생성형 AI 시대의 구조적 사고와 인간-AI 연구 협력

Structural Thinking and Human-AI Research Collaboration in the Generative AI Era: A Framework Proposal, a Case Study, and Educational Implications

저자: 최종훈 (Jonghun Choi)
인하공업전문대학 2학년 재학 · 독립 연구자
연락처: cjh3cjh@naver.com
GitHub: github.com/inha20/8StructuralThinkingAndHumanAICollaboration

---

초록 / Abstract

생성형 AI는 정형화된 문제의 답을 빠르게 산출한다. 이 환경에서 인간 연구자와 학습자에게 남는 역할이 무엇인지, 그리고 AI와 장기 연구를 어떻게 수행할 것인지를 다룬다. 본 논문은 이론 제안 논문이며 다음을 제시한다.

1. **구조 인식 이론(SRT, Structure Recognition Theory).** 인간이 복잡한 표상에서 대칭과 불변량을 발견하여 연구 질문으로 발전시키는 과정에 대한 6단계 흐름 모델, 9단계 질문 계층, 10개 가설(H1~H10)과 일부 가설의 검증 가능한 예측.
2. **인간-AI 연구 연속성 이론(HARCT, Human-AI Research Continuity Theory).** AI 세션의 휘발성과 인간 장기 기억의 비대칭을 보완하는 분산 연구 기억, 핸드오버 프로토콜, 인지적 분업 모델.
3. **단일 사례 연구.** 한 명의 연구자가 AI 어시스턴트와 다중 세션으로 수행한 수학 연구 프로젝트(카르노맵의 대칭 분류)의 세 에피소드. 세 에피소드 모두에서 연구 문서의 오류가 확인되었고, 그중 둘(에피소드 2, 3)의 오류는 별도 프로그램으로의 재계산이나 처음부터 다시 도출하는 작업으로 발견되었다. 이는 HARCT의 이점과 위험을 함께 드러낸다.
4. **교육적 함의와 파일럿 설계.** 결과 중심 교육에서 과정 중심 교육으로의 전환 제안과 이를 검증하기 위한 연구 설계.

본 논문은 가설의 검증 결과가 아니라 가설과 검증 설계를 제시하며, 사례 연구는 연구자 자신의 회고에 기반한 n=1 관찰이다 (7장 한계 참조). 사례에 쓰이는 수학 용어는 3.1절과 4.7절에서 풀어 쓰며, 증명은 자매 논문에 있고 본 논문의 논증은 증명을 요구하지 않는다.

주제어: 구조적 사고, 생성형 AI, 인간-AI 협업, 구조 인식 이론(SRT), 인간-AI 연구 연속성 이론(HARCT), 메타인지, 분산 연구 기억, AI 교육, 카르노맵

---

1. 서론 (Introduction)

1.1 배경

2022년 이후 대규모 언어 모델(LLM)은 문장 생성, 코드 작성, 시험 문제 풀이 등에서 실용적 성능에 도달했다. 문헌 조사, 요약, 초안 작성, 정형화된 계산처럼 과거에 훈련이 필요했던 작업이 짧은 시간에 수행된다. 이는 교육과 연구에 다음 물음을 제기한다.

- 결과물 도출이 쉬워진 환경에서 학습자와 연구자가 거쳐야 할 사고 과정은 어디서 확보되는가?
- AI가 계산과 증명 초안을 대신할 때 인간에게 요구되는 역량은 무엇인가?
- 세션이 끊기는 AI와 몇 년에 걸친 연구를 어떻게 이어갈 수 있는가?

1.2 연구 질문

- **RQ1 (구조 인식):** 인간은 복잡한 표상에서 대칭과 불변량 같은 구조를 어떻게 포착하고 연구 질문으로 발전시키는가?
- **RQ2 (협업 방법론):** 세션 단위로 기억을 잃는 AI와 장기 연구를 수행하려면 어떤 기억 인프라와 분업이 필요한가?
- **RQ3 (교육):** 위 두 물음의 답은 AI 시대의 교육에 무엇을 시사하는가?

1.3 기여와 범위

본 논문은 SRT(3장), HARCT(4장)와 두 이론을 결합한 연구 순환(5장), 교육적 제안(6장)을 제시한다. 제목의 "구조적 사고"는 이 논문에서 변환(예: 변수 이름 바꾸기, 격자 이동)을 가해도 변하지 않는 것을 알아보고 그것을 연구 질문으로 만드는 능력을 뜻하며, 정의 3.1의 구조 인식과 이어진다. 4.7절의 사례 연구는 제안한 틀의 적용 가능성을 보이는 탐색적 관찰이며 가설의 통계적 검증이 아니다.

1.4 AI 사용 공개

본 논문이 다루는 연구 프로젝트는 여러 AI 시스템(저자의 기록에 따르면 ChatGPT, Claude, ANTIGRAVITY. ANTIGRAVITY는 저장소 단위의 작업에 쓰는 AI 코딩 도구이다)과 다중 세션으로 수행되었다. 각 시스템이 맡은 역할은 4.7절에 서술한다. 이와 별도로, 본 원고의 작성과 정리(구조 재편, 오류 정정, 참고문헌 점검)에는 Claude(Anthropic)의 새 세션들이 사용되었다. 이 세션들은 프로젝트 당시의 세션들과 별개이며 그 대화 이력을 갖고 있지 않다. 이론의 제안, 사례의 선택과 해석, 결론의 책임은 저자에게 있다.

참고문헌에 대한 주. 이 논문의 초기 원고의 참고문헌 일부는 AI가 생성한 목록이었고 본문에서 인용되지 않았다. 현재 원고에서는 본문에서 실제로 인용하는 문헌만 남기고, 각 문헌의 존재와 서지 정보를 확인했다.

1.5 관련 연구와 본 논문의 위치

SRT와 HARCT의 구성 요소는 기존 연구와 겹친다. 이 절은 그 관계를 밝히고, 본 논문이 새로움을 주장하는 범위를 한정한다.

- **구조 대 표면 특징.** 숙련자는 문제를 표면 특징이 아니라 깊은 구조(물리 원리)로 분류하고, 초심자는 표면 특징으로 분류한다 (Chi, Feltovich & Glaser, 1981). 관계 구조를 대응시키는 능력을 유추의 핵심으로 보는 이론도 있다 (Gentner, 1983). SRT가 말하는 "구조 인식"은 이 전통과 같은 방향의 개념이며, 본 논문은 이를 부울 함수 같은 형식 대상의 연구 과정에 적용해 6단계 흐름과 가설로 구체화한다.
- **표현의 효과.** 같은 정보라도 도식 표현이 탐색과 지각적 추론을 쉽게 한다 (Larkin & Simon, 1987). 카르노맵 사례(3.1절)는 이 논의의 한 사례로 읽을 수 있다. H2는 이 일반 주장을 진리표와 카르노맵의 비교로 좁힌 가설이다.
- **분산된 인지.** 인지가 개인의 머리 밖의 도구와 환경으로 확장된다는 관점이 있다 (Hutchins, 1995; Clark & Chalmers, 1998). HARCT의 "분산 연구 기억"은 이 관점을 인간-AI 연구의 인계 문서와 저장소에 적용한 것이다.
- **AI의 한계.** 언어 모델은 근거 없는 내용을 생성하며 (Ji et al., 2023), 학술 인용에서는 ChatGPT가 만든 인용의 상당 부분이 실재하지 않았다 (Walters & Wilder, 2023: GPT-3.5 55%, GPT-4 18%). 긴 입력에서는 중간에 놓인 정보의 활용이 떨어진다 (Liu et al., 2024). 컨텍스트 한계를 계층적 메모리로 보완하려는 시스템 연구도 있다 (Packer et al., 2023). HARCT는 같은 문제를 시스템이 아니라 연구자가 쓰는 문서와 절차의 수준에서 다룬다.
- **AI의 자기평가와 동조.** 모델이 외부 피드백 없이 자신의 추론을 교정하는 능력은 제한적이라는 보고가 있고 (Huang et al., 2024), AI 어시스턴트가 사용자의 견해에 맞춰 답하는 경향(sycophancy)도 보고되었다

(Sharma et al., 2023). 사람 쪽에서는 자동화의 출력에 대한 안일함(complacency)과 자동화 편향(automation bias)이 알려져 있다 (Parasuraman & Manzey, 2010). 4.7절에서 그럴듯한 오류가 검토를 통과한 현상은 이 문헌들과 같은 방향의 문제이며, 5.1절의 독립 검증 단계가 필요한 이유이다.

- **인간-AI 상호작용과 정보 관리.** 인간-AI 상호작용의 설계 지침이 제안되어 있다 (Amershi et al., 2019). 개인이 정보를 수집·조직·재발견하는 활동은 개인 정보 관리(PIM) 연구에서 다루어 왔다 (Jones, 2007). HARCT의 인계 문서와 저장소는 이 활동을 연구 프로젝트와 AI 세션 사이의 인계로 옮겨 놓은 것으로 볼 수 있다.
- **AI 리터러시.** AI와 상호작용하고 이를 비판적으로 평가하는 데 필요한 역량의 틀이 제안되어 있다 (Long & Magerko, 2020). 6.3절의 세 단계는 이 틀과의 대응을 확인하지 않았다.
- **교육.** 안내가 적은 문제 중심 수업은 초심자에게 비효율적일 수 있다는 비판이 있다 (Kirschner, Sweller & Clark, 2006; 인지 부하 이론은 Sweller, 1988). 6장의 과정 중심 교육 제안은 이 비판을 받아들여, 초심자에게는 명시적 안내와 단계적 제시를 함께 두어야 한다는 조건을 요구한다.
- **저자성과 공개.** 출판 윤리 지침은 AI를 저자로 올릴 수 없고 AI 사용을 공개하도록 요구한다 (COPE; ICMJE). 1.4절의 공개는 이 요구를 따른다.

본 논문의 한계. 위 조사는 웹 검색에 기반한 제한적 조사이며, 인간-AI 협업, 연구 방법론, 지식 인계의 문헌 전반을 체계적으로 검토한 것이 아니다. 특히 사람 사이의 업무 인계(교대 근무 인계 등)와 컴퓨터 지원 협동 작업(CSCW)의 문헌은 다루지 못했다. 따라서 SRT와 HARCT가 선행 연구에 없다는 주장은 하지 않는다. 본 논문이 제안하는 것은 이 개념들을 하나의 연구 흐름(6단계 모델과 인계 절차)으로 묶고 검증 가능한 가설로 제시하는 일이며, 그 유용성은 7.2절의 검증 설계로 확인해야 한다.

---

2. 생성형 AI와 인간 사고의 구조적 차이

본 장은 이후 이론의 배경으로서 비교 관점을 정리한다. 모델의 내부 상태에 대한 일부 주장은 현재 논쟁 중이며, 아래 서술은 널리 알려진 아키텍처와 관측에 근거한 일반화이다.

2.1 생성형 AI의 동작 원리

현재의 생성형 AI는 대부분 트랜스포머(Vaswani et al., 2017) 기반의 자기회귀 언어 모델이다. 이전 토큰열이 주어졌을 때 다음 토큰의 조건부 확률 분포를 계산한다.

$$P(w_t \mid w_1, \ldots, w_{t-1}; \theta)$$

이 구조에서 다음 세 가지 특성이 논의된다.

1. **통계적 패턴 학습.** 대규모 코퍼스의 동시 출현 구조를 학습한다. 이것이 추론과 구분되는 "단순 모방"인지는 논쟁 대상이다. 본 논문은 모델의 능력이 무엇이든 인간과 작동 방식이 다르다는 점만 전제한다.
2. **외부 세계와의 직접적 접점의 부재.** 텍스트 기반 모델은 신체적 감각이나 공간적 경험과 직접 연결되지 않는다 (Bender & Koller, 2020). 멀티모달 모델은 이 격차를 부분적으로 줄이고 있다.
3. **환각(hallucination).** 유창하지만 사실과 다른 출력은 현재 모델에서 흔하게 관찰되며 연구가 활발하다 (Ji et al., 2023). 완전한 제거가 불가능한지는 열린 문제이다.

2.2 인간 사고의 세 가지 특성

1. **체화된 이해.** 인간은 카르노맵의 체커보드 패턴을 비트열이 아니라 체스판과 같은 시각적 형태(게슈탈트)로 지각한다. 표상에 따라 문제 해결의 난이도가 달라진다는 것은 인지과학에서 오래 논의되어 왔다 (Larkin & Simon, 1987).

2. **메타인지.** 자신의 인지 상태를 감시하고 평가하는 능력이다 (Flavell, 1979). LLM이 자신의 답의 정확성을 어느 정도 추정할 수 있다는 보고도 있다 (Kadavath et al., 2022). 반면 외부 피드백 없이 스스로 추론을 교정하는 능력은 제한적이라는 보고도 있다 (Huang et al., 2024). 본 논문의 주장은 "AI에게 메타인지가 없다"가 아니라, 독립된 검증 없이는 AI의 자기평가를 신뢰하기 어렵다는 것이다. 4.7절의 사례가 이 점을 예시한다.
3. **목적 지향적 질문 생성.** 인간은 "왜 이것이 중요한가"라는 유의성을 부여하며 연구 질문을 정한다.

2.3 비교 요약

비교 영역	인간 사고	생성형 AI
기본 메커니즘	의미, 체화된 직관, 인과 추론	다음 토큰 확률 예측으로 학습된 모델
참조 기반	신체 경험, 시각적 공간 직관	텍스트(및 멀티모달) 데이터의 통계적 구조
자기 오류 감시	회의와 확신도 평가 (불완전)	부분적 자기평가 가능, 독립 검증이 필요
창의성의 동기	목적 의식, 문제 의식	학습된 분포 위의 생성
오류의 양상	인지 부하, 확증 편향	환각, 그럴듯한 오류

2.4 사례: 그럴싸하지만 검증되지 않은 출처

환각의 한 형태는 형식은 정확하지만 실재하지 않거나 관련 없는 출처를 만드는 것이다. 본 연구 프로젝트에서 두 가지 경험이 있었다.

- **존재하지 않는 논문 목록.** 프로젝트 초기에 연구자가 특정 주제의 논문 목록을 요청했을 때, 저자명, 발행 연도, 학술지명까지 갖춘 목록이 제시되었으나 그중 일부는 실제로 검색되지 않았다 (저자의 기록이며 해당 대화는 공개되지 않는다).
- **그럴싸한 참고문헌 목록.** 본 논문의 초기 원고의 참고문헌 일부는 저자가 AI에게 그럴듯한 목록을 만들어 달라고 요청해 얻은 것이었다. 그 항목들은 본문의 주장과 연결되지 않은 채 나열되어 있었다. 현재 원고에서는 본문이 실제로 인용하는 문헌만 남기고, 각 문헌의 존재와 서지 정보를 검색으로 하나씩 확인했다.

이 경험은 두 가지를 시사한다. 첫째, 출처의 형식적 자연스러움은 그 출처의 실재성이나 본문과의 관련성을 보증하지 않는다. 둘째, 출처의 검증은 AI의 답이 아니라 독립된 경로(원문, 학술 데이터베이스)로 해야 한다. 이는 5.1절의 연구 순환에 추가한 독립 검증 단계와 같은 원리이다.

---

3. 구조 인식 이론 (SRT)

3.1 사례 배경: 카르노맵

이후의 예시는 4변수 카르노맵을 쓴다. 카르노맵은 4개의 부울 변수 A, B, C, D의 16가지 입력을 4×4 격자에 Gray code 순서$(00, 01, 11, 10)$로 배열한 도표이며, 인접한 칸은 입력이 한 비트만 다르고 격자의 양끝도 인접하다. 변수를 행과 열에 배치하는 방법은 $4! = 24$가지이고, 교재의 AB/CD 배치는 그중 하나이다. 이후 사례에서 다음 세 가지 수학적 사실이 쓰인다.

1. 격자의 회전, 반사, 순환 이동을 허용하면 24가지 배치는 서로 옮겨지는 3개의 동치류로 나뉜다. 격자의 회전·반사 8개만 허용하면 12개로 나뉜다.
2. 변수의 이름을 바꿔도 같은 함수가 되는 대칭 함수(예: 4변수 XOR, 다수결)의 그림은 어떤 배치에서도 같다.
3. 4변수 부울 함수 65,536개를 변수 순열 아래의 동치류로 세면 3,984개이다.

용어. 부울 함수는 4개의 입력(각각 0 또는 1)에 0 또는 1을 대응시키는 규칙이며, 16가지 입력마다 출력을 적은 표가 진리표이다. 카르노맵의 각 칸에는 그 입력에서의 출력(0 또는 1)을 적고, AB/CD는 행에 A, B를, 열에 C, D를 놓는 배치를 뜻한다. 4변수 XOR(1의 개수가 홀수이면 1)은 이 표에서 0과 1이 번갈아 나타나 체스판 모양, 곧 체커보드가 된다. 동치류는 서로 옮겨질 수 있는 것을 한 묶음으로 센 것이며, 위 1은 서로 옮겨지는 배치를 같은 것으로 보면 3가지만 남는다는 뜻이다. 대칭 함수는 변수의 이름을 어떻게 바꿔도 출력이 변하지 않는 함수이다. 위 3의 "변수 순열 아래의 동치류"는 변수의 이름만 바꾸어 같아지는 함수를 같은 것으로 묶은 개수이다. 입력에서 1의 개수가 같은 입력끼리 묶은 것을 해밍 레이어라 하며, 4변수에서는 1의 개수가 0~4인 다섯 층으로 나뉘고 층의 크기는 1, 4, 6, 4, 1이다.

	00	01	11	10
00	0	1	0	1
01	1	0	1	0
11	0	1	0	1
10	1	0	1	0

	00	01	10	11
00	0	1	1	0
01	1	0	0	1
10	1	0	0	1
11	0	1	1	0

그림 1. 4변수 XOR의 카르노맵(배치 AB/CD). 왼쪽은 Gray 축 순서$(00, 01, 11, 10)$로 그린 체커보드, 오른쪽은 이진 축 순서$(00, 01, 10, 11)$로 그린 이중대각이다. XOR는 대칭 함수이므로 두 그림 모두 24가지 배치에서 같고, 모양의 차이는 변수 배치가 아니라 축 라벨 순서에서 온다.

	00	01	11	10
00	L1	L2	L3	L2
01	L2	L3	L4	L3
11	L3	L4	L5	L4
10	L2	L3	L4	L3

그림 2. 표준 배치 AB/CD에서 각 칸의 해밍 레이어. 그림의 라벨 L1~L5는 입력에서 1의 개수 0~4에 해당하며, 층의 크기는 1, 4, 6, 4, 1이다.

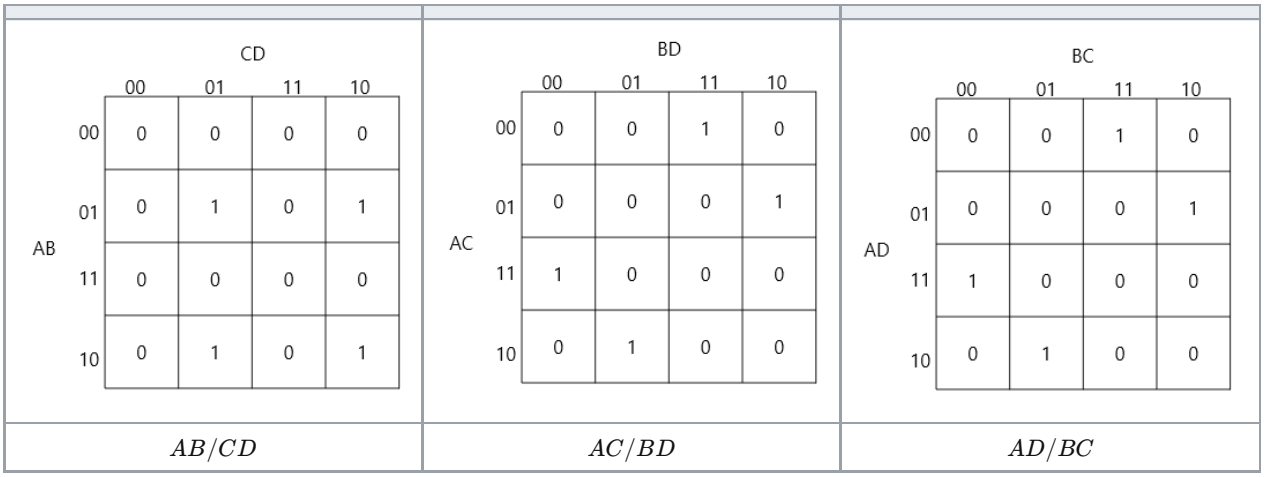


그림 3. 함수 $(A \oplus B)(C \oplus D)$(곧 $m(5,6,9,10)$이며, $A \oplus B$는 A와 B가 다를 때 1이다)를 세 대표 배치로 그린 그림. AC/BD와 AD/BC의 그림은 같고 AB/CD의 그림만 다르다. 이 함수는 대칭 함수가 아니므로 배치에 따라 그림이 달라진다.

(위 사실의 증명은 자매 논문 「카르노맵 위의 4변수 부울 함수 공간의 군론적 분류」(Choi, 2026)에 있으나, 본 논문의 논증은 사실의 내용에만 의존하며 증명을 요구하지 않는다.)

3.2 6단계 인지 흐름 모델

SRT는 연구가 생성되는 과정을 다음 6단계의 흐름으로 모델링한다.

$$\text{Pattern} \to \text{Attention} \to \text{Structure} \to \text{Meaning} \to \text{Question} \to \text{Research}$$

1. **Pattern(패턴 지각):** 표상에서 시각적 규칙성이나 이례를 감지한다. 예: 카르노맵에서 체커보드 형태를 본다.
2. **Attention(주의):** 배경에서 특정 규칙성을 분리한다.
3. **Structure(구조):** 변환 아래에서 보존되는 불변량을 포착한다. 예: 변수 이름을 바꿔도 그림이 같은 함수들.
4. **Meaning(의미):** 발견한 구조가 기존 지식에서 갖는 의의를 평가한다.
5. **Question(질문):** 새로운 연구 질문을 만든다. 예: "배치 24가지는 본질적으로 몇 가지인가?"
6. **Research(연구):** 증명, 구현, 문서화로 질문을 해결한다.

3.4절의 정의는 이 흐름의 단계를 풀어 쓴 것이다. 정의 3.2(주의)는 Attention 단계를, 정의 3.1(구조)은 Structure 단계를, 정의 3.3(설명적 유의성)은 특히 조건 (4)(연구 문턱 통과)가 Meaning에서 Question으로 넘어가는 문턱을 작동적으로 나타낸다.

3.3 질문 계층 (Level 0~8)

연구의 진전을 지식의 양이 아니라 질문의 수준 변화로 기술한다. 3.2절의 6단계가 한 번의 발견이 일어나는 인지 과정이라면, 아래 9수준은 연구가 진행되면서 질문이 바뀌는 단계이다. 두 체계의 일대일 대응은 가정하지 않으며, 예시 질문은 4.7절의 사례에서 가져왔고 수준의 순서가 일반적인지는 검증되지 않았다.

수준	질문의 예
0 Observation	여기에 무엇이 보이는가?
1 Structure	이 패턴을 지배하는 규칙은 무엇인가?
2 Discovery	변수를 재배열해도 패턴이 보존되는가?
3 Significance	이 불변성이 왜 중요한가?
4 Question Generation	이를 설명하려면 어떤 수학적 군이 필요한가?

수준	질문의 예
5 Explanation	궤도의 개수를 계산으로 유도할 수 있는가?
6 Concept	이 발견을 독립된 개념으로 정립할 수 있는가?
7 Expanded Perception	개념을 더 큰 대상(65,536개 함수 전체)으로 확장하면 무엇이 보이는가?
8 New Discovery	새로운 연구 영역이 열리는가?

3.4 10개 가설과 검증 가능한 예측

번호	가설	명제
H1	Structure Discoverer	인간은 구조 발견(구조를 찾고, 질문을 만들고, 의미를 부여하고, 연구 방향을 고르는 일)에서 비교우위를 가질 수 있다. 이는 전체 능력의 차이가 아니라 **역할의 차이**를 뜻한다.
H2	Representation Transformation	어떤 어려운 문제는 계산량을 줄여서가 아니라 표상을 바꿈으로써(예: 논리식에서 카르노맵으로) 다루기 쉬워진다.
H3	Attention Filter	인간은 모든 패턴을 조사하지 않으며, 조사 이전에 선택 과정이 작동한다. 어떤 패턴이 이 필터를 통과하는지는 열린 문제이다(3.5절의 "주의 선택").
H4	Significance Filter	주의 이후에, 구조가 조사할 가치가 있는지를 평가하는 암묵적 필터가 작동한다. 통과 조건의 후보는 압축성, 예측력, 일반화력, 설명 견인력이다.
H5	Explanation Anticipation	인간은 설명을 찾기 전에 설명이 존재하리라고 예상한다. 패턴이 "무작위로 보이지 않는다"는 감각이 탐색에 선행한다.
H6	Explanatory Significance	연구는 구조 발견 자체가 아니라 구조가 지닌 설명적 잠재력의 감지에서 시작된다.
H7	Mature Structure Discoverer (H1의 성숙판)	인간의 핵심 우위는 패턴 인식 자체가 아니라, 어떤 구조가 설명적 유의성을 지니는지 평가하는 능력일 수 있다.
H8	Concept-as-Lens	개념은 관찰의 렌즈로 작동한다. 개념을 얻으면 구조로 인식할 수 있는 대상이 달라질 수 있다.
H9	Concept Evolution	연구는 개념을 낳고, 개념은 확장된 지각을 통해 새로운 연구를 낳는다.
H10	Generative Concept	개념마다 생성력이 다르다. 어떤 개념은 지각을 더 넓고 오래 확장한다.

SRT의 가설은 모두 탐색적(exploratory)이며 통제된 증거가 없다. H2만은 저자의 사례 관찰 몇 건과 부합하지만, 그 관찰은 모두 카르노맵이라는 한 영역에서 나왔고 일화적이므로 체계적 검증이 필요하다. H8~H10은 개념의 역할을 다루는 확장 가설이다. 이 가설들은 현재 실증되지 않았다. H1과 H7, H5와 H6은 서로 가까우며, H7은 H1을 정교화한 것이다. 다음은 검증 가능한 예측의 예이다.

가설	예측	측정
H2	같은 부울 함수를 진리표로 제시한 집단보다 카르노맵으로 제시한 집단이 대칭 함수를 더 빨리, 더 정확히 식별한다.	식별 시간과 정답률 (참가자 간 설계)
H5	대칭 패턴을 본 직후 참가자는 설명 가능한 규칙의 존재를 증명 이전에 높은 확신으로 예측한다.	확신도 평정과 이후 증명 결과의 대응
H8	"해밍 레이어"(3.1절) 개념을 학습한 집단은 다른 논리 문제에서 구조를 더 빨리 찾는다.	학습 전이 과제의 점수 (통제군 비교)

위 예측은 다음과 같은 결과가 나오면 약화된다. H2의 예측은 카르노맵 학습 경험이 없는 집단이 두 표상을 같은 시간 안에 익힌 뒤에도 식별 시간과 정답률에 차이가 없을 때, H5의 예측은 확신도와 이후 증명 결과가 대응하지 않을 때, H8의 예측은 개념을 학습한 집단과 통제군의 전이 과제 점수가 같을 때이다.

형식 정의와 조작화. SRT는 세 개념에 작업 정의를 둔다 (관찰자 O, 관찰 대상인 무늬 C. 3.1절의 변수 배치와 다른 말이다).

- **정의 3.1 (구조).** C가 O에게 구조인 것은 다음 네 조건이 모두 성립할 때이다. (1) **전체적 인식**: O가 C를 셀 단위의 합이 아니라 하나의 통일된 전체로 인식한다. (2) **비자명성**: C가 무작위 분포로 곧바로 설명되지 않는다. (3) **변환 하의 안정성**: C가 적어도 하나의 표상 변환을 거쳐서도 유지된다. (4) **생성적 함의**: C가 이를 만들어 내는 기제의 존재를 함축한다.
- **정의 3.2 (주의).** O가 C에 주의를 기울이는 것은 다음 네 조건이 모두 성립할 때이다. (1) **선택성**: 다른 대안보다 C에 조사 자원을 선택적으로 배분한다. (2) **선행성**: C에 대한 주의가 완전한 구조 분석에 앞서며 그 결과가 아니다. (3) **지속성**: 여러 조사 에피소드에 걸쳐 주의가 유지된다. (4) **기준 불투명성**: 주의를 이끄는 기준이 O 자신에게도 완전히 접근 가능하지 않다.
- **정의 3.3 (설명적 유의성).** 구조 C가 O에게 설명적 유의성을 갖는 것은 다음 네 조건이 모두 성립할 때이다. (1) **비무작위성 지각**: O가 C를 무작위로 생성되지 않았다고 판단한다. (2) **설명 함축**: O가 C가 설명 기제의 존재를 함축한다고 판단한다. (3) **생산적 연구 기대**: O가 C의 조사가 자명하지 않은 발견을 낳을 것이라고 기대한다. (4) **연구 문턱 통과**: (1)~(3)의 판단이 지속적 조사를 시작하기에 충분하다.

위 정의는 네 조건이 모두 성립할 때로 두었으나, 각 조건이 필요한지, 가중합 같은 다른 결합이 더 나은지는 열린 문제이다(3.5절의 "유의성의 경계"와 관련된다). 정의 3.1~3.3은 4.7절의 사례에 적용하지 않았다.

가설과 정의의 연결. 가설 H3~H9는 위 정의의 조건으로 풀어 쓸 수 있다. 이 정의들이 가설과 이어지지 않으면 용어집에 그치기 때문이다.

가설	관련 정의	연결
H3	정의 3.2, 조건 (4)	주의 필터는 선택을 지배하는 기제이며, 그 기준은 연구자 자신에게도 명시되지 않은 채 암묵적으로 작동한다(기준 불투명성).
H4	정의 3.3	유의성 필터는 구조 C가 정의 3.3을 만족할 때에 한해 통과시킨다. 통과 조건의 후보(압축성 등)는 아래의 측정 표가 조작화한다.
H5	정의 3.3, 조건 (1)(2)	설명 예상은 조건 (4)(연구 문턱 통과)에 이르기 전에 조건 (1)(비무작위성 지각)과 (2)(설명 함축)가 먼저 만족된 상태이다.
H6	정의 3.3	연구가 설명적 잠재력의 감지에서 시작된다는 주장이 정의 3.3을 둔 주된 동기이다. 연구로의 전환은 조건 (1)~(3)을 만족하는 구조가 이끌고, 조건 (4)는 문턱을 넘는 일 자체이다.
H7	정의 3.3	인간의 우위는 구조 C가 정의 3.3을 만족하는지를, 네 조건을 일일이 열거하지 않고 여러 영역에서 안정적으로 추정하는 능력이다.
H8, H9	정의 3.1, 조건 (1)	전체적 인식은 관찰자가 가진 개념에 달려 있다. H8은 개념을 얻으면 조건 (1)을 만족하는 구성의 집합이 체계적으로 넓어진다고 보고, H9는 이 집합이 연구 순환을 거치며 시간에 따라 커지는 변수라고 본다.

이 연결은 SRT의 이론적 서술이며, 4.7절의 사례에 적용하여 검증하지 않았다.

H4의 후보 조건에는 다음의 측정이 제안되어 있다 (4.7절의 사례에서는 적용하지 않았다).

기준	조작적 정의
압축성	$\mathrm{comp}(C) = 1 - \lvert\mathrm{description}(G)\rvert / \lvert\mathrm{enumeration}(C)\rvert$ (G는 무늬 C를 만들어 내는 규칙, $\lvert\cdot\rvert$는 기술에 필요한 기호 수)
예측력	보류된 검증 집합에서의 정확도가 영역 기준선을 넘는지
일반화력	$\mathrm{gen}(C) = \lvert\{d : C\text{가 영역 } d\text{에서 구현됨}\}\rvert$
설명 견인력	조사 이전의 확신도 평정 (1~10)

H10의 후보 조작화는 지각 확장(개념 습득 전후로 인식되는 구조의 수), 연구 산출(개념으로 추적되는 질문의 수), 전이성(적용되는 독립 영역의 수), 지속성(생산적 사용 연수), 깊이(전문가 판단)이다. 이 기준들은 아직 검증되지 않았다. SRT는 세 가지 실험 패러다임도 제안한다. H1과 H7을 위해 사람과 AI가 카르노맵에서 구조를 발견하게 한

뒤 유의성을 평정하는 경쟁적 발견 과제, H5를 위해 구조의 설명을 접하기 전과 후에 확신도를 평정하는 과제, H6을 위해 여러 구조에 대해 "조사할 가치가 있는가"를 이분 판단하는 과제이다.

3.5 미해결 과제

SRT는 닫힌 이론이 아니라 열린 연구 프로그램이다. 대표적인 미해결 과제는 다음과 같다.

- **주의 선택:** 인간 인지가 체커보드나 대칭선 같은 특정 구조에 선별적으로 주의를 할당하는 이유는 무엇인가?
- **유의성의 경계:** 자명한 불변량과 새로운 이론으로 이어지는 불변량의 경계는 어디인가?
- **설명 예측:** 증명 없이 "배후에 대칭군이 있을 것"이라는 확신이 생기는 인지적 기제는 무엇인가?
- **표상 설계:** 특정 구조를 가장 투명하게 드러내는 표상 형식을 인지 부하 관점에서 최적화할 수 있는가?
- **AI의 구조 발견:** AI가 인간의 지시 없이 표상만 보고 구조를 발견하여 가설을 낼 수 있는가?
- **질문의 기원:** 수십 년간 교과서에 실려 온 도구에서, 저자가 검색한 범위에서는 선행 논의를 찾지 못한 질문(변수 배치의 동치류)을 묻게 만드는 요인은 무엇인가?
- **학습 전이:** 개념 렌즈를 습득한 학습자는 다른 문제를 어떻게 다르게 보는가?

---

4. 인간-AI 연구 연속성 이론 (HARCT)

4.1 기억의 비대칭

장기 연구에서 인간과 생성형 AI의 기억은 비대칭적이다.

- **AI:** 기본 설정에서는 세션이 끝나면 작업 기억이 사라진다. 컨텍스트 윈도우가 커져도 세션 사이의 연속성은 별도 수단 없이는 유지되지 않는다. 일부 서비스가 제공하는 저장 메모리나 프로젝트 기능의 효과는 이 논문에서 다루지 않는다.
- **인간:** 연구의 지향점, 직관적 확신, 미해결 문제 의식을 장기 기억에 보존한다.

4.2 분산 연구 기억

HARCT는 연구 저장소(Git)를 인간과 AI가 공유하는 외재화된 작업 기억으로 쓴다. 여기서 저장소는 문서와 파일의 변경 이력을 모두 저장하는 폴더(Git 버전 관리)이고, 커밋은 그 이력의 한 시점이다. 세션은 AI와 이어지는 한 번의 대화이며, 컨텍스트 윈도우는 한 세션에서 AI가 한꺼번에 읽을 수 있는 분량, 토큰은 AI가 읽고 쓰는 글자 조각의 단위이다. README는 폴더의 첫 안내문이다.

```
[인간 연구자의 장기 기억]  (지향점 보존, 유의성 판단)
            ▲▼
 분산 연구 기억 (저장소)
   ├─ HANDOVER.md    : 세션 간 인계 사항과 진행 맥락
   ├─ CHANGELOG.md   : 연구 결정의 연대기적 기록
   ├─ TODO.md        : 계층화된 잔여 과제
   ├─ paper/*.md     : 정립된 논문 원고
   └─ index.html     : 결과의 대화형 포털
            ▲▼
 [AI의 세션들]  Session 1 → Session 2 → … → Session N
```

4.3 핸드오버 프로토콜

1. **Context Ingestion:** 새 세션은 README, HANDOVER, 최신 커밋 기록을 먼저 읽고 연구의 현재 상태를 복원한다.

2. **Goal Synchronization:** 인간의 지시를 기록된 현재 목표와 맞춘다.
3. **Execution & Dual Commit:** AI는 결과물과 함께 결정의 이유를 CHANGELOG에 기록한다.
4. **State Sealing:** 세션 종료 전에 미해결 과제와 다음 진입점을 HANDOVER에 기록한다.

4.4 인지적 분업

연구 단계	인간의 역할	AI의 역할
문제 착안	이상함을 감지하고 문제를 제기한다 (유의성 필터)	관련 문헌과 용어 탐색
가설 수립	기하학적·구조적 가설 제시	가설의 형식화 (군의 연산표 구성 등)
검증	논리 정합성 검토, 비약 감지	계산, 증명 초안, **별도 프로그램으로의 전수 재계산**
의미 부여	연구의 가치와 방향 결정	원고 작성, 수식 렌더링, 포털 구현

4.7절의 사례는 표의 "검증" 행에서 인간의 검토와 AI의 계산만으로는 오류가 남을 수 있음을 보여, 별도의 독립 검증 단계가 필요함을 시사한다. 이 표는 이론이 제안하는 분업이며, 4.7절의 프로젝트에서 나타난 실제 분담과 다른 부분이 있다(가설 생성이 그 예이다). 이 분업이 유지되는 이유와 AI의 발전에 따라 바뀔 수 있는지는 열린 문제이다.

4.5 HARCT의 추가 구성요소: 목표 보존, 맥락 전이, 재구성 비용

앞 절들의 내용을 포함해 HARCT는 다음 구성요소로 이루어진다. 모두 제안 또는 가설이며 통제된 검증을 거치지 않았다.

- **중심 주장.** 장기 인간-AI 협력은 근본적으로 지능의 문제가 아니라 **연구 연속성**의 문제이다. 지능은 필요하지만 충분하지 않다.
- **삼요소 아키텍처.** 인간의 목표 기억(방향, 장기 목표), AI의 추론 능력(맥락 재구성, 분석, 문서화), 산출물 기억(맥락, 과거 결정, 구조적 관계)이 상호작용할 때 연속성이 나타난다. 어느 하나로는 충분하지 않다.
- **맥락 전이의 5단계.** 맥락은 참여자 사이를 직접 이동하지 않고, 산출물로 인코딩된 뒤 수신자가 재구성한다. 맥락 원천(Source) → 인코딩(Encoding) → 전이 산출물(Transfer Artifact) → 재구성(Reconstruction) → 수신자(Receiver). 산출물의 품질이 주된 통제 변수이고, 재구성이 가장 자원을 많이 쓰는 단계이다.
- **목표 보존.** 정보가 보존되어도 목표가 사라지면 연구는 방향을 잃는다. 보존된 목표는 손실된 맥락의 재구성을 가능하게 하지만, 보존된 정보만으로는 연속성이 보장되지 않는다는 것이 이 제안의 내용이다. 이는 저자의 인상에 기반하며, 4.7절의 사례는 이를 직접 뒷받침하는 에피소드를 담고 있지 않다. 이 관점에서 미래의 연구자는 프로젝트에 새로 진입하는 참여자와 같으므로(Future-Self Collaboration), 문서는 기록이 아니라 미래 협력자와의 소통 수단으로 설계된다.
- **재구성 비용.** 생산적 연구를 재개할 만큼 맥락을 복원하는 데 드는 노력(시간, 인지적 노력, 토큰)이다. 문서화 품질이 이 비용을 좌우하는 변수라는 가설이며, 새 지식 생성에 쓰인 토큰(생산 토큰)과 맥락 복구에 쓰인 토큰(재구성 토큰)의 비율이 문서화 품질의 지표가 될 수 있다는 제안이 있다. 이 지표는 검증되지 않았다.

4.6 실패 유형, 경계 조건, 검증 가능한 예측

실패 유형. 연속성이 무너지는 방식을 다음 10가지로 분류한다. 인간의 망각, AI의 컨텍스트 손실, 세션 경계, 토큰 고갈, 저장소 단편화, 인계 실패, 목표 표류, Future-Self 소통 실패, 재구성 실패, 그리고 **저장소 표류(Repository Drift)**. 저장소 표류는 문서가 실제 프로젝트 상태와 점점 어긋나 인계 품질이 조용히 저하되고, 후임이 오도된 맥락을 받는 실패이다. 실패는 대개 단일 사건이 아니라 목표 표류 → 맥락 손실 → 재구성 실패 → 연구 중단의 연쇄로 나타난다고 가정한다.

경계 조건. 이 이론의 설명력은 연구 기간이 길고, 중단이 잦고, 여러 AI가 참여하고, 맥락 재구성이 자주 필요할수록 커진다. 짧은 프로젝트, 단일 세션 상호작용, 맥락이 항상 모든 참여자에게 가용한 경우에는 설명력이 제한된다.

검증 가능한 예측과 반증 조건.

1. 저장소 인프라가 좋은 프로그램일수록 복구 속도가 빠르고 재구성 비용과 맥락 손실이 적다.
2. README 품질이 높은 저장소는 재구성 노력이 적다.
3. AI 전환 시 구조화된 인계 문서가 있으면 맥락 손실이 적다.
4. 미래 지향 산출물을 만드는 연구자는 맥락을 더 효율적으로 회복한다.
5. 목표 보존 장치가 있는 프로젝트는 더 긴 중단을 견딘다.
6. 연구 기간이 길수록 재구성 비용이 협력 효율의 지배적 요인이 된다.
7. 지식이 특정 AI가 아니라 지속되는 산출물에 붙어 있을수록 연속성이 높다(AI 교체 가능성).
8. 안정적인 협력 프로토콜은 중단 후 맥락 재확립 시간을 줄인다.

문서화 품질이 재구성 비용에 영향을 주지 않거나, 인계 문서가 AI 전환 후의 맥락 손실을 줄이지 못하거나, 목표 보존 장치가 프로젝트 생존에 도움이 되지 않거나, 저장소 인프라 품질이 맥락 복구 속도와 상관이 없다고 반복 관찰되면 이론의 중심 주장은 크게 약화된다.

맥락의 구성요소와 인계 품질. 이론은 연속성에 필요한 최소 맥락 단위를 목표, 상태, 이력, 다음 행동의 네 가지로 본다. 이 네 가지 구분이 최소이자 충분한지는 검증되지 않았다. 산출물마다 보존하는 요소가 다르다. README는 주로 목표를, 상태 문서는 상태를, 연구 로그는 이력을, 다음 행동 문서는 다음 행동을 보존하고, 인계 문서와 통합된 저장소만 넷을 모두 보존한다. 따라서 어느 한 종류의 산출물만 있는 프로그램은 나머지 요소에서 예측 가능한 취약점을 갖는다. 인계의 품질은 여섯 단계로 평가한다. 0단계(맥락 없음), 1단계(목표만), 2단계(목표와 상태), 3단계(목표, 상태, 이력), 4단계(복구 가능한 맥락: 넷 모두), 5단계(재구성 부담 없이 즉시 생산적 작업이 가능한 연속성)이며, 이론이 권하는 최소 목표는 4단계이다. 매 세션 후의 점검 항목은 다음과 같다. 목표가 최신으로 기록되어 있는가, 현재 상태가, 관련 이력이, 다음 행동이 기록되어 있는가, 산출물이 갱신되었는가, 새 협력자가 산출물만으로 작업을 이어갈 수 있는가. 하나라도 아니오이면 4단계 아래로 떨어진 것이다. 저장소 표류에 대한 이론의 기본 대응은 문서의 갱신을 연구 활동의 일부로 취급하여 이 점검을 매 세션 수행하는 것이다. 저장소 표류가 진행되면 마지막 갱신 시점에 4~5단계였던 저장소가 몇 달 뒤 처음 접하는 협력자에게는 1~2단계처럼 작동할 수 있다.

4.7 사례 연구: 카르노맵 대칭 분류 프로젝트

범위와 자료. 한 명의 연구자(본 논문 저자)가 AI 어시스턴트와 다중 세션으로 4변수 카르노맵의 대칭 분류를 연구했다. 자료는 저장소의 커밋 기록과 인계 문서이며, 대화 로그 원문은 공개하지 않는다. 저장소는 모두 GitHub의 inha20 계정에 공개되어 있다. 이 사례의 기록은 8개 저장소에 걸쳐 있다. 이전 단계의 연구 저장소 여섯 곳(1KMapStructureInvariance, 2SymmetricBooleanFunctionMinorThesis, 3VariableRearrangementInvarianceMinorThesis, 4StructureRecognitionTheory, 5HumanAIResearchCollaboration, 6BooleanFunctionSpaceTheory)과 자매 수학 논문(7KarnaughGroupClassification), 본 논문(8StructuralThinkingAndHumanAICollaboration)이다. 2026-10-17 기준으로 가장 이른 커밋은 2026-05-31이고 커밋은 모두 약 350개이다. 이 저장소들은 이력 보존을 위해 삭제하지 않는다. 아래 서술은 이를 바탕으로 한 저자의 회고적 재구성이며, 방법으로는 연구자 자신이 대상인 자기민족지적(autoethnographic) 사례 연구에 가깝다. 따라서 사례는 틀의 적용 가능성을 보여주는 탐색적 관찰이다. 에피소드 1과 2의 구체적 장면(질문의 문구, 세션 경위)은 저자의 회고이며 공개된 기록으로 교차 확인되지 않았다. 다만 에피소드 1의 가설은 같은 취지의 날짜가 있는 노트가 저장소에 남아 있다. 에피소드 2의 계산 오류와 에피소드 3의 오류들은 저장소 이력과 재계산으로 확인할 수 있다.

프로젝트의 AI 역할 분담 (저자의 기록). ChatGPT는 연구 토론, 이론 개발, 개념 통합, 가설 생성을, Claude는 저장소 조직, 문서화 정제, 구조 편집을, ANTIGRAVITY는 대규모 저장소 분석, 대량 문서 처리, 장문 생성을 주로 맡았다. 이 분담은 사전에 계획된 것이 아니라 장기 협력 중에 나타났다. 4.4절의 표는 이론이 제안하는 분업이고 이 분담은 사례에서 실제로 나타난 것이므로, 둘이 일치하지 않는 부분(가설 생성)이 있다. 협력 중에는 연구 모드로 전환하는 정형화된 문구(예: "지금 내가 놓치고 있는 것은?", "연구 프로그램 관점에서 보면?", "저장할 거 저장해줘.")가 비공식 협력 프로토콜로 자리 잡았다.

에피소드 1: 변수 배치 동치류의 질문과 형식화

- 연구자는 카르노맵을 보다가 "변수 배치 24가지가 가로세로를 바꾸거나 순서를 뒤집으면 사실상 같은 것이 아닌가, 남는 것은 3가지 정도일 것"이라는 직관적 가설을 세웠다 (SRT의 Question 단계의 예이며, 이 장면이 H3~H5를 검증하는 것은 아니다). 저장소 3의 2026-06-07 커밋(4a773d3)에 올라온 노트(future_works/EquivalenceClassAnalysis.md. 이후 삭제되어 커밋 이력에서만 볼 수 있다)에 같은 취지의 가설("24가지 배치는 약 3개의 대표 그룹, 체계적 검증이 필요함")이 적혀 있다. 그 노트가 든 변환은 회전, 반사, 좌표 이름 바꾸기이며 토러스 이동은 언급하지 않았다.
- AI는 이를 군론의 틀(이면체군, 순열군)로 형식화하여 "24가지가 3개의 동치류로 나뉜다"는 증명을 작성했다.
- 후속 점검에서 군의 이름과 범위에 대한 부정확함이 발견되었다. 격자의 회전·반사 8개만으로는 동치류가 12개이고, 토러스 순환 이동까지 포함한 위수 128의 군(원소가 128개인 군)에서 3개가 된다. 결론(3개)은 맞았으나 그 근거가 되는 군의 설명이 틀렸다.
- **관찰:** 인간의 직관적 가설이 질문의 방향을 정했고, AI의 형식화가 증명을 만들었다 (HARCT의 분업). 다만 형식화의 세부는 후속 점검에서 수정되어야 했다.

에피소드 2: 세션 단절, 인계, 그리고 계산 오류

- 이 연구는 여러 세션에 걸쳐 TODO와 HANDOVER 문서로 맥락을 이어 갔다. 이 부분의 세션 경위는 저자의 회고이며 공개 기록으로 교차 확인되지 않았다.
- 확인 가능한 사실(저장소 커밋 이력)은 다음과 같다. 65,536개의 4변수 함수를 변수 이름을 바꿔 같아지는 것끼리 묶으면 정답은 3,984묶음이다(번사이드 보조정리로 센다. 변수 4개를 한 줄로 돌리는 순열 6개 각각이 바꾸지 않는 함수의 수는 $2^6 = 64$이다). 그런데 이를 4,016으로 적은 문서가 있었다. 그 문서의 계산식은 이 항을 6×64가 아니라 6×16으로 적었고, 합계도 항들의 합 95,328이 아니라 96,384로 적어 4,016이 되었다. 항들의 합을 그대로 24로 나누면 3,972이므로 이 값은 두 가지 오류가 겹친 것이다. 이 값은 자매 수학 논문 저장소의 첫 릴리스 README(커밋 d0789d7, 2026-10-07)에 처음 나타났고, 같은 날 수학 원고 본문(커밋 1fd4e01)과 본 논문의 초기 원고(커밋 d2bf1e7)에도 기록되었다. 이전 단계의 연구 저장소 여섯 곳의 이력에는 이 값이 한 번도 나타나지 않았다. 즉 이 오류는 이전에 따로 쓴 연구 저장소들의 결과를 자매 논문 한 편으로 합치는 단계에서 생겼다.
- 오류는 2026-10-08, 자매 논문을 처음부터 다시 쓰면서 65,536개 함수의 궤도를 직접 세는 프로그램으로 발견되어 같은 날 정정되었다(자매 논문 저장소의 커밋 cd4c230, 본 논문 저장소의 커밋 6815851). 기록만으로는 이 오류를 낸 주체가 인간인지 AI인지 구분할 수 없다.
- **관찰:** 인계·요약 문서는 맥락 복원에 효과적이었지만, 문서에 오류가 있으면 그 오류도 함께 보존되고 이후 문서로 전파되었다. 인계 문서 자체는 검증 수단이 아니다.

에피소드 3: 독립적 점검에서 드러난 추가 오류

- 이 논문의 자매 수학 논문을 선행 문헌 없이 읽을 수 있는 형태로 다시 쓰면서, 모든 주장을 본문에서 도출하고 계산으로 재현하도록 요구했다. 그 과정에서 이전 세션과 검토를 통과해 있던 원고의 오류가 추가로 드러났다.

    - 4변수 XOR의 체커보드가 세 배치 동치류에서 서로 다른 모양으로 보인다는 서술은 틀렸다. 완전 대칭 함수의 그림은 24개 배치 모두에서 같다.
    - 4변수 부울 함수 전체의 집합 $\mathcal{F}_4$, 연산 집합 G가 생성하는 함수의 집합 $\mathrm{Space}(G)$(아핀, 단조, 자기쌍대 함수의 집합이 그 예이다), 완전 대칭 함수 32개의 집합 $\mathcal{S}_4$ 사이의 포함 관계 $\mathcal{F}_4 \supset \mathrm{Space}(G) \supset \mathcal{S}_4$는 성립하지 않았다(예: Exactly-2, 곧 1의 개수가 정확히 2일 때 1인 함수는 완전 대칭이지만 아핀, 단조, 자기쌍대 어느 집합에도 속하지 않는다).
    - 한 함수의 안정화군(그 함수를 바꾸지 않는 변수 순열 전체의 군)을 "D_4 (위수 4)"로 잘못 명명했고, 존재하지 않는 크기 24의 궤도(변수 순열로 서로 옮겨지는 함수 24개의 묶음)를 서술했다.
    - 이전 단계의 관찰 논문에는 대칭 함수(Exactly-3, 4변수 XOR)의 그림이 변수 재배열로 이동하거나 변한다는 서술이 있었다. 그중 일부 그림은 변수 재배열이 아니라 축 라벨 순서(Gray 대 이진)를 바꾼 것이었다.
    - 이론 문서의 한 궤도 대표 함수가 잘못 옮겨 적혀 있었고(크기 4 궤도의 대표가 크기 12 궤도의 함수였다), 함수 클론(합성에 닫힌 함수들의 집합, 예: 단조 함수 전체)의 정의 표기(상수 포함 여부)가 문서 간에 일치하지 않았으며, 참고문헌의 호수도 문서 간에 달랐다.
    - 수학 논문의 포털 페이지(저장소의 index.html로 만든 소개 웹페이지)에는 본문과 어긋난 표가 남아 있었다. 그 표는 단조 함수(입력을 늘리면 출력이 줄지 않는 함수)를 만드는 연산을 "AND, NOT"으로, 자기쌍대 함수(입력을 모두 뒤집으면 출력도 뒤집히는 함수)를 만드는 연산을 "NOT, AND, OR"로 적었으나 둘 다 틀렸다. 단조 함수는 AND, OR와 상수로 생성되고, 자기쌍대 함수는 NOT과 3입력 다수결로 생성되며, NOT, AND, OR는 모든 부울 함수를 생성한다. 본문을 고친 뒤 본문과 포털을 처음부터 끝까지 다시 읽는 점검(2026-10-10)에서 발견되었다. 본문과 포털 사이의 불일치도 저장소 표류의 한 형태이다.
- **관찰:** 이 오류들은 다중 세션과 연구자의 검토를 거쳤음에도 남아 있었다. 발견의 계기는 "독자가 선행 지식 없이 따라갈 수 있어야 한다"는 요구였고, 이는 모든 주장을 처음부터 도출하고 재현하도록 강제했다. 이 정정은 3.2절의 예와 모순되지 않는다. 3.2절의 구조 단계 예(변수 이름을 바꿔도 그림이 같은 함수들)는 정정된 서술(완전 대칭 함수의 그림은 모든 배치에서 같다)과 일치한다.

사례의 종합

- **HARCT와 부합하는 관찰 (저자의 회고에 의존하며 교차 확인되지 않음):** 분산 기억이 세션 간 연속성을 가능하게 했다 (에피소드 2). 인간의 직관적 질문과 AI의 형식화·계산이 분업으로 작동했다 (에피소드 1). 세션 수나 인계 문서의 변경 이력 같은 객관적 지표로 연속성의 작동을 측정하지는 않았다.
- **HARCT를 수정하게 하는 관찰:** 공유 기억은 오류도 전파한다 (에피소드 2, 3). 이는 4.6절의 "저장소 표류" 실패 유형에 해당한다. 인간의 검토와 AI의 계산만으로는 부족했고, 독립 재계산과 "처음부터 도출하기"라는 검증 단계가 오류를 찾았다. 이에 5장에서 순환에 독립 검증 단계를 추가한다. 다만 이 사례에서 "처음부터 도출하기"와 재작성에도 같은 계열 AI의 새 세션이 쓰였으므로 완전히 독립이라고 할 수 없다(7.1절).

4.8 HARCT의 한계와 위험

HARCT는 분산 인지(Hutchins, 1995)와 확장된 마음(Clark & Chalmers, 1998)과 개념적 영역이 겹친다. 삼요소 구조는 분산 인지와, 저장소를 인지 인프라로 보는 관점은 확장된 마음과 닮았다. 이 이론이 더하려는 것은 인간의 목표 기억과 AI의 단기 추론 사이의 비대칭, 그리고 산출물을 정보량이 아니라 맥락의 회복 가능성으로 평가하는 기준이다. 기존 문헌과의 구분과 새로움의 정도는 체계적인 문헌 조사 없이는 주장할 수 없다.

- 오류의 전파: 에피소드 2, 3 참조. 인계 문서와 논문 원고에 대한 독립 검증이 필요하다.
- 검토자의 과신: 사람은 자동화의 출력을 점검 없이 받아들이는 경향이 있다 (Parasuraman & Manzey, 2010). 구조화된 인계 문서가 이 경향을 강화할 가능성이 있으며, 독립 검증이 절차에 들어 있어야 하는 이유이다.

- 문서 비대화와 낡은 정보의 누적은 새 세션의 맥락 복원을 방해할 수 있다.
- 이 실천은 일반적인 소프트웨어 연구 관행(버전 관리, 문서화) 및 개인 정보 관리 연구(Jones, 2007)와 유사하며, HARCT가 새롭다고 저자가 보는 점은 AI 세션의 휘발성을 전제로 이를 체계화한 데 있으나, 선행 연구와의 구분은 확인되지 않았다.

---

5. 공생적 연구 순환

5.1 SRT와 HARCT의 결합

SRT는 인간이 구조를 인식하고 질문을 던지는 과정을, HARCT는 그 질문을 AI와 장기 연구로 완성하는 방법을 기술한다. 두 이론은 다음 5단계 순환으로 결합된다.

1. **인간 직관 → AI 형식화:** 인간이 패턴에서 구조를 직관하고 질문으로 제시하면, AI가 이를 수학적 기호 체계로 옮긴다.
2. **AI 연역·계산 → 인간 검토:** AI가 증명과 계산을 수행하고, 인간이 논리의 비약을 검토한다.
3. **독립 검증 (추가):** 핵심 수치와 주장을 원래 결과를 만든 수행자, 세션, 방법과 다른 수단(전수 열거 프로그램, 독립 구현, 선행 지식 없는 독자의 재현)으로 확인한다. 4.7절의 사례에서 도출되었다.
4. **인간 가치 부여 → 연구 확장:** 인간이 결과의 의미를 부여하고 다음 질문을 정한다 (H4, H8).
5. **공동 체계화 → 분산 기억 저장:** AI가 논문과 포털로 정리하고 저장소에 보존하여 다음 순환의 출발점으로 삼는다.

5.2 독립 연구자에게 주는 시사점

4.7절의 사례는 한 명의 연구자가 AI 협업으로 대규모 연구실 없이 일관된 연구 프로그램을 진행할 수 있음을 시사한다. 그러나 이는 단일 사례이며, 같은 효과가 다른 연구자나 분야에서 재현되는지는 검증되지 않았다.

---

6. 교육적 함의

6.1 문제의식

정형화된 계산(예: 진리표에서 최소 논리식 도출)은 AI가 빠르게 수행하지만 항상 정확하지는 않다(4.7절의 에피소드 2). 계산 숙련도 중심의 평가는 학습 성과를 측정하는 능력을 잃을 수 있다. 다만 계산 능력 자체의 가치가 사라지는 것은 아니며, 구조적 이해의 기초로 여전히 필요하다. 또한 표절, 저작권, 개인정보 같은 윤리 문제와, 유료 서비스에 따른 접근 격차가 교육 불평등으로 이어질 수 있다는 점도 함께 고려해야 한다.

6.2 제안: 과정 중심 교육

여기서 과정 중심 교육은 결과물의 정답 여부만이 아니라 질문을 만들고 검증하는 과정을 학습 활동과 평가의 대상으로 삼는 교육을 뜻한다. 형성평가나 문제 제기식 학습 같은 기존 교육 흐름과의 관계는 체계적으로 검토하지 않았으며, 이 논문이 더하려는 점은 AI 출력의 독립 검증을 훈련과 평가의 대상에 포함하는 것이다.

1. **문제 해결자에서 문제 제기자로:** "왜 변수 배치를 바꾸면 그림은 달라지는데 본질은 같은가"와 같은 질문을 만드는 활동을 포함한다.
2. **메타인지적 검증 훈련:** AI의 답과 증명에서 오류를 찾고, 독립적인 방법으로 확인하는 훈련을 포함한다. 4.7절의 사례에서는 그럴듯한 오류가 세 에피소드 모두에서 검토를 통과했다.

3. **구조적 사고 훈련:** 대칭과 불변량을 찾는 활동으로 개념 렌즈(H8)를 형성한다.
4. **과정의 평가와 교사의 역할:** 결과물만이 아니라 가설을 세운 과정과, AI의 출력에서 찾은 오류와 그것을 확인한 방법을 평가에 포함한다. 교사의 역할은 답을 확인하는 사람에서 질문의 틀과 단서를 제공하고 검증 과정을 점검하는 사람으로 옮겨 간다. 이는 제안이며 검증되지 않았다.

6.3 모형: 공생적 문제 중심 학습

1. 학생이 먼저 자신의 직관으로 가설을 세운다.
2. AI와 대화하며 직관을 수학적 기호로 형식화한다.
3. AI의 결과를 비판적으로 점검하고, 독립 계산으로 확인하며, 반박하거나 증명한다.

초심자에게는 1단계 전에 교사가 공통 예시와 질문의 틀을 제시하고, 막힌 학생에게는 정답이 아닌 단서를 주는 규칙을 두어야 한다(1.5절의 조건). 이 규칙의 구체적 설계는 파일럿에서 정한다.

학생의 AI 활용 수준은 세 단계로 구분하여 설계할 수 있다. 1단계는 도구(계산이나 도표 생성의 대행), 2단계는 검증자(결과의 교차 검토와 반례 탐색), 3단계는 공동 탐구자(질문 생성과 가설 정제)이다. 이 구분은 저자의 프로젝트 기록에서 제안된 것이며 검증되지 않았다. AI 리터러시의 역량 틀(Long & Magerko, 2020)과의 대응은 확인하지 않았다.

6.4 검증을 위한 파일럿 설계

이 모형의 효과는 입증되지 않았다. 다음의 파일럿 연구를 제안한다.

항목	설계
대상	디지털 논리 수업을 듣는 학부생, 분반별 배정
조건	(A) 기존 계산 중심 수업, (B) 공생적 PBL 수업
측정	구조 식별 과제(H2), 학습 전이 과제(H8), AI 출력의 오류 탐지율(오류를 심은 AI 답안 몇 개에서 찾아낸 비율), 개념 이해도(도구는 사전에 정한다)
예측	조건 B가 조건 A보다 전이 과제 점수와 오류 탐지율이 높다. 차이가 없으면 모형의 효과 주장은 약화된다
분석	사전·사후 점수 차이의 조건 간 비교, 효과 크기 보고
고려 사항	연구 윤리 승인, 수업 시간 통제, AI 접근 조건의 통일, 분반 배정은 무작위가 아니므로 사전 검사와 교수자 효과의 통제(가능하면 무작위 배정), 표본 크기 목표의 사전 설정

---

7. 한계와 결론

7.1 한계

- **사례의 한계:** 사례는 한 명의 연구자(저자 본인)의 프로젝트이며, 회고적 재구성이고, 대화 로그 원문은 공개되지 않았다. 확증 편향의 위험이 있고 일반화할 수 없다. 연속성이 유지된 것으로 보이는 현상은 HARCT가 아니라 연구자의 높은 동기, 프로젝트 기간 중 AI 모델의 성능 향상, 저장소가 단순한 저장 장치로 쓰인 점, 저자의 문서화 습관 같은 다른 요인으로도 설명될 수 있으며, 이 사례만으로는 이들을 구분하지 못한다.
- **가설의 한계:** SRT의 10개 가설은 실증되지 않았다. 3.4절의 예측은 검증 설계의 예이다.
- **AI 관련 서술의 한계:** 2장의 서술은 현재 모델과 연구 동향에 대한 일반화이며 모델의 발전에 따라 바뀔 수 있다.
- **교육 제안의 한계:** 6장의 제안은 증거가 아닌 설계이다.

- **점검의 한계:** 이 논문의 오류 점검(1.4절, 4.7절의 에피소드 3)에도 같은 계열 AI의 새 세션이 쓰였으므로 완전히 독립된 검증이 아니다.

7.2 향후 과제

1. 3.4절의 예측(H2, H5, H8)을 실험으로 검증한다.
2. 여러 연구자와 분야에서 HARCT를 적용하여 4.7절의 관찰을 일반화하거나 반박한다.
3. 6.4절의 파일럿을 수행한다.
4. HARCT의 독립 검증 단계를 체계화하여 도구화한다.
5. HARCT의 제안 지표(재구성 시간, 목표 복구율, 인계 성공률, AI 교체 가능성 점수, Future-Self 복구 시간 등)로 측정하고, 맥락의 구성요소(목표, 상태, 이력, 다음 행동)를 하나씩 제거했을 때의 복구 가능성을 비교하는 실험, 문서화 수준이 다른 프로그램 사이의 재구성 비용 비교, 성공·실패한 프로젝트 재개의 비교 분석으로 4.6절의 예측을 검증한다.

7.3 결론

본 논문은 인간이 구조를 발견하여 질문을 만들고(SRT), AI와 분산 기억 위에서 장기 연구를 수행하는(HARCT) 틀을 제안했다. 사례는 저자의 회고에 근거하여 이 틀이 세션 간 연속성과 분업에 유효할 수 있음을 시사하는 동시에, 공유 기억이 오류도 전파한다는 위험을 드러냈다. 이에 따라 AI 시대의 연구와 교육에서 중요한 역량은 결과를 생산하는 속도가 아니라 구조를 파악하고 결과를 독립적으로 검증하는 능력일 수 있다는 가설을 제안한다. 학습자를 대상으로 한 증거는 아직 없으며, 이 가설의 검증은 후속 연구의 몫이다.

참고문헌 / References

1. Amershi, S., Weld, D., Vorvoreanu, M., et al. (2019). Guidelines for human-AI interaction. **Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems**. https://doi.org/10.1145/3290605.3300233
2. Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, 5185–5198.
3. Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. **Cognitive Science**, 5(2), 121–152.
4. Choi, J. (2026). 카르노맵 위의 4변수 부울 함수 공간의 군론적 분류 (Group-theoretic classification of 4-variable Boolean function spaces on Karnaugh maps). 자매 논문. https://github.com/inha20/7KarnaughGroupClassification
5. Clark, A., & Chalmers, D. (1998). The extended mind. **Analysis**, 58(1), 7–19.
6. COPE (Committee on Publication Ethics). Authorship and AI tools (position statement); ICMJE. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals (AI-assisted technology section).
7. Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. **American Psychologist**, 34(10), 906–911.
8. Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. **Cognitive Science**, 7(2), 155–170.

9. Huang, J., Chen, X., Mishra, S., et al. (2024). Large language models cannot self-correct reasoning yet. **International Conference on Learning Representations (ICLR 2024)**. arXiv:2310.01798.
10. Hutchins, E. (1995). **Cognition in the Wild**. MIT Press.
11. Ji, Z., Lee, N., Frieske, R., et al. (2023). Survey of hallucination in natural language generation. **ACM Computing Surveys**, 55(12), Article 248.
12. Jones, W. (2007). Personal information management. **Annual Review of Information Science and Technology**, 41(1), 453–504.
13. Kadavath, S., et al. (2022). Language Models (Mostly) Know What They Know. **arXiv:2207.05221**.
14. Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work. **Educational Psychologist**, 41(2), 75–86.
15. Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. **Cognitive Science**, 11(1), 65–100.
16. Liu, N. F., Lin, K., Hewitt, J., et al. (2024). Lost in the middle: How language models use long contexts. **Transactions of the Association for Computational Linguistics**, 12, 157–173.
17. Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. **Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems**.
18. Packer, C., Wooders, S., Lin, K., et al. (2023). MemGPT: Towards LLMs as operating systems. **arXiv:2310.08560**.
19. Parasuraman, R., & Manzey, D. (2010). Complacency and bias in human use of automation: An attentional integration. **Human Factors**, 52(3), 381–410. https://doi.org/10.1177/0018720810376055
20. Sharma, M., Tong, M., Korbak, T., et al. (2023). Towards understanding sycophancy in language models. **arXiv:2310.13548**.
21. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. **Cognitive Science**, 12(2), 257–285.
22. Vaswani, A., et al. (2017). Attention is All You Need. **Advances in Neural Information Processing Systems**, 30.
23. Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. **Scientific Reports**, 13. https://doi.org/10.1038/s41598-023-41032-5