<https://yozm.wishket.com/magazine/detail/2895/>

**뜨거운 감자, ‘AI 에이전트’의 현재와 미래**

AI의 발전으로 인간의 일을 단순 보조 해주는 역할을 넘어서 데이터베이스를 바탕으로 분석하여 논리적으로 추론하고 의사 결정에 참여하는 수준으로 AI에이전트가 발달하고 있다.

사실 이전에도 빅스비나 시리와 같은 AI 에이전트들이 존재했지만, 이들은 음성 인식이 제대로 이루어지지 않고, 스스로 업무를 대신하지도 못하는 등 수준이 낮은 업무 처리만 가능했다. 그렇기에 AI에이전트는 단순 생성형 AI를 보다 잘 활용하게 해주는 보조도구라는 인식이 강했다. 하지만, AI모델의 발전으로 단순 팔다리 역할만 하는 수준을 넘어서는 기술 혁신이 이루어지게 되었다. GPT나 Claude와 같은 AI모델의 혁신적인 발달로, AI는 대량의 데이터를 학습하고, 생성해주고, 추론하는 등 인간의 지적 수준에 버금가는 두뇌의 역할을 할 수 있게 되었다. 이는 천문학적인 양의 매개변수를 바탕으로 텍스트와 이미지 등 다양한 데이터를 학습할 수 있게 되었기 때문이다. 이를 바탕으로 자동화와 지능적 의사결정까지 가능하게 되었다. 이제 사용자의 데이터를 분석하여 사용자의 니즈를 파악하여 그에 맞는 행동까지 할 수 있게 된 것이다.

또한, LLM이나 이미지 생성 AI를 넘어서 텍스트, 이미지, 음성 등 다양한 형식의 데이터를 동시에 처리하고 이해할 수 있는 능력을 가진 멀티 모달 AI 기술이 등장하면서 인간에게 보다 더 직접적이고 적절하고 풍부한 답변을 제공할 수 있게 되었다.

현재 AI에이전트들의 기술 발전은 어디까지 되었는지 살펴보자.

**엔트로픽의 컴퓨터 유즈**

컴퓨터 유즈는 키보드 입력, 버튼 클릭, 마우스 커서 이동 등 컴퓨터 조작에 필요한 모든 작업을 사람처럼 수행할 수 있다. 컴퓨터 유즈는 SoTA모델인 Claude-3.5-Sonnet 기반으로 작동한다. 이전에는 AI에이전트가 규칙 기반으로 제한된 범위 내에서만 명령을 수행할 수 있던 것과 다르게, 컴퓨터 유즈는 컴퓨터의 다양한 소프트웨어와 시스템에 접근하고, 사용자의 명령을 이해하고 실행할 수 있는 능력을 가지고 있다. 문서작성, 스프레드 시트 작업, 인터넷 검색, 프로그램 설치 등도 가능하다. 단적인 예로 사람이 말로 명령하면, 직접 이메일의 제목, 수신자, 내용들을 작성하여 발송해주는 것도 가능하다.

**MS의 에이전트들**

셰어포인트 - 특정 사이트나 파일, 폴더가 있는 곳을 사람처럼 쉽게 찾아줌

직원 셀프 서비스 에이전트 - 직원의 휴가 신청 관리와 급여 및 복지 정보 확인

오픈 AI의 GPT4o 모델 사용.

**구글의 자비스, 오픈 AI의 오퍼레이터**

자비스 - 구글의 웹 브라우저와 연동한 검색, 쇼핑, 호텔, 항공권 예약 등을 처리할 수 있음.

오퍼레이터 - 개발자용 API로 제공, 인터넷에서 사용자의 작업을 대신하거나 코딩하는 능력을 갖춤.

**AI 에이전트의 미래**

AI에이전트의 두뇌 역할을 하는 AI모델의 수준이 높아지면서, 인간의 업무를 직접 수행하고, 보다 효율적인 방법을 제시해주는 등 마치 사람처럼 의견을 제시하고 행동을 수행할 능력을 갖춰 인간 비서 역할을 하는 수준까지 발달했다.

앞으로는 인간의 질문에 답하는 수준을 넘어서 데이터에 기반해 스스로 판단하고 행동하는 능력까지 갖추게 될 것이다. 그렇기에 유발하리는 기업에 법인이라는 지위를 부여했던 것처럼 AI에도 일종의 지위를 부여하고 권한과 책임을 지워야 한다고 말을 하였다. 정보 네트워크 참여자로서 AI도 단순 보조의 역할이 아닌 실질적 의사결정 참여의 가능성이 높다고 보고 의사결정 행위에 대한 권한 및 책임을 부여해야 한다는 의미이다.

앞으로는 단순 1대의 AI에이전트가 의사결정하는 것이 아닌 다양한 AI에이전트들이 의견을 교환하고 결정하는 방식으로 기술이 발달할 것으로 예상된다.

**AI 에이전트의 기술적 한계**

AI는 예기치 못하나 상황에서 오류를 일으킬 가능성도 적진 않다. 아무리 이해와 추론 능력이 발전해도, 특정 업무의 맥락이나 인간의 감정을 이해하는 데에 어려움을 겪을 가능성이 높다. 현재 최첨단 생성형 AI모델은 트랜스포머 아키텍처에 기반하여 작동하는데 이는 인간의 뇌 시냅스 구조에서 착안하여 만들어졌다. 중요한 정보는 강화하고 덜 중요한 정보를 약화시키는 과정을 본 따 만들었다.

이때 추론하는 과정은 인간이 이해할 수 없는 블랙박스 모델이기 때문에, 어떤 정보를 강화했는지 확신할 수 없고, AI에이전트가 스스로 판단하고 실행한 결과가 심각한 문제를 초래했을 때 법적 책임을 묻는 것이 아주 어렵게 된다.