

4. 인간은 에이전트가 아니다

연재 <에이전틱 사고 · 온톨로지> 제4회 — 2부 · 빌려 오되 어디까지

우리는 이미 기계의 말로 자신을 설명합니다

요즘 우리는 자기 자신을 설명할 때 기계에서 쓰는 용어를 자주 빌립니다. 머리가 과부하 걸렸다고 하고 주말 동안 뇌를 리셋하고 오겠다고 하고 바뀐 제도가 아직 입력이 안 됐다고 합니다. 교실도 다르지 않습니다. 학생이 개념을 '저장'했는지, 풀이 과정이 머릿속에서 '처리'되었는지를 묻는 표현이 어색하지 않게 오갑니다.

한때는 방향이 반대였습니다. 컴퓨터가 처음 세상에 나왔을 때 신문과 잡지는 그것을 '전자두뇌'라고 불렀습니다. 말그대로 엄청 똑똑한 사람의 두뇌가 생긴 것 같다는 것을 이야기하는 것이었죠. 낯선 기계를 익숙한 사람의 말로 설명하던 시대에서, 이제는 익숙한 사람을 기계의 말로 설명하는 시대로 넘어온 셈입니다. 말이 먼저 바뀌고, 생각이 그 말을 따라갑니다.

이 연재도 그 흐름에서 자유롭지 않습니다. 앞선 회차들에서 저는 AI 에이전트를 설계하는 쪽에서 정리해 둔 여섯 가지 요건을 빌려 와 학생이 자기 판단을 세울 때 무엇을 명시적으로 갖고 있어야 하는지 묻는 틀로 쓰자고 제안했습니다. 그렇다면 이 연재 역시 사람을 기계의 말로 설명하는 말버릇을 하나 더 보태는 것 아니냐는 물음을 피할 수 없습니다. 가장 짧게 줄이면 이렇습니다.

'그거 또 결국 사람애기 아니고 기계 애기 아닌가요?'

저는 이 물음이 정당하다고 봅니다. 그래서 이번 회는 제 주장을 펼치는 회차가 아니라 제 주장이 어디서 무너지는지를 제가 먼저 말하는 회차입니다. 일종의 자기고백입니다. 이 단계를 하지 않고 넘어가면 앞으로의 연재에서 제가 계속 불편함을 느낄거 같습니다. 칼럼의 순서를 이렇게 잡은 이유가 있습니다. 이 선을 지금 긋지 않으면 남은 열여섯 회 내내 독자가 '그래서 이게 뭘 얘기인가?'를 의심하면서 읽게 됩니다. 그러면 정작 다뤄야 할 것을 못 다룹니다.

예상되는 반박을 제가 먼저 정리하겠습니다

제 논의에 던져질 수 있는 반박은 크게 셋이라고 봅니다. 하나씩 적어 두고, 각각에 대해 제가 물러설 지점과 물러서지 않을 지점을 밝히겠습니다. 첫째, 범주 오류라는 반박입니다. 기계의 설계 명세를 사람의 마음에 갖다 붙이는 것 자체가 어긋난 일이라는 지적입니다. 둘째, 비유의 역사가 반복된다는 반박입니다. 사람은 늘 뇌를 당대의 최신 기술에 빗대어 왔고 그 비유들은 대개 오래가지 못했다는 지적입니다. 셋째, 인간 인지는 그런 식으로 조직되어 있지 않다는 반박입니다. 사람의 앎은 몸과 상황에 묶여 있어서 개체와 관계와 제약의 목록으로 정리되지 않는다는 지적입니다.

첫째 반박에 대하여

이 반박은 정당합니다. 그리고 이 논쟁에는 긴 계보가 있습니다. Newell과 Simon(1976)은 물리적 기호 체계 가설을 내놓았습니다. 기호를 조작하는 물리적 체계는 일반적인 지적 행동에 필요하고도 충분한 수단을 갖는다는 주장입니다. 이 가설을 받아들이면 사람의 사고와 기계의 계산을 같은 틀로 다룰 수 있게 됩니다. 그런데 생각해보면 이런 가정은 성립하지 않는다는 것을 쉽사리 알고 있습니다. '사람의 창의력이 과연 기계로 설명될까?'하는 물음은 오랜 동안 논쟁의 대상이었습니다.

저는 이 노선에 서지 않습니다. 이 연재의 어떤 대목도 인간의 사고가 기호 조작이라는 전제를 필요로 하지 않습니다. 실제로 이 전제의 뿌리는 수학에 닿아 있어서, Hilbert 등의 수학자는 형식주의를 제안하며 수학을 기호와 규칙의 체계로 다시 세우려 했습니다. 점, 직선, 평면 대신 탁자, 의자, 맥주잔이라고 불러도 공리만 지켜지면 기하학은 그대로 성립해야 한다는 그의 말은, 기호의 뜻을 비워 두고 규칙만으로 추론을 끝까지 밀고 가겠다는 선언이었습니다(Hilbert, 1899). 이 기획은 Gödel(1931)의 불완전성 정리로 한계를 드러내고 말았습니다. 하지만 이후 기계적 절차가 무엇인지 정밀하게 정의하려 한 Turing(1936)의 작업으로 이어져 오늘날 컴퓨터의 이론적 토대가 되었습니다. 그렇게 수학에서 태어난 기계가 다시 사람의 마음을 설명하는 틀로 돌아온 것이 Newell과 Simon의 가설인 셈입니다.

하지만 만약 제 논의가 그 전제 위에 서 있다면 반대편의 목소리도 그만큼 길게 이어져 왔습니다. Dreyfus(1972)는 사람의 전문성이 명시적 규칙의 목록이 아니라 몸에 밴 숙련과 상황에 대한 감각에 기대고 있다고 보고, 기호 처리 노선의 인공지능이 바로 그 지점에서 막힐 것이라고 내다봤습니다. 인공지능의 발전사에 있어서 아주 유명한 실험인 Searle(1980)의 ‘중국어 방’ 논증은 다른 곳을 겨누었습니다. 기호를 규칙대로 옮겨 그럴듯한 답을 내놓는 것과 그 기호가 무엇을 뜻하는지 아는 것은 다른 일이라는 것입니다.

(이 부분은 외전으로 집필했던 딸깍 역설을 보면 좋을거 같습니다.)

‘범주 오류’라는 말 자체도 이 논쟁의 뿌리에서 나왔습니다. Ryle(1949)은 대학의 강의동과 도서관과 연구실을 다 둘러본 방문객이 ‘그런데 대학은 어디 있습니까?’라고 묻는 장면으로 이 개념을 설명했습니다. 대학은 건물들 옆에 놓이는 또 하나의 건물이 아니라 그 건물들이 조직된 방식인데, 그것을 같은 범주에서 찾으니 헛걸음이 된다는 것입니다. Ryle이 든 또 다른 예를 우리 식으로 바꾸면, 축구 경기를 끝까지 보고 나서 ‘그런데 팀워크는 몇 번 선수입니까?’라고 묻는 것과 같습니다. 팀워크는 선수 명단에 없지만 선수들이 함께 뛰는 방식 속에 분명히 있습니다. 기계의 부품 목록 옆에서 사람의 마음을 같은 종류의 부품으로 찾으려 하면 똑같은 헛걸음을 하게 됩니다. 에이전트에 목표 모듈과 검증 모듈이 있다고 해서 학생의 머릿속에서도 그런 칸을 찾으려 한다면, 대학 건물을 찾아 헤매는 그 방문객이 되는 셈입니다.

저는 이 경고를 무겁게 받아들입니다. 그래서 제가 에이전트 쪽에서 빌려 오는 것은 부품이 아니라 질문이라는 점을 분명히 하고 싶습니다. ‘이 판단은 무엇을 알아내려는 것인가’, ‘이 결론은 어디까지 적용되는가?’ 같은 질문은 부품이 아니어서 기계에게도 사람에게도 던질 수 있습니다. 달라지는 것은 답이 담기는 형태입니다. 그러니 첫째 반박에 대한 답은 이렇습니다. 저는 사람의 인지 구조가 에이전트 아키텍처와 같다고 주장하지 않습니다. 그 주장은 제 논의에 필요하지도 않습니다.

둘째 반박에 대하여

이 반박도 정당합니다. 뇌를 설명하려고 사람들이 동원한 비유의 목록은 길고, 대체로 당대의 첨단 기술이었습니다. 수력 장치, 시계 장치, 전신망, 전화 교환기, 그리고 컴퓨터. 매번 그럴듯했고 매번 한계가 드러났습니다. 그렇다면 대규모 언어모델 기반 에이전트가 그

목록에 하나 더 얹히는 것 아니냐고 물을 수 있습니다. 아마 그럴 것입니다. 저는 그 가능성을 부정하지 않습니다.

이 목록은 과장이 아닙니다. Daugman(1990)은 뇌 이론에 쓰인 비유의 역사를 되짚으면서, 각 시대가 자기 시대의 가장 인상적인 기술로 마음을 설명해 왔고, 계산 비유 역시 그 계열의 최신판일 수 있다고 경고했습니다. Cobb(2020)은 뇌 연구의 역사 전체를 이 비유들의 교체사로 다시 읽었습니다. 그의 정리에서 각 비유는 새로운 실험과 질문을 열어 주었지만, 동시에 그 비유로는 보이지 않는 것을 가려 왔습니다. 비유가 쓸모없다는 이야기가 아니라, 쓸모와 함정이 한 묶음으로 온다는 이야기입니다.

다만 제가 하려는 일이 비유가 아니라는 점은 분명히 해 두고 싶습니다. 비유는 “A는 B와 같다”는 형태의 주장입니다. 제가 하는 것은 ‘B’를 만드는 사람들이 정리해 둔 목록이 쓸모 있으니 우리 쪽 점검표로 빌려 쓰자”는 제안입니다. 목록의 쓸모는 A와 B가 닮았는지와 별개로 평가받을 수 있습니다.

과학철학에서는 이것을 발견의 맥락과 정당화의 맥락을 나누어 다룹니다(Reichenbach, 1938). 어떤 착상을 어디서 얻었는지와 그 착상이 무엇으로 정당화되는지는 다른 문제라는 것입니다. 케쿨레가 뱀이 제 꼬리를 무는 꿈에서 벤젠 고리를 떠올렸다는 이야기가 사실이든 아니든 벤젠 고리 구조는 꿈이 아니라 증거로 정당화됩니다. 뉴턴의 사과도 마찬가지입니다. 떨어지는 사과를 보고 만유인력을 떠올렸다는 일화는 뉴턴이 말년에 지인에게 들려준 이야기로 전해질 뿐이지만 설령 그 일화가 꾸며진 것이라 해도 만유인력의 법칙은 흔들리지 않습니다. 그 법칙을 받쳐 주는 것은 사과가 아니라 『프린키피아』의 수학과, 그 수학이 행성의 운동과 밀물 썰물까지 설명해 낸 정밀한 예측이기 때문입니다. 착상의 출처가 아무리 극적이어도 정당화는 다른 길로 옵니다.

뉴턴의 사과에서는 한 가지를 더 배울 수 있습니다. 그 착상의 핵심은 사과가 떨어진다는 사실이 아니라 사과를 끌어당기는 힘이 달까지 닿는다면 달도 사실은 지구 쪽으로 끊임없이 떨어지고 있는 것 아니냐는 비유였습니다. 비유는 이렇게 발견의 문을 엽니다. 다만 비유를

쓰되 비유에 먹히지 않으려면 따로 장치가 필요한데, 여기서 Hesse(1966)의 구분이 도움이 됩니다. Hesse는 과학에서 모형을 쓸 때 두 대상 사이의 닮은 점(긍정적 유비), 다른 점(부정적 유비), 아직 닮았는지 모르는 점(중립적 유비)을 나누어 보아야 한다고 했습니다. 사과와 달은 같은 힘에 끌린다는 점에서 닮았지만, 달은 옆으로 빠르게 움직이고 있어서 땅에 닿지 않는다는 점에서 다릅니다. 뉴턴의 계산이 성공한 것은 이 다른 점까지 셈에 넣었기 때문입니다. 모형이 생산적인 것은 중립적 유비를 탐색하게 해 주기 때문이지만, 부정적 유비를 적어 두지 않으면 모형이 대상을 삼켜 버립니다.

이번 회가 하는 일을 이 구분으로 말하면 부정적 유비를 먼저 적는 일입니다. 에이전트와 학생은 다릅니다. 학생에게는 몸이 있고 감정이 있고 옆자리 친구가 있고 틀렸을 때 부끄러워합니다. 목록을 빌려 쓰더라도 이 차이가 작동하는 지점에서는 목록을 그대로 쓰지 않겠다는 것이 이 연재의 약속입니다.

제가 여섯 요건을 어디서 얻었는지는 에이전트 쪽 논의입니다. 그 목록이 교실에서 쓸모가 있는지는 교육학의 개념들과 실제 수업으로 정당화되어야 합니다. 3부에서 여섯 요건을 하나씩 다룰 때 매번 교육학의 기존 개념과 대응시키는 것은 그래서입니다. 대응이 안 되는 요건이 있다면 그 요건은 빼야 합니다.

셋째 반박에 대하여

세 번째가 제일 무겁습니다. 그리고 저는 이 반박에 상당 부분 동의합니다. Varela, Thompson, Rosch(1991)는 인지를 머릿속 표상의 처리로 보는 관점을 비판하면서, 얇이 몸과 환경의 상호작용 속에서 만들어진다는 관점을 제시했습니다. Clark와 Chalmers(1998)는 한 걸음 더 나가 마음이 피부 안쪽에 갇혀 있지 않다고 주장했습니다. 기억이 온전치 못한 오토가 늘 들고 다니는 수첩은 잉가의 생물학적 기억과 같은 역할을 하며, 그렇다면 그 수첩도 오토의 마음의 일부로 보아야 한다는 것입니다. **이 논의들이 옳다면 '학생이 명시적으로 보유해야 한다'는 제 요구는 처음부터 방향이 틀렸을 수 있습니다. 얇이 애초에 개인 안에 담기는 것이 아니라면 말입니다.**

그런데 여기서 조심할 것이 있습니다. 확장된 마음 논의는 제 주장을 무너뜨리는 데 쓸 수도 있지만 떠받치는 데 쓸 수도 있습니다. AI를 오토의 수첩이라고 하면 제 논의가 곤란해지지만, 다르게 볼 수도 있습니다. 오토의 수첩은 오토가 직접 적은 것이고, 무엇이 적혀 있는지 오토가 압니다. 생성형 AI는 학생이 적은 것이 아니고, 무엇이 들어 있는지 학생이 모르며, 물어볼 때마다 다르게 답할 수 있습니다.

1회에서 본 Storm과 Stone(2015)의 단서 조항이 여기서 다시 걸립니다. 저장기 기억을 도운 것은 저장기 믿을 만할 때였고, 믿을 만하지 않다고 판단되면 그 이득은 사라졌습니다. 외부 저장소가 마음의 일부로 편입되려면 그 저장소가 안정적이어야 합니다.

이 조건은 확장된 마음 논의 안에서도 따로 다듬어져 왔습니다. Clark와 Chalmers(1998)는 오토의 수첩이 마음의 일부로 인정받으려면 늘 곁에 있고, 쉽게 꺼내 볼 수 있고, 적힌 내용을 거의 자동으로 믿으며, 그 내용을 과거에 오토 자신이 승인했어야 한다는 조건을 붙였습니다. Heersmink(2015)는 이 조건을 정보의 흐름, 신뢰성, 지속성, 신뢰, 절차적 투명성, 정보적 투명성, 개인화, 변형이라는 여러 차원으로 나누고, 사람과 도구의 결합을 '되느냐 안 되느냐'가 아니라 '어느 차원에서 얼마나 되느냐'의 문제로 보자고 제안했습니다.

이 차원들로 생성형 AI를 보면 흥미로운 그림이 나옵니다. 늘 곁에 있고 쓰기 쉽다는 점에서는 어떤 수첩보다 강하게 결합합니다. 그러나 같은 질문에 늘 같은 답을 주지 않으니 신뢰성과 지속성이 약하고, 그 안에 무엇이 들어 있는지 학생이 들여다볼 수 없으니 정보적 투명성이 낮으며, 학생이 직접 적은 것이 아니니 개인화도 약합니다. 결합은 쉬운데 편입은 어렵습니다. 저는 이 어긋남이 요즘 교실에서 보는 많은 장면을 설명한다고 생각합니다.

흥미롭게도 확장된 마음 논의를 처음 내놓은 Clark 자신이 최근 이 문제를 생성형 AI로 옮겨 다시 다루었습니다(Clark, 2025). 그는 외부 자원에 기대는 것을 곧 정신의 쇠퇴로 보는 시각에 반대하면서도, 무엇을 언제 믿고 기댈지 가늠하는 메타인지적 기술, 그가 '확장된 인지 위생(extended cognitive hygiene)'이라고 부른 습관을 어릴 때부터 길러야 한다고

강조합니다. 확장을 가장 적극적으로 옹호하는 쪽에서도 조건을 이야기하고 있는 셈입니다. 제가 말하는 '안쪽의 자리'는 바로 그 가늠이 일어나는 자리입니다.

최근 한 강연 영상(맨 뒤 참고 영상)은 이 차이를 이렇게 표현했습니다. 사람은 모르는 것을 모른다고 하지만, AI는 자기 저장소를 끝까지 뒤킨 뒤에야 모른다고 하거나 진짜 같은 가짜를 만들어 낸다는 것입니다. 이른바 할루시네이션(hallucination)입니다. 놀라운 사실은 지금은 많이 개선되었지만 처음 Chat GPT-4가 나왔을 때의 Hallucination 비율은 전체 산출물의 7%에 달했습니다.

앞의 절반은 오래된 연구와 맞아떨어집니다. 사람은 모르는 것에 대해서는 기억을 살살이 뒤킨 뒤이지 않고도 빠르게 '모른다'고 판단합니다(Glucksberg & McCloskey, 1981). 그러나 뒤의 절반은 그 자체로 기계를 옛 컴퓨터의 비유로 그린 것입니다. 언어모델은 저장소를 뒤킨다는 것이 아니라 그럴듯한 다음 말을 이어 가며, '모르겠다'보다 추측이 더 보상받는 훈련과 평가 방식이 할루시네이션을 부른다는 분석이 있습니다(Kalai et al., 2025). 그러니까 기계는 모르겠다는 대답을 잘 안합니다.

더 중요한 것은 사람 쪽입니다. 학생들은 모르는 것을 안다고 믿는 일이 흔합니다. 모른다는 것을 아는 능력은 사람에게 거저 주어지는 것이 아니라 길러야 하는 것이고, 이 연재가 학생에게 명시적으로 갖게 하려는 것이 바로 그 능력입니다. 그래서 저는 셋째 반박에 이렇게 답합니다. 얇이 개인 안에만 있다고 주장하지 않습니다. 다만 밖에 있는 것을 자기 것으로 쓰려면 안쪽에 그것을 얹힐 자리가 있어야 하고, 그 자리가 이 연재가 말하는 최소 구조입니다. 확장을 부정하는 것이 아니라 확장의 조건을 말하는 것입니다.

그래서 방어선을 이렇게 긋습니다

지금까지 한 말을 한 문장으로 줄이면 이렇게 됩니다.

온톨로지는 특정 행위주체의 작동 방식에 관한 이론이 아니라 자기 행위를 근거화하고 검증하려는 모든 행위주체에게 제기되는 요건의 목록이다. 이 요건은 행위주체가 기계인지 사람인지와 무관하게 성립한다.

이 문장이 하는 일은 주장의 지위를 낮추는 것입니다. 저는 사람이 어떻게 생겼는지를 말하지 않습니다. 자기 판단을 근거화하고 검증하려는 자리에 서면 무엇이 필요해지는지를 말합니다. 그 자리에 기계가 서든 사람이 서든 필요한 것의 목록은 비슷해집니다. 무엇을 알아내려는지, 그것이 무엇과 어떻게 연결되는지, 어디까지 적용되는지, 얼마나 확신하는지, 무엇으로 확인할 것인지.

이건 인지 이론이 아니라 점검표입니다. 공학 쪽에서 먼저 정련해 둔 목록을, 우리가 학생에게 무엇을 명시적으로 갖게 해야 하는지 점검하는 데 빌려 쓰는 것입니다. 그 이상으로 밀면 무리고, 그 정도만 해도 실무에는 쓸모가 있었습니다.

분산된 인지, 그리고 남는 개인

셋째 반박을 조금 더 밀어 보겠습니다. 이 논쟁은 교육학 안에서도 오래 진행되어 왔습니다.

Salomon과 Perkins(1998)는 학습의 개인적 측면과 사회적 측면을 어떻게 함께 볼 것인지를 정리하면서 둘을 대립시키는 구도 자체를 문제 삼았습니다. 인지가 사람과 도구와 타인에게 분산되어 있다는 것은 맞지만, 그렇다고 개인 안에서 일어나는 학습이 사라지는 것은 아니라는 것입니다.

이들이 강조한 것은 두 측면이 서로를 키운다는 점이었습니다. 개인이 가진 것이 있어야 분산된 체계에 제대로 참여할 수 있고, 그 참여를 통해 개인이 가진 것이 늘어납니다. 어느 한쪽만 보면 다른 쪽이 설명되지 않습니다.

이 구도는 Salomon, Perkins, Globerson(1991)이 지능형 도구를 두고 일찍이 정리한 구분과 이어집니다. 이들은 기술과 ‘함께’ 일할 때 나타나는 효과(effects with)와 기술을 쓰고 난 뒤 기술 없이도 남는 효과(effects of)를 나누었습니다. 도구와 함께라면 학생은 더 복잡한 문제를 풀 수 있지만, 도구를 치운 뒤 무엇이 남는지는 따로 물어야 한다는 것입니다. 앞의 것이 분산된 체계의 성능이라면, 뒤의 것은 그 체계를 거치고 난 뒤 개인에게 남은 인지적 잔여물입니다.

분산 인지의 대표 연구로 꼽히는 Hutchins(1995)의 항해 연구도 같은 방향을 가리킵니다. 군함의 위치를 정하는 일은 어느 한 사람의 머릿속이 아니라 여러 선원과 해도와 계기 사이에서 이루어집니다. 그러나 그 체계가 돌아가는 것은 각 선원이 자기 자리에서 무엇을 읽고 무엇을 넘겨야 하는지 알고 있기 때문입니다. 분산은 개인을 지우지 않습니다. 개인이 맡는 몫의 모양을 바꿀 뿐입니다.

최근 연구들은 이 우려가 추상적이지 않다는 것을 보여 줍니다. 인지적 부담을 바깥으로 넘기는 일(cognitive offloading)은 사람이 원래 써 온 전략이며 그 자체로 나쁘지 않지만, 무엇을 넘기고 무엇을 남길지에 대한 판단이 함께 가야 한다는 정리가 있습니다(Risko & Gilbert, 2016). 튀르키예 고등학교 수학 수업에서 이루어진 대규모 현장 실험(Bastani et al., 2025)에서는 생성형 AI를 제한 없이 쓴 학생들이 연습 문제 성적은 크게 올랐지만, AI 없이 치른 시험에서는 AI를 쓰지 않은 학생들보다 오히려 낮은 성적을 보였습니다. 답을 바로 주지 않고 힌트 중심으로 설계한 AI를 쓴 집단에서는 이 부정적 효과가 대부분 사라졌습니다. '함께일 때의 효과'와 '치운 뒤에 남는 효과'가 서로 다르게 움직인다는 것을 교실 수준에서 확인한 셈입니다.

이 정리가 저에게는 방어보다 경고로 읽힙니다. 이 연재는 개인 쪽에 치우쳐 있습니다. 열다섯 회까지 학생 한 명이 무엇을 갖고 있어야 하는지만 이야기합니다. 5부에서 층위를 바꾸는 것은 그 치우침을 스스로 교정하려는 것입니다. 다만 그 교정이 충분한지는 모르겠습니다.

그리고 이 지점에서 제 주장이 실제로 어디까지 축소되는지 적어 두는 편이 정직하겠습니다. 저는 학생이 모든 것을 안에 담고 있어야 한다고 말하지 않습니다. 밖에 있는 것을 쓰되, 그것이 자기 것으로 편입되려면 편입될 자리가 안에 있어야 한다고 말합니다. 이 정도가 제가 근거를 댈 수 있는 범위입니다.

제가 지키려는 규칙 셋

선을 그어 놓고도 넘기 쉽습니다. 그래서 규칙을 적어 두고 남은 회차에서 지키려 합니다.

하나, 'AI가 이러하니 사람도 이래야 한다'는 형태의 문장을 쓰지 않겠습니다. 이 문장이 등장하는 순간 논증이 아니라 비유가 됩니다. 둘, 여섯 요건을 다룰 때마다 교육학의 기존 개념과 대응시키겠습니다. 대응되는 개념이 없는 요건은 근거가 약한 것이므로, 그때는 근거가 약하다고 적겠습니다. 셋, 에이전트 쪽 용어를 그대로 교실에 들이지 않겠습니다.

그래도 남는 위험

이렇게 방어선을 그어도 위험이 없어지지는 않습니다. 가장 큰 것은 이 목록이 정말 최소이자 충분한지 제가 모른다는 점입니다. 에이전트 쪽에서 목록이 수렴했다는 것은 그쪽 문제를 풀기에 편리하다는 뜻이지, 사람의 학습에 필요한 목록이라는 보증이 아닙니다.

이 문제는 나중에 다시 꺼내겠습니다. 특히 마지막에서는 자기조절학습이나 비판적 사고 같은 기존 틀로 이미 설명되는 것을 굳이 새 이름으로 부르고 있는 것은 아닌지 따져 볼 생각입니다. 지금 미리 말씀드리면, 저는 여섯 가운데 넷은 기존 틀로 설명된다고 봅니다.

생각해볼 문제

1. '그거 결국 기계 얘기 아닌가요?'라는 물음에 여러분이라면 어떻게 답하시겠습니까? 이 글이 제시한 '비유'와 '점검표의 차용' 사이의 구분이 설득력 있다고 보십니까, 아니면 말만 바꾼 것으로 보이십니까?
2. 최근 수업에서 AI나 계산기, 검색 같은 도구를 치웠을 때 학생에게 무엇이 남아 있을지 떠올려 보십시오. '함께일 때의 효과'와 '치운 뒤에 남는 효과'를 구분해 확인하려면 어떤 과제나 평가를 설계할 수 있을까요?

참고 문헌

- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122.
- Clark, A. (2025). Extending minds with generative AI. *Nature Communications*, 16, 4627.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.

- Cobb, M. (2020). *The idea of the brain: The past and future of neuroscience*. New York: Basic Books.
- Daugman, J. G. (1990). Brain metaphor and brain theory. In E. L. Schwartz (Ed.), *Computational neuroscience* (pp. 9–18). Cambridge, MA: MIT Press.
- Dreyfus, H. L. (1972). *What computers can't do: A critique of artificial reason*. New York: Harper & Row.
- Glucksberg, S., & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory*, 7(5), 311–325.
- Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. *Phenomenology and the Cognitive Sciences*, 14(3), 577–598.
- Hesse, M. B. (1966). *Models and analogies in science*. Notre Dame, IN: University of Notre Dame Press.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
- Kalai, A. T., Nachum, O., Vempala, S. S., & Zhang, E. (2025). Why language models hallucinate. arXiv preprint arXiv:2509.04664.
- Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3), 113–126.
- Reichenbach, H. (1938). *Experience and prediction*. Chicago: University of Chicago Press.
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688.
- Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
- Salomon, G., & Perkins, D. N. (1998). Individual and social aspects of learning. *Review of Research in Education*, 23, 1–24.
- Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher*, 20(3), 2–9.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424.
- Storm, B. C., & Stone, S. M. (2015). Saving-enhanced memory: The benefits of saving on the learning and remembering of new information. *Psychological Science*, 26(2), 182–188.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

참고 영상

캐네네 스피치, AI는 흉내 낼 수 없는 인간의 특이한 능력, YouTube Shorts,
<https://www.youtube.com/shorts/BvSwYQodNM8>