

생성형 AI 시대, AI를 어떻게 활용할 것인가?: 인지적 외주화의 위험과 교육적 해법¹⁾

최승락(인문학부 철학전공)

초록

ChatGPT 등장 이후 불과 2년 만에 대학생의 80% 이상이 학업에 생성형 AI를 활용하고 있다. 그러나 이 편익에는 대가가 따른다. 무분별한 AI 사용이 학습 능력 저하, 비판적 사고 약화, 수동적 의존으로 이어질 수 있다는 우려할 만한 연구 결과들이 보고되고 있다.

본고는 이 문제를 인지적 외주화(cognitive offloading) 개념을 통해 분석한다. 생성형 AI는 과업의 효율성을 높이는 수평적 생산의 도구로서는 탁월하지만, 학습자의 사고력을 깊게 만드는 수직적 성장을 대신해줄 수는 없다는 것이 핵심 논지이다. 이를 뒷받침하기 위해 비요크(Bjork)의 바람직한 어려움, 스웰러(Sweller)의 인지 부하 이론, 비고츠키(Vygotsky)의 비계 설정 개념을 검토하고, 대규모 현장 실험 및 상관 연구들을 통해 수행 향상과 실제 학습의 내면화가 일치하지 않을 수 있음을 확인한다. 나아가 알파고 이후 바둑계의 변화를 '먼저 온 미래'의 사례로 분석하여, AI의 효과가 흡수 역량에 따라 달라지며 인지 외주화의 함정이 실재함을 보인다.

이러한 분석을 바탕으로 본고는 학습 과정을 기초 지식 습득, 반복 과업, 아이디어 발산, 비판적 사고 훈련, 피드백 및 교정, 최종 평가의 여섯 영역으로 세분화하고, 각 영역에 적합한 AI 활용·절제 기준을 제안한다. 핵심은 AI를 허용하는가 금지하는가가 아니라, 최종적 이해, 검증, 선택, 책임의 자리를 끝까지 인간에게 남겨 두는 것이다. 이 기준은 한림대학교의 학생 AI 사용 허용 기준 체계에 대응시켜 실제 수업 설계에 적용하는 방식으로 제시되며, 교수자가 수업 활동별로 적절한 기준을 판정할 수 있도록 정량적 진단 도구를 부록으로 함께 제공한다. AI시대 대학 교육의 과제는 기술 활용 지침의 문제를 넘어, 인간성을 보존하기 위한 제도적·철학적 설계의 문제이다.

핵심어: 생성형 AI, 인지적 외주화, 인지적 부채, 바람직한 어려움, 수평적 생산, 수직적 성장, 비판적 사고, 교육적 AI 가이드라인

1) 본고는 정혜선, 박섭형, 김여진, 한수미, 최승락(2025), 『AI 시대, 대학교육을 리디자인하라』 일부 연구를 분석 및 확장한 것이다. 해당 저서의 공동 저자들과 지난 2년간 진행한 연구는 필자의 전문 영역을 넘어 본고의 논의 지평을 넓히는 데 중요한 밑거름이 되었으며 이에 감사드린다. 전공별 AI 활용의 비계 설정 방안과 기준, AI 시대 교수자의 역할, 전공 영역별 AI 사용 제한의 구체적 기준 수립 등은 후속 연구를 통해 계속 발전시켜 나갈 예정이다.

1. 서론: 기술의 가속과 교육의 난제

2022년 말 ChatGPT가 세상에 나온 뒤, 고등교육 현장에서의 AI 도입 속도는 역사상 유례를 찾기 힘든 수준으로 빨라지고 있다. 콘트렉터와 레이에스(Contractor & Reyes, 2025)의 연구에 따르면, 미국의 한 명문대(elite college) 표본에서 2023년 봄 이전에는 10% 미만에 불과했던 학술적 AI 사용률이 불과 2년 만에 80%를 넘어섰다. 이 수치는 단일 기관 표본이므로 모든 대학으로 일반화하기는 어렵지만, 미국 전체 노동자의 AI 채택률인 40%, 성인 전체의 23%를 크게 웃도는 수치다. 인터넷이나 스마트폰의 초기 확산 속도마저 뛰어넘는 이례적인 기술 수용이 대학 캠퍼스에서 일어나고 있는 것이다.

안타깝게도 이러한 급격한 변화에 우리 교육 시스템은 충분히 준비되어 있지 않으며 학생들의 AI 활용이 교육에 부정적 영향을 끼친다는 연구들도 등장하고 있다. 일례로 최근 Bastani et al.(2025)은 무분별한 AI 사용이 학습 능력 저하로 이어질 수 있음을 주장하였으며, OECD(2026)의 “Digital Education Outlook 2026”도 교육적 지침 없이 과업을 외주화하는 방식으로 AI를 사용할 경우 수행 향상과 학습 향상이 일치하지 않을 수 있다고 경고한다. 학생들은 이미 AI를 과제 해결 도구로 광범위하게 활용하고 있고 더 나은 결과물을 제출해 가고 있는 것으로 보이나 이것이 실제 학습으로 이어진다고는 볼 수 없다는 것이다.

페이팔과 팔란티어의 공동 창업자이며, 대학에서 철학을 전공한 소양을 바탕으로 철학적 사유를 주요 경영 수단으로 삼는 피터 틸(Peter Thiel)은 그의 저서 『제로 투 원』(Zero to One)에서 진보를 두 가지 축으로 구분한 바 있다.²⁾ 틸은 이미 효과가 입증된 것을 1에서 n으로 복제·확장하는 것을 ‘수평적 진보’로 규정하며, 존재하지 않던 것을 0에서 1로 새롭게 창조하는 것을 ‘수직적 진보’로 정의한다. 그의 비유를 빌리자면, 타자기 한 대를 100대로 늘리는 것이 수평적 진보이고, 타자기를 워드프로세서로 발전시키는 것이 수직적 진보이다. 거시적 차원에서 수평적 진보는 어딘가에서 작동하는 것을 다른 곳에서도 작동하게 만드는 세계화로, 수직적 진보는 전혀 새롭고 더 나은 방식을 창조하는 기술혁신으로 요약된다.

틸의 진보에 관한 두 가지 구분은 AI시대를 맞이한 대학 교육에도 적용될 수 있을 것으로 보인다. 하나는 도구와 기술을 활용해 과업의 효율을 높이는 수평적 생산 능력을 향상시키는 실용적 기술 교육이며, 다른 하나는 학습자의 인지적 노력을 통해 비판적 사고를 깊게 하고 내적 발달을 이끄는 ‘수직적 성장(vertical growth)’ 교육이다. AI를 활용하여 단순 반복 업무를 자동화하고 이를 통해 동일 시간 내에 더 많은 과업을 처리하거나 결과물의 산출 속도를 높인다면 수평적 생산 능력이 향상되었다고 할 수 있다. 수평적 생산은 효율성을 최우선 가치로 두며, 현재의 생성형 AI 도구들은 방대한 데이터를 분석하여 기존 지식을 빠르게 복제, 재조합, 확장하는 수평적 생산의 극치를 보여준다. 반면 수직적 성장은 외부의 도구에 의존하지 않고 학습자 스스로 지식 체계를 창조해 내는 질적 도약을 의미한다.

본고는 수평적 생산 능력의 향상과 수직적 성장이라는 두 방향의 교육 형태를 문헌 연구를 통하여 실증 자료를 수집하고 ‘먼저 온 미래’라 여겨지는 바둑계의 사례를 통해 AI시대 교육

2) 피터의 『제로 투 원』은 비즈니스와 기술 혁신을 다루는 저서이므로 이를 교육학적 학술 논문의 이론적 기틀로 직접 인용하는 것은 적절하지 않을 수 있다. 다만 그가 제시한 ‘수평적 진보’와 ‘수직적 진보’라는 개념은 생성형 AI 시대의 교육적 난제를 설명하는 데 매우 유용하다고 판단하여, 이를 학습자의 인지적 발달 과정에 맞추어 각각 ‘수평적 생산 능력의 향상’과 ‘수직적 성장’이라는 개념으로 치환하여 전개한다.

방향을 모색한다. 그리고 학습 과정을 6가지 핵심 영역으로 세분화하여, 각 영역에서 AI를 언제 활용하고 언제 절제해야 하는지에 대한 구체적인 기준을 제안한다. 이를 위해 논의는 다음과 같이 전개된다. 2절에서는 수직적 성장이 왜 인지적 어려움을 필요로 하는지 그리고 AI를 교육에 마냥 사용하는 것이 수직적 성장을 왜 방해하는지를 학습 이론의 관점에서 살펴본다.

이를 위해 논의는 다음과 같이 전개된다. 2절에서는 바람직한 어려움, 인지 부하 이론, 비계 설정을 중심으로 수직적 성장의 이론적 조건을 정리한다. 3절에서는 생성형 AI가 연습 단계의 수행을 향상시키더라도 그것이 실제 학습의 내면화로 곧바로 이어지지 않을 수 있음을 실증 연구를 통해 검토한다. 4절에서는 이러한 수행과 학습의 괴리가 교육 현장에서 인지적 외주화, 비판적 사고의 약화, 수동적 의존, 감독 부담의 증가로 어떻게 확장되는지 분석한다. 5절에서는 알파고 이후 바둑계의 변화를 '먼저 온 미래'의 사례로 읽어, AI가 전문성의 형성, 지식의 재편, 교육 시장의 변동에 미치는 영향을 살펴본다. 6절에서는 이상의 논의를 종합하여 학습의 여섯 핵심 영역별 생성형 AI 활용·절제 기준을 제시한다. 7절에서는 이 기준을 한림대학교의 학생 AI 사용 허용 기준 체계에 대응시켜 실제 수업 설계와 평가 운영에 적용하는 방식을 제안한다. 마지막으로 8절에서는 교육의 목적을 AI시대의 인간 형성이라는 관점에서 다시 성찰하고, 대학 교육이 무엇을 끝까지 인간의 몫으로 남겨 두어야 하는지 정리한다.

2. 왜 어려움이 필요하며 어떤 도움이 필요한가?: 수직적 성장의 이론적 조건

생성형 AI 기술을 교육에 도입하려 할 때, 많은 사람들이 간과하는 것이 있다. 대부분의 AI기술은 편리를 목적으로 만들어진 반면, 교육의 목적은 편리가 아니라는 것이다. 오히려 교육은 어려움을 수반하도록 설계해야 할 수 있다. 2절에서는 바람직한 어려움, 인지 부하 이론, 비계 설정에 관한 세 이론을 살펴보고 이를 통해, 수직적 성장이 인지적 노력을 요구하며 노력 없이는 성장이 멈출 수 있음을 살펴본다. 즉, AI를 교육에 적용함에 있어 교육자들은 “교육이란 곧 고통을 제공하는 행위”임을 인지하고 생성형 AI가 교육에 도움이 되는가 해가 되는가를 판단해야 한다는 것이다.

2.1 바람직한 어려움: 느린 학습이 깊은 학습이다.

인지심리학자 로버트 비요크(Bjork, 1994)는 학습 과정의 어려움이 학습을 방해하는 것이 아니라는 흥미로운 주장을 제시한다. 그는 그의 아내 엘리자베스 비요크(Elizabeth Bjork)와 함께 이론을 정교화하여 ‘바람직한 어려움(desirable difficulties)’ 이론을 제시하는데 이에 따르면 학습 과정에서 겪는 특정 수준의 난관이 학습을 방해하는 것이 아니라, 오히려 장기 기억 형성과 학습 전이(transfer)를 촉진한다 (Bjork 1994; Bjork and Bjork 2011; 2020). 여기서 ‘전이’란 한 맥락에서 학습한 지식이나 기술을 새로운 맥락에 적용할 수 있는 능력으로, 교육의 궁극적 목표에 해당한다.

이 이론의 핵심은 ‘학습(learning)’과 ‘수행(performance)’을 명확히 구분하는 데 있다. 비요크 부부는 이를 ‘저장 강도(storage strength)’와 ‘인출 강도(retrieval strength)’라는 두 가지 척도로 설명한다. 저장 강도는 특정 지식이나 기술이 기존의 지식망과 얼마나 깊이 연관되어 굳건하게 내면화되었는지를 나타내며, 한 번 구축된 저장 강도는 시간이 지나도 근본적으로 소실되지 않는다. 인출 강도는 현재 시점에서 그 정보에 얼마나 쉽게 접근할 수

있는지, 즉 순간적인 활성화 정도를 뜻한다.³⁾ 현재 눈에 보이는 수행 능력은 전적으로 인출 강도에 의존하며, 진정한 학습은 지식의 상실을 막아주는 저장 강도의 증가를 의미한다.

비요크 부부는 학습과 수행이 반드시 일치하지 않음을 수십 년간의 실험을 통해 주장한다. 특히 슈미트와 비요크(Schmidt and Bjork 1992)는 세 가지 실험을 통해 훈련 중 인출 강도를 극대화하는 조치들이 저장 강도의 증가와 학습 전이에는 오히려 해로울 수 있다는 점을 발견했다. 첫째, 한 가지 과제를 연속적으로 반복하는 구획 연습은 훈련 중의 수행을 빠르게 향상시켰지만, 여러 과제를 무작위로 섞어 수행 능력을 저하 시킨 연습 방법은 저장 강도를 증가시켜 장기 기억 유지에 훨씬 탁월한 효과를 보였다. 둘째, 기술 습득 과정에서 빈번하고 즉각적인 피드백을 제공하는 것은 당장의 수행 오류를 줄여주었으나, 오히려 피드백의 빈도를 줄이거나 요약해서 지연 제공했을 때 훈련 이후의 장기적인 기억 유지, 즉 파지(retention) 능력은 더욱 향상되었다. 셋째, 동일한 조건에서 반복하는 불변 연습보다 의도적으로 훈련에 변화를 주는 가변 연습이 새로운 상황으로의 일반화 및 전이 능력을 더 크게 향상시켰다. 이처럼 (단기적인) 수행 능력을 떨어뜨리는 의도적인 훈련 상의 어려움들은 학습자가 인출 연습과 같은 더 깊은 정보 처리 과정에 참여하도록 유도하여 결과적으로 저장 강도 향상과 학습 전이를 촉진했다.

학습 간에 진행되는 모든 어려움이 유익한 것은 아니다. 저장 강도를 증가시키는 것으로 비요크 부부가 강조한 것은 단순한 어려움이 아니라 “바람직한” 어려움이다.

“바람직한 어려움은, 일련의 바람직하지 않은 어려움(undesirable difficulties)과 달리, 학습·이해·기억을 지원하는 부호화(encoding) 및 인출 과정을 촉발하기 때문에 바람직하다. 하지만 학습자가 이에 성공적으로 대응할 수 있는 배경 지식이나 기술을 갖추지 못한 경우, 그것은 바람직하지 않은 어려움이 된다.”(Bjork and Bjork 2011, p.59)

위 언급처럼 학습자에게 배경 지식이나 기능이 충분히 갖춰져 있지 않아 성공적으로 대응할 수 없는 어려움은 “바람직하지 않은 어려움”으로, 오히려 학습을 방해한다. 바람직한 어려움이 되려면 도전적이되 달성 가능해야 하며, 학습자가 스스로의 방식으로 배운 것을 이해하여 저장 강도를 증대시키는 부호화 과정과 인출 과정에 능동적으로 참여하도록 유도해야 한다. 대표적인 유형으로는 동일한 내용을 시간을 두고 반복 학습하는 간격 학습(spaced practice), 서로 다른 주제를 섞어 연습하는 혼합 연습(interleaving), 외부 단서 없이 기억에서 정보를 끄집어내는 인출 연습(retrieval practice) 그리고 학습 맥락을 의도적으로 변화시키는 맥락 변화(variation of context) 등이 있다. 이 모든 방법의 공통점은 학습자 스스로 답을 생성하는 과정에서 기억 흔적이 강화되는 ‘생성 효과(generation effect)’를 활용한다는 점이다.

3) 저장 강도와 인출 강도의 차이를 다음과 같은 두 가지 예시를 통해 설명할 수 있다. 먼저, 저장 강도는 높지만 인출 강도는 낮은 경우이다. 예를 들어, 10년 전 매일 치던 피아노 악보나 어릴 적 살던 옛집 전화번호를 지금 당장 떠올리거나 연주하려면 기억이 나지 않아 인출 강도는 낮다고 할 수 있다. 하지만 누군가 첫 마디를 쳐주거나 앞자리를 알려주면 순식간에 전체 기억이 복원되며 다시 익히는 속도가 매우 빠를 것이다. 이는 내면에 깊이 있는 저장 강도가 높기 때문이다. 다른 한편, 벼락치기로 시험 5분 전에 눈으로 훑어본 영어 단어는 방금 보았기 때문에 현재의 인출 강도는 최고조에 달해 당장의 시험 문제에서는 정답을 맞힐 수 있다. 하지만 하루만 지나면 머릿속에서 완전히 사라져 사용할 수 없을 것이다. 이는 지식을 기존의 지식 망과 연결하는 인지적 노력 없이 단순히 단기 기억에 올려둔 상태라 저장 강도가 전혀 형성되지 않았기 때문이다. 즉, 이 경우 저장 강도는 낮지만 인출 강도는 높다고 할 수 있다.

생성형 AI는 즉각적인 정답과 요약을 제공함으로써 바로 이 바람직한 어려움을 제거할 위험이 있다. 학습자가 외부의 단서 없이 기억에서 정보를 스스로 끄집어내는 인출 과정은 학습에 있어 어려움을 유발하지만, 그 자체가 기억을 수정하고 저장 강도를 높이는 교육의 핵심 기제가 될 수 있다. AI를 통해 답을 손쉽게 얻은 학생들은 높아진 인출 강도를 자신의 실제 학습 능력으로 착각하는 능숙함의 환상에 빠질 수 있다. 진정한 수직적 성장을 위해서는 학습자가 스스로 답을 생성해 내는 과정이 필수적이며, AI의 즉각적인 개입은 이 인지적 기회를 박탈한다.

2.2 인지 부하 이론: 무엇을 아끼고 무엇을 써야 하는가

스웰러(Sweller, 1988)의 인지 부하 이론(cognitive load theory)은 인간의 작업 기억(working memory) 용량이 제한적이라는 인지과학의 기본 전제에서 출발한다. 이 이론의 핵심은 학습자의 인지 자원을 무차별적으로 절약하는 데 있지 않고, 학습과 무관한 부담을 줄여 학습에 필요한 처리가 가능하도록 만드는 데 있다.

전통적인 인지 부하 이론은 학습자가 경험하는 인지적 부담을 세 가지로 분류한다. 첫째로 내재적 부하(intrinsic load)는 학습 과제 자체의 본질적 난이도에서 비롯되는 부담이다. 둘째로 외재적 부하(extraneous load)는 학습 내용과 무관한 불필요한 인지적 요구를 말한다. 이는 지나치게 복잡한 수업 교재 구성이나 혼란스러운 수업 설계 등이 여기에 해당한다. 마지막으로 본질적 부하(germane load)는 학습자가 스키마(schema), 즉 지식을 조직화하는 인지적 틀을 형성하는 데 들이는 유익한 노력이다. 다만, 스웰러(Sweller, 2010)는 그의 후속 연구에서 본질적 부하를 독립적인 부하 범주가 아니라 내재적 부하를 처리하는 데 투입되는 작업 기억 자원으로 재해석한 바 있다. 다시 말해, 그는 본질적 부하를 외재적 부하가 감소할 때 학습 관련 처리에 더 많은 작업 기억 자원이 배분되는 방식으로 이해한 바 있어, 이 구분은 고정된 삼분법이라기보다 학습 설계의 방향을 제시하는 개념적 도구로 이해하는 것이 적절하다. 효과적인 학습은 외재적 부하를 최소화하고, 남은 인지 자원을 본질적 부하에 투입할 때 달성된다.

이러한 점을 고려하면, 생성형 AI가 줄여 주는 모든 부담을 곧바로 좋은 절약으로 간주해서는 안 된다. 교육에 있어 무엇이 외재적 부하이든 무엇이 학습에 필요한 핵심 처리인지의 과제의 성격, 교육의 목적, 그리고 학습자의 숙련 수준에 따라 달라지기 때문이다. 즉, 교육에서 AI 활용의 핵심 판단 기준은 “얼마나 많은 인지 자원을 아껴 주는가?”가 아니라 “어떤 종류의 인지 처리를 대신할 수 있는가?”이다.

예컨대 대학교 철학 수업도 그 목적에 따라 AI의 활용을 서로 다르게 생각할 수 있다. 만약 수업의 교육 목적이 전통적인 철학자들의 핵심 주장과 논증 구조를 이해하고 이를 비판적으로 평가하여 자신만의 철학적 관점을 구성하는 것이라면, 철학 원전의 외국어 번역이나 해당 글의 각주 및 참고문헌 형식 정리 그리고 단순한 문법적 오류 탐지 등은 철학적 사유라는 학습 목표와 직접 관련이 적은 외재적 부하가 된다. 따라서 이 수업에서 AI를 활용해 난해한 외국어 문장의 기초 번역과 문서 형식 정리를 맡긴다면, 학습자는 불필요한 인지적 소모를 줄이고 철학자들의 주장과 논증 구조에 대한 이해에 온전히 자신의 인지 자원을 집중할 수 있다.

반면, 철학 원전 번역 실습이나 학술적 글쓰기 연습 수업을 생각해 보자. 원전 번역 수업의 핵심 목적은 철학적 용어의 역사적 맥락과 의미를 파악하여 두 언어 간의 구조적 차이를

이해하고 스스로 번역하는 능력을 기르는 것이다. 학술 글쓰기 수업은 규정된 형식을 정확히 익히고 적용하는 것 자체가 중요한 학습 목표이다. 이러한 수업에서 AI에게 기초 번역이나 형식 정리를 맡기게 되면, 이는 더 이상 외재적 부하를 줄이는 것이 아니다. 오히려 학습자 스스로 단어의 의미를 고민하며 스키마를 형성해야 할 핵심 인지 처리를 AI에게 떠넘기는 셈이 되어, 제대로 된 교육이 이루어질 수 없다. 요컨대 AI는 외재적 부하를 줄이는 도구로서는 유용하지만, 본질적 부하를 대체하는 도구로 사용될 때는 학습을 저해한다. 이 부분이 6절에서 제안할 AI 활용 기준의 핵심적인 판단 근거이다.

2.3 비계 설정: 도움은 언제 거둬야 하는가?

비고츠키(Vygotsky, 1978)의 근접 발달 영역(zone of proximal development, ZPD) 개념은 학습자가 혼자서는 도달하기 어렵지만 유능한 타인의 적절한 도움을 받으면 도달할 수 있는 영역에서 학습이 가장 효과적으로 일어남을 보여준다.⁴⁾ 이때 제공되는 교수적 지원을 흔히 '비계(scaffolding)'라고 부른다. 다만 오늘날 널리 쓰이는 비계라는 용어는 비고츠키 자신이 직접 체계화한 개념이라기보다, 그의 근접 발달 영역 논의를 바탕으로 우드 등(Wood, et al, 1976) 등의 후속 연구에서 정교화된 개념으로 이해하는 편이 더 정확하다.⁵⁾

이 후속 연구들이 강조하는 비계의 핵심은 조건부 조절(contingency), 점진적 제거(fading), 그리고 책임 이전(transfer of responsibility)이다. 먼저, 조건부 조절은 학습자의 현재 이해 수준이나 수행 능력에 맞추어 지원의 유형과 양을 실시간으로 조정하는 것을 의미한다. 이는 단순한 도움의 제공이 아니라, 학습자가 어려움을 겪을 때는 지원을 강화하고 능숙해질 때는 지원을 줄이는 “반응적(responsive)” 상호작용이다. 즉, 학습자의 반응에 조건부로 결합된 조절만이 비계로서 기능한다.

둘째, 점진적 제거는 학습자의 역량이 향상됨에 따라 제공되던 비계를 점차 줄여나가는 과정이다. 비계는 영구적인 지지 구조가 아니라 임시적인 것이며, 학습자가 스스로 과제를 수행할 수 있는 지점에 가까워질수록 AI나 교사의 개입은 점진적으로 소거되어야 한다. 마지막으로 책임 이전은 과제 수행의 주도권과 통제권이 조력자로부터 학습자에게로 옮겨가는 단계이다. 이는 학습자가 단순히 외적인 행동을 성공시키는 것을 넘어, 문제를 진단하고 판단하며 스스로 점검하는 인지적·메타인지적 책임을 온전히 자립적으로 수행하게 되는 비계 설정의 최종 목적이다.

모든 도움이나 편의 제공이 곧바로 비계가 되는 것은 아니며 질문이나 예시를 제시하는 것이 바로 비계가 되는 것도 아니다. 학습자의 수준과 무관하게 언제나 정답을 제공하는 시스템은 겉으로는 도움이 되는 것처럼 보일 수 있으나, 엄밀한 의미의 비계라기보다 과업 대행에 가깝다. 이런 점에서 생성형 AI를 교육에 사용하는 것이 마냥 비계가 되지는 않는다. 만약 AI가 학습자의 수준에 맞는 질문과 암시를 제시하고 설명을 요구하며, 학습자가 어디에서 이해에 어려움을 겪고 있는지를 알게 이끈다면 이는 비계가 될 수 있다. 하지만 AI가 진단, 검토, 판단의 과정을 건너뛰게 한 채 정답이나 완성된 문장을 곧바로 제공한다면,

4) 흔히 비고츠키의 이론을 설명할 때 '사회적 구성주의'라는 표현을 쓰지만, 이는 후대 학자들의 해석적 범주에 가깝다. 비고츠키 자신은 문화역사이론(cultural-historical psychology) 전통에 속하며 해당 용어를 직접 사용하지 않았다.

5) 'scaffolding'이라는 용어는 우드 등(Wood, et al., 1976)의 연구에서 문제 해결 과정 중 교수자의 역할을 설명하며 처음 도입되었다.

그것은 비계가 아니라 과업의 대행이다. 더 나아가 학습자가 성장할수록 AI의 도움은 점차 줄어들어야 하며, 최종적으로는 도움 없이도 과제를 수행할 수 있어야 한다. 자전거를 배우는 아이에게 보조 바퀴는 균형 감각을 익힐 때까지만 필요한 장치이다. 보조 바퀴를 영원히 달고 있으면 넘어지지는 않지만, 결코 균형 잡는 법을 체득하지 못한다.

AI가 영구적인 보조 바퀴로 남는다면, 그것은 비계가 아니라 의존의 장치가 된다. 교육의 최종 목표는 도움을 받는 수행이 아니라, 결국 도움 없이도 스스로 사고하고 판단할 수 있는 자립에 있기 때문이다.

2.4 세 이론의 수렴: AI 활용의 판단 기준

이상의 세 이론을 종합하면 하나의 일관된 원리가 도출된다. 비요크의 이론은 교육에 있어 바람직한 어려움 자체가 수직적 성장의 연료임을 보여준다. 스웰러의 이론은 학습 간에 겪는 어려움 중 본질적 부하와 같은 교육에 필수적인 것은 남기고 외재적 부하와 같은 불필요한 것은 구분해야 함을 설명해 준다. 마지막으로 비고츠키의 이론은 학습에 있어 도움이 어떻게 주어져야 하며 그것이 언제 거둬져야 하는지를 설명한다.

교육에 AI를 사용할 때에는 적어도 이 세 이론이 말하는 바를 고려해야 한다. AI는 바람직한 어려움을 제거하는 도구가 될 수도 있지만, 외재적 부하를 줄여줘 교육을 돕는 도구가 될 수도 있다. 또한, AI는 이상적인 비계가 될 수도 있고 영구적으로 의존하게 만들어 학습을 망치는 원인이 될 수도 있다. AI가 학습에 도움이 되는가 해가 되는가는 기술 자체의 속성이 아니라, 학습의 어떤 단계에서, 어떤 방식으로, 어떤 목적으로 활용되느냐에 달려 있다. 이러한 이론적 틀을 바탕으로 이어지는 3, 4, 5절에서는 현재 교육 현장에서 AI가 바람직한 어려움을 어떻게 방해하고 있는지, 인지적 외주화는 어떠한 방향에서 일어나고 있는지 그리고 이러한 현상이 종합적으로 일어난 바둑계의 상황을 살펴 볼 것이다.

3. AI가 야기한 수행 향상과 학습의 괴리

과업의 표면적 수행 향상과 실제 학습의 내면화가 일치하지 않을 수 있다는 점은 최근의 대규모 현장 실험을 통해서도 확인할 수 있다. 콘트렉터와 레이에스(Contractor and Reyes, 2025)의 연구는 미국 엘리트 대학생들의 AI 수용 실태를 정밀하게 추적하였는데 이는 다음과 같다.

| 지표 | 수치 | 의미 |
|---------------------|--------|---|
| 학업 목적 도입률 | 80% 이상 | ChatGPT 출시 2년 만에 달성. 미국 노동자(40%)의 2배 상회 |
| 사용 목적 (수행 능력 증강) | 61.2% | 학습 보조, 피드백, 튜터링 |
| 사용 목적 (자동화) | 41.9% | 에세이 작성, 문제 해결 자동화 |
| 성적 영향 인식 | 55.4% | 영향 없음. 학습 과정은 편해졌으나, 성적 향상에 대한 확신은 부족. |

[표 1] 명문대 대학생의 생성형 AI 도입 현황 (Contractor and Reyes 2025)

이 자료에 따르면 학생들은 AI를 통해 과제 수행 시간을 단축하고 효율성을 극대화하려

한다. 하지만 성적 영향 인식에서 영향이 없다는 것은 과제 수행 능력의 증가가 학습으로 이어지지 않고 있음을 암시한다. 특히 41.9%의 학생들이 AI를 자동화 목적으로 사용한다는 점은, 학습에 있어 본질적 부하를 줄여 학습 과정에서 필수적인 바람직한 어려움을 유발하는 인지적 노력을 AI에 위임하고 있다고 해석할 수도 있다.

또한 와튼 스쿨의 바스티니 등(Bastani et al., 2025)이 진행한 연구는 기초 지식 습득 단계에서 무분별한 생성형 AI 사용이 초래할 수 있는 위험성을 극명하게 보여준다. 이 연구는 2023년 가을 터키의 한 고등학교에서 9학년부터 11학년까지 약 1,000명의 학생을 대상으로 수행된 대규모 무작위 통제 실험으로, 이미 수업에서 다룬 수학 내용을 복습하는 교내 학습 세션에 GPT-4를 도입하여 진행되었다. 각 세션은 연습 단계와 독립 평가 단계로 구성되었으며, 연습 단계에서 학생들은 AI 없이 전통적 방식으로 학습하는 통제 집단, 정답과 전체 풀이 과정을 직접 제공하는 ‘GPT 베이스(Base)’ 사용 집단 그리고 정답을 알려주는 대신 학생의 사고를 유도하는 암시와 소크라테스식 문답을 제공하도록 특수하게 설계된 맞춤형 교육 방식인 ‘GPT 튜터(Tutor)’ 사용 집단의 세 그룹으로 나뉘었다. 이어지는 평가 단계는 모든 자료와 기기를 제거한 폐쇄형 시험이었으며, 시험 문항은 직전 연습 문항과 개념적으로 매우 유사하게 설계되었다. 따라서 이 연구는 생성형 AI가 장기적·원거리 전이(far transfer)를 촉진하는지까지 검증한다기보다, 적어도 단기적이고 근접한 독립 수행(near transfer)을 훼손하지 않는지를 점검하는 설계에 가깝다는 점에 유의할 필요가 있다.

실험의 연습 단계에서는 AI를 활용한 두 집단 모두 통제 집단에 비해 압도적인 성과 향상을 보였다. 정답을 직접 제공받은 GPT 베이스 그룹은 48%, 힌트를 제공받은 GPT 튜터 그룹은 127%의 놀라운 성적 향상을 기록했다. 그러나 AI의 지원을 완전히 배제하고 치러진 독립적인 시험 단계에서는 충격적인 결과의 역전이 발생했다. 연습 과정에서 AI에 전적으로 의존했던 GPT 베이스 그룹은 AI를 전혀 사용하지 않은 통제 집단보다 오히려 17%나 낮은 성적을 기록하며 심각한 학습 결손을 나타냈다. 반면, 정답 대신 힌트와 질문을 제공받은 GPT 튜터 그룹은 시험에서 통제 집단과 통계적으로 구별되지 않는 성적을 유지하였다. 연구진의 대화 분석에 따르면 GPT 베이스 집단에서는 학생들이 정답을 직접 요구하거나 문제 문장 자체를 그대로 입력하는 방식이 두드러졌다.

| 구분 | 연습 단계(AI 지원) | 시험 단계(AI 제거) | 해석 |
|---------------------|--------------|--------------|----------------|
| 통제 집단 (AI 사용 금지) | 기준점 | 기준점 | 전통적 학습 방식 |
| GPT 베이스 (직접 정답) | +48% 향상 | -17% 하락 | 능숙함의 환상, 학습 결손 |
| GPT 튜터 (힌트 제공) | +127% 향상 | 통제 집단과 유사 | 비계 설정 통한 학습 유지 |

[표 2] AI 튜터 유형에 따른 학습 성과 비교 (Bastani et al. 2025)

이러한 수행과 학습의 괴리는 ‘바람직한 어려움’ 개념을 통해 명확히 해석될 수 있다. 진정한 학습, 즉 지식의 장기적인 저장 강도를 높이기 위해서는 학습자 스스로 기억을 인출하고 지식의 구조를 재구성하는 인지적 난관이 필수적이다. 그러나 즉각적으로 정답을 산출하는 AI는 정답의 탐색이나 오류 점검과 같은 다소 지루하고 부담스러운 인지 과정을 지나치게 단축함으로써 학습 과정에서 겪어야 할 이 필수적인 어려움을 제거해 버린다. 그

결과, 학생들은 스스로 문제를 해결하는 척만 할 뿐 실제로는 개념을 내면화하지 못하며, 당장의 과업 수행인 인출 강도만을 높이는 ‘능숙함의 환상(illusion of competence)’에 빠지게 된다.

반면 정답 대신 힌트를 제공한 GPT 튜터 그룹이 학습 결손을 겪지 않은 이유는 분명하다. 이 경우 AI는 단순한 대행자가 아니라 올바른 비계로 작용하였으며, 정답을 곧바로 제시하는 대신 힌트와 중간 시도를 요구함으로써 적어도 일부의 생성적 어려움과 점검 과정을 남겨 두어 바람직한 어려움을 적절한 수준으로 유지시켰기 때문이다. 다만 한 가지 중요하게 고려할 지점도 존재한다. GPT 튜터 역시 독립 시험에서 통제 집단을 분명히 넘어서는 학습 이득을 보여 주지는 못하였다. 이는 AI 사용에 대한 교육적 통제가 학습에 있어 부정적 효과를 줄일 수는 있어도, 그 자체가 곧바로 깊은 학습을 보장하지는 않는다는 점을 시사한다. 나아가 이 실험에서 학생들은 자신의 AI 의존이 학습을 약화시킬 수 있다는 사실을 충분히 자각하지 못한 것으로 보고되었는데, 이는 문제가 단순히 AI의 설계 방식에만 있는 것이 아니라 학습자의 메타인지(metacognition) 역량에도 있음을 보여준다. 결국 이 연구는 교육 현장에서 AI가 과업 수행을 대신하는 도구가 아니라 학습자의 인지적 노력을 촉발하는 ‘소크라테스적 질문자’로 설계되고 활용되어야 함을 시사한다. 동시에 그러한 설계만으로 충분하지 않으며 학습자 스스로의 성찰적 참여가 함께 요구됨을 분명히 한다.

4. AI가 유발한 인지적 외주화와 비계 설정의 문제

4.1 비판적 사고의 약화와 감독 부담의 증가

거를리히(Gerlich, 2025)는 다양한 연령과 학력 배경의 666명을 대상으로 설문과 심층면담을 결합한 혼합연구를 수행하여 AI 도구 사용, 인지적 외주화, 비판적 사고 사이의 관계를 분석하였다. 이 연구는 교육 현장의 통제 실험이 아니라 일상적 AI 사용의 인지적 양상을 포착한 상관 연구이므로 인과를 단정할 수는 없다. 그럼에도 AI 사용과 인지적 외주화 사이의 강한 정적 상관($r = +0.72$), AI 사용과 비판적 사고 사이의 강한 부적 상관($r = -0.68$), 그리고 인지적 외주화의 부분 매개효과($b = -0.25$)는 AI 의존이 높아질수록 깊은 사고의 실천이 약화되는 경향이 함께 나타난다는 점을 일관되게 시사한다. 특히 이 연구에서 약화된 것으로 보고된 역량은 단순한 사실 회상이 아니라 소스 평가, 정보 검증, 분석, 추론과 같은 비판적 사고의 핵심 요소였다. 따라서 이 경우 외주화된 것은 참고문헌 정리나 문장 다듬기 같은 주변적 마찰이 아니라, 판단 기준을 세우고 근거를 선별하는 본질적 처리였다고 해석하는 편이 타당하다. 다시 말해, AI는 불필요한 외재적 부하를 줄이는 데 그치지 않고, 학습자와 사용자가 직접 수행해야 할 핵심 인지 작업까지 함께 덜어 주는 방향으로 작동했을 가능성이 크다.

| 경로 | 효과 | 해석 |
|--------------------|----------------------------|---|
| AI 사용 빈도 → 인지적 외주화 | $r = +0.72$ | AI를 자주 사용할수록 기억, 판단, 정보 탐색 같은 인지 과정을 외부 도구에 더 많이 위임하는 경향이 강하다. |
| AI 사용 빈도 → 비판적 사고 | $b = -0.42$ $r = -0.68$ | AI 사용이 많을수록 비판적 사고 수준은 유의미하게 낮아지는 방향을 보였 다. 다만 이 연구는 상관-매개 분석 중심이므로 인과를 단정해서는 안 된다. |
| 인지적 외주화 → 비판적 사고 | $b = -0.25$ $r = -0.75$ | 외주화 자체가 비판적 사고 약화와 유의미하게 관련된다. 즉, 핵심 메 커니즘은 단순한 AI 사용 자체보다 ‘사고 위임’일 수 있다. |

| | | |
|-------------------|-----------|--|
| AI 사용 빈도 → 비판적 사고 | b = -0.17 | 인지적 외주화를 통제된 뒤에도 AI 사용은 비판적 사고에 부정적 직접 효과를 남겼다. |
| 소스 평가 능력 | -0.494 | 비판적 사고 하위 요소 중에서도 출처를 가리고 정보의 신빙성을 점검하는 능력이 특히 크게 약화된 것으로 보고되었다. |

[표 3] AI 사용, 인지적 외주화, 비판적 사고의 구조적 관계 (Gerlich 2025)⁶⁾

리 외(Lee et al., 2025)는 지식노동자 319명으로부터 생성형 AI 활용 사례 936건을 수집하여 실제 업무 맥락에서 생성형 AI가 비판적 사고의 실행을 어떻게 변화시키는지 조사하였다. 이 연구의 강점은 AI에 대한 일반적 태도를 묻는 데 머물지 않고, 생성, 정보 탐색, 조언 획득이라는 서로 다른 과업 유형에서 목표 설정, 프롬프트 구성, 응답 검토, 응답 통합이 어떻게 이루어지는지를 구체적으로 포착했다는 점에 있다. 연구 결과, AI에 대한 신뢰가 높을수록 지식, 이해, 분석, 종합, 평가의 여러 차원에서 비판적 사고에 드는 노력이 줄어든다고 지각하는 경향이 확인되었다. 반대로 자기 능력에 대한 자신감이 높은 사용자는 AI 응답을 적용하고 평가하는 데 더 많은 노력을 기울였다. 중요한 것은 이 노력 감소가 언제나 바람직한 부하 감소를 의미하지는 않는다는 점이다. 질적 분석에 따르면, 생성형 AI는 정보 수집과 정리의 일부를 자동화하는 동시에 검증, 통합, 방향 설정, 결과물 감독이라는 새로운 과업을 사용자에게 남겼다. 따라서 사용자가 AI를 지나치게 신뢰할 경우, 줄어드는 것은 단지 외재적 부하만이 아니라 목표 설정, 질문 정교화, 출처 검토, 응답 비판과 같은 본질적 처리일 수 있다. 이 점에서 생성형 AI는 보조 도구와 사고 대행 장치 사이를 오가는 양면적 도구로 보는 편이 적절하다. 다만 이 연구 역시 자기보고 자료에 기초한 조사이므로, 실제 능력 저하의 지속성과 크기는 후속 연구를 통해 더 엄밀하게 검증될 필요가 있다.

| 경로 (Path) | 효과 | 해석 |
|----------------------------|--|---|
| AI 신뢰 → 비판적 사고역이 높음 | $\beta = -0.69$ | AI가 그 과업을 잘할 것이라고 믿을수록, 사용자는 실제로 비판적 사고역이 높은 경향을 보였다. |
| 자기 수행 확신 → 비판적 사고역이 높음 | $\beta = +0.26$ | 자신이 그 과업을 직접 수행할 수 있다고 믿는 사람일수록 AI 사용 중에도 비판적 사고역이 높은 경향을 보였다. |
| AI에 대한 평가 확신 → 비판적 사고역이 높음 | $\beta = +0.31$ | AI 결과를 평가할 수 있다는 확신이 높을수록, 실제 검토와 판단을 더 많이 수행하였다. |
| 반성 경향 → 비판적 사고 실행 | $\beta = +0.52$ | 평소 자신의 일을 되돌아보는 성향이 강한 사람일수록 AI 사용 상황에서도 비판적 사고를 유지하였다. |
| AI 신뢰 → 비판적 사고 저하 | 지식 -0.11/이해 -0.13/분석 -0.15/종합 -0.12/평가 -0.23 | AI를 신뢰할수록 비판적 사고의 여러 하위 활동에 드는 노력이 줄어든다고 지각하였다. 특히 평가 차원 감소가 가장 컸다. |

6) 거를리히(2025)에서 r은 상관계수(correlation coefficient)이다. 이는 두 변수가 함께 움직이는 정도를 보여 주는 값으로, +이면 한 변수가 커질수록 다른 변수도 함께 커지는 경향, -이면 한 변수가 커질수록 다른 변수는 작아지는 경향을 뜻한다. 또 0에 가까울수록 관계가 약하고, 절댓값이 1에 가까울수록 관계가 강하다. 따라서 r=+0.72는 AI 사용이 많을수록 인지적 외주화도 함께 커지는 강한 정적 관계를, r=-0.68은 AI 사용이 많을수록 비판적 사고는 낮아지는 강한 부적 관계를 뜻한다. 표에 제시된 -0.494 역시 같은 종류의 수치로, 소스 평가 능력과 관련하여 AI 사용이 많을수록 정보의 출처를 가리고 검증하는 능력이 더 낮게 나타났다는 뜻으로 읽으면 된다. 한편 b는 회귀분석·매개분석에서 쓰이는 계수로, 다른 요소들을 함께 고려했을 때 한 변수가 증가할수록 결과 변수가 어느 방향으로 움직이는지를 보여 준다. 따라서 총효과 b=-0.42는 AI 사용이 늘수록 비판적 사고가 전반적으로 낮아지는 방향의 효과를, 간접효과 b=-0.25는 그 하락의 일부가 인지적 외주화의 증가를 거쳐 나타난다는 뜻이다.

| | | |
|------------------------|------------------------|--|
| 자기 수행 확신 → 적용/평가 노력 | 적용 +0.08 / 평가 +0.10 | 자기 능력에 대한 확신이 높은 사용자는 AI가 있어도 적용과 평가에 더 많은 인지적 노력을 들었다. |
|------------------------|------------------------|--|

[표 4] 생성형 AI 신뢰, 자기 확신, 비판적 사고 실행의 관계: 리 외(Lee et al., 2025)⁷⁾

4.2 미루기, 수동적 의존, 불평등의 재생산

압바스 외(Abbas et al., 2024)는 AI 사용이 학습자의 심리적 태도와 학업 성과에 미치는 부정적 경로를 탐색하였다. 이 연구는 두 단계로 설계되었다. 첫 번째 단계에서는 대학생 165명을 대상으로, 과제·프로젝트·시험 준비 등 학업 전반에 걸친 ChatGPT 사용 빈도를 측정하는 척도를 개발하고 그 타당성을 분석하였다. 연구진은 전문가 검토와 통계적 분석을 거쳐 8개의 최종 문항을 확정하였으며, 이를 통해 척도의 신뢰도와 타당도를 입증하였다. 두 번째 단계에서는 학생 494명을 대상으로 시간 간격을 둔 종단적 설문을 실시하여, 학업 부담·시간 압박·보상 민감성·품질 민감성이 ChatGPT 사용에 미치는 영향과, 이러한 사용이 미루기 습관, 기억력 저하, 학업 성적에 미치는 영향을 연구하였다. 완전한 무작위 배정 실험이 아니어서 인과관계를 단정하기는 어렵지만, 단일 시점이 아닌 복수 시점에서 변인을 추적했다는 점에서 방법론적 의의가 있다.

분석 결과, 학업 부담이 크고 시간 압박을 크게 느끼는 학생일수록 ChatGPT를 더 많이 사용하는 경향이 나타났으며, 이렇게 증가한 사용량은 학업적 미루기 습관 강화, 기억력 저하, 학업 성취도 하락으로 이어지는 것으로 확인되었다. 여기서 특히 주목할 점은 1단계에서 개발된 척도 문항들이 단순히 문법 교정이나 서식 정리와 같은 주변적 과업을 넘어, 과제 수행·프로젝트 준비·개념 학습 등 학업의 핵심 과정에 대한 AI 의존도를 폭넓게 측정했다는 사실이다. 따라서 이 결과는 3절에서 논의한 수행 향상과 실제 학습의 괴리가 AI를 활용한 교육 현장에서도 일어나고 있음을 보여준다. 학생들이 단지 외재적 부하만을 줄인 것이 아니라, 과제 이해, 개념 정리, 독립적 준비와 같은 학습의 중심 과정을 AI에게 맡겨 AI 의존도가 높아졌을 가능성을 시사하는 것이다. 다시 말해, 교육에 AI를 무분별하게 사용할 경우 인지적 외주화는 편의 제공을 넘어 수동적 의존으로 확장될 수 있다.

한편 AI를 둘러싼 불평등의 문제는 개인 수준에 그치지 않는다. 카르바할, 프랑코, 이삭손(Carvajal, Franco, & Isaksson, 2024)은 학생 595명과 관리자 1,143명을 대상으로 두 차례의 실험을 진행하여, 생성형 AI 수용에서 나타나는 성별 격차와 관련 정책의 효과를 분석하였다. 학생 대상 조사에서는 여성 참여자, 특히 상위권 여학생들이 AI 사용을 더 많이

7) 리 외(2025)에서 β (베타)는 혼합효과 회귀모형의 계수이다. 쉽게 말해, 어떤 설명변수가 한 단계 커질 때 결과가 어느 방향으로 얼마나 움직이는지를 나타내는 값이다. 다만 이것은 퍼센트가 아니므로, 예컨대 $\beta = -0.69$ 를 곧바로 “69% 감소”라고 읽으면 안 된다. 이 논문에서 한 결과변수는 ‘생성형 AI 사용 중 비판적 사고를 실제로 했는가’인데, 여기서 AI에 대한 자신감 $\beta = -0.69$ 는 AI를 더 잘한다고 믿을수록 비판적 사고를 실제로 할 가능성이 낮아지는 방향을 뜻한다. 반대로 자기 능력에 대한 자신감 $\beta = 0.26$, AI 결과를 평가할 수 있다는 자신감 $\beta = 0.31$ 은 비판적 사고를 실제로 할 가능성이 높아지는 방향을 뜻한다. 또 표의 지식, 이해, 분석, 종합, 평가 옆에 붙은 -0.11 , -0.13 , -0.15 , -0.12 , -0.23 같은 값은 각 인지 활동에 드는 노력의 지각이 어떻게 달라졌는지를 뜻한다. 이 논문은 노력을 ‘훨씬 덜 든다(-2)~비슷하다(0)~훨씬 더 든다(+2)’의 척도로 물었기 때문에, 이 값들이 음수라는 것은 생성형 AI를 더 신뢰할수록 그 활동이 덜 힘들게 느껴졌다는 뜻이다. 예를 들어 Knowledge의 $\beta = -0.11$ 은 AI에 대한 자신감이 한 단계 높아질 때 사실을 떠올리거나 찾는 데 드는 노력감이 조금 줄어든다는 뜻이고, Evaluation의 $\beta = -0.23$ 은 여러 하위 활동 가운데 평가 활동에서 그 감소가 더 크게 지각되었다는 뜻이다. 따라서 이 수치들은 “비판적 사고 능력이 그만큼 사라졌다.”는 직접 측정값이 아니라, 사용자들이 그렇게 느꼈다는 자기보고식 노력 변화의 계수라고 이해하는 것이 정확하다.

기피하는 경향이 관찰되었다. 그러나 교수자가 AI 사용을 명시적으로 허용한 환경에서는 이러한 성별 격차가 사실상 소멸하였으며, 반대로 AI 사용을 금지한 환경에서는 여성의 사용 의도가 남성에 비해 훨씬 더 크게 감소하는 것으로 나타났다. 관리자 대상 조사에서는 상위권 여성 지원자가 생성형 AI 활용 능력을 드러낼 때 평가 점수가 더 높아지는 양상도 확인되었다.

이 연구가 제기하는 문제는 단순히 여성이 AI를 덜 사용한다는 표면적 관찰에 있지 않다. 더 본질적인 문제는 학교나 기관의 정책이 이 격차를 확대할 수도, 축소할 수도 있다는 점이다. AI 사용을 전면 금지하는 정책은 단기적으로 부정행위를 억제하는 효과를 낼 수 있으나, 동시에 통제된 환경에서 AI를 다루고 익히는 기회를 불평등하게 박탈함으로써 장기적인 역량 격차를 심화시킬 위험이 있다. 반대로 아무런 조건 없이 사용을 허용하는 것 역시 학습의 핵심 과정을 기계에 전적으로 위임하는 부작용을 초래할 수 있다. 결국 이 연구는 무조건적 허용과 전면적 금지라는 이분법적 접근 모두에 한계가 있음을 보여주며, 어떤 과제에서 AI를 보조 도구로 허용하고 어떤 과제에서 학생의 독립적 수행을 요구할 것인지에 대한 정교한 기준의 필요성을 제기한다. 이러한 관점에서 불평등의 문제는 단순히 도구에 대한 접근 가능성의 문제가 아니라, 학생들이 어떤 교육적 맥락과 규칙 속에서 그 도구를 경험하도록 할 것인지의 문제와 직결된다.

4.3 자동화의 아이러니와 고위험 영역의 검증 문제

베인브리지(Bainbridge, 1983)가 말한 자동화의 아이러니는 생성형 AI 환경에서 다시 중요해진다. 자동화는 표면적으로 인간의 부담을 덜어 주는 듯 보이지만, 실제로는 예외 처리와 감독에서 더 높은 수준의 판단을 요구한다. 생성형 AI의 경우도 마찬가지이다. 마게시 외(Magesh et al., 2025)는 Lexis+ AI, Westlaw AI-Assisted Research, Ask Practical Law AI, GPT-4를 대상으로 사전등록된 법률 질의 평가를 수행하였다. 이 연구는 일반 법리, 최신 판례 변화, 허위 전제 질문, 사실 회상 질문 등 202개의 법률 질의를 바탕으로, RAG 기반 법률 AI 도구가 실제로 환각을 줄였는지를 수작업으로 검증한 최초의 체계적 평가라는 점에서 의미가 크다. 결과적으로 RAG를 적용한 도구들조차 17%에서 33% 수준의 환각을 보였고, 응답의 정확성과 완결성에도 상당한 차이가 남아 있었다. 이는 RAG가 검색의 일부를 도와줄 수는 있어도, 사용자가 출력의 정오성, 근거 적합성, 법원의 위계, 최신성까지 직접 검토해야 한다는 사실을 보여 준다.

법률 영역에서 이러한 검토는 단순한 맞춤법 검사나 형식 점검이 아니다. 어떤 판례가 핵심 판시인지, 인용된 자료가 실제로 그 명제를 지지하는지, 다른 최신 판례에 의해 수정되지는 않았는지를 판단하는 일은 법적 전문성의 심장부에 속한다. 따라서 생성형 AI가 줄여 주는 것은 검색의 일부일 수 있으나, 남은 과업은 오히려 더 높은 수준의 본질적 부하일 수 있다. 문제는 학습자가 아직 그런 감독 역량을 갖추기 전에 AI에 장기적으로 의존할 경우, 표면적 생산성은 높아져도 감독을 수행할 전문성 자체는 취약해질 수 있다는 점이다. 이 점에서 생성형 AI는 단순히 과업을 덜어 주는 도구가 아니라, 숙련이 완성되기 전에는 오히려 더 엄격한 검증 역량을 요구하는 기술이라고 보는 편이 적절하다.

5. 먼저 온 미래: 바둑 사례가 보여주는 것

교육 분야 밖에서 AI가 인간의 지식 체계와 전문성 발달에 미친 영향을 가장 선명하게 보여주는 사례는 바둑이다. 바둑계가 2016년 알파고(AlphaGo) 이후 겪은 변화는 다른 전문 분야들이 머지않아 마주할 미래의 축소판이라 할 수 있다. 기계가 전문가의 권위에 도전하고, 전통적 교육 시장이 흔들리며 AI 기술에 적응한 자와 적응하지 못한 자 사이에 새로운 차이가 생겨나는 바둑계의 현실은 고등교육에 닥칠 변화를 가능하게 해 준다. 특히 중요한 것은 AI가 누구에게나 똑같은 방식으로 작동하지 않았다는 점이다. 최석웅 등(Choi et al., 2025a)은 2015년부터 2019년까지 1,241명의 프로 기사, 24,973국, 749,190수를 분석하여 이 복잡한 변화를 정밀하게 보여준다. 여기서는 이 연구를 중심으로, 바람직한 어려움, 인지 외주화, 비계 설정이 바둑 기사들의 공부법과 대국 현장에서 어떻게 드러났는지 살펴보고, 그것이 교육의 목적을 다시 묻게 만드는 이유를 정리해 보고자 한다.

5.1 지식의 파괴와 창조: 정석의 혁명

최석웅 등(Choi et al., 2025a)에 따르면, AI 바둑 프로그램의 공개 이후 프로 기사들의 수질(手質), 즉 착수의 정확도는 평균 0.756 퍼센트 포인트, 약 30.5퍼센트 향상되었다.⁸⁾ 기사들의 착수는 AI의 최선책과 더 자주 일치했고, 이 변화는 단순한 외적 환경 변화가 아니라 인간이 AI로부터 실제로 학습한 결과였다. 특히 효과는 게임 초반에서 가장 컸다. 이는 가능한 수가 많고 불확실성이 큰 영역일수록 AI의 개입이 기존 직관과 관습적 정석을 더 강하게 흔들었음을 시사한다.

이 변화는 바둑에서 무엇이 좋은 수인가를 판단하는 기준 자체를 바꾸었다. 과거에는 직관과 전통적 감각으로 정당화되던 수들이 AI 앞에서 다시 평가되었고, 대표적으로 초반 3·3 침입 같은 수들이 새로운 표준으로 자리 잡았다는 견해가 있다(이세돌, 서울포럼 2025). 그 결과 기사들은 단지 몇 개의 새 정석을 더 배우는 데 그치지 않고, 자신이 오랫동안 당연하게 여겨 온 판단의 틀을 다시 배워야 하는 상황에 놓이게 되었다는 것이다. 한 세대에 걸쳐 쌓아온 지식과 직관이 극복해야 할 편견이자 장애물이 될 수 있다는 지적도 제기된다(AI타임스 2025). 바둑에서 지식의 축적이 곧바로 권위로 이어지던 시대가 이 지점에서 흔들리기 시작했다는 것이 바둑계의 대체적인 분위기다.

하지만 교육적으로 더 중요한 것은 모든 어려움이 제거되어야 하는 것이 아니라는 사실이다. AI는 낡은 편견과 비생산적 시행착오를 걷어내는 데 큰 힘을 발휘한다. 그러나 스스로 수를 세워 보고, 왜 그 수를 두었는지 설명하고, 패배한 대국을 복기하면서 판단의 구조를 다시 세우는 과정은 여전히 기사 자신의 몫이다. 바로 이런 어려움이 바람직한

8) 최석웅 등(Choi et al., 2025a)의 원문에서는 알파고(AlphaGo) 이후 등장한 AI 바둑 프로그램 일반을 APG(AI-powered Go program)으로 통칭하나, 본고에서는 이를 'AI 바둑 프로그램'으로 표기한다. 알파고(AlphaGo) 이후 현재까지 프로 기사들의 연구 및 대국 중계에 활발히 활용되는 대표적인 AI 바둑 프로그램으로는 카타고(KataGo), 릴라 제로(Leela Zero), 그리고 국내 NHN이 개발한 한돌(HanDol)이 있다. 이 가운데 카타고는 오픈소스 프로그램으로 현재 프로 기사들이 연구 목적으로 가장 널리 사용하며, K바둑과 바둑TV 등 국내 주요 바둑 전문 채널의 생중계에서 실시간 형세 분석 도구로도 활용되고 있다. 한돌은 2019년 이세돌 9단과의 공식 대국에서 2점 접바둑으로 맞붙어 화제를 모은 바 있으며, 한게임 바둑 서비스를 통해 일반인도 접근할 수 있다.

한편 본문의 두 수치는 서로 다른 것을 나타낸다. AI 바둑 프로그램 도입 전 기사들의 수질은 AI 최선책보다 평균 2.47%p 낮았는데, 도입 후 이 격차가 0.756%p 줄었다. 30.5%는 이 절대적 감소량(0.756%p)이 기존 격차(2.47%p)의 몇 퍼센트에 해당하는지를 나타내는 상대적 비율이다. 즉 0.756%p는 승률 격차의 절대적 변화량이고, 30.5%는 그것이 기존 격차 대비 얼마나 개선되었는지를 나타내는 상대적 크기이다.

어려움이다. AI가 초반의 좋은 수를 빨리 보여준다고 해서 그 논리까지 자동으로 내면화해 주는 것은 아니기 때문이다. 바둑계의 정석 혁명은 학생들이 오늘날 던지는 “AI가 더 잘하는데 왜 공부해야 하는가?”라는 질문이 단순한 불안이 아니라, 지식 체계 자체가 재편되는 상황에서 나온 실질적 질문임을 보여준다.

5.2 흡수 역량과 바람직한 어려움: 누가, 어떻게 AI에서 배우는가

AI의 혜택이 모든 기사에게 똑같이 돌아간 것은 아니다. 최석웅 등(Choi et al., 2025a)은 젊은 기사들의 수질 향상 폭이 더 컸고, 동시에 모든 실력대가 도움을 받되 상대적으로 덜 숙련된 기사들이 더 큰 한계편익(限界便益)을 얻었다고 보고한다. 이 결과는 바둑계에서 AI가 기존 장벽을 낮추는 동시에 새로운 차이를 다시 만들어 냈음을 시사한다. 누구나 같은 AI 바둑 프로그램을 사용할 수 있게 되었지만, 그것을 스스로 판단해 얼마나 깊이 흡수하느냐는 또 다른 문제였기 때문이다.

이 차이를 가르는 핵심은 흡수 역량(Absorptive Capacity)이다. 곧 AI가 제시하는 수의 의미를 해석하고, 그것을 자기 전략 체계와 연결하여 다시 자기 것으로 만드는 능력이다. AI는 승률과 승률을 높이기 위한 대안을 즉각적으로 보여줄 수 있지만, 이러한 도움을 받을 수 있다고 해서 바로 실력이 향상되는 것은 아니다. 왜 이 수가 좋은가, 내가 원래 두려웠던 수는 왜 밀리는가, 상대의 응수까지 포함하면 형세 판단이 어떻게 바뀌는가를 스스로 따져 보아야 비로소 학습이 일어난다. 이 지점에서 바람직한 어려움이 다시 중요해진다. AI가 답을 곧장 주더라도, 그 답을 자신의 언어로 다시 설명하고 재구성하는 부담은 학습자가 짊어져야 하는 것이다.

최석웅 등(Choi et al., 2025a)이 인용한 신진서의 말은 이 점을 잘 보여준다. 그는 2017년부터 AI 바둑 프로그램을 사용해 왔고, AI가 제시한 수를 본 뒤 다른 변화도도 검토하며 각 수별 승률을 이해하는 것이 새로운 공부법이 되었다고 밝혔다. 중요한 것은 AI의 수를 베끼는 일이 아니라, 그 수가 왜 나왔는지를 자기 반성을 통해서 다시 묻는 일이다. 신진서가 보여 준 것은 AI가 강한 스승이라는 사실보다, 그 스승에게서 배우는 방식이 수동적 모방이 아니라 능동적 재구성일 때에만 수직적 도약이 가능하다는 사실이었다.

한편, AI의 등장이 바둑계의 성별 역학에도 변화를 가져왔다는 보도도 있다. 과거에는 최고의 지도 사범과 훈련 파트너에 대한 접근성이 실력 향상의 장벽으로 작용했는데, AI가 이 장벽을 허물면서 최정 9단, 김은지 9단과 같은 최상위 여성 기사들이 남녀 통합 랭킹 20~30위권에 진입하는 성과를 거두었다는 것이다(연합뉴스 2025; 서울신문 2025). 2절에서 살펴본 비요크의 이론이 강조하는 바와 같이, 바람직한 어려움은 학습자가 성공적으로 대응할 수 있는 배경 지식과 기술을 갖추었을 때 비로소 제 기능을 한다. 이 사례는 AI가 학습 접근성의 구조적 장벽을 낮춤으로써, 기량 자체에 집중할 수 있는 조건을 새롭게 만들 수 있다는 점을 시사한다.

이는 교육에도 접목이 가능하다. AI를 도입했다고 해서 모든 학습자가 최고 수준의 성취에 도달하는 것은 아니다. 어떤 학생은 AI를 통해 사고의 지평을 넓히지만, 어떤 학생은 AI에게 판단 자체를 넘겨 버린다. 바둑에서 AI의 효과가 기사마다 달랐듯이, 교육에서도 핵심은 접근권 자체보다 그것을 흡수할 수 있는 조건을 어떻게 만들어 주느냐다. “아는 만큼 보인다.”는 격언은 AI 시대에 더욱 냉혹한 현실이 되는 것이다.

5.3 인지 외주화의 함정과 비계 설정의 역설

AI 바둑 프로그램은 기사들에게 막대한 도움을 주었지만, 그만큼 인지 외주화의 유혹도 강하게 만들었다. 바둑판의 어느 장면에서든 최선의 수와 승률 변화를 곧바로 보여주는 도구를 오래 사용하다 보면, 스스로 판단하고 설명하는 과정이 줄어들기 쉽다. 최석웅 등(Choi et al., 2025b)이 보여주듯 AI 도입 이후 프로 기사들의 대국 초반 돌의 배치인 포석(布石)의 다양성은 감소했고, 전략은 점차 일정한 형태(pattern)로 수렴했다. 이것은 단순히 효율적인 전략이 승리했다는 뜻만이 아니다. 기사들이 AI의 수를 이해하기보다 모방하는 방식으로 활용할 때, 사고 전체가 하나의 단일한 형태로 일반화될 수 있음을 말해준다.

비계 설정의 관점에서 보면 이 문제는 더 분명해진다. 좋은 비계는 학습자가 혼자서는 도달하기 어려운 수준까지 올라가도록 돕되, 언젠가는 제거되어야 한다. AI 바둑 프로그램은 조건부 조절의 측면에서는 매우 강력하다. 각 수의 승률 변화, 대안 수, 최선책과의 차이를 정량적으로 보여 주기 때문이다. 최석웅 등(Choi et al., 2025a)에 따르면 AI 바둑 프로그램 도입 이후 오류의 횟수는 33.7% 감소했고, 치명적 실수의 규모는 21.9% 감소했다. AI가 실제로 기사들의 판단을 정교하게 다듬어 주는 훌륭한 학습 도구라는 뜻이다.

하지만 AI가 훌륭한 학습 도구라는 점 때문에 우리는 더 조심해야 한다. 프로 대국에서는 AI 바둑 프로그램을 실시간으로 사용할 수 없다. 기사들은 훈련에서는 도움을 받되, 실전에서는 자기 힘으로 수를 두어야 한다. 따라서 훈련 과정에서 AI가 비계로 작동하려면, 스스로 먼저 읽고 판단한 다음 AI와 비교하고, 마지막에는 AI 없이도 유사한 형세 판단을 해낼 수 있어야 한다. 처음부터 끝까지 AI가 알려 주는 최선의 수와 승률만 따르는 공부법이라면 그것은 비계가 아니라 영구적인 보조 바퀴에 가깝다.

이세돌 9단의 은퇴는 성과 중심의 교육 목적을 경고하는 한 사례로도 생각할 수 있다. 그는 기량의 저하가 아닌 “목적의 상실”로 은퇴를 결심했다고 밝혔다. 인공지능의 등장이 은퇴에 많은 부분을 차지했으며, 정상에 오르더라도 이길 수 없는 존재가 있다는 깊은 무력감을 느꼈다는 것이다(매일경제 2019; YTN 사이언스 2024). 여기서 무력감은 실력의 문제가 아니다. AI 바둑 프로그램이 항상 더 나은 수를 보여주는 환경에서, 기사는 자신의 판단이 아무리 최선이어도 그것이 진정한 최선이 아님을 알게 된다. 이 상황이 지속되면 스스로 수를 생각하는 행위 자체가 의미를 잃는다. 이는 단순한 인지 외주화를 넘어, 존재론적 차원의 목적 상실로 이어진 사례로 볼 수 있다. 이세돌의 은퇴는 단순한 기사의 퇴장이 아니라, AI에 의해 내적 동기를 잃어버린 학습자의 미래를 예견하는 사례로 언급된다.

최석웅 등(Choi et al., 2025a)이 성능이 낮았던 이전 세대 프로그램인 크레이지 스톤(Crazy Stone) 공개 이후에는 바둑계에 유의미한 변화가 나타나지 않았다고 보고한 사실도 흥미롭습니다. 아무 AI 프로그램이나 곁에 두는 것이 곧 좋은 비계가 되는 것은 아니라는 것이다. 중요한 것은 도구의 존재 자체가 아니라, 그 도구가 얼마나 정확한 피드백을 주는가, 그리고 학습자가 그것을 자기 판단력으로 다시 번역하는가이다. 교육에서도 사정은 같다. 학생에게 AI를 붙여 주는 것만으로는 충분하지 않다. 도움이 언제 주어져야 하고 언제 줄어들어야 하는지, 최종적으로 무엇이 학생 자신의 힘으로 남아 있어야 하는지를 세밀하게 설계해야 하는 것이다.

5.4 전통적 교육 시장의 재편과 다양한 미래: 획일적 처방을 경계하며

AI는 바둑계에서 누가 더 강해지는가만 바꾼 것이 아니라, 기사들이 어떤 방식으로 살아가고 어떤 역할을 맡는가도 바꾸었다. 어떤 기사들은 AI를 깊이 흡수하여 최고급 기사로 도약했고, 어떤 기사들은 해설, 교육, 온라인 강의, 유튜브 콘텐츠 제작과 같은 다른 방향으로 활동의 중심을 옮겼다. <김성룡 바둑랩>은 구독자 17만 명, 누적 조회수 3억 회를 넘어섰으며, 한국기원도 프로 기사들의 유튜브 활동을 별도로 조명하고 있다(한국기원 2023). AI의 수를 인간이 이해할 수 있는 언어로 해석하고 학습자의 심리적 동기를 관리하는 영역에서 인간 교육자의 새로운 역할이 모색되고 있다는 견해도 있다(AI타임즈 2025). 중요한 것은 AI가 모두를 같은 종착지로 데려간 것이 아니라는 점이다. 같은 도구를 써도 누군가는 최정상 대국 기사로 성장했고, 누군가는 새로운 형식의 교육자와 해설자로 자리 잡았다.

이 점은 교육의 목적을 생각할 때 매우 중요하다. 바둑 기사들이 모두 신진서가 될 수는 없다. 그리고 모두가 신진서가 될 필요도 없다. 교육도 마찬가지다. 아무리 정교한 AI 기반 제도를 만들더라도, 교육의 목표를 오직 “모든 학습자를 최고 성취자로 만든다.”는 하나의 기준으로 환원할 수는 없다. 어떤 학습자에게 AI는 고난도 문제 해결력을 끌어올리는 도구가 될 수 있지만, 다른 학습자에게는 기초 개념을 자립적으로 이해하게 하는 보조 장치가 될 수도 있고, 또 다른 학습자에게는 진로 탐색과 표현 능력, 혹은 새로운 직업적 방향을 찾게 해 주는 계기가 될 수도 있다.

따라서 교육에서 AI를 무조건적으로 활용해야 한다는 생각도 경계해야 하고, AI를 무조건 부정적으로 보아야 한다는 생각도 경계해야 한다. 최석웅 등(Choi et al., 2025a)은 성능이 낮았던 이전 세대 프로그램에서는 유의한 학습 효과가 나타나지 않았고, AI 바둑 프로그램에 대한 노출이 상대적으로 낮았던 일본의 경우 향상 폭도 더 작았음을 보고한다. 이 두 사실은 동시에 중요한 교훈을 준다. 첫째, 모든 AI가 좋은 것은 아니다. 둘째, 그렇다고 AI를 멀리하는 것이 답도 아니다. 중요한 것은 어떤 AI를 어떤 목적과 단계에서 어떻게 쓰느냐다.

이 때문에 기존의 교육 방식에만 기대는 태도 역시 재고할 필요가 있다. 예전의 강의, 복습, 시험이라는 일률적 구조만으로는 충분하지 않을 수 있다. 먼저 자기 힘으로 답을 만들어 보고, 그다음 AI와 비교하며, 왜 차이가 났는지를 설명하고, 마지막에는 다시 AI 없이 수행해 보는 식의 새로운 설계가 오히려 더 적절할 수 있다. 바둑계가 먼저 보여 준 것은 AI를 도입할 것인가 말 것인가라는 단순한 찬반의 문제가 아니다. 무엇을 지켜야 하고 무엇을 바꾸어야 하며, 각 학습자에게 어떤 성공의 형태가 더 적절한지까지 함께 고민해야 한다는 점이다. AI 시대의 교육은 모든 사람을 하나의 이상형으로 몰아가는 제도가 아니라, 서로 다른 목적과 서로 다른 성장 경로를 정직하게 인정하고 설계하는 제도가 되어야 한다.

6. 해결 전략: AI 활용과 제한의 기준

지금까지 바람직한 어려움이 수직적 성장에 필수적이라는 이론적 배경과, 무분별한 AI 의존이 인지적 외주화를 초래할 수 있다는 여러 연구의 경고를 살펴보았다. 그러나 5절에서 바둑계의 사례를 분석하면서 확인했듯이, AI의 영향은 단선적이지 않다. AI 바둑 프로그램 도입 이후 어떤 기사들은 최고급 대국 기사로 도약했고, 어떤 기사들은 해설·교육·콘텐츠 제작이라는 새로운 방향으로 자신의 역량을 전환했다. 모든 기사가 신진서가 될 수 없고 또 그럴 필요도 없듯이, 고등교육의 목표 역시 최고 성취자를 양성하는 단일한 기준으로 환원될

수 없다. 어떤 학습자에게 AI는 고난도 문제 해결력을 끌어올리는 도구가 되고, 다른 학습자에게는 기초 개념을 자립적으로 이해하게 하는 보조 장치가 되며, 또 다른 학습자에게는 새로운 진로와 표현 능력을 탐색하는 계기가 될 수 있다. 즉, 교육에서 중요한 것은 AI가 학생 대신 답을 산출하는가가 아니라, 각 학습자의 목표와 단계에 비추어 AI가 학습의 연습 방식을 어떻게 바꾸고 무엇을 스스로 판단해야 하는 문제로 남겨 두는가이다.

그렇다면 교육 현장에서 AI를 구체적으로 언제 허용하고 언제 통제해야 하는가? 이 질문에 제대로 답하려면 교육의 목적이 무엇인지, 그 목적에 부합하는 AI는 어떤 것인지, 어떤 방식으로 써야 하는지, 심지어 그 목적을 따를 때 AI를 반드시 써야 하는지조차 먼저 논의해야 한다. 그러나 이 모든 물음을 본고에서 충분히 다루기는 어렵다. 따라서 6절에서는 교육의 목적이나 맥락이 무엇이든 최소한 지켜야 할 제한 조건을 제안하는 데 초점을 맞추고자 한다. 핵심 판단 기준은 앞서 살펴본 세 이론에서 공통으로 도출된다. AI가 학습자의 인지적 부하를 전략적으로 줄여 주는가, 아니면 학습자가 스스로 수행해야 할 본질적 처리까지 대신해 버리는가이다. 이 기준에서 볼 때, 학습은 단일한 사건이 아니라 정보가 처음 입력되어 내면화되고 최종 결과물로 산출되는 일련의 인지적 순환 과정이다. 앞서 살펴본 인지 부하 이론의 절약 원리나 비요크의 바람직한 어려움이 학습의 모든 단계에 동일하게 적용되는 것은 아니기 때문에, 단계별 구분이 필요하다.

본 장에서는 앞선 이론적 논의들을 바탕으로, 대학 수준의 보편적인 학습 과정을 시간의 흐름과 인지적 발달 수준에 따라 여섯 가지 핵심 영역으로 세분화하여 각 영역에 맞는 구체적인 AI활용 지침을 제시하고자 한다. 이 여섯 영역은 새로운 개념과 지식을 처음 마주하는 “기초 지식 습득”, 기본기를 체화하는 “반복 과업”, 사고의 폭을 넓히는 “아이디어 발산”, 정보의 논리와 타당성을 검증하는 “비판적 사고 훈련”, 성장을 위한 조언을 얻고 수정하는 “피드백 및 교정”, 그리고 독립적 역량 또는 실제 협업 역량을 확인하는 “최종 평가”로 구성된다.

6.1 기초 지식 습득: 강력한 절제

첫 번째 영역인 기초 지식 습득에는 강력한 절제가 요구된다. 기초 지식 습득은 새로운 개념을 처음 이해하고, 그것을 장기 기억 속에 구조화하며, 자기 언어로 재구성하는 학습의 출발점이기 때문이다. 비요크 부부(Bjork and Bjork, 2011)와 슈미트와 비요크(Schmidt and Bjork, 1992)가 보여주듯, 학습의 핵심은 훈련 순간의 매끄러운 수행이 아니라 저장 강도의 증가이며, 당장의 인출 편익은 장기적 파지와 전이를 보장하지 않는다. 또한 스웰러(Sweller, 1988; 2010)의 관점에서 보면, 이 단계에서 남겨 두어야 할 것은 학습과 무관한 외재적 부하가 아니라 개념 형성과 스키마 조직화에 필요한 핵심 처리이다. 따라서 개념의 첫 내면화 단계에서 생성형 AI가 정답이나 내용의 요약 혹은 완결된 설명을 직접 제공하는 것은 외재적 부하를 줄이는 행위가 아니라, 학습자가 감당해야 할 수행을 본질적으로 대신하는 행위가 될 수 있다. 이 점은 바스타니 외(Bastani et al., 2025)의 실험이 가장 분명하게 보여준다. GPT 베이스 사용 집단은 연습 단계에서는 높은 수행 향상을 보였지만, AI를 제거한 독립 시험에서는 오히려 성적이 하락하였다. 반면 힌트와 질문 중심으로 설계된 GPT 튜터 집단은 적어도 이러한 학습 결손을 크게 줄였다. 여기서 핵심은 AI의 사용 여부가 아니라, AI가 정답을 대신하는가 아니면 사고를 유도하는가이다. OECD(2026)도 이를 과업 수행과 실제 학습 사이의 불일치라는 방식으로 요약한다. 그러므로 기초 지식 습득 단계에서 굳이 AI를

사용해야 한다면, 이는 원칙적으로 정답을 제공하는 AI가 아니라 힌트를 주거나 이해를 돕기 위해 질문을 하는 혹은 잘못된 개념을 바로 잡아주는 수준의 비계여야 한다. 다시 말해, 이 영역에서 AI는 학습자를 대신해 설명해주는 존재가 아니라, 학습자가 스스로 설명할 능력을 기르도록 밀어붙이는 제한적 도구여야 한다.

6.2 반복 과업: 조건부 활용

두 번째 영역은 반복 과업이다. 다만 여기서 중요한 것은 모든 반복이 동일한 교육적 지위를 갖는 것이 아니라는 점이다. 어떤 반복은 아직 기본기가 체화되지 않은 상태에서 수행되는 핵심 훈련이고, 어떤 반복은 이미 내면화된 기능을 자동화하거나 형식적으로 정리하는 절차에 가깝다. 스웰러(Sweller, 1988; 2010)의 관점에서 보면, 후자의 경우에는 생성형 AI가 일부 부담을 덜어 줄 수 있다. 반면 전자의 경우에는 반복 자체가 스키마를 형성하고 절차적 기억을 안정화하는 과정이므로, 이를 너무 이르게 외주화하면 기본기 형성이 무너질 수 있다.

이 점에서 반복 과업에 대한 생성형 AI 활용은 조건부 활용이 타당하다. 예컨대 이미 숙달된 학습자가 맞춤법 교정, 형식 정리, 자료 정리, 코드 오류 점검, 번역 초안 확인과 같은 기계적 작업에 AI를 활용하는 것은 효율성 향상에 도움이 될 수 있다. 그러나 아직 기본기가 형성되지 않은 학습자가 같은 방식으로 AI를 사용하면, 그것은 효율을 향상시키는 것이 아니라 학습을 대행하는 것이 된다. 콘트랙터와 레이에스(Contractor & Reyes, 2025)가 보여주듯 학생들은 이미 생성형 AI를 증강과 자동화 양쪽 모두에 사용하고 있으며, 압바스 외(Abbas et al., 2024)는 높은 학업 부담과 시간 압박이 ChatGPT 사용을 증가시키고, 그 결과 미루기 습관, 기억력 저하, 성취 하락과 연결될 수 있음을 보고한다. 따라서 반복 과업에서의 AI 허용은 “이미 할 수 있는 사람의 시간 절약”에 한정되어야 하며, “아직 미숙한 사람의 학습 대행”이 되어서는 안 된다.

6.3 아이디어 발산: 하이브리드식 활용

세 번째 영역인 아이디어 발산에서는 생성형 AI를 전면 금지할 필요도, 무제한 허용할 필요도 없다. 오히려 이 단계에서는 하이브리드식 활용이 적절하다. 생성형 AI는 대안 제시, 반례 탐색, 질문 확장, 비유 제안, 관점 전환에 강점을 가지므로, 학습자의 확산적 사고를 넓혀 주는 보조 도구가 될 수 있다. 그러나 이때도 중요한 것은 AI가 첫 발상을 대신해서는 안 된다는 점이다. 학습자가 아무런 자기 초안 없이 곧바로 AI에게 발상을 맡겨 버리면, 그는 탐색을 확장하는 것이 아니라 탐색의 출발점 자체를 외주화하게 된다.

여기서 바둑의 교훈은 다시 중요해진다. 최석웅 외(Choi et al., 2025a)에 따르면, AI 바둑 프로그램 공개 이후 기사들의 수질은 향상되었고 AI 추천수와의 정합성도 높아졌다. 이는 AI가 인간의 선택을 더 효율적이고 정교하게 만들 수 있음을 보여준다. 그러나 이 성과는 대국 중 AI를 사용한 결과가 아니라, 인간이 먼저 사고하고 대국장 밖에서 AI를 활용해 복기하고 훈련한 결과였다. 교육에서도 마찬가지이다. 아이디어 발산 단계에서 AI의 가치는 인간의 탐색을 대체하는 데 있지 않고, 이미 형성된 문제의식과 초안을 더 넓고 더 날카롭게 만드는 데 있다. 따라서 이 영역의 원칙은 분명하다. 먼저 학습자가 스스로 가설, 문제의식, 논지의 초안을 만든 뒤, 그 다음에만 AI를 사용해 대안, 반례, 보완 논점을 탐색하도록 해야

한다. 이때 리 외(Lee et al., 2025)가 보여주듯, 목표 설정, 프롬프트 구성, 응답 검토 및 통합은 여전히 인간의 과업이다. 생성형 AI는 발상을 넓혀 줄 수는 있어도, 어떤 발상을 자기 것으로 채택할 것인지는 학습자 자신이 결정해야 한다.

6.4 비판적 사고 훈련: 적대적 활용

네 번째 영역인 비판적 사고 훈련에서는 생성형 AI를 조력자보다 비판의 대상으로 활용하는 편이 더 교육적이다. 이 단계의 핵심은 정보를 받아들이는 데 있지 않고, 정보를 의심하고 비교하며, 논증의 타당성을 평가하고, 근거의 충분성과 출처의 신뢰성을 판별하는 데 있기 때문이다. 그런데 거를리히(Gerlich, 2025)는 AI 사용이 많을수록 인지적 외주화가 증가하고 비판적 사고가 약화되는 경향을 보여 주었고, 특히 소스 평가 능력이 크게 약화될 수 있음을 시사하였다. 리 외(Lee et al., 2025) 역시 AI에 대한 신뢰가 높아질수록 지식, 이해, 분석, 종합, 평가 전반에서 비판적 사고 노력이 감소한다고 보고한다. 특히 평가 차원에서 그 감소가 가장 크게 나타났다는 점은 중요하다. 생성형 AI를 쉽게 믿을수록, 인간은 가장 마지막에 남아야 할 판단 기능부터 먼저 내려놓기 시작한다는 것이다.

이미 살펴보았듯이 마게시 외(Magesh et al., 2025)는 RAG 기반 법률 AI조차 상당한 환각과 부정확성을 보일 수 있음을 보여준다. 이는 고위험 영역에서 생성형 AI의 출력이 정답이 아니라 “검증을 요구하는 주장”으로 취급되어야 함을 뜻한다. 따라서 비판적 사고 훈련에서 생성형 AI는 “답을 주는 도구”가 아니라 “비판할 텍스트”가 되어야 한다. 예컨대 정혜선 외(2025) 7장에서 소개한 학생에게 AI가 작성한 논문을 제시하고 그 안의 누락, 비약, 출처 문제, 개념 혼동, 허위 인용을 찾아내게 하는 과제는 매우 유효하다. 이 영역에서의 올바른 질문은 “AI가 뭐라고 답하나?”가 아니라 “AI의 답변은 무엇이 틀렸는가?”, “무엇이 검증되지 않았는가?”, “어떤 전제가 숨어 있는가?”여야 한다. 즉, 비판적 사고 훈련에서 생성형 AI는 사용 대상이기 이전에 반박 대상이어야 한다.

6.5 피드백 및 교정: 인간 책임형 하이브리드식 활용

다섯 번째 영역인 피드백 및 교정은 생성형 AI가 가장 교육적으로 유의미하게 기능할 수 있는 영역 가운데 하나이다. 그러나 이 역시 전제조건이 있다. AI가 피드백의 전 과정을 대신해서는 안 되며, 최종 판단과 책임은 인간에게 남아 있어야 한다는 점이다. 비고츠키(Vygotsky, 1978)의 근접발달영역과 우드 외(Wood et al., 1976)의 비계 설정 개념이 시사하듯, 좋은 도움은 학습자의 현재 수준에 조건부로 반응하고, 점차 제거되며, 최종적으로는 책임을 학습자에게 이전해야 한다. 이 기준에 비추어 보면, AI가 곧바로 완성된 수정문을 대신 산출하는 방식은 피드백이 아니라 교정의 외주화에 가깝다.

반면 바스타니 외(Bastani et al., 2025)의 GPT 튜터 사례는 다른 가능성을 보여 준다. 정답 대신 힌트와 질문을 제공하는 방식은 연습 단계의 수행을 돕는 동시에, 적어도 직접 답변형 모델이 낳는 학습 붕괴를 줄였다. 따라서 피드백 및 교정 영역에서는 생성형 AI를 인간 책임형 하이브리드식 활용으로 설계하는 것이 적절하다. 예를 들어 AI는 문장의 모호한 부분을 지적하고, 누락된 논점 후보를 제시하며, 구조적 대안을 추천하고, 질문을 통해 자기 점검을 유도할 수 있다. 그러나 어떤 수정을 채택할지, 어떤 논지를 유지하거나 폐기할지, 피드백을 어떻게 해석할지는 학생과 교사가 책임져야 한다. 이 점에서 OECD(2026)도 일반

목적 챗봇보다 교육적 목적과 인간 감독이 분명한 활용이 더 유망하다고 정리한다. 결국 핵심은 도구의 이름이 아니라, 누가 최종 판단을 하는가이다.

6.6 최종 평가: 목적별 이원화

마지막 영역은 최종 평가이다. 이 영역에서는 무엇보다 평가의 목적을 먼저 구분해야 한다. 모든 평가가 동일한 것을 측정하는 것은 아니기 때문이다. 어떤 평가는 학습자가 “AI 없이도 독립적으로 수행할 수 있는가?”를 검증하기 위해 존재하고, 어떤 평가는 현실과 유사한 환경에서 “도구를 포함한 문제 해결 능력”을 평가하기 위해 존재한다. 이 둘을 구별하지 않으면 채 동일한 정책을 적용하면, 평가는 타당성을 잃는다.

바둑계의 사례는 여기서 매우 직접적인 비유를 제공한다. 최석웅 외(Choi et al., 2025a)가 보여준 개선은 프로 대국 중 AI 사용이 허용된 결과가 아니라, 대국장 밖에서의 학습 효과였다. 다시 말해, 연습과 복기에는 AI를 사용하지만, 승부를 가르는 순간에는 인간의 독립 수행이 남아 있었다. 교육도 마찬가지이다. 기초 역량 검증 시험, 폐쇄형 서술형 시험, 독립적 개념 설명 평가와 같이 인간의 내면화된 지식과 판단을 측정하는 평가는 생성형 AI를 금지해야 한다. 바스타니 외(Bastani et al., 2025)의 연구는 이 구분이 왜 필요한지를 잘 보여 주며, OECD(2026)도 이를 수행 향상과 학습 향상의 불일치라는 방식으로 정리한다.

물론 이것이 평가에 있어 AI를 전면 금지해야 하는 것으로 읽혀서는 안 된다. 실제 직무 상황을 모사한 프로젝트형 평가, 보고서 작성, 설계 과제, 발표 준비, 데이터 분석처럼 현실의 문제 해결에서 AI 활용 능력 자체가 역량의 일부인 경우에는 생성형 AI를 제한적으로 허용할 수 있다. 다만 그 경우 평가 대상은 단순한 결과물이 아니라 “AI 활용 과정 전체”여야 한다. 즉, 어떤 지침(prompt)을 사용했는지, 어떤 제안을 수용 및 기각했는지, 무엇을 직접 검증했는지, 출처를 어떻게 처리했는지까지 함께 평가해야 한다. 또한 카르바할 외(Carvajal et al., 2024)와 콘트랙터와 레이에스(Contractor and Reyes, 2025)가 시사하듯, 정책은 학생 집단에 비대칭적 영향을 줄 수 있으므로, 무차별적 전면 금지보다는 “평가 목적에 따른 정밀한 구획”이 필요하다. 요컨대 최종 평가는 “역량 검증에는 금지하나 실제 협업 능력 평가에는 과정 공개를 전제로 제한적으로 허용한다.”는 이원적 원칙 위에서 설계하는 것이 바람직할 수 있다.

종합하여 생성형 AI는 학습 전 과정에서 동일한 방식으로 허용되거나 금지될 수 있는 도구가 아니다. 기초 개념을 내면화하고, 독립적으로 인출하고, 논증을 검증하고, 최종 판단을 내리는 순간에는 학습자의 인지적 자립성이 보호되어야 한다. 반대로 이미 형성된 기능을 정리하고, 대안을 탐색하고, 수정 방향을 점검하고, 실제적 협업 능력을 시연하는 순간에는 생성형 AI가 제한적 비계로 기능할 수 있다. 따라서 교육의 과제는 AI를 더 많이 쓰게 하는 데 있지 않고, 어떤 인지 처리는 끝까지 인간에게 남겨 둘 것인가를 정교하게 설계하는 데 있다.

| 교육 영역 | 권장 전략 | 핵심 원칙 |
|-----------|-----------|--|
| 기초 지식 습득 | 강력한 절제 | 정답 제공형 AI는 배제하고 힌트·질문형 비계만 제한 허용 |
| 반복 과업 | 조건부 활용 | 이미 숙달된 기계적 절차에 한해 AI 허용, 기본기 형성 단계는 제한 |
| 아이디어 발산 | 하이브리드식 활용 | 자기 초안을 먼저 만든 뒤 AI로 대안·반례·보완 논점 탐색 |
| 비판적 사고 훈련 | 적대적 활용 | AI 답변을 비판·검증의 대상으로 사용 |

| | | |
|----------|---------------------|-------------------------------------|
| 피드백 및 교정 | 인간 책임형 하이브리드식 활용 | AI는 질문·수정 제안 제공, 최종 판단과 책임은 인간이 담당 |
| 최종 평가 | 목적별 이원화 | 역량 검증은 금지, 직무형 과제는 과정 공개를 전제로 제한 허용 |

[표 5] 학습 영역별 생성형 AI 활용 및 절제 매트릭스

7. 이론의 실천: 한림대학교 AI 가이드라인 적용 사례

7절에서는 앞서 논의한 “AI 활용과 제한의 기준”을 2026년부터 한림대학교에서 시행하는 역량중심 Syllabus 시스템의 “학생 AI 사용 허용 기준”에 맞추어 살펴본다.

7.1 한림대학교 학생 AI 사용 허용 기준 체계

한림대학교의 학생 AI 사용 허용 기준은 크게 여섯 가지로 정리할 수 있다. 첫째, 미활용은 해당 교과목에서 AI를 전혀 다루지 않는 경우이다. 둘째, 기준 A(전면 허용)는 과제 및 수업 활동 전반에서 학생의 AI 사용을 허용하되, 출처 명시와 최종 결과물에 대한 책임을 학생에게 귀속시키는 방식이다. 셋째, 기준 B(조건부 허용)는 개념 이해, 구조 정리, 자료 정리 등의 용도로 AI 사용을 허용하되, 핵심 내용의 분석과 판단 대체는 허용하지 않는다. 넷째, 기준 C(도구적 사용 허용)는 맞춤법, 번역, 코딩 검토, 문법 교정, 데이터 정리 등 도구적 기능에 한하여 사용을 허용하되, 아이디어 생성이나 창의적 산출에는 사용을 금지한다. 다섯째, 기준 D(제한적 허용)는 교수자가 지정한 특정 과제나 수업 단계에서만 AI 사용을 허용하고 그 외에는 금지하는 방식이다. 여섯째, 기준 E(전면 금지)는 해당 교과목 또는 과제에서 생성형 AI 사용을 전면 금지하는 것이다.

이 기준 체계는 AI 활용의 스펙트럼을 체계적으로 구조화하였다는 점에서 의의가 있다. 그러나 하나의 교과목 내에서도 학습 단계와 활동 유형에 따라 AI의 적절한 활용 수준이 달라질 수 있다는 점에서, 본 연구의 6대 영역별 매트릭스는 이 기준의 구체적 적용 지침으로 기능할 수 있다.

7.2 6대 교육 영역과 AI 허용 기준의 대응

본 연구에서 제시한 여섯 가지 교육 영역별 권장 전략을 한림대학교의 AI 허용 기준 체계에 대응시키면 다음과 같다.

| 교육 영역 | 권장 전략 | 한림대 AI 허용 기준 | 적용 원칙 |
|-----------|-----------|--------------|---|
| 기초 지식 습득 | 강력한 절제 | 기준 D 또는 기준 E | 기초 개념 내면화 단계시 AI 사용 금지 / 필요한 경우에만 제한된 질문형 비계 허용 |
| 반복 과업 | 조건부 활용 | 기준 C | 이미 숙달된 기계적 절차에 한해 AI를 도구적으로 허용 |
| 아이디어 발산 | 하이브리드식 활용 | 기준 B 또는 기준 D | 자기 발상을 먼저 수행한 뒤 AI를 보조적 확장 도구로 활용 |
| 비판적 사고 훈련 | 적대적 활용 | 기준 D | AI를 비판과 검증의 대상으로 한정하여 |

| | | | |
|----------|---------------------|-----------------------------------|---|
| | | | 허용 |
| 피드백 및 교정 | 인간 책임형 하이브리드식 활용 | 기준 B | AI는 힌트·질문·수정 제안 제공, 최종 판단은 학생·교사가 담당 |
| 최종 평가 | 목적별 이원화 | 역량 검증: 기준 E / 실무 평가: 기준 A 또는 B | 독립 수행 검증은 금지, 실제 협업형 과제는 과정 공개를 전제로 허용 |

* 참고. A: 전면 허용, B: 조건부 허용, C: 도구적 사용 허용, D: 제한적 허용, E: 전면 금지

[표 6] 6대 교육 영역별 권장 전략과 한림대학교 AI 허용 기준 대응표

기준 E(전면 금지)는 학생들을 단순히 통제하기 위한 조치가 아니라, 3절과 4절에서 논의한 인지적 외주화와 학습 결손을 예방하고 인지적 기초 근력을 보호하기 위한 교육적 장치이다. 마찬가지로 기준 A(전면 허용)는 방임이 아니라, 이미 독립적 판단 능력을 갖춘 학습자가 실제 문제 해결 상황에서 AI를 포함한 도구를 책임 있게 활용하도록 허용하는 교육적 판단이다. 6절에서 강조했듯이, 고등교육의 목표는 단일하지 않다. 어떤 학습자에게는 기준 E가 인지적 자립성을 키우는 필수 장치가 되고, 다른 학습자에게는 기준 A가 실제 협업 역량을 기르는 적절한 조건이 된다. 결국 중요한 것은 허용의 폭 자체가 아니라, 어떤 인지 처리를 학생에게 남겨 두는가이다.

7.3 단일 기준의 한계와 복합 적용의 필요성

위 대응표에서 확인할 수 있듯이, 하나의 교과목 내에서도 학습 단계에 따라 적합한 AI 허용 기준은 달라진다. 따라서 한 과목 전체에 기준 A나 기준 E를 일괄 적용하는 방식은 실제 교육의 복합성을 충분히 반영하지 못한다. 보다 타당한 설계는 교과목 전체를 하나의 기준으로 규정하는 것이 아니라, 주차별·활동별·평가별로 서로 다른 기준을 조합하는 방식이다.

예를 들어 철학 교과목을 가정해 보자. 핵심 개념의 첫 학습, 원전 독해, 폐쇄형 개념 확인 퀴즈는 기준 E 또는 D가 적절하다. 반면 인용 형식 정리, 기초 번역 보조, 문법 교정은 기준 C로 둘 수 있다. 토론 질문 발산, 연구 주제 탐색, 논지의 구조 점검은 기준 B가 가능하다. 초안 수정과 피드백 단계에서는 AI가 제시한 조언을 그대로 반영하는 것이 아니라, 수용·기각 이유를 학생이 설명하도록 요구하는 방식의 기준 B가 바람직하다. 그러나 최종 구술시험, 독립 논증문 평가, 폐쇄형 서술시험과 같이 학생의 내면화된 사고를 확인해야 하는 순간에는 다시 기준 E로 돌아가야 한다.

이처럼 한 교과목 안에서 AI 허용 수준을 단계별로 달리하는 복합 적용만이 수평적 생산의 효율성과 수직적 성장의 요구를 동시에 보존할 수 있다. 한림대학교의 기준은 이러한 구획을 가능하게 하는 제도적 언어이며, 본 연구의 6대 영역 매트릭스는 그 언어를 실제 수업 설계로 번역하는 해석 틀이다. 따라서 앞으로의 Syllabus 설계는 "이 과목은 AI를 허용한다/금지한다"라는 이분법에 머물러서는 안 된다. 대신 어떤 과제에서는 왜 허용하고, 어떤 단계에서는 왜 금지하는가를 학생이 납득할 수 있도록 분명히 명시해야 한다. 그래야 학생들은 AI를 편의의 도구로만 사용하는 것이 아니라, 자신의 학습 단계와 목표에 맞추어 책임 있게 활용하게 된다.

8. AI시대 교육의 목적은 무엇이 되어야 하며 AI는 어떻게 활용되어야 하는가?

본고의 논의는 세 층위에서 전개되었다. 첫째, 2절에서는 바람직한 어려움, 인지 부하 이론, 비계 설정을 통해 교육이 본질적으로 일정한 인지적 마찰을 필요로 함을 밝혔다. 학습은 단순히 정보를 빨리 받아들이는 과정이 아니라, 스스로 인출하고, 오류를 경험하고, 다시 구성하는 과정 속에서 저장 강도와 판단력이 형성되는 과정이다. 둘째, 3절과 4절에서는 이러한 이론적 통찰이 생성형 AI 환경에서 어떤 방식으로 시험받고 있는지 살펴보았다. 생성형 AI는 연습 단계의 수행을 향상시킬 수 있지만, 정답 제공형 활용은 학습의 내면화를 약화시키고 능숙함의 환상을 낳을 수 있으며, 반복적 의존은 인지적 외주화, 비판적 사고의 약화, 수동적 의존, 감독 부담의 증가로 이어질 수 있었다. 셋째, 5절에서는 바둑계를 통해 AI가 지식의 재편, 전문성의 재구성, 교육 시장의 변화까지 동반한다는 사실을 확인하였다. 알파고 이후의 바둑은 AI가 단순한 계산 도구가 아니라 인간의 판단 기준 자체를 다시 쓰는 매개라는 점을 보여주었다. 그리고 6절과 7절에서는 이러한 분석을 바탕으로, 기초 지식 습득에서 최종 평가에 이르는 학습의 여섯 영역별로 AI 활용과 절제의 기준을 제시하고, 이를 한림대학교의 제도적 언어에 대응시켰다.

이상의 논의를 종합하면, 생성형 AI는 수평적 생산의 도구로서 정보의 탐색, 정리, 재조합, 초안 산출의 속도를 비약적으로 높여 준다. 그러나 교육이 보호해야 하는 것은 속도 자체가 아니라, 속도를 넘어서는 인간의 내적 성장이다. 비판적 사고, 개념의 내면화, 오류를 견디는 힘, 판단의 책임, 자기 수정의 능력은 대체로 학습의 부담 속에서 이해의 지연과 실패를 반복하며 형성된다. 따라서 생성형 AI가 교육에 유익한가 해로운가는 기술의 성능 일반이 아니라, 그 기술이 학습자에게서 무엇을 덜어 주고 무엇을 남겨 두는가에 의해 판정되어야 한다. 바둑계의 사례가 보여주듯, AI는 준비된 전문가에게는 새로운 가능성을 여는 스승이 되지만, 준비되지 않은 자에게는 사고의 책임을 넘겨 버리게 만드는 편리한 대행 장치가 될 수 있다.

물론 AI시대 교육의 목적을 수평적 생산 능력의 향상과 수직적 성장이라는 두 개념만으로 완전히 환원할 수는 없다. 본고가 이 구분을 도입한 이유는 생성형 AI가 교육에 제기하는 핵심 긴장을 선명하게 드러내기 위해서이지, 교육의 목적 전체를 두 축으로 폐쇄하기 위해서가 아니다. 교육은 오래전부터 백년지대계라 불려 왔듯이 단기적 산출만이 아니라 인간의 장기적 형성과 사회의 지속 가능성을 함께 겨냥해 왔다. 어떤 교육은 직업적 역량을, 어떤 교육은 시민적 판단을, 어떤 교육은 도덕적 성숙을, 또 어떤 교육은 자기 이해와 표현의 형식을 더 중시한다. 그러므로 AI시대의 교육 역시 단일한 성취 모델로 획일화될 수 없다. 중요한 것은 모든 교육을 같은 방식으로 자동화하는 일이 아니라, 서로 다른 교육 목적들 가운데 무엇이 AI와의 협업을 통해 증진될 수 있고 무엇이 끝내 인간 자신의 몫으로 남아 있어야 하는지를 분별하는 일이다.

특히 스마트 기기와 초단기 영상 매체의 확산은 즉시적 보상과 짧은 주기 순환에 익숙한 인지 환경을 강화하고 있다. 최근 연구들은 과도한 스마트폰 사용이 비판적 사고 수행과 관련된 일부 능력 저하와 연결될 수 있고, 숏폼(short-form) 영상 시청이 자기통제와 주의집중을 약화시킬 수 있음을 보여준다. (참고. Fabio & Suriano, 2023; Yan et al., 2024). 생성형 AI는 이러한 흐름을 여가 영역에 머무르게 하지 않고, 읽기, 쓰기, 요약, 판단, 설계와 같은 핵심 인지 영역으로 끌어들이고 있다. 그 결과 인간은 단순히 빠르게 반응하는 데에는

익숙해지지만, 오래 붙들고 따져 보고 스스로 결론을 형성하는 합리적 훈련에서는 점차 멀어질 위험이 있다. 3절과 4절에서 검토한 능숙함의 환상, 인지적 외주화, 미루기, 수동적 의존의 문제는 바로 이 위험이 교육 현장에서 이미 구체적 징후로 나타나고 있음을 보여 준다. 따라서 AI시대 대학 교육의 과제는 AI 활용을 무조건 장려하거나 일괄 금지하는 데 있지 않다. 오히려 대학은 어떤 단계에서 AI가 외재적 부하를 줄이는 유익한 비계가 될 수 있는지, 또 어떤 단계에서 AI가 본질적 인지 처리를 대행하여 학습을 훼손하는지를 더 정교하게 가려내야 한다. 기초 지식 습득, 독립 수행형 최종 평가, 비판적 사고 훈련처럼 인간의 내면화와 판단 책임이 핵심인 영역에서는 강한 절제와 제한이 필요하다. 반면 아이디어 발산, 피드백 및 교정, 일부 실제 협업형 과제에서는 하이브리드식 활용과 인간 책임형 하이브리드식 활용이 더 타당할 수 있다. 핵심은 AI를 허용하는가 금지하는가가 아니라, 최종적 이해, 검증, 선택, 책임의 자리를 인간에게 남겨 두어야 한다는 것이다.

칸트(Kant, 1904)의 말처럼, “인간은 교육을 통하지 않고는 인간이 될 수 없는 유일한 존재”이다. 이는 AI시대에 오히려 더 무겁게 다가온다. 인간이 인간으로 남기 위해서는 단순히 정보를 더 빨리 처리하는 능력만으로는 충분하지 않다. 스스로 생각하고, 근거를 따지고, 유혹적인 지름길을 거절하고, 자기 판단에 책임지는 훈련이 필요하다. 이런 의미에서 AI시대 대학 교육의 목표와 방안에 대한 연구는 단순한 기술 활용 지침의 문제가 아니라, 인간성을 보존하기 위한 제도적·철학적 과제라 할 수 있다. 앞으로 대학은 AI를 잘 쓰는 법만이 아니라, AI를 쓰더라도 무엇보다 스스로 해야 하는지를 가르쳐야 한다. 바로 그 지점에서 교육은 여전히, 그리고 앞으로도 인간성을 지키는 최후의 보루일 것이다.

부록 A. 학생 AI 사용 허용 기준 진단 도구

본 부록은 교수자가 수업 활동, 과제 수행, 평가 상황별로 적절한 AI 허용 기준을 판정할 수 있도록 설계한 운영 도구이다. 핵심 원칙은 단순하다. AI가 외재적 부하를 줄이는 유익한 비계로 기능하는가, 아니면 학습자가 수행해야 할 본질적 인지 처리를 대행하는가를 가려내는 것이다. 따라서 본 도구는 과목 전체가 아니라 활동 단위로 적용하는 것을 원칙으로 하며, 총점뿐 아니라 고위험 항목의 분포를 함께 보아야 한다.

사용 방법은 다음과 같다. 첫째, 수업 활동·과제·평가를 각각 하나의 판정 단위로 삼는다. 둘째, 아래 8개 항목에 대해 1점부터 5점까지 부여한다. 점수가 높을수록 AI 제한 필요성이 크다는 뜻이다. 셋째, 총점을 산출한 뒤 제2부 대응표와 빠른 판정 규칙을 함께 적용한다. 넷째, 최종적으로는 교과목 전체가 아니라 주차별·활동별·평가별로 세부 기준을 공지한다.

빠른 판정 규칙은 다음과 같다. 1번, 2번, 3번, 7번 항목 가운데 하나라도 5점이면 기준 A는 적용하지 않는 것이 원칙이다. 이 네 항목 가운데 둘 이상이 5점이면 기준 D 이상을 우선 검토한다. 2번, 3번, 7번이 모두 5점이면 기준 E를 우선 검토한다. 5번 항목이 5점이면 허용이 필요하다더라도 기준 A나 기준 B보다 기준 C 또는 기준 D가 안전하다.

A.1. 정량적 진단 체크리스트

아래 표의 각 항목에 1점(제한 불필요), 2점, 3점(중간), 4점, 5점(제한 필요) 중 하나를 부여한다. 표에는 1점, 3점, 5점의 기준만 제시하였으며, 2점과 4점은 인접 기준 사이의 중간

판단에 해당한다. 총점은 최소 8점, 최대 40점이다.

| 번호 | 평가 항목 | 판단 질문 | 1점 | 3점 | 5점 |
|----|----------------|-----------------------------|------------------------|-------------------------|------------------------------|
| 1 | 핵심 인지 처리 대체 위험 | AI가 핵심 사고를 대신할 수 있는가? | 형식 정리나 기계적 보조에 머문다. | 일부 핵심 처리까지 보조할 수 있다. | 핵심 분석·논증·문제 해결을 거의 대신할 수 있다. |
| 2 | 학습 단계 | 지금 활동이 최초 학습 또는 기초 형성 단계인가? | 이미 숙달한 기능을 효율화하는 단계이다. | 학습이 진행 중인 단계이다. | 처음 배우는 개념·기능을 내면화하는 단계이다. |
| 3 | 독립 수행 필요도 | 독립 수행 자체가 학습 목표인가? | 결과물 완성이 더 중요하다. | 과정과 결과가 모두 중요하다. | 독립 수행 자체가 핵심 학습 목표이다. |
| 4 | 대행 유혹 수준 | AI를 질문·힌트형 비계로 제한하기 쉬운가? | 질문·힌트형으로 통제하기 쉽다. | 통제 가능하지만 완성본 제공 위험도 있다. | 완성본 제공이 쉬워 과업 대행이 되기 쉽다. |
| 5 | 오류 검증 가능성 | 학습자가 AI 출력을 스스로 검증할 수 있는가? | 학습자가 쉽게 검증할 수 있다. | 일부만 검증할 수 있다. | 높은 전문성이 없으면 검증하기 어렵다. |
| 6 | 결과물의 고유성 요구 | 자기 논지와 자기 판단이 필수인가? | 표준화된 산출물로도 충분하다. | 개인적 판단이 일부 필요하다. | 고유한 사고와 판단의 증거가 필수이다. |
| 7 | 평가 부담과 공정성 | 이 활동의 평가 부담은 어느 정도인가? | 연습 또는 형성평가이다. | 수행평가 수준이다. | 총괄평가 또는 고부담 평가이다. |
| 8 | 인지 의존성 위험 | 반복 허용 시 사고 위임이 굳어질 위험이 큰가? | 도구적 보조에 머문다. | 활용 범위 관리가 필요하다. | 사고 위임이 습관화될 가능성이 높다. |

[부록 표 1] AI 허용 기준 정량적 진단 체크리스트

A.2. 합산 점수와 AI 허용 기준 대응표

8개 항목의 점수를 합산하면 최소 8점에서 최대 40점의 범위를 갖는다. 다음 표는 합산 점수 구간별 권장 AI 허용 기준, 기본 판단, 권장 운영 방식, 대표 활동을 제시한다. 교수자는 Syllabus 작성 시 교과목의 전반적 성격을 고려하여 대표 기준을 선택하되, “수업운영방법”란에 활동별 세부 기준을 명시함으로써 학생들에게 일관되고 투명한 AI 활용 가이드라인을 제공할 수 있다.

| 점수 구간 | 권장 기준 | 기본 판단 | 권장 운영 방식 | 대표 활동 |
|--------|---------------------|--------------------------------|--|---|
| 8~12점 | 기준 A (전면 허용) | AI 사용이 학습 목표를 거의 훼손하지 않는다. | 출처 표시와 결과 책임을 전제로 폭넓게 허용한다. | 이미 숙달된 기능의 정리, 실제 협업형 프로젝트에서의 도구 활용 |
| 13~18점 | 기준 B (조건부 허용) | AI를 쓰되 학생 스스로가 검증·재구성·선택해야 한다. | 하이브리드식 활용 또는 인간 책임형 하이브리드식 활용으로 운영한다. | 자기 초안 작성 후 아이디어 확장, 자기 글에 대한 AI 피드백 후 수용·기각 사유 제출 |
| 19~24점 | 기준 C (도구적 사용 허용) | 내용 생성보다 기계적 보조에 한정해야 한다. | 번역, 문법, 서식, 데이터 정리, 코드 점검 등 도구 기능만 허용한다. | 번역 보조, 문법 교정, 서식 정리, 오류 탐지 |

| | | | | |
|--------|------------------|--------------------------------------|-------------------------------|------------------------------------|
| 25~32점 | 기준 D (제한적 허용) | 핵심 학습을 보호하기 위해 교수자 지정 단계에서만 허용해야 한다. | 적대적 활용 또는 특정 단계 한정 허용으로 운영한다. | AI 답변의 오류 찾기, AI 초안 비판하기, 비교·검증 과제 |
| 33~40점 | 기준 E (전면 금지) | AI 사용이 학습 목표의 본질을 훼손한다. | 독립 수행을 요구하고 AI 사용을 금지한다. | 기초 개념 학습, 폐쇄형 시험, 독립 논증문, 구술 평가 |

[부록 표 2] 합산 점수별 AI 허용 기준 대응표

위 대응표를 6절의 여섯 영역과 연결하면 다음과 같다. 기준 A는 주로 숙달된 반복 과업이나 일부 실제 협업형 최종 평가에 대응한다. 기준 B는 아이디어 발산과 피드백 및 교정 영역에 가장 잘 맞는다. 기준 C는 반복 과업 가운데 도구적 보조가 가능한 활동에 대응한다. 기준 D는 비판적 사고 훈련이나 일부 기초 지식 단계처럼 허용 범위를 정밀하게 제한해야 하는 활동에 대응한다. 기준 E는 기초 지식 습득과 독립 수행형 최종 평가처럼 인간의 내면화와 판단 책임을 보호해야 하는 활동에 대응한다. 따라서 교수자는 과목 전체에 하나의 기준을 일괄 부여하기보다, 활동 단위 점검 결과에 따라 복합 기준을 설계하는 것이 바람직하다.

참고문헌

- Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21(1), 10. <https://doi.org/10.1186/s41239-024-00444-7>
- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775-779.
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakci, O., & 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. <https://doi.org/10.1073/pnas.2422633122>
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). Worth Publishers.
- Bjork, E. L., & Bjork, R. A. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, 9(4), 475-479.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). MIT Press.
- Carvajal, D., Franco, C., & Isaksson, S. (2024). Will artificial intelligence get in the way of achieving gender equality? (NHH Department of Economics Discussion Paper No. 3/2024).
- Choi, S., Kim, N., Kim, J., & Kim, H. (2025a). How does artificial intelligence improve human decision-making? Evidence from the AI-powered Go program. *Strategic Management Journal*. <https://doi.org/10.1002/smj.3694>
- Choi, S., Kim, N., Kim, J., & Kim, H. (2025b). The dual edges of AI: Advancing knowledge while reducing diversity. *PNAS Nexus*, 4(5), pgaf138. <https://doi.org/10.1093/pnasnexus/pgaf138>
- Contractor, Z., & Reyes, G. (2025). Generative AI in higher education: Evidence from an elite college (IZA Discussion Paper No. 18055). IZA Institute of Labor Economics.
- Fabio, R. A., & Suriano, R. (2023). The influence of smartphone use on tweens' capacity for complex critical thinking. *Children*, 10(4), 698. <https://doi.org/10.3390/children10040698>
- Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), 6. <https://doi.org/10.3390/soc15010006>
- Kant, I. (1904). On pedagogy. In E. F. Buchner (Ed. & Trans.), *The educational theory of Immanuel Kant* (p. 107). J. B. Lippincott Company.

- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. *Proceedings of CHI 2025*. <https://doi.org/10.1145/3706598.3713778>
- Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., & Ho, D. E. (2025). Hallucination-free? Assessing the reliability of leading AI legal research tools. *Journal of Empirical Legal Studies*, 22(2), 216-242. <https://doi.org/10.1111/jels.12413>
- OECD. (2026). *OECD digital education outlook 2026: Exploring effective uses of generative AI in education*. OECD Publishing. <https://doi.org/10.1787/062a7394-en>
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, 13(3), 273-304. https://doi.org/10.1207/s15327809jls1303_2
- Scale AI. (2025). Remote labor index leaderboard & analysis. Scale AI Research. <https://scale.com/research/rli>
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3(4), 207-217.
- Shukla, P., Bui, P. N. N., Levy, S., Kowalski, M., Baigelenov, A., & Parsons, P. (2025). De-skilling, cognitive offloading, and misplaced responsibilities: Potential ironies of AI-assisted design. *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3706599.3719931>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2), 123-138. <https://doi.org/10.1007/s10648-010-9128-5>
- Thiel, P., & Masters, B. (2014). *Zero to one: Notes on startups, or how to build the future*. Crown Business. (한국어판: 피터 틸, 블레이크 매스터스. (2014). 제로 투 원: 스탠퍼드 대학교 스타트업 명강의 (이지연 역). 한국경제신문.)
- Yan, T., Su, C., Xue, W., Hu, Y., & Zhou, H. (2024). Mobile phone short video use negatively impacts attention functions: An EEG study. *Frontiers in Human Neuroscience*, 18, 1383913. <https://doi.org/10.3389/fnhum.2024.1383913>
- van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational Psychology Review*, 22(3), 271-296. <https://doi.org/10.1007/s10648-010-9127-6>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100.

정혜선, 박섭형, 김여진, 한수미, 최승락. (2025). AI 시대, 대학교육을 리디자인하라. 학지사.

인터넷 자료

AI타임즈. 2025. "AI 사범 시대 -- 바둑계의 훈련법이 바뀌었다."
<https://www.aitimes.com/news/articleView.html?idxno=170578>

매일경제. 2019. "AI, 은퇴에 큰 영향...다음 생엔 바둑은 취미로."
<https://www.mk.co.kr/news/it/10968504>

서울신문. 2025. "김은지, 최정 제치고 두 달 만에 여자바둑 1위 탈환."
<https://www.seoul.co.kr/news/sport/2025/05/05/20250505500089>

연합뉴스. 2025. "최정, 김은지 제치고 한 달 만에 여자바둑 랭킹 1위 복귀."
<https://www.yna.co.kr/view/AKR20250205157500007>

이세돌. 2025. "알파고 이후 프로바둑 비약적 성장...활용능력이 실력 갈라." 서울포럼 2025.
서울경제. <https://www.sedaily.com/NewsView/2GSZ21W9R1>

한국기원. 2023. "특집/대세는 바둑 유튜버."
https://m.baduk.or.kr/news/B03_view.asp?news_no=85

YTN 사이언스. 2024. "'AI 시대' 바둑계 떠난 이세돌이 찾은 '신의 한 수.'"
<https://science.ytn.co.kr/program/view.php?mcd=0082&key=202408161209316574>