LLM 기반 합성 범죄수사 데이터 생성 시스템의 설계

김희두*, 우병관, 문병훈**, 김지온

*고려대학교, 한림대학교, **수사연수원

heedou123@korea.ac.kr, byeongkwan.woo@gmail.com, mbh010@police.go.kr, jion972@hallym.ac.kr

Design of an LLM-Based Synthetic Criminal Investigation Data Generation System

Hee Dou Kim*, Byeong Kwan Woo, Byung Hoon Moon**, Jion Kim

*Korea Univ., Hallym Univ., **Korean Police Investigation Academy

요 약

본 논문은 실제 수사 데이터를 활용하기 어려운 법적·윤리적 제약을 극복하기 위해, 멀티에이전트 기반 합성 데이터 생성 파이프라인과 전문가 참여형 시드데이터 설계를 제안한다. 제안된 방법은 사건 시나리오와 수사 기록물을 실제와 유사하게 합성하여 LLM 학습용 데이터로 활용 가능하며, 정합성과 품질을 검증할 수 있는 체계를 포함한다. 이를 통해 안전하고 신뢰성 있는 LLM 기반 수사지원시스템 개발의 기초를 마련하고자 한다.

Ⅰ. 서 론

최근 대규모 언어모델(LLM, Large Language Model)의 활용은 경찰 실무 영역에서도 점차 확산되고 있다[1]. 범죄 수사는 사건을 둘러싼 법률적 해석과 쟁점에 대한 추론, 피의자 신문, 공소 유지에 필요한 보고서 작성, 방대한 사건 기록 분석 등 높은 수준의 전문성을 요구하는 영역이다. 그러나 SNS나 온라인 커뮤니티를 기반으로 조직적·대규모로 발생하는 사이버 범죄 피해가 급격히 증가는 현상이 지속되면서, 수사관들은 수사의 전문성을 유지하면서도 복잡하고 방대한 사건을 신속하고 효율적으로 처리하는 데 큰 어려움을 겪고 있다. 이러한 상황 속에서 LLM은 데이터 분석을 도와주거나 법률적 쟁점을 신속히 분석해주고, 수사 보고서 초안을 자동으로 작성해주는 등 업무 효율을 높이는 핵심 도구로 주목받고 있으며, 이에 따라 범죄 수사 자동화 시스템의 필요성이 높아지고 있다[2].

그러나 범죄 수사 지원을 위한 LLM 기반 시스템 개발의 큰 난제는 수사 데이터 확보의 어려움이다. 실제 사건 기록물은 영장, 진술서, 수사보고서 작성 시 가장 유사한 예제가 될 수 있는 가치 있는 원천 데이터이지만, 개인정보보호법, KICS 운영상의 법·보안적 제한, 그리고 AI 활용 시의 감시·편향 등 윤리적 문제로 인해 직접 활용은 사실상 불가능하다[3].

이러한 배경 속에서 LLM 기반 합성 데이터(synthetic data)는 데이터 부족을 보완하면서 안전한 수사지원시스템 개발의 중요한 대안으로 주목받고 있다. 합성 데이터는 실제 데이터를 완전히 대체할 수는 없고, 생성 과정에서 편향 증폭이나 성능 저하, 새로운 윤리적 위험을 초래할 수 있는 문제가 제기되기도 하지만, 최근 발전한 LLM의 정교한 문장 생성 능력에 체계적 검증 도구와 필터링 절차를 결합한다면 특히 데이터가 희소한 경찰 AI 영역에서 학습 모델 보강 수단으로 큰 효과를 기대할 수 있다[4].

이에 본 연구는 멀티에이전트 기반 합성 수사데이터 설계 프레임워크를 제안한다. 이를 통해 국내의 수사 지원 시스템 개발 환경에서 활용 가능한, 안전하고 검증된 데이터 생성 시스템의 구축 방안을 제시하고자 한다.

그림 1과 같이 제안하는 합성 수사데이터 생성 기술은 사건 데이터를 크게 두 가지 층위로 나눈다. 첫째, 실제 수사에서 다루는 죄종별 시나리오를 실제 발생 사건과 유사하게 재현한 이벤트 정보를 생성한다. 둘째, 이러한 이벤트 전개에 따라 실제 수사 과정에서 작성되는 수사문서를 생성하여, 마치 실제 사건의 수사 기록물처럼 여겨지는 데이터를 만들어낸다. 이를 통해 LLM의 수사 추론이나 자동 수사보고서 작성을 위한 학습용·예제 데이터로 활용할 수 있다. 본 연구에서 제안하는 데이터 생성 기술을 적용하면, 법적 제약으로 인해 직접 활용할 수 없었던 실제 수사 기록물을 대신해 실제에 가까운 합성 수사데이터를 활용할 수 있으며, 이를 기반으로 LLM 기반 수사지원 시스템 개발에 기여할 수 있을 것으로 기대한다.

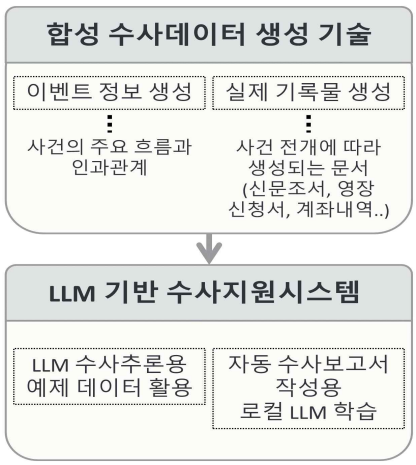

그림 1. 합성 수사데이터 생성 기술의 기대효과

Ⅱ. 기술의 개요 : 합성 수사데이터의 타당성 확보를 위한 전제조건

합성 수사데이터를 LLM 수사지원시스템 개발에 활용하기에 앞서 판단해야 할 가장 중요한 고려 요소는 생성된 데이터가 실제 세계의 사건 데이터와 얼마나 유사한가, 그리고 데이터 내 정합성이 얼마나 타당하게 검증되었는가에 있다. 이러한 요건을 충족하여 품질 높은 합성 데이터셋을 확보한다면, 실제 수사 데이터를 사용하지 않더라도 안전하게 사건 기록물을 활용한 LLM 학습을 시도해볼 수 있다. 본 연구는 합성 수사데이터의 타당성 검증이 유효하도록 하기 위해, 다음과 같은 구조를 필수적으로 따를 것을 전제 조건으로 설정한다. 첫째, 데이터 생성 단계를 세분화한 멀티에이전트 기반 파이프라인을 적용하여, 각 에이전트가 독립적으로 작동하

면서도 순차적으로 정합성을 유지하도록 설계한다. 이는 실제 사건 데이터의 내용이 아주 복잡하기 때문에 단순화된 LLM 프롬프트 설계만으로는 복잡하고 방대한 사건 데이터 내의 논리적 흐름을 유지하고, 이와 유사하게 변형시키는 작업이 어려울 것이라는 점에 착안하였다. 이와 같은 독립된 역할을 담당하는 멀티에이전트 기반의 설계를 통해 단일 문서 단위가 아닌, 사건 시나리오별로 기록이 묶여 제공되는 체계를 구현할 수 있을 것이다. 둘째, 수사 전문가가 직접 참여하여 합성 데이터 생성을 위한 품질 높은 시드 데이터를 제작한다. 이를 통해 합성 데이터 생성에 필요한 필수적인 LLM 프롬프트용 예제를 만들고, 실제 세계의 사건 데이터와 유사한 합성을 유도하는 전략을 취할 수 있다. 마지막으로, 합성된 생성 데이터의 타당성을 검증하는 전문가 기반 절차를 도입한다.

Ⅲ. 멀티에이전트 기반 합성 파이프라인

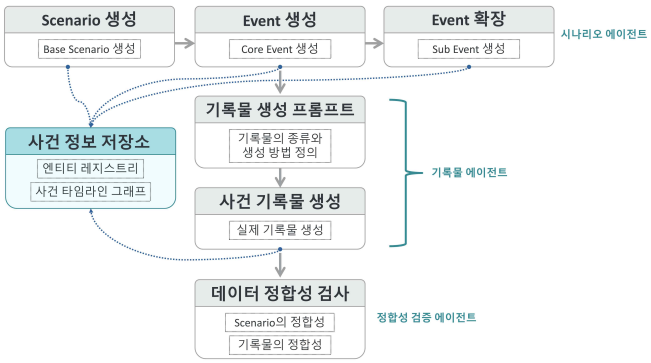

그림 2 합성 수사데이터 생성 파이프라인 구조

그림 2는 합성 데이터 파이프라인의 구조를 보여준다. 제안하는 파이프라인은 LLM의 생성 기술을 결합한 멀티에이전트 기반 시스템으로 설계되었으며, 각 에이전트가 독립적으로 작동하면서도 순차적으로 협력하는 생성 - 검증-패키징의 구조를 따른다. 또한 각 에이전트는 GPT5, Gemini 2.5와 같은 최신 상용 언어모델의 API만을 활용하되, 각 단계에서의 결과물의 완성도를 스스로 판단하여 다음 단계의 에이전트에 결과를 전달할 수 있는 기능을 지니고 있다. 먼저, 시나리오 에이전트에서는 사건의 이벤트 정보를 정교하게 생성하여 사건의 시나리오를 완성한다. 이를 위해 이벤트 정보의 속성을 Base Scenario, Core Event, Sub Event라는 하위 범주로 구분한다. 각 이벤트는 다음의 의미를 지니고 있다.

· Base Scenario : 전체 사건의 뼈대를 이루는 기본 정보
· Core Event : 10개 이내로 압축된 핵심 사건
· Sub Event : Core Event를 기준으로 세분화된 사건

이처럼 이벤트를 구분함으로써 범인·피해자·관련자 등의 사건 발생과 관련된 일련의 행위와 결과들을 독립적으로 생성할 수 있다. 또한 생성된 이벤트 정보는 공유된 사건 정보 저장소에 저장되며, 이후 정합성 검증 에이전트를 통해 논리적 오류, 예를 들어 동일 범인이 같은 시간대에 서로 다른 행위를 한 경우 등을 점검한 뒤, 재생성이나 수정 과정을 거칠 수 있다. 다음으로, 기록물 에이전트는 완성된 이벤트를 토대로 실제 기록물을 작성하여 최종 패키징을 하여 사용자에게 제공하며, 아래 두 단계로 나뉜다.

· 기록물 생성 프롬프트 작성 단계: 어떤 기록물을 생성해야 하는지 규정
· 기록물 작성 단계: 해당 프롬프트에 따라 실제 기록물을 생성

이는 기록물의 종류, 예를 들어 압수영장, 수사보고서, 신문조서 등에 따라 서술 방식과 필수 기재 항목이 모두 다르기 때문에, 기록물별로 별도의 LLM 지침이 필요하기 때문이다. 기록물 생성 과정 역시 사건 정보 저장소의 엔티티와 사건 타임라인을 기준으로 오류 검증 절차를 거친다. 최종적으로 완성된 기록물들은 시간적 순서와 정합성이 유지된 상태로 하나의 사건에 대한 수십 건의 기록물 묶음이 되며, 이를 JSON이나 마크다운(markdown) 형식 등으로 패키징하여 산출물로 저장할 수 있다.

Ⅳ. 전문가 참여형 시드데이터 제작 및 품질 평가

다음으로, 합성 수사데이터의 품질을 높이기 위해 본 연구는 전문가가 주도하여 제작한 시드 데이터셋을 도입한다. 여기서 시드데이터란 실제 사건 기록물을 바탕으로 수사 전문가가 작성한 시나리오 데이터를 의미한다. 시드데이터는 그림 2에서 제시한 합성 데이터 파이프라인과 유사한 구조를 따른다. 즉, 사건 기록물을 선정한 뒤 사건의 기본 시나리오와 이벤트 타임라인을 재작성하고, 해당 이벤트에 따라 분류된 기록물에서 핵심 내용을 추출·정리한다. 이렇게 재작성된 데이터는 합성 데이터와 동일한 형식을 갖춘 정답 데이터셋이 된다. 이때 시드데이터의 제작은 다음과 같은 의의를 지닌다. 첫째, 시드데이터는 실제 사건 기록물에서 LLM 학습에 필요한 핵심 정보만 추출한 예제 데이터로서, 합성 데이터 생성에 필수적인 참고 자료가 된다. 둘째, 시드데이터는 합성 과정에서 LLM이 강력하게 의존할 수 있는 구체적 예시를 제공한다. 셋째, 모든 세분화된 생성 단계에서 시드데이터를 기반으로 LLM의 생성을 유도함으로써, 실제 사건 데이터와 구조적으로 유사한 합성 데이터를 생성할 수 있으며, 이는 곧 few-shot 프롬프트 활용을 가능하게 한다.

마지막으로, 전문가는 합성 데이터의 품질을 평가하여 이를 LLM 기반 수사지원시스템의 학습 재료로 활용할 수 있는지 여부를 판단하는 절차에 참여한다. 이 과정에서 완성된 시드데이터는 정답셋(Golden Answer) 또는 참고용 데이터(Reference Answer)로 사용될 수 있으며, 합성 데이터와 비교하여 종합적인 품질 점수를 산출할 수 있다. 또한 LLM-as-a-Judge 기법을 활용하여 활용도, 가독성, 논리적 일관성, 사건의 다양성의 측면을 지표로 사용하여 인간 평가와의 상관성을 측정하고, 나아가 평가 과정의 자동화를 시도할 수 있다.

Ⅴ. 결론

본 연구는 실제 수사 데이터를 직접 활용하기 어려운 제약을 극복하기 위해, 멀티에이전트 기반 합성 파이프라인과 전문가 참여형 시드데이터 설계를 제안하였다. 이를 통해 실제성과 정합성이 확보된 합성 수사데이터를 제공함으로써, 향후 안전하고 신뢰할 수 있는 LLM 기반 수사지원시스템 개발에 기여할 수 있을 것으로 기대된다. 향후 연구에서는 실제 프로토타입 구현과 다양한 범죄 유형에의 적용, 성능 평가 실험을 통해 제안 시스템의 실효성을 검증할 계획이다

Acknowledgment

이 논문은 2025년도 경찰청의 재원으로 과학치안진흥센터 사이버범죄 수사단서 통합분석 및 추론시스템 개발 사업사업의 지원을 받아 수행된 연구임(No. RS-2025-02218280)

참 고 문 헌

[1] Adams, I. T. (2024). Large language models and artificial intelligence for police report writing. CrimRxiv.

[2] Kim, H., Kim, D., Lee, J., Yoon, C., Choi, D., Gim, M., & Kang, J. (2024, October). LAPIS: language model-augmented police investigation system. In Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (pp.

4637-4644).

[3] 김희두, & 김대호. (2025). 대규모 언어모델의 치안 분야 도입을 위한 기술 및 정책전략 수립 연구. 치안정책연구, 203-242.

[4] Schmidhuber, M., & Kruschwitz, U. (2024). Llm-based synthetic datasets: Applications and limitations in toxicity detection. LREC-COLING 2024, 37.