좌석 위치 기반 자율주행 로봇을 이용한 LLM 부정행위 탐지 및 경고 시스템

차서연, 박호성
전남대학교

cwkite6x2@jnu.ac.kr, hpark1@jnu.ac.kr

Detection and Warning System against LLM Cheating Using Seat Location-based Autonomous Robot

Seoyeon Cha, Hosung Park
Chonnam National University

요 약

본 논문은 인터넷은 허용하되 AI 도구는 제한하는 프로그래밍 시험 환경에서 LLM 기반 부정행위를 탐지하고 좌석 위치 기반 자율주행 로봇과 연동해 즉시 대응하는 시스템을 제안한다. 학생 PC 측 에이전트는 1 초 주기로 화면을 캡처하고, 사전 저장한 LLM 웹 서비스 UI 템플릿과 정규화 상관계수 기반 템플릿 매칭을 통해 LLM 사용 여부를 판별한다. 탐지 시 좌석 번호, 탐지 시각, 서비스명을 포함하는 MQTT 메시지가 브로커를 통해 교수자 PC 와 터틀봇 3 Waffle 로봇으로 전송되며, 교수자 PC 는 해당 시점의 화면만 selective logging 방식으로 저장하여 프라이버시 침해를 최소화하면서도 부정행위 판단에 필요한 최소한의 증거만 남긴다. 터틀봇 3 는 수신된 좌석 정보를 강의실 좌표와 매핑하여 자율주행으로 해당 좌석까지 이동한 후, USB 스피커를 통해 음성 경고를 출력함으로써 현장에서 즉각적인 제지를 수행한다. 강의실 규모에서 구현한 프로토타입 실험 결과, 주요 LLM 서비스에 대해 약 95%의 탐지율과 낮은 오탐률을 보였으며, 탐지에서 로봇 경고까지의 전체 지연은 약 1 초 이내로 유지되었다.

Ⅰ. 서 론

최근 대형 언어모델(LLM)이 빠르게 확산되면서, 프로그래밍 교육·온라인 강의·개인 학습 환경에서 코드 자동 생성과 디버깅, 개념 설명을 LLM 에 의존하는 사례가 크게 증가하고 있다. LLM 은 학습 도구로서는 유용하지만, 평가 상황에서는 수험자의 실제 역량과 LLM 이 생성한 답안을 구분하기 어렵다는 점에서 시험의 공정성과 신뢰도에 대한 우려를 낳고 있다. 특히 텍스트나 코드 결과물을 사후에 분석해 LLM 사용 여부를 추정하는 기존 AI 탐지 기반 접근은 정확도의 논란이 크고, 부정행위가 발생하는 순간에 실시간 제지나 현장 대응을 수행하기 어렵다는 한계를 가진다.

이러한 문제를 완화하기 위해 일부 대학의 프로그래밍 과목, 기업 코딩 테스트, 부트캠프 평가 등에서는 공식 문서·예제 코드 검색은 허용하되 LLM 기반 답안 생성은 금지하는 '인터넷 허용·AI 제한형' 시험이 도입되고 있다. 그러나 수십 명이 동시에 응시하는 실습실 환경에서 소수의 감독자가 각 수험자의 화면을 실시간으로 확인해 문서 검색과 LLM 사용을 구분하는 것은 현실적으로 어렵다. 방화벽·프록시·DNS 차단을 이용해 LLM 사이트 접속을 막는 방식이 사용되지만, 이는 개인 노트북이나 모바일 핫스팟을 통한 우회 사용을 탐지하지 못하고, 언제 어떤 방식으로 LLM 을 사용했는지에 대한 증거를 남기지 못한다는 한계를 가진다. 또한 인터넷 검색은 허용하되 LLM 만 제한해야 하는 환경에서는 과도한 도메인 차단이 정상적인 문서·예제 코드 검색까지 방해할 위험도 존재한다.

본 논문에서는 이러한 한계를 보완하기 위해, 시험 중 학생 PC 화면에서 LLM 웹 서비스 UI 를 직접 인식해 사용 여부를 실시간에 가깝게 판별하고, 탐지 시점의 화면만 선택적으로 저장하는 selective logging 구조를 적용하며, 탐지 결과를 좌석 정보와 연동하여 자율주행 로봇이 해당 좌석까지 이동해 음성 경고를 수행하는 통합 시스템을 제안한다. 제안 시스템은 인터넷 허용·AI 제한형 시험 환경을 대상으로, 화면 기반 LLM 사용 탐지, 프라이버시를 고려한 최소 증거 기록, 좌석 위치 기반 자율주행 로봇을 이용한 즉각적 현장 대응을 통해 감독자의 부담을 줄이고 평가 공정성을 높이는 것을 목표로 한다.

Ⅱ. 본론

제안 시스템은 크게 세 가지 구성 요소로 이루어진다. 첫째, 학생 PC 모듈은 시험 중 화면을 주기적으로 캡처하고, LLM 웹 서비스 UI 와의 유사도를 계산하여 LLM 사용 여부를 판별하는 역할을 수행한다. 둘째, 브로커/교수자 PC 모듈은 학생 PC 로부터 부정행위 탐지 이벤트를 수신하여 좌석 번호·탐지 시각·서비스명을 실시간으로 표시하고, 동일 시점의 화면을 좌석별로 선택 저장(selective logging)한다. 셋째, 터틀봇 3 Waffle 로봇

모듈은 수신된 좌석 ID 를 강의실 좌표계와 매핑하여 해당 좌석까지 자율주행하고, 도착 시 음성 경고를 출력함으로써 현장에서 즉각적인 제지를 담당한다.

학생 PC	->	브로커 / 교수자 PC	->	터틀봇 3 Waffle
화면 캡처		MQTT 브로커		좌석 ID 수신
템플릿 매칭 (OpenCV)		좌석/시간/서비스명 표시		좌표 매핑 후 자율주행
LLM 탐지 이벤트 생성		Selective logging		음성 경고 출력

학생 PC 모듈에서의 화면 캡처 주기는 탐지 지연과 시스템 부하 간의 균형을 고려하여 1 초로 설정하였다. 주기를 0.5 초 이하로 줄이면 CPU 사용률과 MQTT 메시지 발생 빈도가 급격히 증가하는 반면, 3 초 이상으로 늘리면 수험자가 LLM 창을 짧게 열었다 닫는 형태의 부정행위를 놓칠 가능성이 커졌기 때문이다. 캡처는 mss 라이브러리를 사용해 OS 의 저수준 API 로 수행하며, 캡처된 이미지는 디스크에 저장하지 않고 OpenCV 기반 템플릿 매칭 단계로 바로 전달된다. matchTemplate() 함수의 정규화 상관계수(NCC) 결과가 0.8 이상일 때를 탐지로 판정하였는데, 이는 GitHub 페이지 등 일반 웹 화면에서 발생하는 오탐 사례를 분석한 뒤 설정한 값으로, 주요 LLM UI 는 안정적으로 검출하면서 일반 웹사이트에 대한 오탐은 억제하는 임계값이다. 전 과정이 CPU 기반으로 동작하므로, 고사실 수준의 일반 PC 환경에서도 1 초 주기의 실시간 처리가 무리 없이 가능하였다[1].

브로커/교수자 PC 모듈은 학생 PC 에서 발생한 LLM 탐지 이벤트를 MQTT 로 수신한다. MQTT 는 TCP 기반 publish/subscribe 구조의 경량 메시징 프로토콜로[2], 다수의 학생 PC 가 동시에 전송하더라도 오버헤드가 작아 지연이 거의 없다는 장점이 있다. 본 구현에서는 이벤트 누락을 방지하기 위해 최소 한 번 전달을 보장하는 QoS 1 을 사용하였고, 중복 수신 가능성은 좌석·시간·서비스명을 기준으로 교수자 UI 단계에서 제거한다. 수신된 이벤트는 즉시 교수자 화면에 표시되며, 동시에 해당 시점의 스크린샷과 메타데이터를 저장하여 사후 판정을 위한 최소한의 증거가 남도록 효율적으로 설계하였다.

이를 바탕으로 터틀봇 3 모듈에서는 좌석 위치 기반 자율주행을 위해 강의실 바닥 평면에 대한 2 차원 좌표계를 정의하고, 각 좌석 ID 를 이 좌표계 상의 위치와 매핑하였다. 구체적으로 터틀봇을 수동 조작하여 각 좌석 앞 통로 지점까지 이동시킨 뒤 ROS 에서 제공하는 자세 정보를 확인하여 좌표를 측정하고, 이를 좌석 ID 와 함께 설정 파일 내 테이블 형태로 저장하였다. 운영 시에는 MQTT 로 전달된 좌석 ID 를 키로 좌표를 조회하여 해당 좌석 앞을 목표 위치로 지정하며, 별도의 추가 연산 없이 좌석 앞까지 자율주행이 가능하도록 구성하였다.

이와 같이 제안 시스템은 학생 PC 의 화면 기반 LLM 탐지 모듈, MQTT 기반 알림 및 selective logging 모듈, 좌석 좌표계와 연결된 자율주행 로봇 모듈로 구성되며, 각 모듈 간 데이터 흐름을 통해 탐지-알림-현장 경고까지의 종단 동작을 수행한다.

제안 시스템의 구현 환경과 성능을 평가하기 위해, 실제 프로그래밍 시험을 모사한 강의실 환경에서 프로토타입을 구축하였다. MQTT 브로커는 Mosquitto 기반으로 교수자 PC 에 설치하였고, 학생 PC 에이전트 및 교수자 모듈은 Python 으로 구현하였다. 터틀봇 3 Waffle 은 ROS 기반으로 제어하였다[3].

주요 LLM 웹 서비스(예: ChatGPT, Claude, Gemini 등)를 대상으로 총 20 회의 LLM 사용 시나리오를 구성하여 탐지 성능을 측정한 결과, 이 중 19 회를 성공적으로 탐지하여 탐지율은 95%를 기록하였다. 일반 웹사이트(문서·예제 코드·검색엔진 등)에 대해서는 총 20 회 접속 시나리오 중 1 회에서만 오탐이 발생하여, 오탐률은 5.0% 수준으로 나타났다. 또한, LLM 창을 매우 짧게 띄웠다가 닫는 특수한 경우에는 20 회 중 5 회 정도 일부 미탐이 관찰되었는데, 이는 1 초 캡처 주기 사이의 짧은 구간에서 LLM UI 가 포함된 프레임이 생성되지 않은 경우로, 캡처 주기와 탐지 윈도우 길이에 의한 구조적 한계로 볼 수 있다. 다만 실제 프로그래밍 시험에서 LLM 을 부정 사용하려는 경우 대부분 일정 시간 이상 창을 띄워 두는 패턴을 보이므로, 이러한 미탐 사례는 전체 시험 관리 시나리오에서 수용 가능한 수준이라고 판단된다.

Ⅲ. 결론

본 논문에서는 LLM 기반 부정행위를 화면 기반으로 탐지하고, 좌석 위치 기반 자율주행 로봇을 통해 현장에서 즉시 경고까지 수행하는 시험 관리 시스템을 구현하고 그 동작을 검증하였다. 실제 강의실을 모사한 환경에서 프로토타입을 구축한 결과, 1 초 주기의 화면 캡처와 템플릿 매칭 방식으로 주요 LLM UI 를 실시간에 가까운 속도로 탐지할 수 있었으며, 이후 좌석 ID 를 받아 좌표 매핑을 통한 터틀봇 3 의 자율주행 및 음성 경고가 실제로 정상 동작함을 확인했다. 또한 전체 시험 과정을 녹화하는 것이 아닌 탐지 시점 화면만 selective logging 방식으로 저장함으로써, 부정행위 판단에 필요한 최소한의 화면 증거는 남기면서도 프라이버시 침해와 저장·관리 부담을 줄일 수 있음을 보였다.

이와 같은 결과는 제안 시스템이 감독자의 실시간 모니터링 부담을 경감하면서도 LLM 기반 부정행위를 일정 수준 견제하고, 평가 공정성과 신뢰도를 높이는 보조 수단으로 활용될 수 있음을 시사한다. 향후에는 좌석 테이블 기반 이동 방식을 SLAM 기반 자율주행으로 확장하여, 좌석 배치 변경이나 예기치 못한 장애물이 존재하는 경우에도 안정적으로 좌석 위치까지 접근할 수 있도록 고도화할 예정이다.

ACKNOWLEDGMENT

본 논문은 마이크로프로세서캡스톤프로젝트 교과목의 일환으로 수행되었다. 연구 진행에 걸쳐 지도해 주신 전남대학교 전자공학과 김진영 교수님께 감사드립니다.

본 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(지역지능화혁신인재양성사업, IITP-2026-RS-2022-00156287, 30%)과 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 첨단분야 혁신융합대학사업(차세대통신)의 연구 결과임.

참 고 문 헌

[1] G. Bradski, “The OpenCV Library”, Dr. Dobb’s Journal of Software Tools, 2000.

[3] ROBOTIS, “TurtleBot3 Waffle Pi e-Manual”, ROBOTIS, 2017.

[2] A. Banks and R. Gupta, “MQTT Version 3.1.1," OASIS Standard, 2014.