

138 사용자 인터페이스 제어형 에이전트

User Interface (UI) Manipulation Agent

화면에 보이는 UI 요소를 인식·조작하며 작업을 수행하는 AI 에이전트

- 내부 API 연동에 의존하지 않고 화면 기반 정보 인식으로 클릭·입력·전환 등을 실행해 다양한 서비스를 자동화
- 비정형 UI와 변화하는 화면에서도 적응적으로 행동 절차를 조정한다는 점이 특징

사용자 인터페이스 제어형 에이전트의 개념

사용자 인터페이스(UI) 제어형 에이전트는 AI가 화면을 직접 읽고 조작해 애플리케이션을 사용하는 방식으로 설계된 에이전트입니다. 웹, 모바일, 데스크톱 등 다양한 환경에서 버튼·입력창·메뉴 같은 UI 요소를 시각적으로 인식하고, 목표 달성을 위한 조작 절차를 스스로 계획합니다. 내부 API나 전용 스크립트가 없어도 기존 소프트웨어를 그대로 활용할 수 있어 적용 범위가 넓고, 사람의 사용 행위를 모사한다는 점에서 전통적 자동화와 구별됩니다. 단순 지시 반복이 아니라 화면 상태를 이해하고 작업 경로를 구성하는 자율적 인터페이스 제어 능력이 핵심입니다.

사용자 인터페이스 제어형 에이전트의 부상

UI 제어형 에이전트는 UI 변경에 취약하고 비정형 화면에서는 규칙 기반 접근이 어려웠던 기존 자동화 기술이 가진 한계를 AI 기반의 시각·추론·행동 결합 구조로 극복하고 있다는 점에서 주목받습니다. 시각 모델이 화면 요소를 인식하고, 언어모델이 작업 목표를 문맥적으로 판단하며, 행동 모듈이 이를 실제 조작으로 실행해 인식-판단-행동이 하나의 루프로 작동합니다. 이 때문에 화면 배치가 바뀌거나 예외 팝업이 등장해도 목표를 다시 해석해 절차를 조정할 수 있습니다. 또한 UI 기반 도구 사용이 가능해 문서 정리, 웹 탐색, 데이터 입력 등 서로 다른 작업을 하나의 연속된 흐름으로 처리할 수 있어 범용 자동화의 기반 기술로 평가됩니다. 특히 API가 없거나 변경이 어려운 레거시 시스템에서도 바로 적용할 수 있어 기업 환경에서 활용도가 크게 확대되고 있습니다.

사용자 인터페이스 제어형 에이전트의 한계

UI 제어형 에이전트는 화면 기반 행동이라는 특성상 작은 UI 변화에도 취약합니다. 해상도 차이, 로딩 지연, 우발적 팝업만으로도 인식 오류가 발생해 잘못된 행동을 수행할 수 있습니다. 내부 API 기반 연동에 비해 보안 통제가 상대적으로 어려울 수 있어 민감 정보 노출, 계정 오작동, 권한 오남용 같은 위험도 존재합니다. 작업 단계가 길어질수록 행동 추적과 검증이 어려워지고, 비의도적 반복 조작이 누적될 수 있는 점도 실사용의 부담입니다. 이러한 한계에도 불구하고 UI 제어형 에이전트는 기존 시스템을 변경하지 않고도 폭넓은 자동화를 구현할 수 있어, 범용 에이전트와 기업용 AI 자동화의 핵심 기반 기술로 자리 잡고 있습니다.