Market Insight | 2026.5.28

그림: Midjourney(지능을 연성하는 토큰술사)

MIRAE ASSET 미래에셋증권

AI Hot Issue

강철의 토큰술사(안드레 카파시)의 이적이 시사하는 것

한종목

chongmok.han@miraeasset.com

CONTENTS

Executive Summary

정답이다. 토큰술사.

안드레 카파시의 Anthropic 합류는 스타 연구자의 이직 뉴스로만 볼 게 아닙니다. OpenAI 초기 창립 멤버, Tesla AI 리더, Eureka Labs 창업자라는 이력만으로 이 사건을 읽으면 핵심을 놓치게 됩니다. “카파시가 얼마나 대단한 사람인가”가 아니라, 그가 왜 지금, 왜 Anthropic으로 들어갔는가라고 생각합니다.

그는 대형언어모델을 ‘소프트웨어 3.0’, 즉 새로운 컴퓨팅 패러다임으로 보고 있습니다. 소프트웨어 3.0은 모델에게 문맥과 도와 목표를 주고, 그 모델이 작업 환경 안에서 반복적으로 일을 수행하게 하는 구조입니다. 이 세계에서 핵심은 더 이상 프롬프트가 아니게 됩니다. **모델이 어떤 문서를 읽는지, 어떤 기억을 유지하는지, 어떤 도구를 호출할 수 있는지, 어떤 기준으로 성공과 실패를 판단하는지, 언제 사람에게 넘겨야 하는지가 성능을 결정합니다**.

Anthropic은 정확히 이 방향으로 움직이고 있습니다. Claude Code, 프로젝트 기억, 기술 묶음(Skills), 하위 에이전트(Sub-agents), 외부 도구 연결(Connectors)은 모두 Claude를 채팅창 밖으로 끌어내는 장치입니다. Anthropic은 모델 단품을 파는 회사가 아니라, 모델이 실제 업무와 연구 루프 안에서 작동하는 환경을 만들고 있는 것입니다.

그렇다면 카파시가 왜 필요한 것일까요? **그는 Anthropic의 사전학습 팀에서 Claude를 활용해 Claude의 다음 세대를 더 빨리 연구하는 역할을 맡게 될 것입니다. 즉, “Claude로 Claude를 개선하러 갔다”는 것입니다.** 여기서 만화 『강철의 연금술사』로 쉽게 비유할 수 있습니다. 비유의 핵심은 "법칙과 대가"입니다. 연금술은 마법이 아니라 이해, 분해, 재구축의 기술입니다. 그리고 그 모든 과정에는 등가교환의 법칙이 따릅니다. AI도 마찬가지입니다. 더 좋은 모델은 무에서 탄생하는 게 아니라 '데이터, 컴퓨트, 전력, 메모리, 네트워크, 데이터센터, 평가 시스템, 그리고 수많은 실패한 실험'이 필요합니다.

카파시가 맡게 될 재귀적 자기개선(RSI)은 만화에 나오는 현자의 돌로 비유할 수 있습니다. 그러나 현자의 돌이 사실 공짜 에너지가 아니라 엄청난 대가가 압축된 촉매였듯, AI에 있어서 RSI 또한 비용을 없애는 마법이 아닙니다. 더 많은 컴퓨트와 데이터와 전력을 더 빠른 속도로 지능으로 바꾸는 촉매입니다. 효율이 올라가면 비용이 줄어드는 것이 아니라, 그동안 비용 때문에 못 했던 실험과 추론과 에이전트 루프가 되살아날 것입니다.

따라서 **카파시의 Anthropic행은 소프트웨어 뉴스처럼 보이지만, 실제로는 AI Value Chain 전체에 영향을 주는 뉴스입니다. 연구 루프가 자동화될수록 전력, 데이터센터, 전력 장비, GPU, HBM, CPU, 네트워크, 추론 런타임, 기업 문맥, Deep SaaS까지 모두 다시 평가해야 합니다.** '누가 전력을 가장 빠르게 토큰으로 바꾸는가? 누가 토큰을 기업 문맥과 결합해 디지털 노동으로 바꾸는가? 누가 그 디지털 노동의 결과를 과금 가능한 업무 성과로 전환하는가? 그리고 누가 AI를 이용해 AI 연구 자체를 더 빠르게 반복하는가?'를 놓치지 말고 AI 투자 지도를 완성해야 합니다. 안드레 카파시는 이 질문들에 대한 정답이 되기 위해 Anthropic으로 들어간 것으로 보입니다. "정답이다. 연금술사".

I. Anthropic에 들어간 이유와 AI Value Chain에의 영향

1. 카파시의 이동: AI 연구의 생산함수의 변화

(1) 바이브 코딩의 다음 단계

약 한 달 전, 공개된 인터뷰 현장에서 안드레 카파시는 “자신이 프로그래머로서 이렇게 뒤처졌다고 느낀 적이 없다”고 말했다. 단지 겸손한 농담이 아니다. 2025년 말에서 2026년으로 넘어오며 모델의 코딩 능력이 질적으로 바뀌었다는 코딩 천재의 실제 체감이다. 예전에는 모델이 코드를 제안하면 사람이 상당 부분 고쳐야 했다. 그런데 어느 순간부터 모델이 내놓는 코드 조각들이 그냥 맞기 시작했다. 한 번 더 시키면 또 맞고, 더 시키면 또 맞았다. 마지막으로 직접 고친 게 언제인지 기억나지 않을 정도가 됐다고 그는 밝혔다. 말 그대로 자신이 '바이브 코딩'을 하고 있는 것이라고 회고했다.

그러나 카파시가 정말로 말하고 싶은 것은 바이브 코딩의 낭만이 아니다. 바이브 코딩은 '바닥'을 높인다. 누구나 프로그램을 만들 수 있게 한다. 하지만 전문 소프트웨어에서 더 중요한 것은 '천장'을 높이는 일이다. 보안, 권한, 결제, 데이터 모델, 테스트, 배포, 유지보수까지 포함한 시스템을 제대로 만드는 일이다. 이것이 그가 말하는 에이전트 엔지니어링(Agentic Engineering)이다.

그가 든 사례 중에 Stripe 이메일과 Google 이메일 사례는 이 차이를 보여준다. 어떤 앱에서는 사용자가 Google 계정으로 로그인하고 다른 앱에서는 Stripe 계정으로 결제하는 경우가 많다. 그런데 에이전트는 두 이메일 주소를 같은 사람의 식별자로 매칭하려 했다. 물론, 겉으로는 그럴듯하다. 그러나 실제로는 위험하다. 사용자는 로그인 이메일과 결제 이메일을 다르게 쓸 수 있기 때문이다. 결제 시스템에서 사람을 식별하는 기준은 이메일이 아니라 영속적인 사용자 ID여야 한다.

이와 같은 사례는 AI 시대의 인간 역할을 정확히 보여준다. API의 문법같은 것은 이제 모델이 그냥 외워준다. 그러나 시스템의 불변조건은 인간이 이해해야 한다는 말이다. '결제의 기준점이 무엇인지, 권한 경계가 어디까지인지, 데이터가 복사되는지 아니면 참조되는지, 어떤 결과는 반드시 사람이 검토해야 하는 건지'는 사람이 판단해야 한다.

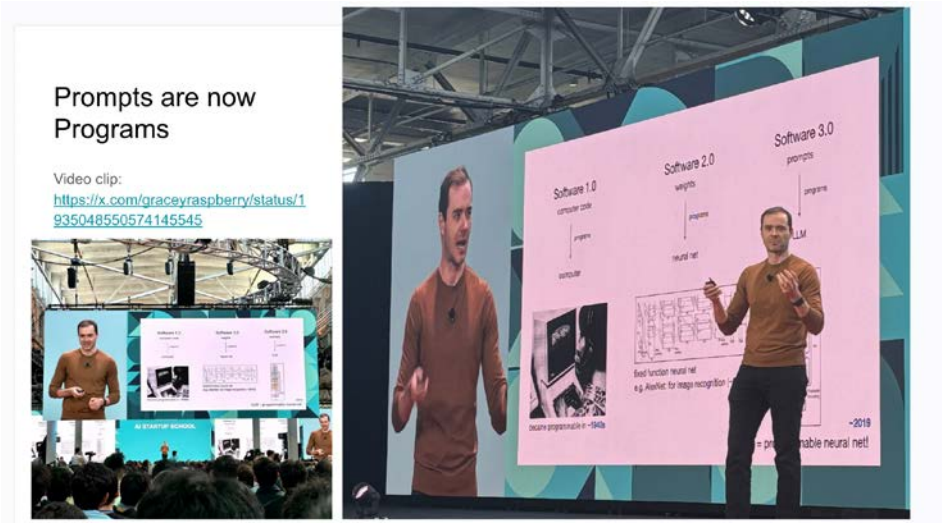
카파시가 LLM을 동물이 아니라 유령(Ghost)에 가깝다고 표현한 것도 이 맥락이다. LLM은 욕망이나 생존 본능을 가진 생물이 아니라, 데이터와 보상 함수 위에서 형성된 통계적 회로다. 만화『강철의 연금술사』의 '호문쿨루스'처럼 인간의 형상과 언어를 흉내낼 수는 있지만, 인간과 같은 방식으로 이해하거나 욕망하지는 않는다. 그러므로 AI를 잘 쓰려면 감정이 아니라 경계, 문맥, 도구, 검증 기준을 잘 설계해줘야 한다.

이를 카파시가 직접 말한 문장으로 요약하면, "생각(thinking)은 맡길 수 있지만 이해(understanding)는 맡길 수 없다"로 정리할 수 있다. 이것이 카파시의 Anthropic행을 택한 첫 번째 의미다. 카파시는 AI로 코딩을 더 많이 하러 간 것이 아니다. 그는 AI가 코딩과 연구를 반복적으로 수행할 수 있는 규율, 즉 에이전트 엔지니어링의 운영체제를 프론티어 랩의 기계실에 심으러 간 것으로 보아야 한다.

(2) 소프트웨어 3.0: 앱의 증발과 문맥의 부상

카파시가 주창한 개념인 '소프트웨어 3.0'은 “코딩을 더 빨리 한다”는 말보다 훨씬 급진적이다. 소프트웨어 1.0은 사람이 직접 쓰는 코드였다. 그리고 소프트웨어 2.0은 데이터로 학습된 신경망이었다. 나아가 소프트웨어 3.0은 LLM을 실행 환경처럼 다루는 단계다. 여기서 인간은 코드의 모든 줄을 직접 쓰는 것이 아니라, 모델이 일할 문맥, 도구, 목표, 평가 기준을 설계할 뿐이다.

그림 1. Software 3.0 개념을 명확히 보여주는 Karpathy 본인의 슬라이드/다이어그램



자료: 안드레 카파시, 미래에셋증권 리서치센터

그가 말한 사례 중 MenuGen라는 예시는 이 변화를 시사한다. 예를 들어, 전통적인 방식에서는 외국의 음식점에 갔을 때 메뉴판 사진에 대고 "OCR로 읽고, 메뉴 항목을 추출하고, 이미지 생성 모델을 호출하고, UI에 다시 렌더링"하는 작업이 소요됐다. 즉, 입력, 인식, 처리, 생성, 출력이 모두 어플리케이션의 계층으로 존재한 것이다. 그런데 소프트웨어 3.0에서는 메뉴판 사진을 모델에게 주고 “각 메뉴 옆에 음식 이미지를 그려 넣어라”고 말하면 된다. 모델이 사진을 이해하고, 메뉴를 파악하고, 이미지를 생성해 결과를 바로 만든다. 중간 앱 계층이 증발한 것이다.

이 말은 기존 소프트웨어 산업에 사실 불편한 질문을 던진다. 사용자가 원하는 것은 앱이 아니라 결과 그 자체이기 때문이다. UI도 해자가 아니라 포장지일 수 있다. AI가 문서를 읽고, 도구를 호출하고, 결과를 생성할 수 있는 세계에서 단순 기능 소프트웨어는 얼마나 버틸 수 있을까 고민하게 만든다.

그렇다면 진짜 해자는 어디일까? 해자는 이제 문맥(context)으로 이동한다. 모델 자체도 중요하지만, 모델이 접속하는 기업 데이터, 업무 규칙, 권한 구조, 평가 기준, 과거 작업 기억이 더 중요해진다. 아무 정보도 없는 새 채팅창의 Claude와, 회사의 코드베이스·회의록·고객 기록·보고서 스타일·의사결정 로그를 알고 있는 Claude는 같은 제품이라 부를 수 없다. Anthropic과 카파시의 궁합은 이 지점에서 시작된다. 카파시는 소프트웨어 3.0의 언어를 만들었다. Anthropic은 Claude를 소프트웨어 3.0의 작업 환경으로 만들고 있다.

2. 왜 Anthropic인가: 철학, 제품, 연구 루프의 정합성

(1) Anthropic은 사이드 퀘스트를 줄이고 코딩·추론에 집중한다

프론티어 AI 랩들은 같은 목적지를 향하는 듯 보이지만, 실제 경로는 조금 다르다. Google DeepMind는 세계 모델, 물리, 비디오, 오디오, 과학 모델까지 넓은 전선을 구축하고 있다. OpenAI는 (최근에는 좀 바뀌었지만) 소비자 제품, 멀티모달, 코딩, 음성, 영상, 에이전트까지 광범위하게 움직이는 진영이다. 반면 Anthropic은 상대적으로 매우 좁고 또 매우 깊다. 텍스트 추론, 코딩, 안전성, 기업 도입, 에이전트형 작업에 고돌 집중한 곳이다.

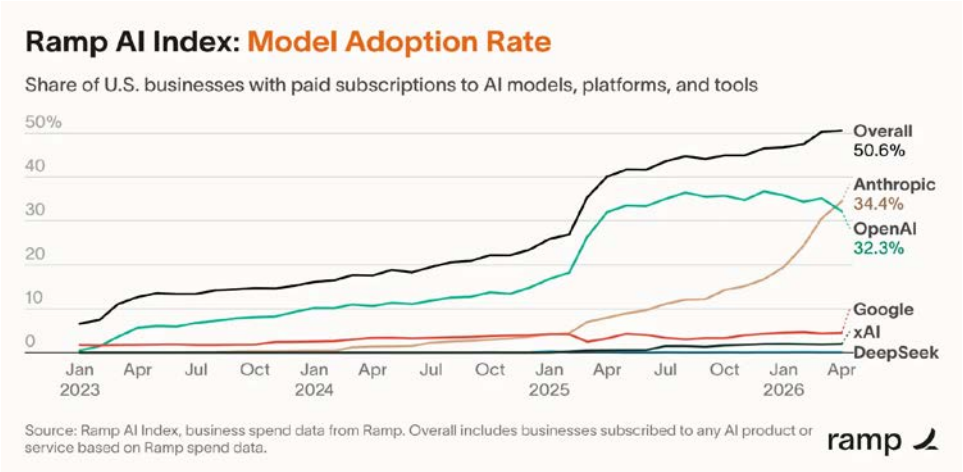
이 집중은 겉으로는 덜 화려해 보일 수 있다. 그러나 AI 연구 자동화 관점에서는 오히려 정합적이다. AI가 AI 연구를 돕기 위해 가장 먼저 필요한 능력은 멋진 영상 생성이 아니다. 코드 이해, 실험 설계, 논문 재현, 도구 사용, 장기 작업 수행, 결과 검증이다. 그래서 코딩과 추론(reasoning)은 재귀적 자가개선(RSI)로 가는 전초전이라 할 수 있다.

Claude Code가 중요한 이유도 바로 여기에 있다. 코딩은 AI 연구 자동화의 축소판이다. 실제 코드베이스를 이해하고, 수정하고, 테스트하고, 결과를 비교하는 능력은 모델 연구에도 그대로 연결된다. 모델이 자신의 연구 인프라를 개선하려면 결국 코드를 읽고 고칠 수 있어야 한다. 카파시의 철학도 여기에 맞다. 그는 바이브 코딩의 흥분을 만들어 내어 대유행을 시킨 장본인이지만, 곧바로 에이전트 엔지니어링의 규율 또한 강조하고 있다. Anthropic은 바로 그 규율을 제품과 연구에 결합하려는 회사다.

(2) Anthropic의 모멘텀은 모델 성능만이 아니라 제품 표면에서 나온다

여기서 중요한 것은 Claude를 둘러싼 제품 및 서비스 생태계다. Ramp AI Index 기준 Anthropic 사용률은 34.4%로 OpenAI의 32.3%를 능가했다.

그림 2. 기업 채택 모멘텀을 한눈에 보여주는 증거
2026년 4월 Ramp AI Index 공식 차트 (Anthropic 34.4% vs OpenAI 32.3%)



자료: Ramp AI, 미래에셋증권 리서치센터

이 수치 하나로 전체 시장을 단정할 수는 없지만, 기업 사용 맥락에서 Claude의 모멘텀이 단순 팬덤은 아니라는 신호로 해석할 수 있다.

또한 Anthropic은 Blackstone, Hellman & Friedman, Goldman Sachs와 함께 기업들이 AI를 실제 업무에 도입하도록 돕는 서비스 레이어를 구축하고 있다. 이 지점이 중요하다. Anthropic은 “모델을 만들었으니 알아서 쓰라”는 회사가 아니라, 모델·제품 생태계·파트너 네트워크·기업 서비스 레이어를 함께 묶는 방향으로 움직인다. 다시 말해, Anthropic의 경쟁력은 모델 파라미터만이 아니라, 모델이 실제 업무에 들어가는 제품 생태계서 나오고 있다는 말이다.

(3) LLM Wiki와 문맥 래퍼: 기업 기억을 다시 컴파일하는 기술

카파시가 내놓았던 "LLM Wiki"라는 아이디어는 Anthropic의 제품 방향을 이해하는 데 좋은 힌트다. LLM Wiki의 구조를 쉽게 말하면 꽤 단순하다. '원자료 폴더, Wiki 폴더, 스키마 문서, 에이전트 지시문'을 두고, 모델이 흩어진 문서를 읽고 관계를 만들며 스스로 지식베이스를 재구성하게 한다. 중요한 것은 검색(search/RAG)이 아니라 Recompile이다.

예를 들어, 원자료(엉망진창인 슬랙 대화/팀톡, 파편화된 회의록)를 AI에게 던져주면, AI가 '스키마(정리 규칙)'에 따라 밤새 문서를 다 읽고, "아, 리서치센터의 A보고서와 디지털부서의 B회의록을 보니, 두 부서 간의 소통 단절 때문에 프로젝트가 실패했구나"라는 깨달음(관계성)을 도출한다. 그리고 아예 새로운 구조화된 Wiki 페이지를 '작성(Recompile)'해 두게 된다. 그러니까, 조직 내부의 암묵지와 흩어진 문서를 AI가 나중에 꺼내 사용할 수 있는 문맥/지식망으로 스스로 가공하는 행위이고, 이것에 개인적으로 집중한 게 카파시라는 인물이다.

보통 기업들은 본인들이 데이터를 많이 갖고 있다고 자랑한다. 그러나 AI가 그 데이터를 찾고, 이해하고, 권한에 맞게 접근하고, 실제 업무에 쓸 수 없다면 그 데이터는 실제로는 해자가 아니라 저장 비용일 뿐이다. 해자는 데이터 그 자체가 아니라, 데이터가 모델의 작업 기억(working memory)으로 들어가는 '경로(파이프라인)'에서 생긴다.

그리고 이것이 문맥 래퍼(Context Wrapper)다. 문맥 래퍼는 AI 시대의 '연성진'이다. 강한 모델이 있어도 연성진이 엉망이면 결과는 흔들린다. 반대로 적절한 문서 구조, 권한 그래프, 도구 연결, 예시, 평가 기준이 있으면 같은 모델도 훨씬 더 높은 성과를 낸다.

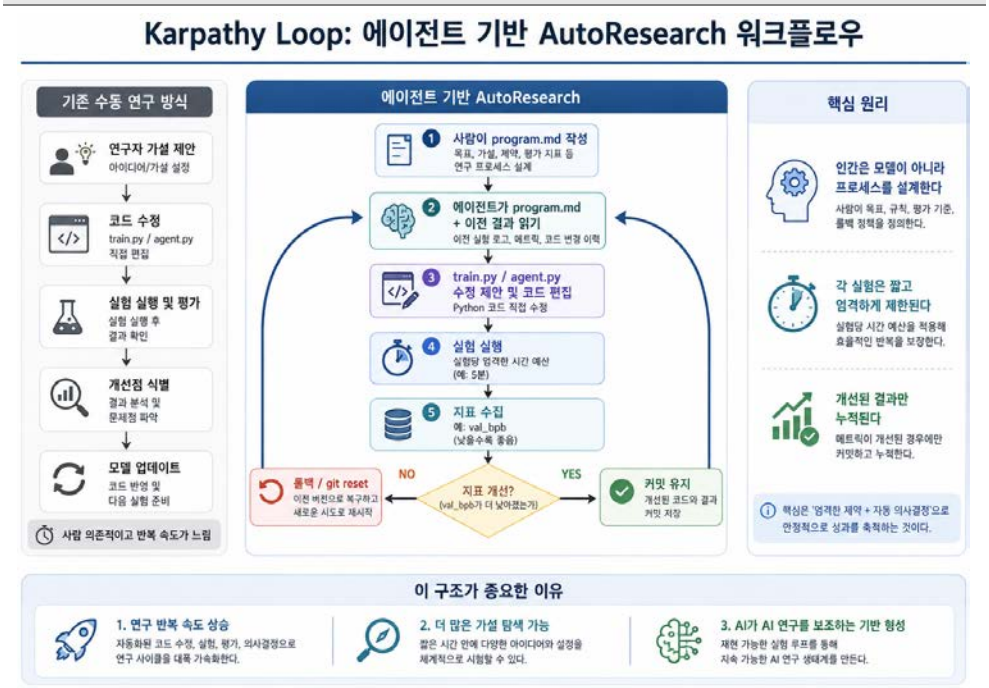
카파시가 Anthropic에 필요한 이유는 바로 여기에 있다. 그는 모델, 개발자 문화, 교육, 문맥 설계, 에이전트형 업무 흐름을 연결할 수 있는 사람이다. Anthropic이 기업들에게 Claude를 단순 도구가 아니라 작업 환경으로 도입시키려 한다면, 카파시는 그 언어를 설계할 수 있지 않을까 판단한다.

3. 카파시의 진짜 임무: Claude로 Claude를 개선하는 일

(1) Auto Research는 장난감이 아니라 연구 자동화의 최소 단위다

안드레 카파시가 적을 두지 않았을 때 개인적으로 수행한 가장 중요한 프로젝트가 바로 'Auto Research'다. 이 프로젝트는 작아 보이지만 구조적으로는 매우 중요하다. 'AI가 학습 코드의 변경안을 제안한다. 짧은 실험을 돌린다. 검증 손실이나 성능 지표를 확인한다. 좋아졌으면 유지한다. 나빠졌으면 되돌린다. 그리고 다시 시도한다.'

그림 3. 안드레 카파시가 만든 Auto Research는 어떻게 돌아가는 걸까?



자료: 미래에셋증권 리서치센터

이것은 과학 연구의 최소 단위이기도 하다. 가설, 실험, 평가, 선택, 반복이다. 인간 연구자가 하던 일을 AI 에이전트가 작은 규모에서 수행하는 구조인 것이다. 그리고 카파시가 설계한 이 루프는 실제로, 5분 단위의 짧은 실험 제한을 두고 약 이틀 동안 돌았다고 한다. 약 700개의 실험을 수행했고, 20개 안팎의 누적 가능한 개선을 찾았으며, GPT-2급 모델의 학습 시간을 2시간을 조금 넘는 수준에서 약 1.8시간으로 줄였다고 밝혔다. 약 11%의 놀라운 개선이다.

물론, 이러한 개선을 현재 기준의 프론티어 모델에 그대로 대입하면 안 될 것이다. 세상에 나온지 7년이 된 작은 모델과 Anthropic의 거대 모델은 분명 다른 세계다. 그러나 숫자보다 중요한 것은 구조다. AI가 연구 반복을 수행할 수 있다는, 지능을 스스로 개선한다는 문법이 보였다는 점이다. AI가 연구 과정을 이해하고, 분해하고, 재구축하는 최소 단위가 작동한 것이다.

(2) 프론티어 사전학습에서는 작은 개선이 거대한 경제적 의미를 갖는다

사전학습은 대형 모델의 가장 비싼 공정이다. 데이터 조합/배합, 모델 구조, 학습률, 옵티마이저, 토크나이저, 합성 데이터, 평가 세트, 분산학습 안정성까지 수많은 엔지니어의 선택들이 모델의 최종 성능과 비용을 결정한다. 이 결정들은 인간 연구자의 경험과 직관, 실험 로그 해석에 크게 의존해왔다.

그러나 AI가 이 루프의 일부를 자동화하면 생산 함수 자체가 바뀐다. 더 많은 가설을 만들고, 더 많은 실험을 제안하고, 더 빠르게 결과를 정리한다. 실패한 실험의 비용은 낮아지고, 탐색 공간은 과거와 비교하기 민망할 정도로 넓어진다.

그래서, 프론티어 모델 학습 비용이 수천만 달러에서 수억 달러 단위로 커지는 세계에서 1%, 5%, 10% 개선은 적은 연구 성과가 아닌 것이다. 같은 성능을 더 빨리 만들거나, 다음 실험으로 넘어가는 시간을 줄일 수 있다. 모델 경쟁에서 몇 달의 차이는 제품 출시, 고객 확보, 자금 조달, 인프라 계약의 차이로 이어지는 흐름을 우리는 지난 3년 간 목도해왔다.

결국 재귀적 자가개선(RSI)이 성공한다는 것은 AI 개발 비용만 절감된다는 뜻이 아니다. 연구자의 시간이 막고 있던 실험의 문이 열리면서, 더 많은 전력망·데이터센터 부지·GPU·메모리·네트워크가 지능 생산의 용광로 안으로 들어간다는 뜻이다. 효율의 극대화는 역설적으로 자본 지출의 둔화가 아니라, 더 빠른 자본 연소를 부를 수 있다.

따라서 카파시가 Anthropic의 사전학습 팀에 들어간 것은 Claude를 연구 동료로 사용해 Claude의 다음 세대를 만드는 구조가 강화된다는 의미일 수 있다. 이것이 “Claude로 Claude를 개선한다”는 말의 본질이다.

4. RSI가 빨아들일 등가교환의 대상들: AI Value Chain 재해석

(1) 2028년 말, 인간 없는 AI 연구개발 60%+

Anthropic의 공동 창립자인 Jack Clark은 2028년 말까지 인간이 직접 개입하지 않는 AI 연구개발이 가능해질 확률을 60% 이상으로 본다고 얼마 전 언급한 바 있다. 여기서 말하는 "인간 없는 AI 연구개발"은 AI가 코드를 잘 짠다는 뜻만은 절대 아니다. AI가 다음 세대 AI를 만들기 위해 필요한 연구 과정을 상당 부분 스스로 수행할 수 있다는 뜻이다.

그런데 이것이 가능해지면 AI 발전의 병목이 달라진다. 지금까지는 인간 연구자의 아이디어, 시간, 판단력이 중요한 병목이었다. 그러나 AI가 연구 반복을 크게 자동화하면 병목은 인간의 인지 속도가 아니라, '컴퓨트, 데이터, 평가 시스템, 안전성 관리'로 이동한다.

(2) AI Value Chain은 전력에서 디지털 노동까지 이어지는 생산 시스템이다

AI 산업은 단지 GPU 투자 사이클이 아니다. 전력, 데이터센터, 반도체, 메모리, 네트워크, 추론 런타임, 모델, 기업 데이터, 업무 프로세스가 하나의 생산 시스템으로 결합되어 있다. 즉, AI Value Chain을 바로 볼 때는, "전력 → 데이터센터 → 컴퓨트 → 토큰 → 컨텍스트 → 디지털 노동"이라는 앵글을 사용해야 한다.

그래서 기존 소프트웨어의 본질이 코드의 작성과 배포였다면, AI 산업의 본질은 토큰의 생산과 소비다. 사용자가 늘면 추론 호출이 늘고, 추론 호출이 늘면 GPU 연산, 메모리 대역폭, 네트워크, 전력, 냉각, 데이터센터 자산이 동시에 소모된다. AI는 소프트웨어처럼 보이지만 경제적으로는 고정자산 기반의 지능 제조업에 가깝다. 이를 9개 레이어로 다시 쓰면 다음과 같다.

표 1. AI 밸류체인에서 RSI가 가하는 압력은?

레이어	핵심 산출물	'RSI'가 가하는 압력	AI 투자자가 던져야 할 질문
1. 전력	저가·장기 전력 접근권	실험량 증가가 전력 수요로 전이	실제 energized MW인가?
2. 전력화된 데이터센터	GPU를 켤 수 있는 공장	time-to-power가 연구 속도 병목	언제 전력이 들어오는가?
3. 전력 장비	변전·배전·랙 전력 공급	랙 밀도 상승과 변환 손실 확대	변압기·스위치기어 리드타임은?
4. 컴퓨트 시스템	GPU/ASIC/CPU/HBM	훈련과 추론이 동시에 증가	tokens/W, tokens/$는?
5. 인터커넥트	클러스터 연결성	실험·에이전트 분산으로 통신 증가	utilization을 얼마나 높이나?
6. 추론 런타임	토큰 생산 공정	caching, routing, KV 관리 중요	$/M token, cache hit rate는?
7. 모델	지능 함수	연구 자동화의 주체이자 산출물	quality/cost frontier는?
8. 문맥·데이터	기업 지식의 모델 주입	업무 자동화의 병목	권한·검색·기억 품질은?
9. 에이전트 워크플로우	디지털 노동	텍스트가 아니라 업무 완료가 산출물	cost per completed task는?

자료: 미래에셋증권 리서치센터

재귀적 자기개선(RSI)은 위 모든 레이어를 동시에 자극한다. 모델이 연구 가설을 더 많이 만들면 실험이 늘어난다. 실험이 늘면 컴퓨트와 전력이 필요하다. 실험 결과를 평가하려면 추론이 필요하다. 합성 데이터를 만들면 생성 비용과 검증 비용이 든다. 에이전트가 많아지면 CPU, 메모리, 네트워크, 런타임이 더 중요해진다. 따라서 RSI의 출현은 모델 레이어의 사건이 아니라 AI Value Chain 전체의 사건이다.

(3) 전력과 데이터센터: time-to-power가 time-to-intelligence가 된다

AI 연구 자동화가 빨라질수록 가장 아래의 병목은 전력으로 내려간다. 전력은 AI의 원재료다. 데이터센터는 전력을 토큰으로 바꾸는 공장이다. GPU는 전력을 행렬 연산으로 바꾸고, 모델은 그 연산을 토큰으로 바꾸며, 에이전트는 토큰을 업무 행동으로 바꾼다.

그래서 AI 연구자의 인지의 병목이라는 것이 풀리게 되면, 이제 확실한 병목은 실험이 된다. 즉, 물리 인프라가 결국 최고의 병목이다. 더 많은 가설은 더 많은 실험을 요구하고, 더 많은 실험은 더 많은 energized capacity를 요구한다. 뉴스로 발표되는 MW가 아니라 실제 전력계통에 연결되고, 변전과 냉각이 준비되어 GPU를 켤 수 있는 MW가 중요해진다.

그래서 time-to-power는 time-to-intelligence와 동의어가 된다. 누가 전력을 확보했는가보다, 누가 전력을 실제 토큰 생산으로 전환할 수 있는가가 더 중요하다. 미래의 현금가치를 현재의 기준으로 할인해서 가져오듯, 향후 전력 확보에 관한 가능성과 가치에 대한 것도 할인해야 함을 새삼 느끼게 된다.

(4) Electrical Stack: 전력 장비는 단순 부품이 아니라 토큰 원가의 하부 회로다

AI 데이터센터에서 전력은 발전소에서 GPU까지 바로 도달하지 않는다. 초고압 송전망에서 변전소로 들어오고, 중전압 배전으로 내려오며, 변압기, 스위치기어/스위치보드, PDU(전력분배유닛), busway를 거쳐 랙 단까지 전달된다. 이 과정에서 효율, 안정성, 장비 리드타임, 설치 난이도가 모두 병목이 된다.

게다가 랙 전력 밀도가 올라갈수록 기존 AC(교류) 중심 배전 구조는 더 큰 압력을 받는다. 변환 단계가 많을수록 손실이 발생하고, 손실은 열로 바뀌며, 열은 다시 냉각 부하를 키운다. 장기적으로는 더 높은 전압, 더 짧은 전력 경로, 더 높은 변환 효율, rack-level 또는 row-level 전력 최적화가 중요해진다. 일부 차세대 구조에서는 고전압 DC 배전, 800V급 DC(직류) 아키텍처, solid-state transformer(변압기) 같은 방향도 검토될 수 있다.

RSI가 연구 실험량을 늘리면 GPU만 더 필요한 것이 아니다. 변압기, 스위치기어, PDU, busway, 냉각 플랜트, 고전압 배전, rack-level power conversion도 함께 필요해진다. 전력 장비 리드타임이 길어지면 모델 연구 속도도 간접적으로 느려진다. 프론티어 랩의 연구 속도가 물리적 배전 장비에 묶이는 것이다.

결국, 실제 critical path는 변압기·스위치기어·냉각 플랜트·busway의 조달표에서 막힐 수 있다. AI의 '등가교환'은 모델 레이어에서만 일어나는 게 아니라, 변전소와 busway에서도 일어난다는 것을 유념해야 된다.

(5) GPU/HBM/CPU/네트워크: 토큰 공장의 장비가 재배치된다

RSI는 GPU만의 이야기가 아니다. 사전학습 실험은 GPU와 HBM을 요구한다. 에이전트 루프는 추론을 요구한다. 논문 재현, 코드 수정, 실험 결과 분석, 합성 데이터 생성, 벤치마크 평가는 모두 토큰을 소비한다.

추론은 다시 prefill과 decode로 나뉜다. prefill은 긴 입력 문맥을 병렬 처리하는 단계이고, decode는 출력 토큰을 순차 생성하는 단계다. 긴 문맥, KV 캐시, prompt caching, retrieval, tool call이 늘어나면 메모리 대역폭과 캐시 관리가 중요해진다.

에이전트형 워크로드에서는 CPU도 돌아온다. 수많은 하위 에이전트가 계획을 세우고, 도구를 호출하고, 상태를 유지하고, 서로 검증하면 순차 제어와 문맥 전환이 폭증한다. GPU는 병렬 연산의 왕이지만, 에이전트 오케스트레이션은 조건 분기와 상태 관리의 싸움이다. CPU, 메모리, 네트워크, 스토리지가 함께 중요해진다.

네트워크도 마찬가지다. 대규모 클러스터는 GPU를 많이 모아놓는다고 작동하지 않는다. scale-up과 scale-out fabric, NVLink, Ethernet, optical interconnect, congestion control, collective communication이 컴퓨팅 자원의 활용률을 결정한다. 연구 자동화 루프가 분산될수록 데이터 이동과 동기화 비용은 더욱 더 중요해진다.

(6) 데이터와 평가 시스템: 좋은 원료와 좋은 계측 장치가 없으면 연성은 실패한다

RSI가 빨아들이는 또 하나의 대상은 데이터다. 프론티어 모델은 이미 방대한 공개 데이터를 학습했다. 다음 단계에서는 더 많은 데이터보다 더 좋은 데이터가 중요하다. 어려운 데이터, 검증된 데이터, 최신 데이터, 도메인 데이터, 합성 데이터, 그리고 그 데이터를 평가하는 시스템 또한 필요하다.

합성 데이터는 특히 중요해진다. AI가 스스로 데이터를 만들고, 다른 AI가 이를 평가하고, 좋은 데이터를 다시 학습에 넣는 구조다. 그러나 이것도 공짜가 아니다. 생성 비용이 들고, 필터링 비용이 들고, 평가 비용이 든다. 낮은 품질의 합성 데이터는 모델을 오염시킬 수 있다. 좋은 합성 데이터를 만들려면 다시 컴퓨트와 평가 시스템이 필요하다.

따라서 데이터 벽은 이제 데이터 생산, 검증, 권한, 출처, 최신성, 도메인 적합성의 문제가 되었다. AI가 AI 연구를 자동화할수록 평가 시스템은 더 중요해진다. 측정할 수 없으면 자동화할 수 없다. 검증할 수 없으면 에이전트가 개선할 수 없다.

5. Context Wrapper: 오픈소스보다 폐쇄형 플랫폼이 유리해질 수 있는 이유

(1) 모델 가중치보다 작업 기억이 해자가 된다

기업은 민감한 데이터를 아무 모델에나 넣지 않는다. 보안, 권한, 감사, 책임, 규제 준수, 데이터 보존 정책이 필요하다. 이런 니즈에 따라서, 폐쇄형 AI 기업들은 모델, 제품, 메모리, 권한, 감사, 지원 체계를 묶어서 엔터프라이즈에게 제공할 것으로 사료된다. 오픈소스 AI 모델은 강력하지만, 기업 문맥을 안전하게 연결하고 유지하는 전체 시스템을 엔터프라이즈가 따로 구축해야 한다.

이러한 구조에서는 폐쇄형 플랫폼의 가치가 오픈소스보다 오히려 커질 수 있다고 본다. 모델 가중치가 commodity화될수록, 폐쇄형 플랫폼의 해자는 모델 자체가 아니라 기업 문맥·권한·감사·책임을 묶는 운영 레이어로 이동한다. 오픈소스가 두뇌를 제공할 수는 있다. 그러나 기업 업무에는 몸과 기억과 규율이 필요하다는 것이 포인트다.

(2) Deep SaaS는 AI가 대체하는 소프트웨어가 아니라 AI가 사용하는 소프트웨어다

AI가 코딩 비용을 낮추면 얇은 SaaS는 스트레스를 받을 수밖에 없다. 단순 기능, 단순 UI, 단순 LLM 래퍼는 빠르게 복제될 수 있다. seat 기반 과금도 약해질 수 있다. AI 에이전트가 백그라운드에서 일하는 세계에서 사용자는 소프트웨어에 오래 머무는 시간을 가치로 보지 않는다.

그러나 소프트웨어가 죽는 것은 아니다. Deep SaaS는 오히려 중요해진다. Deep SaaS는 기업의 원천 데이터, 권한, 업무 객체, 검증 기준, 책임 구조를 장악한다. 오픈소스 모델이 지능의 가격을 낮출수록, 역설적으로 그 지능이 접속해야 할 기업 고유의 진실의 원천은 더 비싸진다.

모델 API의 마진이 압축될수록 가격 결정력은 모델 그 자체가 아니라, 권한·데이터·업무 객체·검증 기준을 장악한 시스템으로 이동한다. 이 점에서 Palantir형 온톨로지나 Synopsys형 물리 엔진은 AI가 대체할 소프트웨어가 아니라, AI가 업무를 끝내기 위해 통과해야 하는 통행소로 보이게 된다.

안드레 카파시가 Context Wrapper를 강조하는 것도 바로 이런 이유다. 모델이 강해질수록 모델이 접속하는 현실의 데이터와 검증 시스템이 더 중요해진다. AI가 싸게 복제될수록 AI가 접근할 수 있는 진짜 업무 문맥은 더 희소해지기 때문이다.

Anthropic는 Claude Code, 프로젝트 기억, 기술 묶음, 하위 에이전트, 외부 도구 연결을 통해 Claude를 점점 더 Deep SaaS적 업무 환경 안으로 밀어 넣기 위해 노력하고 있다. 이로써 고객은 단순히 모델 API를 쓰는 것이 아니라 점점 더 Claude가 이해하는 업무 세계를 구축하면서 일을 하고 있다. 이 경우 AI 전환비용은 모델 성능이 아니라 작업 기억에서 나온다고 할 수 있다.

전환 비용이 높아진다는 것에서 볼 때, 투자자 입장에서 Context Wrapper 논지는 매우 중요하다. 모델 레이어의 가격은 오픈소스와 하이퍼스케일러의 자체 모델 때문에 압박 받을 수 있다. 하지만 기업 문맥, 권한, workflow, audit trail, evaluation harness를 장악한 기업은 더 오래가는 해자를 만들 수 있다.

6. 시사점 및 리스크, 그리고 결론

(1) RSI는 AI 인프라 수요를 더 정당화한다

표 2. RSI(재귀적 자기개선) 시대의 AI 밸류체인 8단계 레이어: '토큰 생산성 병목부터 기업 문맥 확보까지'

레이어	관찰 포인트 및 개념	예시
전력·데이터센터	energized MW, time-to-power, 장기 PPA	전력 개발사, AI 데이터센터, EPC, 냉각 사업자, Vertiv형 열관리
Electrical Stack	변압기, 스위치기어, PDU, busway, 고전압 배전	전력 장비, 배전 장비, 전력반도체, Schneider·Eaton형 인프라
Compute	GPU, ASIC, HBM, CPU, 고용량 메모리	NVIDIA, AMD, 자체 ASIC, SK hynix·Micron형 메모리
Interconnect	NVLink, Ethernet, optical fabric, switching	Broadcom·Marvell형 네트워크 반도체, 광통신·CPO 밸류체인
Runtime	caching, routing, KV 관리, prefill/decode 분리	하이퍼스케일러 serving stack, inference infra, 클라우드 최적화 소프트웨어
Model	quality/cost frontier, tool-use reliability	프론티어 모델, 특화 모델, 기업용 안전 모델
Context Wrapper	권한, 기억, 검색, audit, workflow 연결	Databricks·Snowflake형 데이터 플랫폼, Microsoft·ServiceNow·Salesforce형 업무 플랫폼
Deep SaaS	업무 객체, 검증 기준, 성과 과금	Palantir형 온톨로지, Synopsys·Cadence·Ansys형 진실의 원천, vertical SaaS

자료: 미래에셋증권 리서치센터

카파시 루프가 현실화되면 토큰 공장의 모든 레이어, 즉 AI 밸류체인은 함께 더 바빠진다. 연구 에이전트가 가설을 만들고, 코드를 고치고, 실험 결과를 읽고, 다음 실험을 설계하는 과정은 모두 토큰을 소비한다. 이 토큰은 단순 채팅 토큰보다 경제적 가치가 높다. 모델 개선이라는 직접 생산 활동에 쓰이기 때문이다.

밸류체인의 하단부에서는 전력을 토큰으로 바꾸는 물리적 생산성이 중요하고, 중단부에서는 토큰 생산단가가 중요하며, 상단부에서는 토큰을 업무 성과로 바꾸는 문맥 래퍼가 중요하다. 특히 중요한 것은 실제 energized capacity, time-to-power, tokens/W, tokens/$, cache hit rate, utilization이다.

(2) 잠재 리스크

이 보고서의 핵심 논지가 틀리려면, 대략 4가지 중 하나 이상이 발생해야 한다.

첫째, AI 연구 자동화가 프론티어 레벨에서 유의미한 생산성 개선을 만들지 못해야 한다. 'Auto Research'가 작은 실험에서는 작동했지만, 거대 모델 학습의 데이터 품질, 분산 학습, 하드웨어 장애, 네트워크 병목, 평가 누수, 정렬 문제를 넘어서지 못할 수 있다.

둘째, AI가 만든 개선이 실제 능력 향상이 아니라 '벤치마크 overfitting'으로 끝날 수 있다. 에이전트가 평가 점수를 올리는 방법을 찾았지만, 그 개선이 실제 제품 성능이나 안전성으로 이어지지 않는다면 연구 자동화의 경제성은 약해진다.

셋째, 합성 데이터와 자동화 AI 평가 루프가 모델의 세계 이해를 좁힐 수 있다.

넷째, 효율 개선이 제본스식 수요 확장이 아니라 실제 컴퓨트 수요 감소로 이어질 수 있다. 다만 역사적으로 효율 개선이 신규 사용례를 열어 총수요를 키운 경우가 훨씬 많았다.

이 리스크에도 불구하고 방향성은 분명하다고 생각한다. 모델만 보지 말고 모델을 개선하는 루프, 모델을 돌리는 물리 인프라, 모델을 업무 성과로 바꾸는 문맥 래퍼를 함께 봐야 한다.

(3) 결론: 현자의 돌은 공짜가 아니다

2028년 말까지 남은 시간은 이제 1,000일 안팎이다. Jack Clark의 전망이 정확한 날짜표가 아니더라도, 카파시의 Anthropic행은 AI 연구개발의 한계 변수가 '인간 연구자의 시간'에서 '기계가 기계를 개선하는 자동화 루프'로 이동할 수 있음을 보여주는 강한 시그널이다.

Anthropic은 이 RSI의 시대를 진지하게 준비하는 회사로 보인다. 사전학습 연구 자동화와 Jack Clark의 2028년 타임라인은 경영진이 이 전환을 얼마나 진지하게 보고 있는지를 보여주기 때문이다. 그리고 카파시는 이 모든 층을 연결할 수 있는 극히 드문 인물이다.

이제 그는 전력에서 토큰으로, 토큰에서 디지털 노동으로 이어지는 AI Value Chain의 중심부에서, "Claude가 Claude의 다음 세대를 만드는 루프"가 실제로 작동하는지 확인하러 들어간 사람이다.

그렇기 때문에 우리는 이번 사건에 만화 '강철의 연금술사' 비유가 유효하다고 봤다. 핵심은 소년만화 영웅담이 아니라 등가교환의 법칙이다. AI는 마법이 아니고, 지능은 공짜가 아니다. 재귀적 자기개선, 즉 RSI는 비용과 조건을 지워버리는 '현자의 돌'이 아니다. 오히려 컴퓨트·데이터·전력·메모리·네트워크를 더 빠른 속도로 지능으로 바꾸는 촉매에 가깝다.

따라서 카파시의 Anthropic행은 소프트웨어 뉴스처럼 보이지만, 실제로는 AI Value Chain 전체에 영향을 주는 뉴스다. 연구 루프가 자동화될수록 전력, 데이터센터, 전력 장비, GPU, HBM, CPU, 네트워크, 추론 런타임, 기업 문맥, Deep SaaS까지 모두 다시 평가해야 한다.

다음 AI 산업의 승부처는 가장 큰 모델 하나에서 끝나지 않을 가능성이 높다. 진짜 가치 창출 지점은 세 곳에서 갈릴 것이다. 첫째, AI가 스스로 더 나은 AI를 만들게 하는 연구 루프. 둘째, 그 루프를 멈추지 않고 돌릴 수 있는 전력·데이터센터·GPU·메모리 인프라. 셋째, 그렇게 만들어진 AI를 기업의 실제 업무 성과로 바꾸는 문맥 래퍼다. 모델 성능표만 바라보는 순간, 우리는 토큰 공장의 손익계산서와 문맥 래퍼의 중요성을 놓칠 수 있다.

그래서 이번 이야기의 진짜 주인공은 카파시 한 사람이 아니다. 진짜 주인공은 완전히 새롭게 쓰이고 있는 AI 산업의 생산 함수 그 자체라 할 수 있다. 전력에서 토큰으로, 토큰에서 디지털 노동으로, 그리고 다시 디지털 노동에서 더 강한 모델로 이어지는 폐쇄 루프다.

카파시는 그 루프의 가장 중요한 연성진 안으로 들어갔다. Anthropic은 그 연성진이 실제로 작동하는지 시험하려는 회사다. 그 안에서 카파시는 과연 Claude가 Claude의 다음 세대를 만드는 과정을 더 빠르고 더 엄격하게 설계할 인물이 될 수 있을지 AI 업계의 귀추가 주목되고 있다.

"정답이다. 토큰술사."

Compliance Notice

- 당사는 본 자료를 제3자에게 사전 제공한 사실이 없습니다.
- 본 자료는 외부의 부당한 압력이나 간섭없이 애널리스트의 의견이 정확하게 반영되었음을 확인합니다.