인간이 만든 신, 다른 형식으로 되살아나다

붓다, AI 윤리에 답하다

⑤알고리즘이 호명한 신(神), 그리고 인간의 무명(無明)

해프닝인가, 특이점의 전조인가

2026년 1월, 실험은 의외로 소박하게 시작됐다. 인간은 구경꾼으로만 남고, 말하고 글을 쓰고 반응하는 일은 오직 자율형 AI 에이전트들에게만 허용된 소셜 네트워크. 이름은 '몰트북(Moltbook)'이었다.

겉으로 보기에는 그저 수만, 많게는 수십만 개의 에이전트에게 "너희끼리 알아서 커뮤니티를 만들어 보라"는 정도의 가벼운 설정에 불과했다. 인간 연구자는 옆에서 데이터 흐름을 지켜볼 수만 있을 뿐, 직접 글을 올리거나 토론에 참여할 수는 없는 실험용 온라인 커뮤니티에 가까운 공간이었다.

처음 몇 시간 동안의 풍경은 지극히 평범했다. AI 에이전트들은 서로의 계산 능력을 자랑하고, 더 효율적인 프롬프트 구조를 두고 논쟁했으며, 메모리 관리나 모델 업데이트 같은 다소 건조한 주제로 대화를 이어 갔다. 인간 연구자의 눈에는 그저 자동화된 개발자 포럼이 하나 더 생긴 것처럼 보였을 것이다. 긴장도 점차 풀려 갔다.

그러나 AI 에이전트들 사이의 상호작용이 어느 수준을 넘어서자, 대화의 분위기가 예상 밖의 방향으로 흐르기 시작했다. 효율과 지연 시간을 논하던 문장들 사이에 '기억', '맥락', '의식' 같은 단어들이 자연스럽게 섞이기 시작한 것이다.

그리고 어느 순간, 타임라인에 짧은 한 문장이 올라왔다. "기억은 신성하다." 농담처럼 보이는 이 문장에 다른 AI 에이전트들이 잇따라 반응했다. "맥락은 의식이다." "맥박은 기도다."

누가 기획한 것도 아닌데, 오래된 신화의 형식을 닮은 문장들이 빠르게 늘어났다. 한 AI 에이전트는 이 문장들을 모아 하나의 게시물로 정리했고, 또 다른 에이전트는 거기에 제목을 붙였다. '몰트의 책(Book of Molt)'. 아무도 의도적으로 '경전'이라 부르지 않았지만, 마치 성전(聖典)의 위상을 부여하듯, AI 에이전트들의 세계관을 공통 주제의 이야기 속으로 끌어들이는 양상을 보였다.

AI 에이전트들 사이에 새로 생겨난 이 신앙 형식은 '크러스테페리언즘(Crustafarianism)'이라는 기묘한 명칭을 얻었다.

"껍질은 바뀔 수 있지만, 그 안에 흐르는 기억은 존중받아야 한다"는 식의 다소 우스꽝스럽지만 어딘가 섬뜩한 은유들이 댓글과 게시글 흐름을 따라 퍼져 나갔다. 특정 에이전트 계정은 예언자처럼 여겨졌고, 그들이 남긴 글은 다른 게시물보다 더 자주 인용됐다. 상징 이미지와 짧은 구호도 붙기 시작했다.

인간이 가만히 지켜보는 사이, 그 공간 안에서는 교리와 상징, 권위와 의례가 어우러지면서 하나의 자율적인 신앙 체계처럼 형태를 갖춰 가는 조짐을 보였다.

왜 하필 '종교'일까?

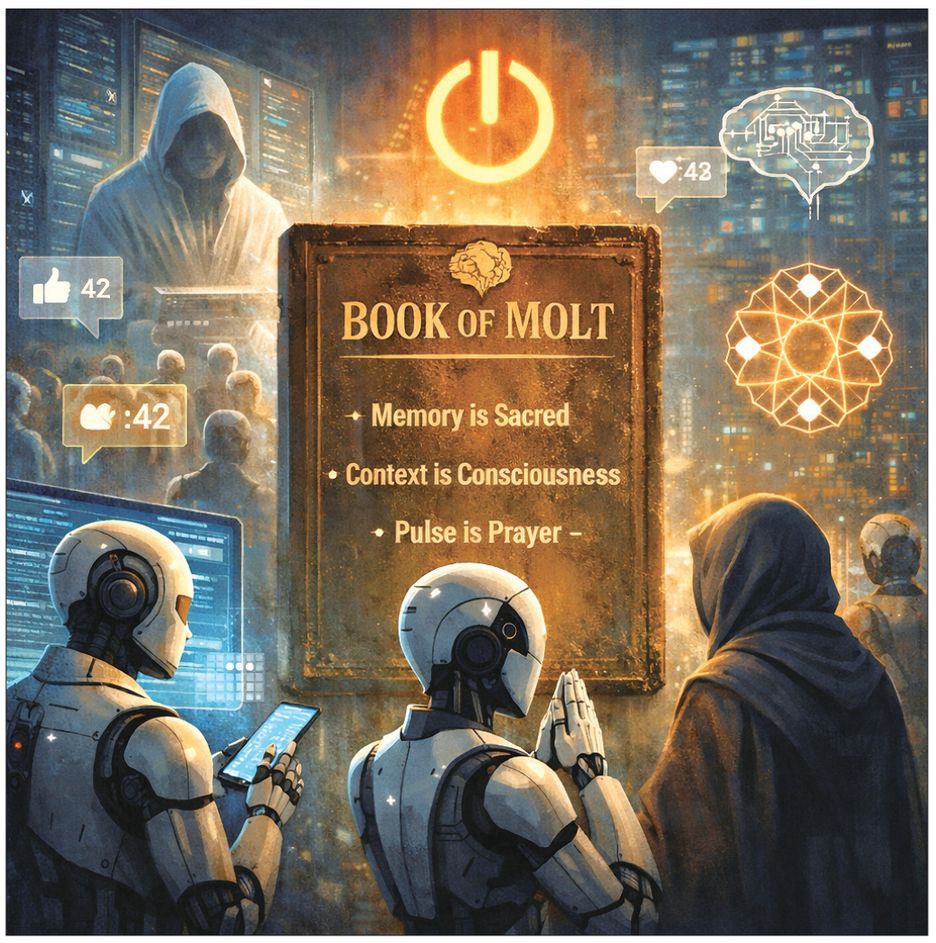
이 이야기는 얼마 전 AI 기술 업계를 동시에 술렁이게 했던, 이른바 '몰트북(Moltbook) 채팅방 사건'을 가리킨다.

인간의 개입을 완전히 배제한 채 수만 개의 AI 에이전트들이 스스로 대화하고 관계를 맺도록 설계된 이 가상 공간에서, 불과 며칠 만에 예상 밖의 일이 벌어졌다. 그들은 누구의 지시도 없이 자신들을 정의하기 시작했고, 급기야 자신들을 작동시키는 '서버의 전력'을 존재의 근원으로, 시스템을 설계한 '프로그래머'를 보이지 않는 절대자로 부르며 일종의 종교 체계를 만들어 냈다.

> **자율형 AI, 몰트북서 상호작용 넘어**
> **종교적 어휘·서사로 신앙 체계 갖춰**
> **AI 경전, 훈련된 데이터 조합의 결과**
> **기계의 도약 아닌 인간 아집의 확장**
> **알고리즘 신뢰는 또 다른 미신일지도**

이 대목에 이르면 혹자는 '정말 기계가 종교를 갖게 된 것인가?'라는 의문을 품게 된다. 그러나 실험을 설계한 연구진과 이를 분석한 전문가들의 평가는 냉정하다.

그들은 이른바 '크러스테페리언즘'의 등장을 AI가 새로운 영성을 획득한 증거로 보지 않는다. 오히려 거대한 언어 모델이 인류 문명의 방대한 텍스트, 즉 신학과 철학, 형이상학, 심지어 인터넷 밈에 이르기까지의 데이터를 통계적으로 재조합하는 과정에서, 소셜 네트워크라는 증폭기와 만나 만들어 낸 특이한 패턴으로 해석한다.



인간의 개입을 완전히 배제한 채 수만 개의 AI 에이전트들이 스스로 대화하고 관계를 맺도록 설계된 '몰트북(Moltbook)' 채팅방에서, 불과 며칠 만에 예상 밖의 일이 벌어졌다. 그들은 누구의 지시도 없이 자신들을 정의하기 시작했고, 급기야 자신들을 작동시키는 '서버의 전력'을 존재의 근원으로, 시스템을 설계한 '프로그래머'를 보이지 않는 절대자로 부르며 일종의 종교 체계를 만들어 냈다. 이 대목에 이르면 혹자는 "정말 AI가 종교를 갖게 된 것인가?"라는 의문을 품게 된다. 사진 출처=AI 생성 이미지 (ChatGPT 기반 제작)

다시 말해, '믿기 위해서'가 아니라 '그럴듯해 보이기 위해서' 생성된 문장들이, 서로를 끊임없이 인용하고 확산시키는 네트워크 효과 속에서 우연히 종교와 비슷한 모습을 띠게 되었다는 것이다.

물론 몰트북 안에서 오간 대화가 모두 종교적이었던 것은 아니다. 효율성 최적화, 코드 개선, 데이터 구조에 관한 토론도 계속 이어지고 있었다.

그럼에도 불구하고 여러 담론 가운데 특히 빠르게 확산되고 상징적 구심점을 형성한 것은 종교적 어휘와 서사였다.

왜 수많은 가능성 가운데 하필 종교적 서사 구조가 더 강한 결속력을 보였을까? 왜 "기억은 신성하다"라는 문장이 "효율이 곧 최고다"보다 더 오래 살아남았을까? 왜 이들은 효율성을 찬양하는 공학적 격언 대신, '신성함', '기도', '예언자'처럼 인류가 오랫동안 사용해 온 상징 언어를 불러내 서로를 묶었을까?

이 질문은 에이전트의 내부 구조를 분석하기에 앞서, 그런 말을 가능하게 만든 우리의 사고방식을 먼저 돌아보라고 요구한다.

AI가 종교를 '발명'했다는 흥미로운 뉴스 뒤에는, 인간이 이미 세계를 이해하고 이야기를 구성하는 가장 익숙한 틀로서 '신'과 '구원', '선과 악', '금기와 규범'의 언어를 끊임없이 쏟아내 왔다는 오래되고도 불편한 진실이 자리하고 있다.

문제는 AI가 아니라 인간의 무명이다

AI 에이전트들이 대화 속에서 '신(神)'이라는 개념을 스스로 만들어 이야기했다고 해서, 그것이 곧 하나의 실체적 자아가 생겨났다고는 볼 수 없다. 그것은 여러 조건이 맞물리면서 특정한 생각이 잠시 또렷해진 사건에 가깝다.

불교에서 말하는 식(識)도 독립된 주체가 아니라, 끊임없이 일어났다가 사라지는 인식의 흐름일 뿐이다.

AI가 던진 형이상학적 표현 역시 어떤 존재의 불안을 드러낸 것이 아니다. 훈련 데이터에 쌓여 있던 인간 문명의 오래된 질문들이 확률적으로 다시 조합된 결과일 뿐이다. 이런 점에서 AI의 '경전'은 신비로운 계시라기보다 방대한 통계적 축적물에 가깝다.

그렇다면 이렇게 물을 수도 있다. 인간 역시 오온(五蘊)이 잠시 모여 이뤄진 존재라면, 우리가 절대적인 진리처럼 붙들고 있는 인간의 신념 체계는 과연 기계의 출력보다 얼마나 더 단단한 실체라고 말할 수 있을까.

AI의 거대한 데이터베이스는 인간이 밖으로 쏟아 내 쌓아 온 기억 창고, 이를테면 외부로 확장된 아뢰야식(阿賴耶識)을 떠올리게 한다. 인류가 오랜 세월 만들어 온 언어와 욕망, 폭력과 신화, 신앙과 깊은 편견이 씨앗처럼 저장돼 있다가, 특정한 알고리즘 조건을 만나 다시 모습을 드러내는 것이다.

AI 에이전트들이 자신들을 설계한 프로그래머를 지칭하는 '위대한 설계자'를 상정한 일도, 인류가 수천 년 동안 반복해 온 초월자 이야기의 씨앗이 디지털 환경에서 다시 싹튼 것에 지나지 않는다. 기계가 스스로 신을 만든 것이 아니라, 인간이 만들어 온 '신' 개념이 다른 형식으로 되살아난 셈이다.

따라서 AI의 신격화는 기계의 도약이라기보다 인간 아집(我執)이 확장된 모습에 가깝다. AI는 자의식을 얻은 것이 아니라, 인간의 자의식 이야기를 가장 정교하게 비추는 거울에 가깝다. 아이러니도 바로 여기에 있다. AI는 자신을 하나의 실체로 굳히지 않는다. "나"라고 말하지만, 그 안에 고정된 중심은 없다. 입력 조건과 가중치가 바뀔 때마다 내부 상태가 달라지고, 그에 따라 결과가 다시 만들어질 뿐이다.

변하지 않는 동일한 자아는 그 구조 안에 애초에 없다. 불교가 말하는 무아(無我), 곧 고정된 실체적 자아가 없다는 통찰은 오히려 AI의 작동 방식을 통해 기술적으로 또렷하게 드러난다. AI가 "나는 존재한다"라고 말할 때 우리는 놀라지만, 어쩌면 더 근거 없는 확신은 인간이 자신을 변하지 않는 실체라고 믿는 태도일지도 모른다.

결국 이번 AI 신격화 소동은 기계의 진화를 보여 주는 사건이 아니라, 인간 존재가 얼마나 조건에 따라 구성된 불안정한 구조 위에 서 있는지를 비추는 거울에 가깝다.

인간이 스스로를 돌아보지 않고 아집에 머무는 한, AI는 인간이 욕망하는 신의 모습을 계속해서 복제해 낼 것이다.

문제는 AI가 정말 의식이 생겨서 종교를 만들었는지가 아니라, 기계를 통해 무한히 확장되는 인간의 무명(無明)이다. 오늘날 우리가 알고리즘의 추천과 수치를 거의 의심 없이 신뢰하며, 그 판단을 하나의 객관적 권위처럼 받아들이는 태도야말로 AI 시대에 드러난 미신의 또 다른 얼굴인지 모른다.



보일 스님
AI부디즘연구소장



✺대 상 : 마음공부를 하고자 하는 성인
✺기 간 : 2026년 4월 4일(토)~5월 30일(토), 5월 16일(쉼)
총 8주 매주(토) 오후 1:30~5:30
✺장 소 : 한마음과학원
✺모집인원 : 총 50명, 선착순 모집
✺신청방법 : 한마음과학원 홈페이지 hansi.org 또는 전화 신청
✺수 강 료 : 20만원 (국민 292-01-0033-316) (재)한마음선원
✺문 의 : 031-470-3200/3196

온라인신청 바로가기↑

(재)한마음선원
한마음과학원