei University, Republic of Korea
[3]Professor, Department of Urban Big Data Convergence, University of Seoul, Republic of Korea
[4]Professor, Department of Applied Statistics and Department of Statistics and Data Science, Yonsei University, Republic of Korea

Corresponding Author: Hosik Choi, choi.hosik@uos.ac.kr; Kyungwoo Song, kyungwoo.song@yonsei.ac.kr

ABSTRACT

Cybercrime has become increasingly sophisticated, leveraging multimodal data such as text, images, and PDFs. This trend has significantly increased the manual data processing burden on investigators. This study proposes an intelligent Data Agent designed to autonomously process this complex cyber investigation data, thereby assisting investigators and maximizing investigative efficiency. The proposed agent features an integrated pipeline consisting of a Vision Language Model(VLM) that autonomously identifies the input data type, a multimodal parser that analyzes document structure to extract key information, and a module that refines data by de-duplicating text and images based on semantic similarity. The core modules of the system were selected through rigorous comparative experiments using public benchmark datasets, including DP-Bench, KorSTS, and INRIA Holidays. Finally, the agent was applied to real-world cybercrime sample data, successfully validating the practical feasibility of its end-to-end process for automatically handling complex multimodal data. This research presents a blueprint for the automation and intellectualization of data processing in the field of cyber investigation and is expected to contribute to enhancing future investigative capabilities.

KEYWORDS

Cyber Investigation, Data Agent, Multimodal Data, Document Parsing, Deduplication

I. 서론

1.1. 디지털 대전환 시대의 사이버 범죄와 수사 환경의 변화

인터넷의 발전으로 디지털 대전환은 우리 사회에 전례 없는 편의를 가져왔지만, 동시에 범죄의 양상 또한 근본적으로 바꾸어 놓았다. 사이버 공간은 더 이상 현실 세계의 보조적인 공간이 아닌, 경제 활동, 사회적 교류, 그리고 범죄 발생의 핵심 무대가 되었다. 통계청의 사이버 범죄 통계 현황[1]에 따르면 사이버 범죄 발생 건수는 해마다 급증하고 있으며, 그 수법 또한 날로 지능화, 고도화되고 있다. 과거의 단순 해킹이나 바이러스 유포를 넘어, 오늘날의 사이버 범죄는 금융 시스템을 교란하는 피싱, 개인의 심리를 교묘하게 파고드는 스미싱, 온라인 플랫폼의 신뢰를 악용하는 중고 거래 및 쇼핑몰 사기, 그리고 마약 유통이나 성 착취물 제작과 같은 중대 범죄의 온상으로까지 확장되고 있다.

이러한 범죄들은 사이버 공간의 익명성과 빠른 전파성을 무기 삼아 단시간에 막대한 피해를 야기한다. 범죄자들은 법망을 피하기 위해 대포폰, 가상 자산, 다크웹 등 추적이 어려운 기술을 적극적으로 활용하며, 범죄 증거 또한 전통적인 물리적 증거가 아닌 디지털 데이터 형태로 남긴다. 이로 인해 현대의 수사 환경은 과거와 비교할 수 없을 정도로 복잡해졌다. 수사관들은 이제 물리적 현장뿐만 아니라, 방대한 양의 디지털 데이터가 얽혀 있는 가상의 현장을 동시에 분석해야 하는 도전에 직면해 있다. 디지털 증거의 신속하고 정확한 확보 및 분석 능력은 이제 사이버 범죄 수사의 성패를 가르는 핵심적인 요소가 되었다.

1.2. 멀티모달 증거 데이터의 폭증과 수사관의 당면 과제

사이버 범죄 현장에서 수집되는 디지털 증거는 단일한 형태를 띠지 않는다. 범죄자들은 자신의 신원을 숨기고 피해자를 기만하기 위해 다양한 형태의 멀티모달 데이터를 복합적으로 사용한다. 예를 들어, 기관 사칭형 스미싱 범죄에서는 공신력을 위장하기 위해 공식 로고가 포함된 PDF 형태의 위조 공문이 사용되며, 중고 거래 사기에서는 피해자를 안심시키기 위해 정교하게 조작된 안전결제 페이지 스크린샷 이미지가 활용된다. 또한, 범죄 조직 내부에서는 메신저를 통해 텍스트로 범행을 모의하고, 계좌번호나 연락처와 같은 핵심 정보를 이미지 파일로 공유하여 추적을 피하기도 한다.

이처럼 텍스트, 이미지, PDF 등 다양한 형식으로 파편화된 멀티모달 데이터의 폭증은 일선 수사관들에게 심각한 실무적 부담을 안겨주고 있다. 현재 수사관들은 압수된 수백, 수천 개의 파일을 처리하기 위해 다음과 같은 수작업을 반복하고 있다.

- 수동 분류 및 열람: 각 파일의 확장자를 확인하고, 형식에 맞는 뷰어(이미지 뷰어, PDF 리더, 텍스트 편집기 등)를 사용하여 내용을 일일이 열람하고 사건과의 관련성을 판단한다.
- 수동 정보 추출: 스크린샷이나 이미지 파일에 포함된 중요한 텍스트 정보(계좌번호, URL, 연락처 등)를 눈으로 확인하고 직접 타이핑하여 옮겨 적는다. 이 과정에서 오타나 누락이 발생할 위험이 상존한다.
- 수동 중복 검사: 내용이 유사하거나 동일한 파일들을 식별하기 위해 파일들을 번갈아 가며 육안으로 비교한다. 이는 파일의 수가 많아질수록 기하급수적으로 어려워지며, 미세한 차이를 가진 중요한 변형 증거를 놓칠 가능성이 크다.

이러한 과정은 사건의 초기 단계에서 막대한 시간과 노력을 소모시켜 신속한 대응을 저해하는 병목 현상을 유발한다. 이는 결국 수사관이 용의자 추적, 증거 간의 연관 관계 분석, 범죄 스

키마 파악 등 본질적인 분석 및 추론 업무에 집중할 시간이 부족해지는 결과를 초래한다.

1.3. 기존 데이터 처리 기술의 한계와 특화된 접근의 필요성

수사관의 업무 부담을 덜기 위해 기존의 범용 데이터 처리 기술을 도입하려는 시도가 있었지만, 이는 몇 가지 본질적인 한계에 부딪힌다. 첫째, 대부분의 범용 자연어 처리나 컴퓨터 비전 모델들은 일상적인 언어나 표준적인 이미지에 대해 학습되어, 사이버 범죄 현장에서 사용되는 특수한 은어, 비속어, 비정형적인 텍스트 레이아웃을 정확하게 이해하지 못한다.

둘째, 기존의 문서 파싱 도구들은 정형화된 보고서나 논문 형식에 최적화되어 있어, 채팅 대화 캡처 이미지나 여러 요소가 무질서하게 혼합된 증거 자료의 구조를 올바르게 분석하지 못하는 경우가 많다. 이로 인해 중요한 정보가 누락되거나 잘못된 형태로 추출될 수 있다.

마지막으로, 가장 중요한 한계는 '보수적 처리' 원칙의 부재이다. 앞서 언급했듯이, 수사 데이터는 단 하나의 단서도 소실되어서는 안 된다. 기존의 중복 제거 기술은 저장 공간 효율화나 데이터 클리닝을 목표로 하므로, 중복의 기준이 비교적 관대하여 의미적으로 중요한 차이가 있는 데이터마저 동일한 것으로 처리하고 삭제할 위험이 있다. 예를 들어, Charikar (2002)[2]가 제안한 SimHash와 같은 전통적인 해시 기반 방식은 데이터의 일부만 변경되어도 해시 값이 크게 달라져 유사성을 탐지하지 못하거나, 반대로 다른 내용의 데이터가 우연히 유사한 해시 값을 가질 수 있다. 따라서, 사이버 수사 데이터의 특수성을 이해하고, 데이터의 무결성을 최우선으로 고려하며, 동시에 자동화의 이점을 제공할 수 있는 특화된 접근 방식이 반드시 필요하다.

1.4. 지능형 데이터 에이전트의 필요성 및 연구 목표

본 연구는 위에서 제기된 문제들을 해결하기 위한 방안으로, 지능형 데이터 에이전트의 도입을 제안한다. 여기서 에이전트란 Russell & Norvig (2020)[3]의 정의에 따라, 주어진 목표를 달성하기 위해 환경을 인식하고 스스로의 판단에 따라 행동하는 자율적 시스템을 의미한다. 본 연구에서 제안하는 데이터 에이전트는 '파편화된 멀티모달 수사 데이터를 분석 가능한 형태로 정제한다'는 명확한 목표를 가지며, 이를 달성하기 위해 분류, 파싱, 중복 제거 등의 작업을 인간의 개입 없이 자율적으로 수행한다.

이러한 목표를 달성하기 위해, 본 연구는 다음과 같은 세부 목표를 설정하였다.

■ 자율적 처리 파이프라인 아키텍처 설계: 최신 VLM을 의사결정의 첫 단계에 배치하여 입력 데이터의 특성에 따라 처리 흐름을 동적으로 제어하고, 각 단계별로 특화된 전문 AI 모듈들을 유기적으로 연결하는 새로운 방식의 자율 에이전트 아키텍처를 설계한다.

■ 핵심 모듈의 객관적 성능 검증 및 선정: 다양한 오픈소스 AI 모듈들을 한국어가 포함된 공개 벤치마크 데이터셋을 통해 객관적으로 평가한다. 이 실증적 결과를 바탕으로, 제안하는 에이전트의 각 기능(파싱, 중복 제거 등)에 가장 적합한 최적의 모듈을 선정하고 그 근거를 명확히 제시한다.

■ 통합 시스템의 실용적 타당성 검증: 최종적으로 통합된 데이터 에이전트를 실제 사이버 범죄와 유사한 데이터 시나리오에 적용하여, 자동화가 의도대로 작동하는지, 그리고 복잡한 멀티모달 데이터를 효과적으로 처리할 수 있는지를 보여줌으로써 시스템의 실용적 가치와 가능성을 입증한다.

II. 관련연구

2.1. Vision-Language Model의 발전과 응용

Vision-Language Model(VLM)은 시각 정보(이미지)와 언어 정보(텍스트)를 동시에 이해하고 처리하는 인공지능 모델로서, 최근 몇 년간 괄목할 만한 발전을 이루었다. VLM의 발전은 크게 세 단계로 구분할 수 있다.

첫 번째 Representation Learning 단계이다. Radford et al. (2021)[4]이 제안한 CLIP(Contrastive Language-Image Pre-training)은 이 단계의 대표적인 모델이다. CLIP은 수억 개의 이미지-텍스트 쌍 데이터를 활용하여, 이미지 인코더와 텍스트 인코더가 각각의 정보를 동일한 차원의 벡터 공간에 투영하도록 학습한다. 이 과정에서 서로 연관된 이미지와 텍스트 쌍의 벡터는 가깝게, 연관 없는 쌍은 멀어지도록 하는 Contrastive Learning 방식을 사용한다. 이를 통해 모델은 별도의 라벨링 없이도 이미지와 텍스트 간의 복잡한 의미적 관계를 학습하게 되며, 이는 제로샷 이미지 분류와 같은 새로운 응용 가능성을 열었다.

두 번째 단계는 VLM 기술을 특정 도메인에 적용하고 정교화하는 단계이다. 특히 문서 이해 분야에서 VLM은 핵심적인 역할을 수행했다. Xu et al. (2020)[5]이 제안한 LayoutLM은 BERT와 같은 언어 모델에 텍스트의 시각적 위치 정보와 레이아웃 정보를 추가하여, 표와 같이 구조가 중요한 문서의 이해도를 획기적으로 높였다. Kim et al. (2022)[6]이 제안한 Donut 모델은 한발 더 나아가, 별도의 광학 문자 인식(OCR) 엔진 없이 문서 이미지를 직접 입력받아 내용과 구조를 텍스트 시퀀스로 변환하는 자율처리 방식을 구현하였다. 이러한 연구들은 VLM이 단순한 이미지 분류를 넘어 복잡한 구조를 가진 정보를 처리할 수 있는 잠재력을 보여주었다.

세 번째 단계는 대규모 멀티모달 모델(LLM)과 지시 이행 능력의 등장이다. GPT-4V, Gemini와 같은 최신 LLM들은 사전 훈련된 LLM을 '뇌'로, Vision Transformer와 같은 이미지 인코더를 '눈'으로 결합한 구조를 가진다. 이를 통해 모델은 이미지를 '보고' 그 내용에 대해 인간과 자연어로 대화하거나, 복잡한 지시를 수행하는 것이 가능해졌다. Tu et al. (2024)[7]이 제안한 MiniCPM 역시 이러한 지시 이행 LLM의 한 종류이다. 본 연구는 이러한 최신 LLM의 지시 이행 능력을 사이버 수사 데이터 처리라는 특수 목적에 응용하여, MiniCPM을 '이미지 내 텍스트 유무를 판단하는 눈을 가진 의사결정자'로 활용하고 이를 에이전트의 핵심 제어 요소로 삼았다. 이는 VLM을 데이터 처리 워크플로우의 동적 제어 요소로 통합했다는 점에서 기존의 정적인 파이프라인과 차별화된다.

2.2. AI 에이전트의 진화와 자율적 워크플로우

2.2.1. AI 에이전트의 개념과 도구 사용 패러다임

AI 에이전트는 LLM의 등장과 함께 새로운 국면을 맞이하였다. LLM의 강력한 언어 이해 및 추론 능력을 바탕으로, 복잡한 목표를 달성하기 위해 스스로 계획을 세우고 외부 도구를 자율적으로 호출하여 사용하는 '도구 사용 에이전트' 패러다임이 AI 연구의 핵심 조류로 부상했다.

이 패러다임의 핵심은 LLM이 단순히 응답을 생성하는 것을 넘어, 행동의 주체로 작동한다는 점이다. ReAct(Reasoning and Acting)와 같은 프레임워크는 LLM이 목표를 달성하기 위해 '생각→ 행동 → 관찰'의 순환 구조를 반복하도록 설계되었다. 예를 들어, "오늘 서울 날씨를 검색해서 알려줘"라는 요청에 대해 에이전트는 (1) 생각: 날씨를 알려면 검색 도구가 필요함을 인지, (2) 행동: 검색 API에 '서울 날씨'를 쿼리로 호출, (3) 관찰: API로부터 날씨 정보를 반환받

음, (4) 최종 응답 생성의 과정을 거친다. 이러한 능력은 에이전트가 LLM 내부의 지식만으로는 해결할 수 없는 최신 정보 검색, 복잡한 계산, 외부 시스템과의 상호작용 등 다양한 문제를 해결할 수 있게 한다.

2.2.2. 데이터 과학 및 수사 분야의 에이전트

AI 에이전트의 활용은 범용 작업을 넘어 특정 전문 분야로 확장되고 있다. Cao et al. (2024)[8]는 Spider2-V를 통해 멀티모달 에이전트가 데이터 사이언스 및 엔지니어링 워크플로우 자동화에 얼마나 근접했는지를 평가했다. 이 연구는 전문적인 데이터 분석에 초점을 맞춘 최초의 멀티모달 에이전트 벤치마크를 제시하며, 데이터 처리와 분석을 자동화하는 데이터 파이프라인에 대한 새로운 관점을 제공했다.

본 연구에서 제안하는 '데이터 에이전트'는 이러한 연구 흐름과 맥을 같이 한다. 특히, 범용 데이터 과학이 아닌 '사이버 수사'라는 특수하고 중요한 도메인에 초점을 맞추고, 데이터 분석의 전 단계인 '데이터 전처리 및 정제' 과정을 자동화하는 에이전트를 제안한다는 점에서 차별성을 가진다. 본 에이전트는 VLM(MiniCPM), 파서(MinerU), 중복 제거기(Unisim, Fiftyone)라는 명확한 '도구'들을 정해진 파이프라인에 따라 순차적으로 호출하여, 멀티모달 수사 데이터를 분석 가능한 형태로 가공하는 구체적인 목표를 달성한다. 이는 AI 에이전트 기술을 실제 사회 문제 해결에 적용한 실용적인 응용 사례라 할 수 있다.

2.3. 문서 파싱 및 중복 제거 기술

2.3.1. 문서 파싱 기술 및 벤치마크

문서 파싱은 문서 이미지에서 텍스트, 표, 그림 등의 논리적 구성 요소를 식별하고 그 내용을 추출하는 기술이다. 앞서 언급한 LayoutLM[5]이나 Donut[6]과 같은 모델들의 성능을 객관적으로 평가하기 위해 DP-bench, OmniDocBench와 같은 표준화된 벤치마크가 제안되었다. 이러한 벤치마크들은 TEDS(Tree-Edit-Distance-based Similarity)나 NID(Normalized Intersection-over-Union for Detections)와 같은 지표를 사용하여 파싱의 정확도를 측정한다. 본 연구에서 핵심 모듈로 채택한 MinerU[9]는 이러한 벤치마크들에서, 특히 표와 이미지 등 구조적 요소들을 분리하는 능력에서 높은 성능을 입증받은 오픈소스 솔루션이다.

2.3.2. 데이터 중복 제거 기술

데이터 중복 제거는 방대한 데이터셋에서 동일하거나 유사한 데이터를 식별하여 정제하는 기술이다.

텍스트 중복 제거 분야에서는, Charikar (2002)[2]가 제안한 SimHash나 Broder (1997)[10]의 연구에 기반한 MinHash와 같은 해시 기반 기법이 속도의 이점으로 인해 오랫동안 사용되어 왔다. 그러나 이러한 방법들은 의미적 유사성을 포착하는 데 한계가 있었다. Devlin et al. (2018)[11]이 제안한 BERT와 같은 문맥 기반 언어 모델은 텍스트를 의미가 풍부한 고차원 벡터로 변환하는 것을 가능하게 했고, 이를 통해 텍스트 간의 의미적 유사도를 정교하게 측정할 수 있게 되었다. 본 연구에서 사용하는 Unisim[12]은 이러한 의미 기반 임베딩을 활용하여, 표현은 다르지만 의미가 같은 문장들을 효과적으로 찾아낸다.

이미지 중복 제거 분야 역시 지각 해시 방식에서 딥러닝 방식으로 발전해왔다. 지각 해시는 이미지의 전반적인 시각적 특징을 요약한 해시 값을 생성하여 비교하는 방식이지만, 회전이나

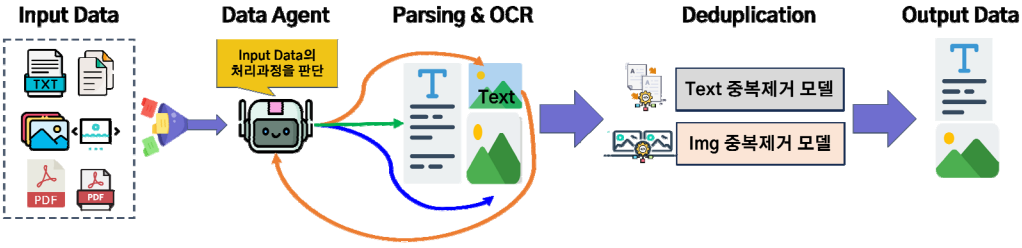
잘라내기 등 변형에 취약할 수 있다. 반면 딥러닝 방식은 사전 훈련된 CNN 모델을 사용하여 이미지로부터 불변의 특징을 추출하고, 이 벡터들 간의 거리를 측정한다. Fiftyone[13]은 이러한 딥러닝 임베딩 방식을 채택하여, 다양한 시각적 변형에도 강건하게 유사 이미지를 탐지하는 능력을 제공한다. 본 연구는 수사 데이터의 특수성을 고려하여, 두 모달리티 모두에서 더 정교한 의미/시각 기반의 최신 방법론을 채택하고 그 성능을 검증하였다.

III. 지능형 데이터 에이전트 설계

본 장에서는 앞서 제기된 사이버 수사 데이터 처리의 한계점들을 극복하기 위해 제안하는 지능형 데이터 에이전트의 전체 아키텍처와 각 핵심 모듈의 상세한 설계 내용을 기술한다. 본 에이전트는 자율성, 모듈성, 확장성을 핵심 설계 원칙으로 삼았다.

3.1. 제안하는 데이터 에이전트의 전체 시스템 아키텍처

제안하는 데이터 에이전트는 복잡하고 다양한 멀티모달 데이터를 인간의 개입 없이 자동으로 처리하는 것을 목표로 설계되었다. 이를 위해 <Figure 1>과 같이 '판단 → 추출 → 정제'의 3단계로 구성된 논리적 파이프라인 아키텍처를 채택하였다. 이 구조는 AI 에이전트가 주어진 데이터를 인식하고, 목표(데이터 정제)를 달성하기 위해 순차적으로 최적의 도구(모듈)를 사용하는 '도구 사용 에이전트'의 개념을 구체화한 것이다.



<Figure 1> Intelligent Data Deduplication Architecture: Multimodal Content Processing Pipeline for Text and Image Duplicate Detection

■ 입력: 수사 과정에서 수집된 다양한 형식(TXT, JPG, PNG, PDF 등)의 원본 데이터는 에이전트의 입력 전처리 단계를 거친다. 이 단계에서 모든 텍스트 기반 파일(TXT 등)은 텍스트 스트림으로, 모든 시각 기반 파일(JPG, PNG 등)은 이미지 스트림으로 통합된다. 특히, PDF 파일의 경우, 각 페이지가 별도의 이미지 파일로 변환된 후 이미지 스트림에 합류한다. 결과적으로 에이전트는 이후 모든 처리를 '텍스트'와 '이미지'라는 두 가지 표준화된 데이터 형태를 대상으로 시작한다.

■ 판단 : 에이전트의 첫 번째 자율적 의사결정 단계이다. 입력된 파일이 이미지일 경우, VLM 기반 분류 모듈이 해당 이미지 내에 분석 가치가 있는 텍스트의 포함 여부를 판단한다. 이 결과를 바탕으로 데이터는 'Pure Image' 또는 'Mixed (Image with Text)'로 분류되어 다음 단계로의 처리 경로가 결정된다. 이 판단 단계는 계산 비용이 높은 OCR 처리를 모든 이미지에 무분별하게 적용하는 비효율을 방지하는 핵심적인 역할을 한다.

■ 추출: 판단 단계의 결과에 따라 데이터에서 핵심 정보를 추출한다. 'Mixed'로 분류된 이미지는 멀티모달 파싱 모듈로 전달되어, 문서 내의 텍스트, 표, 이미지 등의 구조적 요소들이 분리

및 추출된다. 순수 텍스트 파일은 별도의 파싱 없이 이 단계의 출력으로 통합된다.

■ 정제: 추출 단계에서 생성된 모든 텍스트 데이터와 이미지 데이터는 각각의 중복 제거 모듈로 전달된다. 이 단계에서는 의미적, 시각적 유사도를 기반으로 중복되거나 매우 유사한 데이터들이 식별되고 그룹화된다.

■ 출력: 최종적으로 중복이 제거되고 정제된 텍스트 데이터와 고유한 이미지 데이터셋이 구조화된 형태로 출력된다. 이 결과물은 수사관이 다음 단계의 분석 및 추론에 즉시 활용할 수 있는 형태를 가진다.

3.2. 핵심 모듈별 상세 설계 및 구현

3.2.1. VLM 기반 분류 모듈 (판단)

■ 역할 및 목표: 본 모듈은 에이전트의 '눈'과 '초기 판단 두뇌' 역할을 수행한다. 입력된 이미지 데이터가 후속 OCR 처리가 필요한 'Mixed' 데이터인지, 아니면 이미지 자체로만 의미가 있는 'Pure' 데이터인지를 분류하는 것을 목표로 한다. 이는 전체 파이프라인의 효율성을 결정하는 중요한 관문이다.

■ 모델 선정 및 이유: 분류기로는 경량화된 고성능 LLM인 MiniCPM-V2.6[7]을 채택했다. 이 모델은 비교적 적은 리소스로도 우수한 이미지 이해 및 지시 이행 능력을 보여주어, 에이전트 시스템에 통합하기에 적합하다고 판단했다.

■ 프롬프트 엔지니어링 과정: LLM의 출력을 안정적으로 제어하기 위해 프롬프트 엔지니어링은 필수적이다. 초기 실험에서 "Does this image contain any text? Answer with only one word: 'pure' or 'mixed'"와 같은 단순한 프롬프트를 사용했을 때, 모델은 PDF에서 변환된 이미지와 같이 명백히 텍스트가 포함된 이미지도 'Pure'로 오분류하는 경향을 보였다. 이는 모델이 'text'라는 단어의 개념을 인간의 의도와 다르게 해석했기 때문으로 분석되었다.

```
# 개선된 프롬프트
question = {
    "You are an image classifier that only answers 'pure' or 'mixed'.₩n"
    "- pure: no readable text in the image₩n"
    "- mixed: any readable text present₩n"
    "Answer exactly one word."
}
```

이 문제를 해결하기 위해, 모델에게 명확한 역할과 출력 형식, 그리고 각 용어에 대한 구체적인 정의를 제공하는 few-shot 방식의 프롬프트를 새롭게 설계했다.

이처럼 역할(image classifier), 출력 단어('pure' or 'mixed'), 그리고 각 단어의 명확한 정의(no readable text / any readable text)를 제공함으로써, 모델은 높은 정확도로 의도한 분류 작업을 수행할 수 있었다. 이는 복잡한 AI 모델을 실제 시스템에 통합할 때, 명확한 지시와 제어가 얼마나 중요한지를 보여준다.

3.2.2. 멀티모달 파싱 모듈 (추출)

■ 역할 및 목표 :본 모듈은 에이전트의 '손' 역할을 하며, PDF와 같은 복합 문서나 텍스트가

포함된 이미지에서 구조화된 정보를 추출하는 것을 목표로 한다. 즉, 문서 내에서 텍스트 블록, 표, 그림 등의 논리적 단위를 정확하게 식별하고 분리하여 다음 단계에서 처리하기 용이한 형태로 변환한다.

■ 모델 선정 및 이유 : MinerU[9]는 특히 이미지와 표를 원본 형태 그대로 정확하게 분리해내는 능력에서 타 모델 대비 압도적인 성능을 보였다. 수사 증거로서 이미지나 표의 시각적 레이아웃 자체가 중요한 정보를 담고 있을 수 있다는 점을 고려할 때, 이는 매우 중요한 장점이다.

3.2.3. 데이터 중복 제거 모듈 (정제)

■ 설계 원칙

• 보수적 접근: 본 모듈의 최우선 설계 원칙은 '보수성'이다. 즉, 중복이 아닌 데이터를 중복으로 판단하는 오류(False Positive)를 최소화하는 것을 목표로 한다. 이를 위해 각 하위 모듈의 유사도 임계값을 비교적 높게 설정하여, 명백하게 동일하거나 매우 유사한 콘텐츠만을 대상으로 정제를 수행하도록 설계하였다.

■ 텍스트 중복 제거

• 알고리즘: 텍스트 중복 제거 모듈은 파싱 및 OCR 단계를 거쳐 여러 원본 파일로부터 생성된 전체 텍스트 데이터 집합을 입력으로 받아, 파일의 경계를 넘어 의미적으로 중복되는 내용을 식별하고 정제하는 것을 목표로 한다. 이 과정은 다음과 같이 진행된다. 첫째, 전체 텍스트 집합 내의 각 문장 또는 문단은 Unisim[12]의 언어 모델을 통해 고차원의 의미 벡터로 변환된다. 둘째, 생성된 모든 벡터들은 서로 간의 코사인 유사도를 아래 수식에 따라 계산한다.

$$similarity = \cos(\theta) = \frac{A \cdot B}{\| A \| \ \| B \|}$$

셋째, 계산된 유사도 값이 사전에 설정된 임계값 0.85을 초과하는 텍스트들은 '의미적으로 동일한' 콘텐츠로 판단되어 하나의 중복 클러스터로 그룹화된다. 이 클러스터링 과정은 서로 다른 파일에서 추출되었더라도 내용이 유사하다면 하나의 그룹으로 묶어준다. 마지막으로, 각 클러스터에서는 대표 텍스트로 선정하고 나머지는 중복으로 처리함으로써, 전체 파일셋에 대한 정제된 고유 콘텐츠 목록을 최종적으로 생성한다.

• 임계값 조정: 4.2.3절의 오류 사례 분석에서, Unisim이 한국어의 유연한 어순이나 미묘한 의미 차이를 완벽하게 구분하지 못하는 경우가 관찰되었다. 이에 따라, 본 연구에서는 임계값을 0.85 이상으로 설정하여, 표현이 거의 동일한 수준이 아닌 이상 중복으로 판단하지 않도록 하였다. 이는 잠재적 단서가 될 수 있는 유사 문장을 보존하기 위한 보수적 선택이다.

■ 이미지 중복 제거

• 알고리즘: 파싱된 모든 이미지와 원본 이미지 파일들은 Fiftyone[13] 프레임워크를 통해 처리된다. Fiftyone은 내부적으로 ImageNet 등으로 사전 훈련된 CNN 모델을 사용하여 각 이미지로부터 2,048차원 등의 특징 벡터를 추출한다. 이후 텍스트와 마찬가지로 이미지 벡터들 간의 코사인 유사도를 계산하여 중복 여부를 판단한다.

• 임계값 조정: 이미지 역시 배경이나 구도가 비슷하다는 이유로 다른 피사체를 담은 이미지가 중복으로 처리되는 오류를 방지하기 위해, 유사도 임계값을 0.98 이상으로 엄격하게 설정하였다. 에이전트는 중복으로 판단된 이미지를 로그데이터로 기록하며, 중복된 이미지를 제거하여 관리한다.

이처럼 각 모듈은 뚜렷한 목표와 엄격한 기준 하에 설계 및 구현되었으며, 이들의 유기적인 결합을 통해 데이터 에이전트는 복잡한 멀티모달 데이터를 자율적으로 처리하는 능력을 갖추

게 된다.

IV. 실험 및 평가

본 장에서는 3장에서 설계한 지능형 데이터 에이전트의 핵심 모듈들의 성능을 객관적으로 검증하고, 최종 통합된 시스템의 실용적 타당성을 확인하기 위해 수행한 일련의 실험과 그 결과를 상세히 기술한다. 본 장의 모든 실험 설계, 과정 및 결과는 선행 연구에 기반한다.

4.1. 실험 설계

4.1.1. 평가 목표

본 연구의 실험은 다음과 같은 세 가지 주요 목표를 달성하기 위해 설계되었다.

■ 최적 파싱 모듈 선정: 다양한 형태의 한국어 멀티모달 문서에 대해, 여러 오픈소스 파서들의 구조적 정보(이미지, 표) 추출 능력과 OCR 정확도를 객관적으로 측정하고 비교하여, 에이전트의 '추출' 단계에 가장 적합한 모듈을 선정한다.

■ 최적 중복 제거 모듈 선정: 텍스트와 이미지 데이터 각각에 대해, 전통적인 해시 기반 알고리즘과 최신 의미/시각 기반 알고리즘의 성능을 정량적으로 평가한다. 이를 통해 '보수적 처리' 원칙에 부합하면서도 가장 정확도가 높은 중복 제거 모듈을 선정한다.

■ 통합 에이전트 기능 검증: 최종 선정된 모듈들을 통합한 데이터 에이전트가 실제 사이버 범죄 데이터 시나리오에서 의도한 대로 자율적으로 작동하는지, 그리고 복잡한 멀티모달 데이터를 효과적으로 처리할 수 있는지 그 기능적 타당성을 검증한다.

4.1.2. 평가 데이터셋

각 평가 목표를 달성하기 위해 다음과 같이 특성이 다른 데이터셋을 활용하였다.

■ 문서 파싱용 데이터셋: 텍스트 중심, 표 중심, 그리고 텍스트/이미지/표가 복합적으로 구성된 5종의 실제 한국어 문서를 자체적으로 구축하여 사용하였다. 여기에는 2025학년도 대학수학능력시험 문제지(국어, 수학), 네이버 감사 보고서, KCI 등재 논문 등이 포함되어, 다양한 레이아웃에 대한 파서의 강건성을 평가하고자 하였다.

■ 텍스트 중복 제거 데이터셋: 대규모 영어 데이터에 대한 성능을 측정하기 위해 Wiki-40B 데이터셋을, 한국어 문장의 미묘한 의미 차이를 평가하기 위해 KakaoBrain KorSTS[14] 데이터셋을 사용하였다.

■ 이미지 중복 제거 데이터셋: Jégou et al. (2008)[15]이 제안한 INRIA Holidays 데이터셋을 사용하였다. 이 데이터셋은 실제 촬영된 사진들로 구성되어 있어, 조명, 시점, 크기 변화 등 현실적인 변형에 대한 알고리즘의 성능을 평가하는 데 적합하다.

■ 종합 시나리오 데이터셋: 실제 사이버 범죄 신고 및 증거 자료를 익명화하여 제공하는 수사관련 샘플 데이터를 활용하였다. 여기에는 스미싱 문자 내용, 피싱 사이트 스크린샷, 사기 유형 안내 PDF 등 본 연구가 해결하고자 하는 문제와 직접적으로 관련된 데이터들이 포함되어 있다.

4.1.3. 평가 지표

본 연구에서는 각 모듈의 성능을 객관적으로 평가하기 위해 다음과 같은 표준 지표들을 사용하였다.

■ 문자 오류율 (Character Error Rate, CER):파서의 OCR 성능, 즉 텍스트 인식 정확도를 측정하기 위한 지표이다. 원본 텍스트와 OCR 결과물을 비교하여 잘못 인식되거나(치환), 누락되거나(삭제), 불필요하게 추가된(삽입) 문자의 수를 전체 원본 문자 수로 나눈 값이다. 수치가 낮을수록 OCR 성능이 우수함을 의미한다. (S: 치환된 문자 수, D: 삭제된 문자 수, I: 삽입된 문자 수, N: 전체 문자 수)

$$CER = \frac{S + D + I}{N}$$

■ 이미지/테이블 정확도 (Image/Table Accuracy):문서 내에 포함된 전체 이미지 또는 테이블의 수 대비, 파서가 해당 객체를 올바르게 검출하고 성공적으로 추출한 객체의 비율로 측정하였다. 이는 문서의 구조적 요소를 얼마나 잘 이해하고 분리하는지를 평가하는 지표이다.

■ 정밀도 (Precision), 재현율 (Recall), F1-score:데이터 중복 제거 알고리즘의 성능을 종합적으로 평가하기 위한 표준 분류 지표이다. 각 지표의 의미는 다음과 같다. (TP(True Positive): 실제 중복을 중복으로 올바르게 탐지, FP(False Positive): 중복이 아닌 것을 중복으로 오탐, FN(False Negative): 실제 중복을 놓침)

- 정밀도 (Precision):모델이 '중복'이라고 판단한 데이터 중에서, 실제로 중복인 데이터의 비율이다. 이 지표는 모델의 예측이 얼마나 정확한지를 나타내며, 특히 오탐(False Positive)을 줄이는 것이 중요할 때(예: 고유한 증거를 중복으로 잘못 판단하는 경우) 핵심적인 역할을 한다.

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP}$$

- 재현율 (Recall):실제 전체 중복 데이터 중에서, 모델이 '중복'이라고 올바르게 찾아낸 데이터의 비율이다. 이 지표는 모델이 얼마나 빠짐없이 중복을 찾아내는지를 나타낸다.

$$Recall = \frac{TP}{TP + FN}$$

- F1-score:정밀도와 재현율의 조화 평균으로, 두 지표가 상충 관계일 때 모델의 전반적인 성능을 평가하는 데 유용하다.

$$F1 - score = 2 \times \frac{\mathrm{Pr}ecision \times Recall}{\mathrm{Pr}ecision + Recall}$$

4.1.4. 실험 환경

모든 실험은 NVIDIA RTX 4090 (24GB VRAM) GPU가 탑재된 서버에서 진행되었으며, Python 3.10, PyTorch 등의 라이브러리를 활용하였다.

4.2. 핵심 모듈 성능 비교 평가

4.2.1. 멀티모달 파서 성능 평가

에이전트의 정보 추출 성능을 결정하는 파서 모듈 선정을 위해, 주요 오픈소스 파서 5종에 대한 비교 실험을 수행했다. <Table 1>은 실험 결과이다.

<Table 1> Performance comparison of multimodal parsers

모델	처리시간	GPU메모리	Text CER	Image Acc.	Table Acc.
Unstructured	40.8s	-	70.7%	100%	77.78%
Maker	48s	19.56GiB	23.6%	36.4%	88.89%
MinerU	50.76s	7.14GiB	34.4%	100%	100%

실험 결과, Wang et al. (2024)[9]이 제안한 MinerU는 이미지 및 테이블 분리에서 100%의 완벽한 정확도를 보여 가장 뛰어난 구조 분석 성능을 나타냈다. Unstructured는 가장 낮은 텍스트 인식률을 보였으며, GPU를 지원하지 않아 대용량 데이터 처리 시 병목 현상이 우려되었다. Marker는 텍스트 인식에서 가장 높은 성능을 보였지만, Image Acc가 저조했다. 따라서, 텍스트 인식률은 Marker보다는 낮지만 준수하고 구조적 무결성 보존이라는 측면에서 가장 뛰어난 성능을 보인 MinerU를 최종 파서로 선정하였다.

4.2.2. 데이터 중복 제거 성능 평가

■ 텍스트 중복 제거: 텍스트 중복 제거 성능은 대규모 영어 데이터셋과 한국어 데이터셋을 사용하여 각각 평가하였다. 먼저, 대규모 영어 데이터셋인 Wiki-40B[16]를 이용한 실험에서는 <Table 2>와 같이, 의미 기반의 Unisim[12]이 F1-score 0.9685라는 압도적으로 높은 성능을 기록했다. 이는 규칙 기반 방법론 중 가장 성능이 좋았던 Minhash[10]의 F1-score(0.6766)를 크게 상회하는 수치로, 대규모 일반 텍스트 환경에서 의미 기반 접근법이 훨씬 효과적임을 명확히 보여준다. KorSTS[14] 한국어 데이터셋을 이용한 실험에서, <Table 3>와 같이 의미 기반의 Unisim[12]이 F1-score 0.8095를 기록하며, 해시 기반의 Simhash보다 우수한 성능을 보였다. 이는 단어의 순서나 표현이 달라도 문맥적 의미를 파악하는 능력의 중요성을 보여준다.

<Table 2> Text (Wiki-40B) deduplication performance comparison

방법론	Precision	Recall	F1-score	Accuracy
Simhash	0.6634	0.6405	0.6518	0.5642
Minhash	0.6670	0.6865	0.6766	0.5822
Semhash	0.6726	0.5946	0.6312	0.5576
Unisim	**0.9562**	**0.9812**	**0.9685**	**0.9594**

<Table 3> Text (KorSTS) deduplication performance comparison

방법론	Precision	Recall	F1-score	Accuracy
Simhash	0.6869	0.8047	0.7411	0.6730
Minhash	0.9600	0.2840	0.4384	0.5766
Semhash	0.7032	0.8343	0.7632	0.6988
Unisim	**0.8083**	**0.8107**	**0.8095**	**0.7780**

■ 이미지 중복 제거: INRIA Holidays[15] 데이터셋을 이용한 실험에서도, <Table 4>과 같이 딥러닝 임베딩 기반의 Fiftyone[13]이 F1-score 0.8139를 기록하며, 지각 해시 기반의 ImageHash(0.6776)를 크게 상회했다. 이는 복잡한 시각적 변형에 대한 강건성 측면에서 딥러닝 방식이 훨씬 우월함을 입증한다.

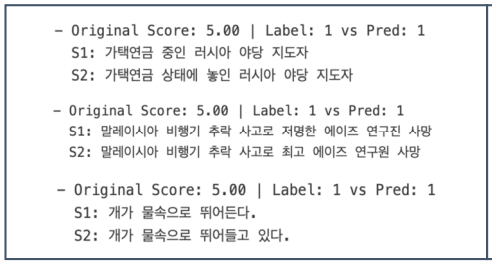
<Table 4> Image deduplication performance comparison

방법론	Precision	Recall	F1-score	Accuracy
Czkawka	0.6667	0.0020	0.0040	0.3360
ImageHash	0.6863	0.6690	0.6776	0.5768
imagededup(CNN)	0.6796	0.6700	0.6748	0.5708
Fiftyone	**0.8008**	**0.8274**	**0.8139**	**0.7485**

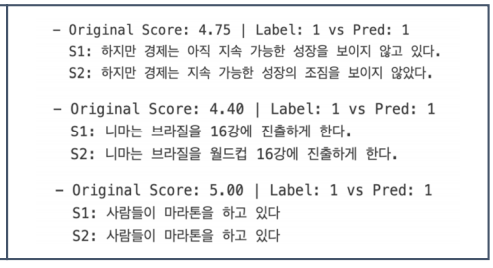
4.2.3. 성공 사례 심층 분석

선정된 핵심 모듈인 Unisim과 Fiftyone이 왜 최종적으로 채택되었는지를 뒷받침하기 위해, 각 모델이 성공적으로 중복을 탐지한 대표적인 사례를 분석하였다.

Unisim은 단순한 키워드 매칭을 넘어 문장의 의미적 유사성을 정확하게 포착하는 능력을 보여주었다. 예를 들어, <Figure2>에서 보듯이, 모델은 '가택연금 중인 러시아 야당 지도자'와 '가택연금 상태에 놓인 러시아 야당 지도자'처럼 일부 조사나 표현 방식이 다른 경우에도 두 문장의 핵심 의미가 동일함을 인지하고 정확히 중복으로 판단했다. 또한, <Figure3> 에서 보듯이 '하지만 경제는 아직 지속 가능한 성장을 보이지 않고 있다'(현재형)와 '하지만 경제는 지속 가능한 성장의 조짐을 보이지 않았다'(과거형)처럼 시제가 다르거나, '저명한 에이즈 연구진'과 '최고 에이즈 연구원'처럼 유의어가 사용된 경우에도 의미적 동질성을 성공적으로 포착하였다. 이는 Unisim이 텍스트의 표면적인 형태가 아닌, 임베딩을 통해 추출된 의미적 핵심을 기반으로 유사도를 판단하기에 가능한 결과이며, 본 에이전트의 신뢰성을 뒷받침하는 중요한 근거가 된다.



<Figure 2> Examples of true positive classification 1



<Figure 3> Examples of true positive classification 2

이미지 중복 제거에서도 Fiftyone은 단순 픽셀 비교를 넘어선 시각적 이해 능력을 보여주었다. <Figure 4>에서처럼, 촬영 각도나 조명이 미세하게 달라진 동일한 과일 그릇 이미지를 성공적으로 중복 처리하였으며, <Figure 5>와 같이 원본에서 일부가 잘려나가거나 각도가 틀어진 그림 이미지의 경우에도 동일한 대상으로 정확하게 인식하였다. 이는 모델이 이미지의 핵심적인 시각적 콘텐츠와 구성을 이해하고, 외형적 변화에 대한 강건성을 갖추고 있음을 입증한다. 이러한 능력은 다양한 환경과 기기에서 촬영될 수 있는 수사 증거 이미지를 효과적으로 처리하는 데 필수적이다.

<Figure 4> Image examples of true positive classification 1

<Figure 5> Image examples of true positive classification 2

4.2.4. 오류 사례 심층 분석

선정된 모델들의 한계를 명확히 파악하기 위해 텍스트, 이미지 데이터셋 각각에 대한 오류 사례를 심층 분석했다

Unisim의 경우, <Figure6>과 같이 “한 남자가 기타를 치고 있다”와 “한 남자가 파스타를 먹고 있다”처럼 주어(“한 남자”)는 같지만 핵심 서술어(“기타를 치다”, “파스타를 먹다”)가 달라 의미가 완전히 다른 문장을 중복으로 오인하는 경향(False Positive)을 보였다. 이는 임베딩 모델이 문장의 특정 요소에 과도하게 집중하여 발생하는 문제로 분석된다. 반대로 <Figure7> 과 같이 ‘식당 식탁에 둘러앉은 사람들 무리.’와 ‘한 무리의 사람들이 식당 테이블에 앉아 있다.’처럼 의미적으로는 완전히 동일하지만, 어순이나 문장 구조(명사구 vs 서술형 문장)가 다른 경우 모델이 두 문장의 유사성을 제대로 포착하지 못하고 어순이 다른 동일 의미의 문장을 비중복으로 판단하는 오류(False Negative)도 관찰되었다.

- Original Score: 0.53 | Label: 0 vs Pred: 1
  S1: 한 남자가 기타를 치고 있다.
  S2: 한 남자가 파스타를 먹고 있다.

- Original Score: 0.75 | Label: 0 vs Pred: 1
  S1: 자살 폭탄 테러범, 나이지리아 교회에서 12명 살해
  S2: 자살 폭탄 테러범, 파키스탄 북서부에서 21명 살해

- Original Score: 0.80 | Label: 0 vs Pred: 1
  S1: MIER, 2012년 성장 전망치를 4.2%로 올린다.
  S2: CMIE, 2012-13년 성장 전망치를 6.3%로 내린다.

<Figure 6> Examples of false positive classification

- Original Score: 4.80 | Label: 1 vs Pred: 0
  S1: 식당 식탁에 둘러앉은 사람들 무리.
  S2: 한 무리의 사람들이 식당 테이블에 앉아 있다.

- Original Score: 4.20 | Label: 1 vs Pred: 0
  S1: 한 노인이 집에서 노트북으로 일한다.
  S2: 노트북을 들고 소파에 앉아 있는 노신사.

- Original Score: 4.00 | Label: 1 vs Pred: 0
  S1: 의자가 위에 있는 중고 목제 건축 자재 더미.
  S2: 나무 파편 더미 위에 놓여 있는 의자.

<Figure 7> Example of false negative classification

<Figure 8> Image examples of false positive errors

Fiftyone의 경우, <Figure 8>과 같이 유사한 사막 배경과 피라미드라는 구도를 가진 두 개의 다른 장소 이미지를 중복으로 오인하는 사례가 있었다. 이는 모델이 이미지의 핵심 객체보다 전역적인 배경이나 구도 특징에 더 큰 가중치를 두었기 때문으로 추정된다.

이러한 오류 분석은 실제 시스템 적용 시, 임계값을 보수적으로 설정하고 필요에 따라 도메인 특화 파인튜닝을 고려해야 함을 시사한다.

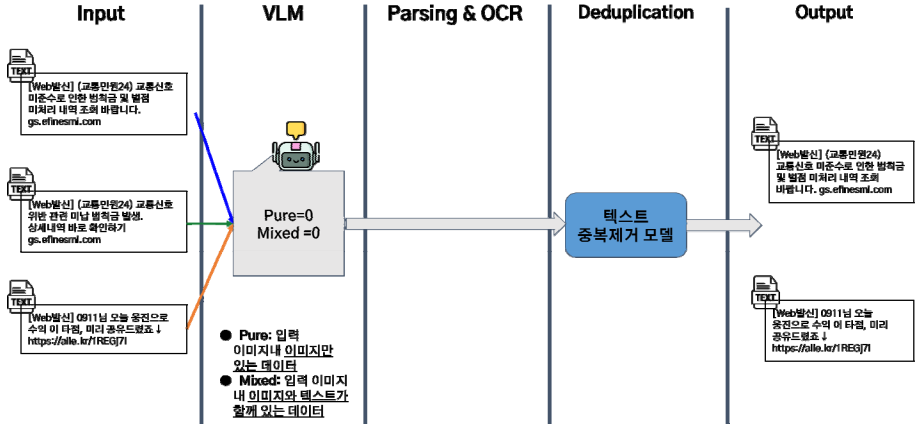
4.3. 통합 데이터 에이전트 시나리오 테스트

4.3.1. 테스트 시나리오 및 과정

선행 연구를 통해 선정된 최적의 모듈들인 MiniCPM-V2.6, MinerU, Unisim, Fiftyone을 통합하여 구축한 데이터 에이전트의 기능적 타당성과 실제 데이터 처리 능력을 검증하기 위해, 엔드투엔드 시나리오 테스트를 수행했다. 테스트에는 더치트에서 제공받은 실제 사이버 범죄 샘플 데이터를 활용하였으며, 입력 데이터의 유형을 (1) 텍스트 전용, (2) 이미지 전용, (3) PDF 전용, (4) 복합 데이터의 네 가지 시나리오로 나누어 각 상황에서 에이전트가 어떻게 자율적으로 작동하는지를 관찰하고 그 결과를 상세히 분석하였다.

4.3.2. 텍스트 데이터 처리 시나리오

첫 번째 시나리오에서는 스미싱 의심 문자 메시지와 같은 순수 텍스트 파일들만 에이전트에 입력되는 상황을 가정했다.



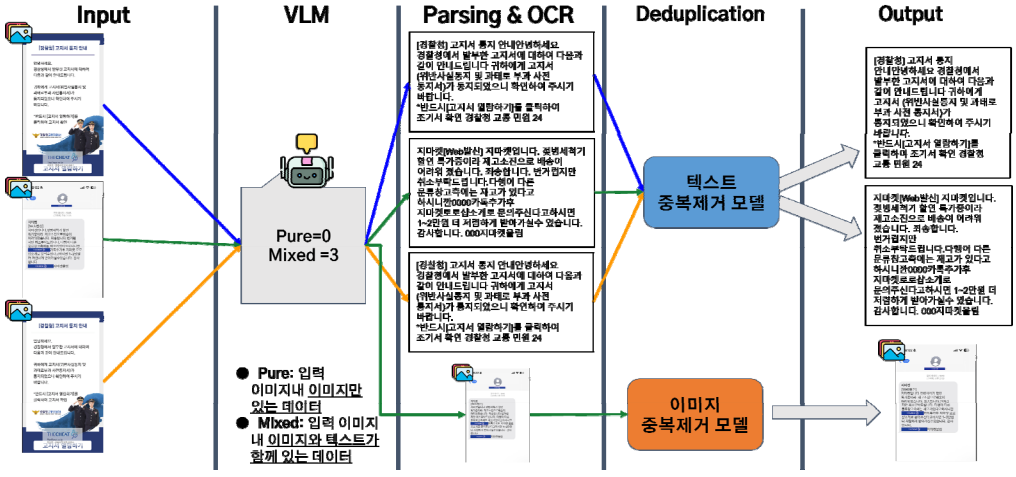
<Figure 9> Text-only scenario

■ 처리 과정: 입력된 데이터가 텍스트 파일이므로, 에이전트의 첫 단계인 VLM 분류 모듈은 비활성화된다. 데이터는 즉시 텍스트 중복 제거 모듈인 Unisim으로 전달된다. Unisim은 입력된 텍스트들의 의미적 유사도를 계산하여 중복 여부를 판단하고, 동일한 의미를 가진 텍스트들을 하나의 클러스터로 그룹화한 뒤 대표 텍스트만을 남긴다.

■ 결과 및 분석: <Figure 9>에서 보듯이, 유사한 내용의 스미싱 문자 3건이 입력되었을 때, Unisim은 내용이 거의 동일한 2건을 성공적으로 중복 처리하여 최종적으로 고유한 의미를 가진 2건의 텍스트만 출력하였다. 이는 에이전트가 동일한 내용으로 대량 발송되는 스미싱 문자나 사기 안내글들을 효과적으로 정제하여, 수사관이 분석해야 할 정보의 양을 줄여줄 수 있음을 보여준다.

4.3.3. 이미지 데이터 처리 시나리오

두 번째 시나리오는 피싱 사이트 스크린샷과 같은 이미지 파일들만 입력되는 상황이다.



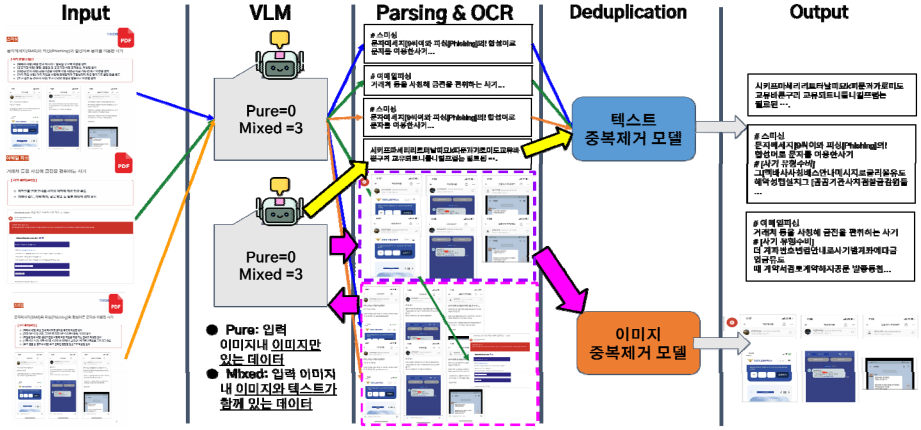
<Figure 10> Image-only scenario

■ 처리 과정: 모든 입력 데이터는 먼저 VLM 분류 모듈로 전달된다. MiniCPM은 각 이미지에 텍스트가 포함되어 있는지 판단하여 모두 'Mixed'로 분류한다. 'Mixed'로 분류된 이미지들은 파싱 및 OCR 모듈인 MinerU로 전달되어, 이미지 내의 모든 텍스트 정보가 추출된다. 이후 추출된 텍스트는 Unisim으로, 원본 이미지는 Fiftyone으로 각각 전달되어 텍스트와 이미지의 중복이 동시에 제거된다.

■ 결과 및 분석: <Figure 10>는 이 처리 과정을 보여준다. 동일한 피싱 사이트를 캡처한 중복 이미지 2장과 다른 사기 관련 이미지 1장이 입력되었을 때, 에이전트는 먼저 각 이미지에서 텍스트를 성공적으로 추출했다. 이후 Unisim은 중복된 텍스트 내용을, Fiftyone은 중복된 이미지를 각각 식별하여 최종적으로 고유한 텍스트 2건과 고유한 이미지 2장을 출력하였다. 이는 에이전트가 단일 유형의 입력으로부터 여러 모달리티의 정보를 자율적으로 추출하고 각각에 맞는 정제 작업을 수행하는 복합 처리 능력을 갖추었음을 입증한다.

4.3.4. PDF 데이터 처리 시나리오

세 번째 시나리오는 사기 수법을 정리한 PDF 문서와 같이 PDF 파일들만 입력되는 상황이다.



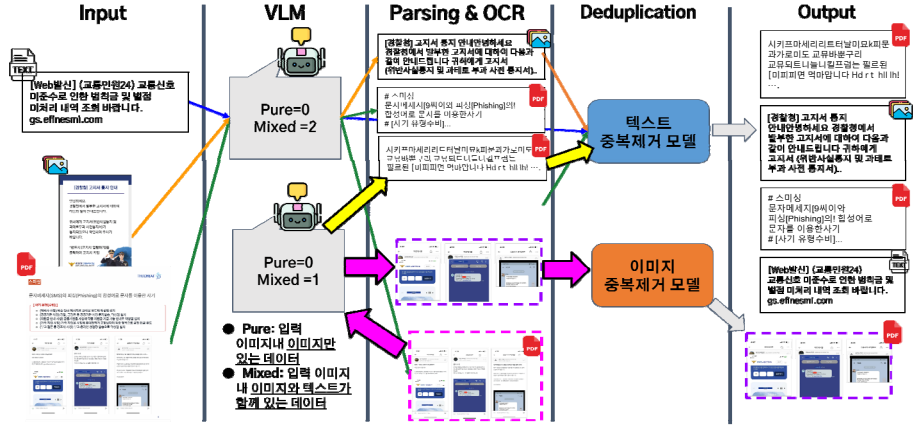
<Figure 11> PDF-only scenario

■ 처리 과정: 본 에이전트의 설계에 따라, PDF 파일은 입력 전처리 단계에서 각 페이지가 개별 이미지로 변환된다. 이후의 처리 과정은 4.3.3의 이미지 데이터 처리 시나리오와 동일하다. 즉, 이미지로 변환된 PDF 페이지들은 VLM에 의해 'Mixed'로 분류되고, OCR을 통해 텍스트가 추출된 후, 텍스트와 이미지 각각에 대한 중복 제거가 수행된다.

■ 결과 및 분석: <Figure 11>에서처럼, 에이전트는 PDF 문서를 성공적으로 이미지로 변환하고 내용을 추출하여 중복을 제거하는 기본 워크플로우를 완수하였다. 하지만 이 과정에서 다음과 같은 중요한 한계점이 관찰되었다. 우선, 에이전트는 PDF 페이지를 이미지로 변환하여 1차 파싱을 진행할 때, 페이지 내의 텍스트와 이미지를 성공적으로 추출했다. 이후, 에이전트는 추출된 이미지를 'Mixed'로 올바르게 판단하고 2차 OCR을 시도 했으나, 바로 이 1차 추출 및 저장 과정에서 이미지의 해상도가 원본 PDF에 포함되었을 때보다 저하되는 현상이 발생했다. 결과적으로, 이 해상도 저하로 인해 2차로 수행된 OCR의 텍스트 인식 정확도가 크게 떨어지는 문제가 나타났다. 따라서 향후 PDF를 파싱할 때 이미지 추출에 있어 품질 저하를 최소화 하는 방안을 마련하는 개선이 필요하다.

4.3.5. 복합 데이터 처리 시나리오 및 종합 평가

마지막으로, 실제 수사 환경과 가장 유사하게 텍스트, 이미지, PDF 파일이 무작위로 혼합되어 입력되는 복합 시나리오를 테스트했다.



<Figure 12> Composite data scenario

■ 처리 과정: 복합 시나리오에서 에이전트는 먼저 입력된 이기종 데이터들을 '텍스트'와 '이미지' 스트림으로 표준화하는 전처리 단계를 수행한다. 이 과정에서 PDF 파일은 각 페이지가 이미지로 변환되어 이미지 스트림에 포함된다. 이후, 텍스트 스트림의 데이터는 곧바로 Unisim모듈로 전달되고, 이미지 스트림의 데이터는 VLM을 시작으로 하는 시각 데이터 처리 경로로 자율적으로 라우팅된다. 모든 경로를 거쳐 처리된 데이터는 최종 정제 단계에서 통합되어, 전체 입력 파일셋에 대한 포괄적인 중복 제거가 이루어진다.

■ 결과 및 종합 평가: <Figure 12>는 이 복합 처리 과정을 요약해서 보여준다. 에이전트는 다양한 형식의 데이터가 혼합된 상황에서도 각 데이터의 특성에 맞는 파이프라인을 동적으로 적용하여, 최종적으로 정제되고 고유한 단서인 텍스트 및 이미지의 집합을 성공적으로 생성하였다. 이는 제안하는 데이터 에이전트가 이기종 멀티모달 데이터를 처리하는 통합 허브로서 기능할 수 있음을 실증적으로 보여준다. 비록 PDF 처리 과정에서 OCR 성능 저하라는 개선점이 발견되었지만, 전반적인 시나리오 테스트를 통해 본 에이전트의 아키텍처가 실제로 작동하며, 복잡한 사이버 수사 데이터를 자동 정제하는 데 높은 실용적 가치와 가능성을 가지고 있음을 명확히 확인할 수 있었다.

수사관련 샘플 데이터의 규모가 작아 처리 시간 단축 등의 대규모 성능 지표를 측정하는 것은 한계가 있었지만, 이 PoC(Proof-of-Concept) 실험은 제안하는 에이전트의 아키텍처가 실제로 작동하며, 복잡한 멀티모달 데이터를 별도의 수동 개입 없이 자율적으로 처리할 수 있다는 기능적 정확성과 실용적 가능성을 명확히 보여주었다는 점에서 큰 의의를 가진다.

V. 논의

본 연구는 사이버 수사 데이터 처리를 위한 지능형 데이터 에이전트를 성공적으로 설계하고 그 기능적 타당성을 검증하였다. 4장에서 제시된 실험 결과들은 제안하는 에이전트의 효용성과 기술적 선택의 타당성을 뒷받침하며, 동시에 미래 연구를 위한 중요한 시사점을 제공한다. 본 장에서는 실험 결과에 대한 심층적인 해석을 통해 연구의 의의를 논하고, 연구의 한계점과 윤리적 고려사항을 고찰하고자 한다.

5.1. 실험 결과에 대한 심층 해석

5.1.1. 파서 선정의 함의: 구조적 무결성과 텍스트 정확도의 상충 관계

문서 파서 선정 실험(4.2.1절)에서 나타난 결과는 본 연구가 지향하는 '사이버 수사 데이터 처리'의 특수성을 명확히 보여준다. 실험에서 Marker 모델이 23.6%로 가장 낮은 오류율을 보였고, MinerU는 34.4%로 그 뒤를 이었다. 반면 Unstructured는 70.7%로 가장 낮은 텍스트 인식 성능을 기록했다. 본 연구에서는 최종적으로 MinerU를 채택하였다. 이러한 결정은 텍스트 인식 정확도와 구조적 무결성이라는 두 성능 지표 간의 상충 관계에 대한 깊은 고민의 결과이다.

일반적인 문서 정보 추출 환경에서는 텍스트 정확도가 가장 중요한 지표일 수 있다. 그러나 법적 증거를 다루는 수사 환경에서는 데이터의 구조적 무결성이 그에 못지않게, 혹은 그 이상으로 중요하다. 예를 들어, 범죄에 사용된 계좌 이체 내역을 담은 '표'나, 피싱 사이트를 캡처한 '이미지'가 파싱 과정에서 누락되거나 깨진다면, 이는 단순히 일부 텍스트를 오인식하는 것보다 훨씬 치명적인 정보 손실에 해당한다. 훼손된 텍스트는 수사관이 원본을 대조하여 일부 교정할 수 있지만, 누락되거나 훼손된 시각적 증거는 복구가 불가능하기 때문이다.

따라서 본 연구는 '일단 모든 구조적 증거를 100% 온전히 보존하고, 텍스트 인식률은 향후 개선 과제로 삼는다'는 실용적인 전략을 선택하였다. 즉, 텍스트 인식 성능이 가장 뛰어났던 Marker가 아닌, 이미지와 표 추출에서 100%의 완벽한 정확도를 보인 MinerU를 채택한 것이다. 이는 제안하는 에이전트가 단순히 기술적 성능 수치를 높이는 것을 넘어, 실제 수사 현장의 요구사항, 즉 '증거의 보존'이라는 최우선 원칙을 시스템 설계에 반영했음을 의미한다.

5.1.2. 중복 제거 방식의 패러다임 전환: 정제에서 분석으로

중복 제거 실험(4.2.2절) 결과는 데이터 처리 방식의 패러다임 전환 가능성을 시사한다. 텍스트와 이미지 모두에서, 의미/시각적 유사도에 기반한 Unisim과 Fiftyone이 전통적인 해시 기반 방식을 압도하는 성능을 보였다. 이 결과는 단순한 성능 우위를 넘어, '중복 제거'라는 작업의 개념을 확장할 수 있음을 의미한다.

기존의 해시 기반 방식은 '동일한 파일'을 찾아 제거하는, 즉 데이터를 '정제'하는 데 목적이 있었다. 그러나 의미 기반 방식은 단순히 동일한 데이터를 넘어, 내용적으로 관련이 깊은 데이터들을 하나의 그룹으로 '클러스터링'하는 분석적 기능을 제공한다. 예를 들어, 에이전트는 표현이 조금씩 다른 여러 스미싱 문자들을 '동일 계좌번호 언급 그룹'으로 묶거나, 각도나 해상도가 다른 피싱 사이트 스크린샷들을 '동일 피싱 사이트 그룹'으로 묶어 수사관에게 제시할 수 있다.

이는 수사관이 개별 증거들을 하나씩 분석하는 것에서 벗어나, 관련 증거 묶음을 단위로 사건의 패턴을 파악하고 용의자들의 관계망을 추적하는 등 더욱 효율적이고 거시적인 분석을 수행할 수 있게 함을 의미한다. 즉, 중복 제거 모듈은 데이터를 삭제하는 소극적 역할을 넘어, 숨겨진 연관성을 드러내는 적극적인 분석 지원 도구로 기능할 수 있는 잠재력을 가진다.

5.1.3. 오류 사례 분석을 통한 모델의 현재와 미래

오류 사례 분석(4.2.3절)은 현재 AI 모델들의 명확한 한계와 미래 발전 방향을 동시에 보여준다. Unisim이 한국어의 유연한 어순과 문맥을 완벽히 이해하지 못하거나, Fiftyone이 이미지의 핵심 객체보다 배경에 더 큰 영향을 받는 모습은 현재의 사전 훈련된 범용 모델들이 특정 전문 도메인의 미묘함까지 포착하는 데는 어려움이 있음을 드러낸다.

그러나 이는 동시에 도메인 특화 파인튜닝의 중요성을 역설한다. 만약 익명화된 실제 사이버 범죄 데이터(예: 사기 유형별 문자 메시지, 범죄 도구 이미지 등)를 활용하여 이 모델들을 추가로 학습시킨다면, 에이전트의 성능은 비약적으로 향상될 수 있다. 예를 들어, 금융 사기 관련 텍스트로 파인튜닝된 언어 모델은 계좌번호나 금액과 같은 핵심 엔티티를 더 정확하게 인식할 것이며, 범죄 도구 이미지로 파인튜닝된 비전 모델은 배경의 방해 없이 특정 객체를 더 강건하게 식별할 것이다. 따라서 본 연구에서 관찰된 오류들은 기술의 실패가 아니라, '범용 AI'를 '전문가 AI'로 진화시키기 위한 구체적인 연구 개발 로드맵을 제시하는 중요한 진단 데이터라 할 수 있다.

5.2. 제안된 데이터 에이전트의 의의 및 독창성

5.2.1. 자율적 워크플로우의 실용성

본 연구의 핵심적인 기여는 개별 AI 모듈의 성능을 단순히 나열하는 것을 넘어, 이들을 '자율적 워크플로우'로 성공적으로 통합했다는 점에 있다. PoC 테스트(4.3절)는 비록 소규모 데이터

로 진행되었지만, 제안된 아키텍처가 실제로 작동함을 입증했다. VLM을 이용한 동적분류 메커니즘을 통해 데이터가 적절한 처리 경로로 자동 분기되고, 각 단계의 출력이 다음 단계의 입력으로 끊김 없이 연결되는 과정은, 수사관의 반복적인 수작업을 자동화할 수 있다는 실질적인 가능성을 보여준다. 이는 AI 에이전트가 복잡한 추론뿐만 아니라, 명확하지만 소모적인 절차를 자동화하는 데 매우 강력한 도구가 될 수 있음을 시사한다.

5.2.2. '보수적 처리' 원칙의 시스템적 구현

본 에이전트는 '보수적 처리'라는 추상적인 원칙을 구체적인 시스템 설계로 구현했다. 이는 특히 중복 제거 모듈의 유사도 임계값 설정에서 명확히 드러난다. 본 연구에서는 중복이 아닌 데이터를 중복으로 오인하는 오류(False Positive)를 최소화하기 위해, 의도적으로 임계값을 높게 설정하였다.

이러한 선택은 일부 유사한 중복 데이터를 놓치는 결과(False Negative 증가)를 낳을 수 있지만, 단 하나의 고유한 증거라도 유실될 경우 사건 전체에 미칠 수 있는 치명적인 영향을 고려할 때, 이는 법률적, 수사적 관점에서 올바르고 책임감 있는 설계 방향이다. 이는 제안하는 에이전트가 단순한 기술적 성능 지표를 넘어, 실제 사용될 도메인의 특수성과 철학을 깊이 있게 고려했음을 보여주는 독창적인 지점이다.

5.3. 연구의 한계점 및 윤리적 고려사항

5.3.1. 연구의 한계점

본 연구는 다음과 같은 명확한 한계를 가진다.

■ 데이터셋의 대표성 문제: PoC에 사용된 더치트 샘플 데이터는 실제 수사 현장에서 다루는 데이터의 양과 다양성, 그리고 예측 불가능한 노이즈 수준에 비해 매우 정제되고 제한적이다. 따라서 본 에이전트의 성능과 안정성은 더 방대하고 다양한 실제 사건 데이터에 대한 추가적인 검증을 통해 입증될 필요가 있다.

■ 처리 모달리티의 한계: 현재 시스템은 텍스트와 이미지를 주로 다루고 있어, 사이버 범죄에서 점차 중요해지는 음성(보이스피싱 녹취), 동영상(딥페이크 사기), 네트워크 트래픽 데이터 등은 처리 범위에 포함하지 못했다.

■ OCR 성능의 병목 현상: 파서로 MinerU를 선택함에 따라, 한국어 텍스트 인식률이 현재 시스템의 성능을 저해하는 가장 큰 병목 지점으로 남아있다. 에이전트가 추출하는 정보의 품질은 OCR의 정확도에 크게 의존하므로, 이는 시급히 개선되어야 할 과제이다.

■ 동적/실시간 처리 미고려: 현재 에이전트는 사전에 수집된 정적인 파일을 처리하도록 설계되었다. 실제 수사 환경에서는 데이터가 실시간 스트리밍 형태로 유입될 수 있으므로, 이를 처리하기 위한 아키텍처 변경이 요구된다.

5.3.2. 윤리적 고려사항 및 책임

본 연구와 같이 실제 수사 데이터를 다루는 시스템은 엄격한 윤리적 고려사항을 수반한다.

■ 개인정보보호: 처리되는 모든 데이터는 사건 관계인의 민감한 개인정보를 포함할 수 있다. 따라서 시스템의 설계, 개발, 운영 전반에 걸쳐 강력한 데이터 보안, 접근 제어, 그리고 법률에 의거한 익명화 조치가 필수적으로 전제되어야 한다.

■ 자동화 편향의 위험: 수사관이 AI 에이전트의 처리 결과를 맹신하고 비판적 검토 없이 수용할 경우, '자동화 편향'에 빠질 위험이 있다. 이로 인해 AI의 오류가 수사 방향에 치명적인 영향을 미칠 수 있다. 따라서 에이전트는 인간의 판단을 대체하는 것이 아닌, 판단을 돕는 의사결정 지원 도구로서의 역할을 명확히 해야 한다. 시스템의 모든 처리 결과는 그 근거와 함께 제시되어야 하며, 최종적인 증거 채택 및 해석의 책임은 항상 인간 수사관에게 있음을 주지시켜야 한다.

■ 기술의 오용 가능성: 본 기술이 범죄 조직이나 다른 악의적인 주체에 의해 악용될 경우, 증거 인멸이나 여론 조작 등 사회에 큰 해를 끼칠 수 있다. 따라서 기술의 접근과 사용에 대한 엄격한 통제와 관리 감독이 필요하다.

Ⅵ. 결론 및 향후 연구

6.1. 연구 요약 및 의의

본 연구는 날로 고도화되고 복합적인 양상을 띠는 사이버 범죄에 효과적으로 대응하기 위해, 수사관의 과중한 수동 데이터 처리 업무를 자동화하는 지능형 데이터 에이전트를 성공적으로 설계하고 그 실용적 가능성을 입증하였다. 현대 사이버 범죄 수사 환경은 텍스트, 이미지, PDF 등 다양한 멀티모달 데이터의 폭증으로 인해 심각한 병목 현상을 겪고 있으며, 기존의 범용 데이터 처리 기술은 수사 데이터의 특수성을 반영하지 못하는 한계를 가지고 있었다.

이러한 문제를 해결하기 위해, 본 연구는 '자율성'과 '보수적 처리'를 핵심 원칙으로 하는 데이터 에이전트 아키텍처를 제안했다. 제안된 에이전트는 VLM을 활용한 동적 분류를 시작으로, MinerU을 통한 정보 추출, 그리고 Unisim, Fiftyone을 통한 데이터 정제에 이르는 엔드투엔드 파이프라인을 갖는다. 객관적 벤치마크를 통해, 사이버 수사 환경에서는 텍스트 인식률보다 구조적 무결성을 보존하는 MinerU가 더 적합함을 실증했으며, Unisim과 Fiftyone으로 대표되는 의미/시각 기반 중복 제거가 기존 해시 방식보다 월등히 효과적임을 입증했다.

최종적으로 구축된 통합 에이전트는 실제 사이버 범죄와 유사한 데이터 시나리오 테스트를 통해 제안된 아키텍처의 자율 처리 가능성을 확인하는 한편, PDF 처리 과정에서의 OCR 성능 저하와 같은 구체적인 기술적 한계점을 발견하고 이를 향후 과제로 제시했다. 결론적으로, 본 연구는 사이버 수사 현장의 반복적이고 소모적인 데이터 전처리 업무를 자동화하고, 수사관이 고차원적인 분석과 추론에 집중할 수 있도록 지원하는 지능형 시스템의 구체적인 청사진과 그 기술적 타당성을 제시했다는 점에서 중요한 의의를 가진다.

6.2. 연구의 학술적 및 실용적 기여

본 연구의 기여는 다음과 같이 요약할 수 있다. 첫째, AI 에이전트 기술을 사이버 수사 데이터 전처리라는 특수 도메인에 적용한 구체적이고 실증적인 사례를 제시했다는 학술적 의의를 가진다. 둘째, 다양한 오픈소스 도구에 대한 한국어 멀티모달 벤치마크 결과를 제공하여 후속 연구의 발판을 마련했다. 셋째, 수사관의 수작업을 획기적으로 줄여줄 수 있는 자동화 시스템의 청사진을 제시함으로써, 수사 자원의 효율적 배분과 분석 데이터의 품질 향상이라는 실용적 가치를 입증했다.

6.3. 향후 연구 방향

본 연구는 지능형 수사 지원 시스템의 가능성을 열었지만, 동시에 더 많은 발전의 여지를 남겼다. 향후 다음과 같은 방향으로 연구를 확장하고 고도화할 필요가 있다.

6.3.1. 단기적 개선 과제

■ OCR 성능 최적화: 본 연구에서 발견된 가장 시급한 개선 과제는 PDF 처리 시 발생하는 OCR 성능 저하 문제이다. 이를 해결하기 위해 (1) MinerU의 내장 OCR을 Naver CLOVA OCR과 같은 고성능 한국어 OCR 엔진으로 교체하거나, (2) PDF-to-Image 변환 시 해상도 손실을 최소화하는 라이브러리를 도입하는 방안을 우선적으로 연구해야 한다.

■ 중복 제거 임계값 최적화: '보수적 처리' 원칙을 더욱 정교하게 구현하기 위해, 사건의 종류나 데이터의 특성에 따라 최적의 유사도 임계값을 자동으로 추천하거나, 수사관이 직접 '보수성' 수준을 조절할 수 있는 사용자 설정 기능을 개발할 필요가 있다.

6.3.2. 중기적 기능 확장

■ 다중 모달리티 지원 확대: 현재의 텍스트, 이미지 중심에서 나아가, 보이스피싱 녹취 파일을 처리하기 위한 음성-텍스트 변환(STT) 모듈, 딥페이크 영상 분석을 위한 비디오 처리 모듈을 통합하여 에이전트가 다룰 수 있는 데이터의 범위를 넓혀야 한다.

■ 지능형 분석 및 추론 기능 연동: 현재의 '전처리 에이전트'를 '분석-추론 에이전트'로 확장하는 연구가 필요하다. 정제된 데이터를 기반으로 다음과 같은 고차원적인 분석 기능을 수행할 수 있다.

■ 자동 개체명 인식 및 관계망 분석: 텍스트에서 용의자명, 계좌번호, 연락처, 암호화폐 지갑 주소 등을 자동으로 추출하고, 이 개체들 간의 관계를 지식 그래프로 구축한다. 이를 통해 수사관은 복잡한 범죄 네트워크를 직관적으로 시각화하고 핵심 인물을 식별할 수 있다.

■ 사건 보고서 초안 생성: 추출 및 정제된 핵심 정보를 바탕으로, LLM을 활용하여 사건 개요나 경과 보고서의 초안을 자동으로 생성하여 수사관의 문서 작성 업무를 경감시킬 수 있다.

이러한 단기적 개선과 중기적 기능 확장을 통해, 본 연구에서 제안한 데이터 에이전트는 단순한 자동화 도구를 넘어, 수사관의 지능을 증강시키는 필수적인 파트너로서 사이버 범죄 대응 역량 강화에 실질적으로 기여하게 될 것이다.

참고문헌(References)

[1] Statistics Korea. 2024. Statistics on the incidence of cybercrime and detection/arrest rates [사이버범죄 발생 및 검거 현황]. e-National Index [e-나라지표]. Available at: https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx_cd=1608 accessed on 2025. 6. 13.

[2] Charikar MS. 2002. Similarity estimation techniques from rounding algorithms. Proceedings of the 34th annual ACM symposium on Theory of computing, pp. 380-388. https://doi.org/10.1145/509907.509965

[3] Russell S, Norvig P. 2020. Artificial Intelligence: A Modern Approach, 4th ed. Pearson.

[4] Radford A, Kim J W, Hallacy C, et al. 2021. Learning transferable visual models from natural language supervision. Proceedings of the 38th International Conference on Machine Learning, pp. 8748-8763.

[5] Xu Y, Li M, Cui L, et al. 2020. LayoutLM: Pre-training of text and layout for document image understanding. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. https://doi.org/10.1145/3394486.3403172

[6] Kim G, Hong T, Yim M, et al. 2022. OCR-free document understanding transformer. Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel. https://doi.org/10.1007/978-3-031-19815-1_29

[7] Hu S, Tu Y, Han X, et al. 2024. Minicpm: Unveiling the potential of small language models with scalable training strategies. arXiv:2404.06395. https://doi.org/10.48550/arXiv.2404.06395

[8] Cao R, Lei F, Wu H, et al. 2024. Spider2-v: How far are multimodal agents from automating data science and engineering workflows? Advances in Neural Information Processing Systems 37: 107703-107744.

[9] Wang B, Xu C, Zhao X, et al. 2024. MinerU: An open-source solution for precise document content extraction. arXiv:2409.18839. https://doi.org/10.48550/arXiv.2409.18839

[10] Broder AZ. 1997. On the resemblance and containment of documents. Proceedings of the Compression and Complexity of Sequences 1997, Salerno, Italy, pp. 21-29. https://doi.org/10.1109/SEQUEN.1997.666900

[11] Devlin J, Chang M W, Lee K, et al. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805. https://doi.org/10.48550/arXiv.1810.04805

[12] Google Research. 2022. Unisim: A universal-schema-based model for text-to-SQL. Google AI Blog. Available at: https://ai.googleblog.com/ accessed on 2025. 6. 13.

[13] Voxel51. 2020. FiftyOne: The open-source tool for building high-quality datasets and computer vision models. GitHub. Available at: https://github.com/voxel51/fiftyone accessed on 2025. 6. 13.

[14] Ham J, Choe YJ, Park K, et al. 2020. KorNLI and KorSTS: New benchmark datasets for Korean natural language understanding. Findings of the Association for Computational Linguistics: EMNLP 2020. pp. 422-430.

[15] Jégou H, Douze M, Schmid C. 2008. Hamming embedding and weak geometric consistency for large scale image search. Computer Vision – ECCV 2008. https://doi.org/10.1007/978-3-540-88682-2_24

[16] Raffel C, Shazeer N, Roberts A, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21, 1-67.