AI가 발견한 것, 우리가 이미 알고 있던 것

Chungil Chae

Fri, 3 April 2026

AI가 발견한 것, 우리가 이미 알고 있던 것

AI-Scientist가 데이터로 자율 생성한 논문의 해석

작성일: 2026-04-03

들어가며: 기계에게 데이터를 줬다

우리는 36개의 모형을 돌렸다. 5가지 방법론 패러다임으로, 16개 이론을 동원해서, 282건의 데이터에 세 번 질문했다. 그 결과를 두 개의 보고서(전략편, 해석편)로 정리했다.

그리고 같은 데이터를 AI-Scientist에게 줬다. "네가 알아서 분석하고, 논문을 써라."

AI는 스스로 K-means 클러스터링을 선택했다. 5번의 실험을 자율 수행했다. 8페이지 논문을 작성했다. 비용은 $8, 시간은 20분.

그 AI가 발견한 것은, 우리가 이미 도달했던 결론의 한 조각과 **정확히 겹친다**.

1장. AI가 본 것

클러스터링이라는 선택

AI에게는 아무런 이론적 지침이 없었다. TOE도, P-E Fit도, 흡수역량도 모르는 상태. seed_ideas 두 개(TOE 비선형 확장, 용량-반응 분리)를 힌트로 줬을 뿐이다.

AI는 스스로 세 번째 아이디어를 만들었다: **"K-means로 기업을 분류하고, 클러스터별로 따로 분석하자."** 점수는 Interestingness 9, Feasibility 9, Novelty 8. 네 개 아이디어 중 가장 높은 평가를 자기 스스로에게 줬다.

이것은 우리가 1차 분석에서 T2(LPA)로 했던 것과 같은 직관이다. 데이터 안에 **숨겨진 유형**이 있다는 감각. 평균의 이면에 이질성이 있다는 의심.

세 개의 프로파일

AI가 분류한 3개 클러스터:

클러스터	n	DT인식	스마트시스템	기업규모(log)	Q3_차이
0 (저준비도)	122	1.84	16.1	2.51	1.525
1 (고준비도)	51	3.24	25.4	3.34	1.498
2 (중준비도)	70	1.89	21.3	3.73	1.247

여기서 첫 번째 이상한 점이 보인다. **고준비도 기업(클러스터 1)의 Q3_차이가 저준비도(클러스터 0)와 거의 같다.** 1.498 vs 1.525. DT인식은 2배, 스마트시스템은 60% 더 많은데, 교육 효과는 같다.

이것은 우리가 1차 분석 T2(LPA)에서 발견한 것과 똑같다: "더 준비된 기업이 반드시 더 잘 배우지는 않는다." 해석 보고서에서 "대기업 역설"과 "천장효과"로 설명했던 그 패턴이다.

R^2 .002 → .336의 극적 격차

AI의 가장 임팩트 있는 발견:

- **클러스터 0** (저준비도, n=122): R^2 = .002. 어떤 변수도 교육효과를 설명하지 못한다.
- **클러스터 1** (고준비도, n=51): R^2 = .336. 스마트시스템(+)과 기업규모(−)가 강력한 예측변수.
- **클러스터 2** (중준비도, n=70): R^2 = .018. 역시 아무것도 설명 못한다.

전체 데이터에서 R^2 = .029였던 것이, 고준비도 기업만 분리하면 .336으로 뛴다. **12배.** AI는 이것을 "dramatic heterogeneity"라고 불렀다.

2장. 우리가 이미 알고 있던 것

AI의 발견을 우리의 36개 분석과 대조하면, 퍼즐 조각이 맞아떨어진다.

대조 1: "고준비도에서만 예측변수가 작동한다"

AI의 발견: 클러스터 1에서만 R^2 가 유의미하다.

우리의 발견: - T2(LPA): 프로파일 간 교육효과 차이가 유의하지 않았지만, 프로파일별 **메커니즘**이 달랐다. - T3/T20(QCA): 등결과성. 높은 교육효과로 가는 경로가 여러 개. ~DT_AWARE(낮은 인식)도 유효 경로. - combo1(RSM): P-E Fit 교차항(p=.002)은 "인식과 인프라가 **동시에** 높을 때"에만 유의.

AI가 K-means로 발견한 것은, 우리가 QCA의 등결과성과 RSM의 교차항으로 도달했던 것의 **다른 표현**이다. "고준비도에서만 작동한다"와 "인식×인프라 교차항이 유의하다"는 같은 현상의 다른 각도다.

대조 2: "기업규모의 부정적 효과"

AI의 발견: 클러스터 1에서 기업규모 $\beta = -0.839$ ($p<.001$). 고준비도 기업 중에서도 규모가 클수록 교육효과가 **떨어진다**.

우리의 발견: - T14(ANOVA): 대기업 역설. 규모가 큰 기업에서 역효과. - T15: SF+복수참여 역시너지. 이미 많이 가진 기업에서 추가 투입의 한계. - RSM(B1): 기업규모 역U자 (p=.029), 최적 ~13명.

AI의 $\beta = -0.839$는 **역U자의 우측 하강 구간**에 해당한다. 고준비도(클러스터 1) 기업은 평균 규모가 크기 때문에($\log 3.34 \approx 28$명), 역U자의 최적점(~13명)을 이미 넘어선 상태. 수확체감의 영역에 있는 것이다.

해석 보고서에서 "Liability of Smallness + 천장효과의 이중 메커니즘"으로 설명했던 것의 **한쪽 면**을 AI가 포착한 것이다.

대조 3: "절편이 아니라 기울기의 차이"

AI의 가장 정교한 발견: 클러스터 더미 변수를 넣으면 비유의(p=.856), 하지만 클러스터×변수 교차항을 넣으면 유의(p=.043).

이것은 세 그룹의 **평균 교육효과는 비슷**하지만(절편 차이 없음), **교육효과를 결정하는 메커니즘이 다르다**는 것(기울기 차이)을 의미한다.

해석 보고서의 통합 명제가 정확히 이것이었다:

> "중소기업 디지털 전환 교육의 효과는 '있거나 없거나'가 아니라, '인식과 인프라가 만날 때' 나타난다."

"있거나 없거나"(절편 차이) 아님 → AI의 더미 비유의(p=.856)와 일치. "만날 때"(조건부) → AI의 교차항 유의(p=.043)와 일치.

3장. AI가 놓친 것

AI의 논문이 Reject (내부리뷰 지가 쓴걸 지가 스스로 리뷰한다 CS쪽 페이퍼들로 만들어진 모델기반으로)을 받은 이유는 단순히 저널 미스매치가 아니다. **사회과학 논문에 필수적인 것들이 빠져있다.**

빠진 것 1: 이론

AI의 논문에는 이론이 없다. "이질성이 있다"는 **기술**(description)은 있지만, "왜 이질성이 생기는가"라는 **설명**(explanation)이 없다.

우리는 16개 이론을 동원했다. P-E Fit이 교차항을, 수확체감이 역U자를, 등결과성이 다경로를, 경험학습이 용량-반응을 설명했다. **이론 없는 발견은 패턴에 불과하다.** AI가 R^2 .002 → .336을 발견한 것은 인상적이지만, "왜 고준비도에서만 예측변수가 작동하는가?"에 대한 이론적 답이 없으면 그것은 **미완의 발견**이다.

답: P-E Fit이다. 인식(P)과 인프라(E)의 적합도가 높은 기업에서만, 추가적인 조직 특성(스마트시스템, 기업규모)이 교육효과를 예측하는 메커니즘이 활성화된다. 적합도가 낮은 기업에서는 메커니즘 자체가 "꺼져 있다."

빠진 것 2: 삼각검증

AI는 K-means 하나로 결론을 냈다. 우리는 같은 질문에 5가지 방법론을 적용했다: - LPA (인원중심) → 프로파일 유형 - QCA (집합론) → 조건 조합 - RSM (비선형) → 교차항/역U자 - DID/PSM (인과추론) → 처치효과 - CLPM (종단) → 인과 방향

K-means 하나의 실루엣 .293은 "합리적"이라고 할 수 있지만, **그 패턴이 다른 방법론으로도 확인되는가?** 우리의 답은 "예, 5가지 모두에서 수렴한다"였다. AI의 답은 "모르겠다, 시드를 바꿔봤더니 비슷하다" 수준에 머물렀다.

빠진 것 3: "측정의 한계" 통찰

우리의 가장 중요한 발견 중 하나: **교육은 작동하고 있다. 다만 Q3_차이가 그것을 포착하지 못할 뿐이다.**

- 스마트시스템 DV → R^2 = .517
- Q3_차이 DV → R^2 = .136

같은 IV인데 DV만 바꾸면 설명력이 **4배** 차이. 이것이 5-방법론 삼각검증(OLS, PSM, DID, CLPM, RSM)으로 수렴한 "용량-반응 분리" 현상이다.

AI는 이 질문 자체를 하지 못했다. Q3_차이만 종속변수로 사용했고, "혹시 다른 DV로 바꾸면 어떨까?"라는 발상이 없었다. 이것은 **도메인 지식의 부재**다. 데이터를 안에서 최적화하는 능력은 뛰어나지만, 데이터 **밖을 상상하는 능력**은 아직 없다.

빠진 것 4: Discussion

사회과학 논문의 핵심은 Discussion이다. "이 발견이 기존 이론에 무엇을 추가하는가?" "실무자에게 무엇을 의미하는가?" "정책적 함의는?"

AI의 Conclusion은 결과 요약 + 한계 나열에 그쳤다. 우리의 해석 보고서가 도달한 "맞춤형이라는 단어" — 기업의 현재 위치를 진단하고 부족한 차원을 보완하는 개입을 설계해야 한다는 **처방적 결론**은 AI에게 없었다.

4장. 그래서 이것은 무엇을 의미하는가?

AI-Scientist의 진짜 가치

AI가 20분과 $8로 발견한 R^2 .002 → .336은, 우리가 3개월과 36개 모형으로 도달한 결론의 한 조각이다. 전체의 약 1/5. 하지만 **방향은 정확**했다.

이것이 의미하는 것: 1. **탐색의 가속기로서의 가치**: AI가 먼저 클러스터링으로 이질성을 스캔하고, 연구자가 이론을 입히고 삼각검증하는 워크플로우가 가능하다. 2. **가설 생성 도구로서의 가치**: AI의 "고준비도에서만 작동한다"는 발견은, "왜?"라는 다음 질문을 유발한다. 그 질문에 P-E Fit으로 답하는 것은 인간의 몫이다. 3. **패턴 확인 도구로서의 가치**: 우리가 36개 분석으로 도달한 결론을, AI가 독립적으로 확인해준다. 이것은 일종의 **방법론적 삼각검증**이다.

AI-Scientist의 한계

1. **이론 부재**: 패턴은 찾되, 설명하지 못한다.
2. **DV 상상력 부재**: 주어진 DV 안에서만 최적화한다.
3. **문헌 부재**: 선행 연구와의 대화가 없다.
4. **Discussion 부재**: "so what?"에 답하지 못한다.
5. **삼각검증 부재**: 단일 방법론에 의존한다.

이 다섯 가지는 모두 **도메인 지식**의 문제다. 통계적 패턴 인식 능력은 이미 인간에 근접하거나 초과한다 (20분 vs 3개월). 부족한 것은 "이 분야에서 무엇이 중요한가"를 아는 것이다.

다음 단계: AI-Scholar

이 Gap을 메우는 것이 AI-Scholar의 개발 목표다.

- Theory Engine: 이론 DB + 패턴-이론 매칭
- Literature Engine: SSCI 문헌 검색 + 인용 생성
- Analysis Engine: 삼각검증 자동화 (QCA + LPA + RSM + SEM)
- Writing Engine: Theory → Hypotheses → Discussion 구조

AI-Scientist가 보여준 것은 **가능성**이다. 20분과 $8로 방향이 맞는 발견을 할 수 있다. 여기에 이론과 맥락을 입히면, 사회과학 논문의 자동 생성이 불가능하지 않다.

마치며: 기계의 눈과 연구자의 눈

AI는 R^2 .002 → .336을 봤다. 우리는 "인식과 인프라가 만날 때"를 봤다.

AI는 "dramatic heterogeneity"라고 썼다. 우리는 "맞춤형"이라는 단어에 도달했다.

AI는 패턴을 발견했다. 우리는 그 패턴에 이름을 붙이고, 의미를 부여하고, 다음 행동을 제안했다.

둘 다 같은 데이터를 봤지만, **보는 깊이가 다르다**. AI의 깊이는 빠르게 깊어지고 있다. 하지만 "왜?"와 "그래서?"라는 질문은 아직 인간의 영역이다.

이 프리프린트는 AI-Scientist의 첫 번째 사회과학 테스트이자, AI-Scholar 개발의 출발점이다. 기계의 눈과 연구자의 눈이 협력하는 방법을, 이제부터 설계해야 한다.

