AI 에이전트에서 에이전트 AI로

권오현 계간 스켑틱 편집자

건강 관리를 위해 AI의 도움을 받으려 할 때 AI가 자율적으로 몇 달에 걸쳐 나의 수면 패턴을 분석하고, 스트레스 수준을 파악하며, 점진적으로 운동 루틴을 조정하고, 심지어 나도 모르게 건강에 해로운 습관들을 하나씩 바꿔나가도록 환경을 설계한다면 어떨까? 이것이 바로 차세대 '에이전트 AI(Agentic AI)'의 모습이다. 기존 'AI 에이전트(AI Agent)'가 우리가 묻는 질문에 답하는 반응형 도구였다면 새로운 에이전트 AI는 스스로 판단하고 행동하는 능동형 동반자라 할 수 있다.

에이전트란 무엇인가

에이전트(Agent)란 환경을 인식하고 목표를 달성하기 위해 자율적으로 '행동'하는 컴퓨터 프로그램을 말한다. 그렇지만 챗GPT(ChatGPT)가 등장하기 이전 에이전트는 자율적이라고 하기는 어려웠다. 그 시절 컴퓨터는 주로 규칙 기반 및 전문가 시스템의 형태였다. 이는 미리 정의된 규칙과 조건문(만약 무엇이라면 무엇을 해라)을 통해 작동했으며 특정 영역에서 인간 전문가의 지식을 모방하는 데 불과했다.

따라서 이 시대의 에이전트는 결정론적이고 예측 가능한 특징을 가졌다. 체스 프로그램, 금융 분석 도구 등은 모두 사전에 프로그래밍된 로직 내에서만 동작했다. 새로운 상황이나 예외 상황에 직면하면 인간이 직접 새로운 규칙을 추가해야 했고, 복잡하고 모호한 문제에 대해서는 대처 능력이 떨어졌다. 또한 자연어 이해 능력이 떨어져 인간과의 소통도 단순하게만 이뤄졌다.

제한적 자율성을 갖춘 LLM 기반 AI 에이전트

챗GPT라는 LLM(거대언어모델, Large Language Model) 모델 출현 이후 비로소 자율적 의미의 AI 에이

전트와 한층 가까워졌다. 기존 규칙 기반 시스템과 달리 LLM 기반 AI 에이전트는 자연어를 통해 목표를 이해하고, 목표 달성을 위해 외부 도구를 끌어 들여 활용한다. 물론 다소 제한적이기는 하지만 말이다.

현재 최신 LLM을 바탕으로 삶의 많은 영역을 자동화하려는 AI 에이전트 프로그램이 출시돼 있다. 예를 들어 2023년 초 등장한 AutoGPT는 사용자가 정의한 목표를 일련의 하위 작업으로 분해하고 챗GPT를 사용해 이를 완수하는 방식으로 작동한다. '스타트업 사업계획서 작성'이라는 목표를 받으면 시장 조사부터 경쟁 분석, 재무 계획, 마케팅 전략, 최종 문서 작성까지 순차적으로 수행한다.

LLM 에이전트의 핵심은 'ReAct(Reasoning+Acting)', 즉 '추론+행위' 접근법이다. 복잡한 질문에 답하기 위해 LLM은 작업을 하위 부분으로 나눈 뒤 도구를 사용해 원하는 최종 응답으로 이어지는 일련의 작업 흐름을 수행한다. 의료 상담 에이전트를 예로 들면, '당뇨병 환자에게 안전한 운동 방법'을 문의했을 때 최신 의학 정보 검색이 필요하다고 추론하고, 의학 논문 데이터베이스를 검색한 후, 개별 환자 상태를 고려한 추가 질문을 통해 맞춤형 운동 계획을 제시한다.

지금까지의 설명을 보면 AI 에이전트가 완전히 자율적으로 행동하는 능동적 행위자라고 보이지만 엄밀히는 그렇지 않다. LLM의 추론 능력이 완벽하지 않아서 잘못된 정보를 제시하는 환각 현상, 목표에 맞지 않는 도구를 호출하는 현상, 지속적인 메모리의 부재로 주어진 임무가 증발되는 것이 문제다. 더 근본적으로, AI 에이전트는 여전히 작업 수행자에 머물러 있다. 즉 사용자가 "보고서를 써라", "호텔을 예약해라" 등 목표를 정해줘야 하며 이렇게 주어진 임무를 완수하면 그 역할은 끝이 난다.

완벽한 자율적 행위자를 꿈꾸다

에이전트 AI는 한발 더 나아가 완벽한 의사 결정자의 역할을 수행하고자 한다. 더 높은 수준의 자율성을 바탕으로 상황을 인식하고, 환경에 대한 적응력을 갖추며, 윤리적 사항을 포함해 복잡한 결정을 내리는 것이 진정한 의미의 행위자라 할 수 있기 때문이다.

자율주행차로 예를 들면, 에이전트 AI가 탑재된 자동차는 환경에 따라 주행을 조정하는 것은 물론이고, 개인 맞춤 비서이자 운전사가 될 수 있다. 사용자가 "오늘 하루 일정 관리해줘"라고 한 번만 말하면, AI가 사용자 캘린더를 보고 오전에 병원, 오후에 회의, 회의 장소 근처 주차장 예약 등을 스스로 판단해 하루 종일 최적의 동선을 계획하고 실행한다. 이를 위해선 스케줄 관리, 운전, 예약, 검색, 에너지 공급 등 수많은 임무를 수행하는 에이전트가 필요하다. 그렇기에 에이전트 AI는 특화된 여러 에이전트들이 목표를 나눠 협업하며 문제를 해결하는 '다중 에이전트 시스템'과 같다. 에이전트 AI의 이런 능력은 시간이 지날수록 진보한다. 자기학습 능력이 있어 임무 개선이 가능하며 결과를 최적화하기 위해 자신의 행동을 재적응시키기 때문이다.

에이전트 AI가 그려내는 미래는 단순히 기술의 진보를 넘어 인간과 AI의 관계를 근본적으로 재정의하는 새로운 시대다. 앞으로 5년 내에 우리는 진정한 의미의 AI 파트너와 함께 살게 될 것이다. 이 AI는 우리가 명확히 표현하지 못한 필요까지도 미리 파악해 해결해 주고, 우리가 미처 생각지 못한 기회들을 발견해서 제안해줄 것이다. 하지만 이러한 변화는 새로운 도전도 함께 가져올 것이다. AI가 우리 대신 많은 결정을 내리게 되면서 인간의 자율성과 창의성을 어떻게 보장할 것인지, 개인 정보와 프라이버시를 어떻게 보호할 것인지에 대한 근본적인 질문들이 제기된다.