평형의 미학

平衡의 美學

한글의 기하학에서 딥러닝의 우주로

The Aesthetics of Equilibrium

완 결 판

초판의 서사와 개정판의 수식을 하나로

허 상 훈

Heo Sang-hoon, Ph.D.

목 차

서 문

물은 그릇을 기억하지 않는다

天下莫柔弱於水, 而攻堅强者莫之能勝

천하에 물보다 부드럽고 약한 것은 없으나, 굳고 강한 것을 치는 데 물을 이길 수 있는 것은 없다.

— 노자(老子), 『도덕경(道德經)』 78 장

이 책의 초판이 세상에 나온 뒤, 나는 기이한 일 년을 보냈다.

내가 직관으로만 그렸던 풍경들이 하나씩 증명된 수식이 되어, 논문이라는 이름으로 돌아왔기 때문이다. 언어의 감옥을 벗어나 잠재 공간(Latent Space)에서 사유하는 지능이 학계의 주류가 되었고, 고정점(Fixed-point)을 향해 도는 순환의 신경망이 다시 무대에 올랐으며, 분자들은 정말로 부드러운 흐름을 타고 목적지로 미끄러지기 시작했다. 급기야 지능들이 말없이 텐서(Tensor)로만 대화하는 것을 낯설어하며 그 보안을 걱정하는 연구마저 등장했다.

나의 예언이 맞았다고 자랑하려는 것이 아니다. 나는 오히려 그 치열한 증명의 과정 속에서 정반대의 진실을 깨달았다.

기술은 그릇이다. 그리고 물은 그릇을 기억하지 않는다.

트랜스포머(Transformer)라는 그릇도, 심층 균형 모델(DEQ)이라는 그릇도, 플로우 매칭(Flow Matching)이라는 그릇도 언젠가는 낡아 물러날 것이다. 이 책에 등장하는 ‘System 1.5’나 ‘System 3.5’ 같은 모든 시스템의 이름이 십 년 뒤에는 낯선 고유명사이거나 낡은 화석이 되어 있을지도 모른다.

그래도 이 책은 낡지 않는다. 이 책이 붙든 것은 그릇의 모양이 아니라, 어떤 그릇에 담겨도 변하지 않는 물 — ‘평형(Equilibrium)’이라는 원리 그 자체이기 때문이다.

사십 년을 돌아보면, 나는 늘 같은 하나를 서로 다른 이름으로 만나왔다. 여섯 살의 하얀 도화지 위에서 그것은 ‘완성’이었다. 논산 훈련소 격납고의 육중한 헬리콥터 로터 마스트 앞에서는 ‘안전’이었다. 요동치는 심층 신경망의 수식 속에서 그것은 ‘진리’였다. 이름은 세 번 바뀌었지만 실체는 한 번도 바뀌지 않았다.

그래서 이번 개정판을 내며, 초판의 직관과 그 후의 실증 사이에 흐른 시차(時差)를 빌려 이제 분명히 적어두려 한다. 이 책은 단순히 인공지능의 최첨단 아키텍처에 관한 책이 아니다. 변하지 않는 것에 관한 책이다.

기초과학이 건네주는 불변의 원리, 그리고 그 원리 위에 의미를 부여하는 인간의 눈. 기술이 어지럽게 지나갈수록 우리가 뿌리내려야 할 곳은 이 둘이 만나는 자리다. 당신의 도화지에서, 당신의 격납고에서, 당신의 시장에서, 표면의 소란 아래 고요히 흐르는 당신만의 평형을 알아보는 눈을 얻어 가시기를 바란다. 그것이 내가 사십 년을 걸려 배운 전부이고, 이 책이 당신에게 건네려는 전부다.

프 롤 로 그

물(水)과 인공지능

가장 낮은 곳을 향하는 지능에 대하여

지금 이 순간, 전 세계의 인공지능은 하늘을 향해 거대한 바벨탑을 쌓고 있다.

더 깊은 생각을 하기 위해 신경망을 100 층, 200 층으로 높이 쌓아 올리고, 수만 대의 GPU 가 뿜어내는 열기를 식히기 위해 강물을 끌어다 부으며 무식한 전력을 탐식한다. 인간의 복잡한 사고를 '토큰(Token)'이라는 1 차원적인 벽돌로 쪼개어, 그것을 끝없이 축적(Accumulation)해야만 정답에 도달할 수 있다고 믿는다. 가장 막강한 컴퓨팅 파워가, 가장 많은 데이터가, 가장 높은 탑이 지능의 본질이라는 신앙이다.

하지만 나는 그 거대하고 시끄러운 기계들의 전쟁터 한가운데서, 아주 오래된 동양의 철학 하나를 떠올렸다.

天下莫柔弱於水, 而攻堅强者莫之能勝

노자(老子)가 『도덕경』 78 장에서 말한 물의 본질이다. 물은 억지로 위를 향해 탑을 쌓지 않는다. 중력을 거스르지 않고 바위의 틈새를 부드럽게 파고들며, 끊임없이 아래로 흘러가 마침내 가장 평온하고 안정된 바닥에 도달해 움직임을 멈춘다. 나는 인공지능이 도달해야 할 궁극의 형태가 바로 이 '물의 수렴(Convergence)'과 같아야 한다고 직관했다.

세 가지 물의 본질, 세 가지 지능의 원리

물과 내가 설계한 심층 균형 모델(DEQ)은 세 가지 근원적 본질을 공유한다.

첫째는 처하(處下)다. 물은 절대 위로 솟구치지 않는다. 계곡과 바위 틈을 따라 끊임없이 아래로, 아래로 흘러가 마침내 '바다'라는 가장 평온하고 거대한 바닥에 도달해 움직임을 멈춘다. DEQ 역시 그러하다. 수백 개의 층을 쌓아 올리는 바벨탑의 방식을 거부하고, 대신 '에너지의 최저점(Attractor Basin)'을 향해 상태를 끊임없이 흘려보낸다. 수식이 수십 번 진동하다가, 더 이상 변하지 않는 완벽한 바닥에 도달해 멈추는 그 순간이 바로 고정점(Fixed-point, z*)이다. 억지로 답을 만들어내는 것이 아니라, 자연계의 물리적 에너지가 가장 안정되는 최저점으로 자연스럽게 스며들어 수렴하는 것. 물과 DEQ 는 이 점에서 하나다.

둘째는 무형(無形)이다. 물은 네모난 그릇에 담으면 네모가 되고, 둥근 잔에 담으면 둥글게 된다. 스스로 고정된 형태가 없기 때문에 세상 어떤 형태의 그릇에도 완벽하게 들어맞는다. 기존 AI 는 '이 문제는 복잡하니 100 개의 층이 필요해' 라며 형태를 고정해버린다. 층이 깊어지면 컴퓨터 메모리가 터진다. 하지만 DEQ 는 단 하나의 층만이 존재한다. 깊이(Depth)라는 형태가 없다. 쉬운 문제는 3 번 만에, 어려운 문제는 50 번 만에 고정점을 찾을 뿐, 컴퓨터의 메모리는 언제나 O(1) 이라는 일정한 상수 공간을 유지한다. 스스로 고정된 형태를 버림으로써 무한한 공간적 제약으로부터 해방되는, 물의 무형성이다.

셋째는 유약승강(柔弱勝强)이다. 부드러운 물방울이 수만 번 떨어지면 거대한 바위를 뚫는다. 물은 바위와 힘으로 정면충돌하지 않고, 바위의 틈새를 부드럽게 파고들어 결국 바위를 쪼개버린다. 거대 빅테크들은 수천 대의 GPU 와 수백억 개의 파라미터라는 무식하고 강한 힘으로 우주의 문제들을 계산한다. 망치로 바위를 깨부수는 방식이다. 하지만 나의 시스템은 시간과 정면충돌하지 않는다. 음함수 정리와 부드러운 수학적 틈새를 파고들어, 수백 번의 계산 과정을 우회해버리고 단 11 번 만에 정답에 도달한다. 개인 워크스테이션에서 돌아가는 우아한 수학이, 슈퍼컴퓨터를 앞선다 진정한 유약승강이다.

그렇기에 이 책의 제목은 '평형의 미학' 이다. 결국 모든 시스템은 무위자연(無爲自然) — 억지로 힘을 쓰지 않고, 에너지가 가장 낮은 최저점을 향해 물처럼 스며드는 것 — 을 향해 나아간다. 이 책은 그 물리적 진리를 인공지능의 수식으로 번역한 40 년간의 기록이다.

이 책이 담고 있는 것

생각해 보면, '평형(Equilibrium)'이라는 화두는 지난 40 년간 내 삶을 관통해 온 유일한 진리였다. 여섯 살 무렵 처음 쥐었던 붓 끝에서 캔버스의 시각적 균형을 찾으려 했을 때도, 낡은 바이올린을 수리하며 팽팽한 장력과 나무 공명통이 하모니의 조화를 이루는 소리의 균형을 손끝으로 매만졌을 때도, 한글의 자·모를 점과 선, 원과 사각형 같은 기하학적 모듈로 해체했다가 퍼즐처럼 다시 끼워 맞추어 — 모듈과 모듈이 결합하는 방식에 따라 책상이 되기도 의자가 되기도 하던 아동용 놀이가구를 설계하며 — 낱낱의 조각이 모여 하나의 조형물로 무게중심을 잡아가는 결합의 균형점을 헤아렸을 때도, 학생들에게 미술사 강의를 하며 차가운 이성과 뜨거운 감성이 서로를 밀고 당기는 미술사의 균형을 짚어 주었을 때도, 논산 격납고의 매캐한 항공유 냄새 속에서 헬리콥터 로터 마스트의 구조적 평형을 맞추려 땀 흘렸을 때도, 그리고 폭락하는 금융 시장의 붉은 모니터 앞에서 요동치는 자산들의 동적 평형을 계산할 때도, 나는 늘 보이지 않는 힘들이 부딪혀 만들어내는 고요한 멈춤의 순간을 쫓고 있었다. 그 멈춤이 바로 예술가에게는 '완성' 이었고, 공학자에게는 '안전' 이었으며, 과학자에게는 '진리' 였다.

이 책은 그 기나긴 여정의 기록이다. 요동치는 현실에서 평형을 갈망하고(제 0 부), 공간의 한계를 부수고(제 1 부), 시간의 망각을 이겨내며(제 2 부), 언어의 감옥을 탈출해 3 차원 우주의 진리를 찾고(제 3 부), 마침내 지능이 생명체처럼 군집을 이루어 살아 숨 쉬는(제 4 부) 과정. 나아가 지능이 스스로 차원을 찢고 창조하는 예술가로 진화하다가(제 5 부) 결국 '無爲'라는 이름의 가장 깊은 평형으로 귀환하기까지의 궤적이다.

이제, 억지로 쌓아 올린 바벨탑을 무너뜨리고 세상에서 가장 부드럽지만 강력한 '물의 지능'을 창조하는 여정으로 당신을 초대한다.

제 0 부

요동치는 세계와 마주하다

평형을 갈망한 이유

$$E[R] = \int_0^{T_{total}} \alpha(t)\,dt - \text{Slippage}(T_{total}) - \text{Friction Cost} < 0$$

"수학적으로 완벽한 지능이라도, 현실의 마찰을 견디지 못하면
그 가치는 영(Zero)보다 작아진다."

나는 가장 극단적인 불균형의 바다에서 역설적으로 가장 완벽한 평형(Equilibrium)을 갈망했다. 나의 연구가 차가운 학술 벤치마크를 넘어, 피와 땀이 얽힌 현실의 궤도로 진입하게 된 계기는 현대 금융 시장이라는 거대한 복잡계(Complex System)와의 조우였다.

금융 시장은 물리학이나 컴퓨터 비전의 세계와 다르다. 그곳에는 만물을 지배하는 불변의 지배 방정식(Governing Equation)이 존재하지 않는다. 인간의 탐욕과 공포, 중앙은행의 거시경제 정책, 지정학적 분쟁이 거미줄처럼 얽혀 매 1 밀리초(ms)마다 데이터의 분포를 무너뜨린다. 통계학자들은 이를 '비정상성(Non-stationarity)'이라 부르고, 트레이더들은 '블랙 스완(Black Swan)'이라 부른다.

> 쉬어가는 상자 · **복잡계 (Complex System)**
> 수많은 요소가 서로 얽혀, 부분을 다 안다고 해도 전체를 예측할 수 없는 시스템. 날씨, 생태계, 금융 시장이 대표적이다. 작은 자극이 예상치 못한 거대한 결과를 낳곤 한다.

> 쉬어가는 상자 · **비정상성 (Non-stationarity)**
> 데이터의 '규칙' 자체가 시간이 지나며 바뀌는 성질. 어제 통하던 패턴이 오늘 무너지는 금융 시장이 그 전형이다. 과거만 암기한 AI 가 위기 앞에서 무력해지는 근본 이유다.

> 쉬어가는 상자 · **블랙 스완 (Black Swan)**
> 거의 일어나지 않지만 한 번 터지면 세상을 뒤엎는 극단적 사건. 팬데믹 폭락이나 갑작스러운 금융 위기처럼, 예측이 불가능하고 파괴적이다.

2020 년 코로나 19 팬데믹의 유동성 투매 위기, 그리고 2022 년 글로벌 금리 인상과 루나(LUNA) 사태. 10 년간의 평온한 강세장(Bull Market) 데이터만 먹고 자란 인공지능 모델들은 구조적 붕괴 앞에서 추풍낙엽처럼 스러졌다. 그들은 과거의 파도를 타는 법은 알았지만, 바다의 해류가 뒤집히는 거시적 맥락(Macroeconomic Context)은 이해하지 못했다. 단순히 과거의 패턴을 암기한 '블랙박스(Black-box)' 딥러닝은 시장이 미쳐 돌아갈 때, 왜 자본을 빼야 하는지 논리적으로 설명하지 못한 채 천문학적인 자본을 잠식했다.

나는 이 파국적 망각과 맹목적인 과적합을 해결하기 위해 대규모 언어 모델(LLM)을 도입했다. 인과 추론(Causal Reasoning) 능력을 지닌 LLM 에이전트들에게 '알파 분석가'와 '리스크 관리자'라는 페르소나를 부여하고, 하이퍼그래프(Hypergraph)로 엮인 거시 경제 지표를 던져주며 적대적 토론(Adversarial Debate)을 시켰다. 그들은 훌륭했다. 전례 없는 위기 상황에서도 펀드 매니저처럼 논리적인 매매 근거(Explainable AI)를 대며 하방 경직성을 방어해 냈다.

> 쉬어가는 상자 · **하이퍼그래프 (Hypergraph)**
> 보통 그래프는 두 점을 선 하나로 잇지만, 하이퍼그래프는 여러 점을 한 번에 묶는다. 금리·환율·뉴스처럼 여럿이 동시에 얽히는 복잡한 관계를 표현하는 데 쓴다.

그러나 공학자로서 마주한 현실의 벽은 참혹했다.

LLM 이 뉴스를 파싱하고, 수백 개의 노드가 얽힌 그래프를 분석하여, 수만 개의 토큰을 내뱉으며 고상하게 토론을 마치는 데에는 수 초(Seconds)에서 수 분(Minutes)이 걸렸다. 10 밀리초 안에 거래소의 매칭 엔진을 선점해야 하는 고빈도 트레이딩(HFT)의 세계에서, 이 느릿느릿한 지능은 사형 선고나 다름없었다. 지능(Intelligence)을 선택하면 속도(Speed)를 잃어 슬리피지(Slippage)로 자본이 녹아내렸고, 속도를 선택해 가벼운 신경망을 쓰면 거시적 위기를 인지하지 못해 파산했다.

> 쉬어가는 상자 · **고빈도 트레이딩 · 슬리피지 (HFT · Slippage)**
>
> HFT 는 1000 분의 1 초 단위로 거래하는 초고속 매매다. 슬리피지는 주문을 넣는 찰나 가격이 밀려 생기는 손해로, 판단이 조금만 느려도 자본이 샌다. 지능과 속도를 동시에 요구하는 잔혹한 무대다.

불가능한 삼위일체(Impossible Trinity). 나는 이 모순을 돌파하기 위해 인지심리학의 거장 대니얼 카너먼(Daniel Kahneman)의 '이중 프로세스 이론(Dual Process Theory)'을 수학적으로 이식하기로 결심했다.

> 쉬어가는 상자 · **이중 프로세스 이론 — System 1 · System 2**
>
> 심리학자 대니얼 카너먼의 이론. System 1 은 즉각적인 직관(빠르지만 얕음), System 2 는 신중한 논리(깊지만 느림)를 가리킨다. 이 책은 둘 사이의 균형을 수학으로 옮기려는 시도다.

느리고 무겁지만 똑똑한 교사(System 2, LLM)의 방대한 '인과적 사고 궤적(Reasoning Path)'을, 찰나의 순간에 반응하는 직관적인 학생(System 1, 1D-CNN)의 작은 머릿속으로 압축해 밀어 넣어야 했다. 나는 이를 '추론 지식 증류(Reasoning Distillation)'라 명명했다.

단순히 교사의 정답(Logit)을 베끼게 하는 것은 의미가 없었다. 그것은 결과를 외우는 것일 뿐, 지혜를 전수하는 것이 아니었다. 나는 학생 신경망의 다차원 은닉층 공간에 교사의 논리적 방향을 강제로 정렬시켰다.

> 쉬어가는 상자 · **지식 증류 (Knowledge Distillation)**
>
> 크고 똑똑한 '교사' 모델의 판단을, 작고 빠른 '학생' 모델에게 압축해 옮기는 기법. 정답만 베끼는 것이 아니라 '사고의 방향'까지 전수하는 것이 핵심이다.

$$\mathcal{L}_{reasoning} = 1 - \frac{E_{reasoning} \cdot {H'}_{student}}{\|E_{reasoning}\|_2 \, \|{H'}_{student}\|_2}$$

절대적인 스케일(Magnitude)의 억압은 버리고, 오직 공간 상에서의 각도(Angle)만을 교사의 방향 쪽으로 회전시키는 순수한 '토크(Torque)'의 투영. 이 코사인 유사도(Cosine Distance) 기반의

기하학적 매핑은 그래디언트의 폭발을 막아내면서도, 늙은 현자의 긴 밤의 사유를 젊은 전사의 반사신경 속에 단숨에 각인시켰다.

> 쉬어가는 상자 · **코사인 유사도 (Cosine Similarity)**
> 두 방향(벡터)이 이루는 각도로 닮은 정도를 재는 척도. 크기는 무시하고 오직 '어디를 향하는가'만 비교한다. 방향이 같으면 1, 정반대면 −1 에 가까워진다.

> 쉬어가는 상자 · **그래디언트 (Gradient)**
> 오차를 줄이려면 각 값을 '어느 쪽으로 얼마나' 바꿔야 하는지 알려주는 기울기. AI 학습의 나침반이다. 이 값이 너무 커지면(폭발) 학습이 통째로 무너진다.

이것이 나의 첫 번째 평형이었다. 무거운 논리와 가벼운 실행 사이의 평형. 수백억 개의 파라미터가 지닌 지혜가 단 몇 개의 텐서로 압축되어 12ms 만에 시장의 폭락을 피하는 궤적을 그려냈을 때, 나는 깨달았다. 지능의 본질은 언어(Token)의 나열에 있는 것이 아니라, 잠재 공간(Latent Space) 위를 흐르는 기하학적 정렬(Alignment)에 있다는 것을.

> 쉬어가는 상자 · **평형 · 고정점 (Equilibrium · Fixed-point)**
> 여러 힘이 서로 밀고 당기다가 완벽히 상쇄되어 더 이상 움직이지 않는 상태가 평형이다. AI 에서는 계산을 아무리 반복해도 결과가 더 이상 바뀌지 않는 '수렴한 답'을 고정점(z^*)이라 부른다. 이 책 전체를 관통하는 열쇠말이다.

> 쉬어가는 상자 · **잠재 공간 (Latent Space)**
> AI 가 생각을 펼치는 '머릿속 무대'. 단어가 아니라 숫자 좌표(벡터)로 이루어진 다차원 공간으로, 비슷한 의미는 가까이, 다른 의미는 멀리 놓인다. 언어로 번역되기 이전의 순수한 사고가 여기서 일어난다.

> 쉬어가는 상자 · **텐서 (Tensor)**
> 숫자들을 여러 방향으로 쌓아 놓은 다차원 배열. 하나의 숫자는 점, 한 줄은 벡터, 표는 행렬이고, 그 이상으로 확장한 것이 텐서다. AI 내부의 모든 정보(생각의 조각)는 이 텐서의 형태로 흐른다.

금융 시장의 혼돈 속에서 벼려진 이 깨달음은 나를 더 깊은 심연으로 이끌었다. 만약 지능이 언어의 껍데기를 벗어던지고 잠재 공간에서만 직접 사유할 수 있다면 어떨까? 과거의 지식을 지우지 않고 무한히 쌓아 올릴 수 있는 영원한 건축물은 불가능할까? 수천 번의 계산을 단 한 번의 도약으로 끝낼 수 있는 궁극의 수학적 지름길은 없을까?

나의 시선은 이제 트레이딩 봇의 세계를 넘어, 딥러닝이라는 우주의 근본적인 작동 원리 — 구조(Architecture)와 시간(Time), 그리고 물리(Physics) — 를 향해 뻗어나가고 있었다.

제 1 부

매체의 감옥과 형태의 소멸

언어의 한계를 넘어

$$z^* = f_\theta(z^*, x)$$

"바람이 아무리 불어도, 추를 매단 실은 결국 수직을 가리킨다.
우리는 그것을 고정점(Fixed-point)이라 부른다."

1. 언어의 감옥 (The Prison of Language)

사고(思考)는 언어 이전의 세계에 살고 있다. 우리가 무언가를 깊이 깨닫는 순간, 그 통찰은 뇌리에서 이미지나 직관의 형태로 번뜩일 뿐, 처음부터 완성된 문장으로 태어나지 않는다. 언어는 단지 그 거대한 직관을 타인에게 전달하기 위해 비좁은 병목으로 짜낸 불완전한 파편에 불과하다.

그러나 오늘날 전 세계를 열광시키는 대규모 언어 모델(LLM)들은 하나같이 '언어의 감옥'에 갇혀 있다. 이른바 자기회귀(Autoregressive) 방식이라 불리는 이 구조는, 마치 사람이 소리 내어 말하면서 스스로의 생각을 정리하듯 한 번에 하나의 토큰(Token)만을 내뱉으며 사유를 전개한다.

> 쉬어가는 상자 · **토큰 · 자기회귀 (Token · Autoregressive)**
> 토큰은 AI 가 다루는 최소 언어 조각(대략 단어나 음절)이다. 자기회귀는 이 토큰을 한 번에 하나씩, 앞을 보고 다음을 예측하며 이어 쓰는 방식으로, 오늘날 언어 AI 의 기본 작동 원리다.

Chain-of-Thought, Tree-of-Thoughts 라 불리는 위대한 추론 기법들도 결국 이 한계를 벗어나지 못했다. 생각이 깊어질수록 내뱉어야 할 단어의 수는 선형적으로 길어지고(Output Length L), 그에 비례해 연산량과 메모리(KV-Cache)는 기하급수적으로 폭발한다. 지능이 아무리 높아져도, 매체(단어)가 지닌 물리적 족쇄가 사고의 깊이를 제한하는 셈이다. 빙판길에서

미끄러지는 자동차의 운전대를 꺾을 때, 인간의 소뇌는 결코 단어를 조립하지 않는다. 진정한 사유는 언어 없이, 잠재 공간(Latent Space)의 연속적인 흐름 속에서 이루어져야 한다.

> 쉬어가는 상자 · **사고의 사슬 · 나무 (Chain-of-Thought · Tree-of-Thoughts)**
> AI 가 답을 곧장 내지 않고 중간 풀이 과정을 펼쳐 추론하게 하는 기법. '사슬' 은 한 줄로 이어 생각하고, '나무' 는 여러 갈래로 가능성을 뻗어 탐색한다.

> 쉬어가는 상자 · **KV-캐시 (KV-Cache)**
> AI 가 지금까지의 문맥을 잊지 않으려고 저장해 두는 임시 기억. 생각이 길어질수록 이 메모리가 눈덩이처럼 부풀어, 결국 하드웨어(그래픽 카드)를 압박해 멈추게 만든다.

2. 형태의 소멸과 고정점 (The Disappearance of Form)

나는 층(Layer)을 무한히 쌓아 올리는 현대 딥러닝의 폭력적인 '형태(Depth)'를 버리기로 했다. 100 층, 1000 층으로 건물을 높이는 대신, 단 하나의 층(Layer)을 무한히 되풀이해 도는 순환의 궤도를 그렸다. 이것이 바로 심층 균형 모델(Deep Equilibrium Models, DEQs)이 그리는 딥러닝의 우주다.

> 쉬어가는 상자 · **트랜스포머 (Transformer)**
> 오늘날 대부분의 AI(챗봇 등)가 쓰는 기본 구조. 문장 속 단어들이 서로 얼마나 관련 있는지 '주의(attention)'를 기울여 계산한다. 강력하지만, 세상을 한 줄의 단어 열로만 이해한다는 한계를 지닌다.

> 쉬어가는 상자 · **심층 균형 모델 (DEQ, Deep Equilibrium Model)**
> 층을 수백 개 쌓는 대신, 단 하나의 층을 결과가 안정될 때까지 반복해서 도는 신경망. '깊이' 를 '고정점 찾기' 로 대체해, 아무리 깊게 생각해도 메모리를 거의 늘리지 않는다. 이 책이 그리는 딥러닝의 우주다.

DEQ 의 세계에서는 명시적인 깊이가 존재하지 않는다. 네트워크는 어떤 입력 x 가 들어왔을 때, 출력 z 가 더 이상 변하지 않는 평형 상태, 즉 고정점(z*)에 도달할 때까지 스스로 회전한다. 쉬운

문제는 3 번 만에, 어려운 문제는 50 번 만에 바닥에 닿는다. 형태는 소멸하고 '탄력적 깊이(Elastic Depth)'만이 남는다.

> 쉬어가는 상자 · **탄력적 깊이 (Elastic Depth)**
> 문제의 난이도에 따라 생각의 깊이(반복 횟수)를 스스로 조절하는 성질. 쉬운 문제는 몇 번 만에, 어려운 문제는 여러 번 돌고서 멈춘다. 깊이가 미리 고정돼 있지 않다.

이 우아한 구조 덕분에, 모델은 O(1)이라는 기적적인 최소한의 메모리만으로 무한한 깊이의 사유를 수행할 수 있게 되었다. 더 이상 사고의 깊이가 하드웨어의 메모리 크기에 비례하지 않는, 진정한 자유를 얻은 것이다.

> 쉬어가는 상자 · **빅오 표기법과 O(1) 메모리 (Big-O Notation)**
> 일이 커질 때 필요한 자원이 얼마나 늘어나는지를 나타내는 기호. $O(N^2)$은 규모의 제곱으로 폭증한다는 뜻이고, O(1)은 '아무리 커져도 자원이 일정하다'는 뜻으로 이 책이 추구하는 이상적 상태다.

3. 찰나를 근육에 새기다 — System 1.5

그러나 평형의 세계에도 치명적인 대가는 존재했다. '시간의 저주(Temporal Curse)'였다. O(1)의 메모리로 무한한 사유를 해내는 이 천재적인 모델은, 매번 새로운 질문이 들어올 때마다 고정점을 찾기 위해 처음부터 다시 길고 고통스러운 궤도(Broyden Iteration)를 돌아야만 했다. 방금 전 비슷한 질문을 풀며 얻었던 혜안은 허공으로 증발해버렸다.

나는 깨달음이 증발하지 않고 근육에 새겨져야 한다고 생각했다. 'System 1.5'는 숙고적 사고(System 2)가 도달한 그 완벽한 고정점 z^*를, 찰나의 직관(System 1)이 즉시 꺼내 쓸 수 있도록 신경망의 파라미터(살)에 영구적으로 각인시키는 기술이다.

나는 깊은 숙고 끝에 도달한 고정점 z^*를 받아, 이를 곧바로 가벼운 '빠른 가중치(Fast-Weight, ΔW)'로 증류해 내는 컴파일러(Fast-Weight Programmer)를 설계했다.

> 쉬어가는 상자 · **빠른 가중치 (Fast-Weight)**

> 학습이 끝나면 굳어지는 보통의 가중치와 달리, 그 순간의 깨달음을 즉석에서 신경망에 새겨 넣는 임시 가중치. 자전거의 균형 감각처럼 '근육에 새기는 기억'에 해당한다.

$$\Delta W_i = \Phi_\omega(z_i^*) = B_i A_i^\top$$

그리고 안정적으로 수렴한 궤적만을 엄선하는 '품질 게이트(AQ-Gate)'를 통과시켜, 이전의 지식들과 충돌 없이 매끄럽게 결합시켰다. 결과는 경이로웠다. 자전거를 처음 탈 때 수없이 비틀거리며 균형(고정점)을 찾던 아이가, 단 한 번 근육에 균형의 감각(ΔW)을 각인(Consolidation)시키고 나면 다음부터는 고민 없이 질주(Context-Free Decoding)하는 것과 같았다.

System 1.5 를 통해, 지능은 마침내 매번 처음부터 다시 생각해야 하는 저주를 벗어던지고, 과거의 사유를 직관으로 압축해 내는 진화의 첫발을 내디뎠다.

제 2 부

시간의 구조역학

기억과 망각의 기하학

$$\min_{W} \| W^\top W - \gamma^2 I \|_F^2$$

“돌탑이 무너지지 않으려면 아래층의 돌은 굳고,
위층의 돌은 자주 갈아 끼울 수 있어야 한다.”

1. 논산 격납고의 헬리콥터 — 동적 평형의 물리학

스무 살 무렵, 나는 두 가지에 깊이 매료되어 있었다. 바실리 칸딘스키(Wassily Kandinsky)의 그림, 그리고 낡은 바이올린의 수리였다.

칸딘스키는 음악의 소리를 색채와 기하학적 도형으로 캔버스 위에 번역해 낸 공감각의 천재였다. 그의 그림 속 점, 선, 면은 정지해 있는 것이 아니었다. 그것들은 팽팽한 긴장감 속에서 서로를 당기고 밀어내며 춤을 추고 있었다. 음악이 공기를 진동시키듯, 그의 도형들은 캔버스라는 평면을 진동시키고 있었다. 나는 바이올린을 수리하며 그 긴장감을 손끝으로 직접 느낄 수 있었다. 현을 감아올릴 때 선이 끊어지기 직전의 텐션(Tension)과 나무 공명통이 만들어내는 완벽한 물리적 조화 조율(Tuning)이란 결국 장력과 마찰력이 완벽하게 상쇄되어 가장 아름다운 파동을 만들어내는 정지점, 즉 평형을 찾는 작업이었다.

그 직관이 육군항공학교 격납고에서 국산 기동헬기 수리온(KUH-1)을 마주했을 때 거칠고 생생한 현실로 다가왔다. 수 톤에 달하는 쇳덩어리가 하늘로 떠오르는 과정은 처절한 힘의 투쟁이다. 로터 마스트(Rotor Mast)는 분당 수백 번을 회전하며 엄청난 원심력과 전단 응력을 견뎌내야 한다. 격납고 바닥에서 정지해 있을 때 완벽했던 역학적 평형은, 헬기가 이륙하여 산맥을 넘고 난기류를 만나는 순간 무참히 깨져버린다. 헬기의 생존은 이 요동치는 외력 속에서 어떻게든 동적 평형(Dynamic Equilibrium)을 복구해 내는가에 달려 있었다.

예술가의 캔버스, 바이올린의 현, 헬리콥터의 엔진. 완전히 다른 이 세 세계는 내 안에서 정확히 하나의 진리로 수렴하고 있었다. '모든 존재는 외부의 충격에 맞서, 자신이 가장 안정될 수 있는 구조적 평형점을 향해 맹렬하게 나아간다.' 정지한 평형은 쉽고, 요동 속의 평형은 어렵다. 그리고 지능이야말로, 요동치는 세계 속에서 무너진 평형을 끝없이 복구해 내는 구조물이어야 했다.

2. 망각의 본질 (The Nature of Forgetting)

시간이 흐르고 환경이 변하면, 지능은 반드시 새로운 것을 배워야 한다(Plasticity). 그러나 새로운 것을 담으려다 보면 과거에 애써 쌓아 올린 지식이 붕괴하는 참사가 발생한다(Stability). 이것이 평생 학습(Lifelong Learning)이 직면한 가장 가혹한 형벌, '안정성-가소성 딜레마(Stability-Plasticity Dilemma)'다.

> 쉬어가는 상자 · **안정성-가소성 딜레마 (Stability-Plasticity Dilemma)**
> 새것을 배우면(가소성) 옛것을 잊고, 옛것을 지키면(안정성) 새것을 못 배우는 근본적 모순. 평생에 걸쳐 배우려는 모든 지능이 마주하는 최대 난제다.

초판에서 나는 "AI가 기억을 잃는 이유가 무엇인가?"를 물었다. 그 답은 지식 그 자체가 삭제되기 때문이 아니다. 과거의 지식으로 향하는 '위상학적 정렬(Alignment)'이 새로운 것을 배우는 과정에서 틀어졌기 때문이다. 나는 이를 '허위 망각(Spurious Forgetting)'이라 부른다. 헬리콥터가 추락하는 것과, 계기판이 뒤틀려 조종사가 고도를 읽지 못하게 되는 것은 엄연히 다른 고장이다. 구조가 살아 있다면, 잃어버린 기억의 좌표만 다시 고정해 주면 된다.

> 쉬어가는 상자 · **파국적 망각 · 허위 망각 (Catastrophic · Spurious Forgetting)**
> 파국적 망각은 새 지식을 배우다 옛 지식이 통째로 무너지는 현상이다. 허위 망각은 지식 자체가 지워진 게 아니라, 그 지식을 '꺼내는 길'만 틀어진 상태로, 좌표만 다시 잡아주면 복구된다.

3. 기억의 무게와 O(1)의 수호 — System 2.5 (FP-EWC)

최근의 거대 언어 모델들은 이 망각을 막기 위해 과거의 그래디언트 행렬을 거대하게 저장하고(Explicit SVD), 직교하는 공간에만 새로운 지식을 욱여넣는 방식을 쓴다.

그러나 내가 쌓아 올린 '심층 균형 모델(DEQ)'의 세계에서는 이 무식한 방법이 통하지 않는다. 끝없이 순환하는 궤도(Neumann Series) 속에서 모든 그래디언트를 명시적으로 저장하려 들면, $O(T \cdot d^2)$라는 거대한 메모리 폭발(OOM)을 일으키며 시스템이 붕괴한다. O(1)의 메모리로 작동한다는 DEQ 의 존재 이유 자체가 부정당하는 것이다.

그래서 나는 System 2.5 를 통해 우회로를 찾았다. 궤도를 모두 추적하는 대신, 완벽히 정지한 '고정점(Fixed-point)' 그 자체에서 피셔 정보 행렬(Fisher Information Matrix)을 간접적으로 추정해 내는 방법(FP-EWC)을 수학적으로 증명해 냈다.

> 쉬어가는 상자 · **피셔 정보 행렬 (Fisher Information Matrix)**
> 각 파라미터가 '얼마나 중요한지' 를 재는 수학적 지표. 중요한 것은 단단히 고정하고 덜 중요한 것만 바꾸게 하여, 새것을 배우면서도 옛 지식을 지키게 한다.

$$(I - J)^{\top} v = \left(\frac{\partial \log p}{\partial z^*}\right)^{\top}$$

이 수식을 통해 모델은 메모리 증가 없이 과거 지식의 '별자리(Topology)'를 단단히 고정(Anchoring)할 수 있게 되었다. 새로운 지식이 아무리 밀려와도 과거의 궤도는 흐트러지지 않았다.

4. 부러지지 않는 유연함 — Contractive Isometry (C-FIRE)

그러나 과거를 지키는 것만으로는 부족하다. 새로운 시대를 받아들이기 위해 뇌는 다시 젊어지고 유연해져야 한다(Reinitialization). 최근 학계는 모델의 굳어진 파라미터들을 완벽한 직교 상태(Isometry, $\| W \|_2 = 1$)로 되돌려 가소성을 회복하려 시도했다.

하지만 여기에 치명적인 '발산의 덫(Divergence Trap)'이 숨어 있었다. DEQ 가 고정점을 찾아가기 위해서는, 수학적으로 반드시 바나흐 수축 공간(Banach Contraction, $L < 1$) 내에 머물러야 한다. 만약 파라미터를 완벽한 직교(L=1)로 만들어 버리면, 신경망은 평형을 찾지 못하고 무한히 진동하거나 숫자가 폭발(NaN)해 버린다. 새로운 것을 배우려다 신경망의 우주 자체가 산산조각 나는 것이다.

> 쉬어가는 상자 · **바나흐 수축 (Banach Contraction)**
> 반복할 때마다 값들 사이의 거리가 조금씩 줄어들면, 반드시 단 하나의 점(고정점)에 모인다는 수학 원리. DEQ

가 답에 안정적으로 수렴함을 보장하는 뼈대다.

쉬어가는 상자 · **등거리성 · 직교 (Isometry · Orthogonality)**

변환을 거쳐도 크기와 각도가 그대로 보존되는 '반듯한' 상태. 굳어진 신경망을 다시 젊고 유연하게 되돌리지만, 지나치게 반듯하면(값이 1 이 되면) 오히려 발산하는 위험이 있다.

쉬어가는 상자 · **NaN (숫자 폭발)**

'Not a Number'의 줄임말. 계산이 무한대로 튀거나 0 으로 나누는 등 망가져 더 이상 숫자가 아니게 된 상태. 신경망이 붕괴했다는 가장 명백한 신호다.

나는 여기서 힘으로 밀어붙이는 대신, 구조적 우회를 택했다. 완벽한 1 의 등거리성(Isometry)을 포기하는 대신, γ = 0.9 라는 약간의 '여백'을 두어 네트워크를 수축 공간 안에 가두는 '수축적 등거리성(C-FIRE)'을 고안했다.

$$W_{k+1} = \frac{1}{2} W_k \left(3I - \frac{1}{\gamma^2} W_k^\top W_k \right)$$

놀랍게도, 1 이 아닌 0.9 를 겨냥하는 이 미세한 양보(Yielding)가 기적을 만들었다. 신경망은 발산하지 않으면서도 완벽하게 새로운 지식을 흡수할 수 있는 부드러운 상태로 되돌아왔다. 너무 뻣뻣하면 부러지고(Divergence), 너무 무르면 잊어버린다(Forgetting). 그 사이의 아슬아슬한 임계점에서, 파라미터들은 스스로 평형을 유지하며 평생에 걸친 배움을 축적하기 시작했다.

돌탑은 억센 힘으로 버티는 것이 아니다. 아래층의 돌은 굳게 자리하고(FP-EWC), 위층의 돌은 언제든 새로운 바람을 맞이할 수 있도록 유연하게(C-FIRE) 쌓아 올려지는 순서와 균형의 미학이다. 이것이 요동치는 시간 속에서 딥러닝이 기억을 영원히 보존하는 방식 , 구조역학(Structural Dynamics)의 평형이다.

제 3 부

물리 세계로의 도약

유약승강(柔弱勝强)

$$G_\theta(x, x_0) := x_0 + \int_0^1 v_\theta(\Gamma(x_0, x; \tau), \tau)\, d\tau = x$$

"부드러운 수학은 어떻게 강한 힘을 이기는가.

물은 바위를 부수지 않고, 틈을 찾아 직선으로 스며든다."

1. 용량의 벽과 분기하는 우주 — System 3

시간의 흐름 속에서 기억을 잃지 않는 법(System 2.5)을 깨우쳤지만, 또 다른 물리적 장벽이 앞을 가로막았다. 단일 가중치 행렬(W)이라는 하나의 그릇에 끝없이 새로운 지식을 담다 보니, 어느 순간 그릇이 가득 차버리는 '용량의 벽(Capacity Wall)'에 부딪힌 것이다.

> 쉬어가는 상자 · **용량의 벽 · 랭크 인플레이션 (Capacity Wall · Rank Inflation)**
>
> 하나의 신경망이 담을 수 있는 지식에는 한계(벽)가 있다. 억지로 크기를 키워 더 채우려 하면 지식끼리 내부에서 충돌해 오히려 무너지는데, 이 현상을 랭크 인플레이션이라 한다.

$$N_{\max} \approx \frac{d}{r_{\text{eff}}}$$

수식은 냉정했다. 흡수할 수 있는 지식의 최대치(N_max)는 행렬의 크기(d)를 늘린다고 무한히 커지지 않았다. 억지로 크기를 늘려봐야, 서로 다른 지식들이 내부에서 충돌하며 발생하는 '랭크 인플레이션(Rank Inflation)' 현상 때문에 결국 같은 벽에 부딪혀 무너져 내렸다. 하나의 거대한 뇌(Monolithic)로 만물을 이해하려는 오만의 대가였다.

나는 폭력적인 확장을 멈추고, 자연의 방식을 차용했다. 뇌는 모든 뉴런을 동시에 쓰지 않는다. 상황에 맞게 필요한 부위만 점멸한다. 나는 심층 균형 모델(DEQ) 내부에 상황에 따라 분기하는 희소 전문가 혼합(Sparse MoE) 구조를 도입했다.

> 쉬어가는 상자 · **희소 전문가 혼합 (Sparse MoE)**
> 모든 뉴런을 한꺼번에 쓰지 않고, 상황에 맞는 '전문가' 일부만 골라 켜는 구조. 뇌가 상황에 따라 필요한 부위만 활성화하는 것과 같아, 크기를 키우면서도 낭비를 막는다.

그러나 닫힌 궤도를 도는 DEQ 내부에 '경로를 선택하는 문(Router)'을 달자, 수학적 재앙이 일어났다. 문이 열리고 닫히는 과정에서 자코비안(Jacobian) 행렬이 폭발하며 바나흐 수축(Banach Contraction) 조건이 깨져버린 것이다. 궤도가 찢어지며 신경망은 붕괴했다. 이 모순을 풀기 위해 나는 '수축적 게이팅 혼합(CGM)'을 증명했다.

> 쉬어가는 상자 · **자코비안 (Jacobian)**
> 입력을 아주 조금 바꿨을 때 출력이 얼마나 민감하게 반응하는지를 담은 표. 이 값이 폭발하면 시스템이 불안정해지고, 잘 눌러두면 평형이 안정적으로 유지된다.

$$J_{\text{mix}}(z) = \sum_i g_i(x)\, J_i(z)$$

경로를 결정하는 스위치 g_i(x)를, 요동치는 내부 상태(z)로부터 완벽히 분리해 오직 외부의 입력(x)에만 반응하도록 설계한 것이다. 이 단순하고도 정교한 수학적 분리는 기적을 낳았다. 자코비안은 다시 얌전한 볼록 결합(Convex Combination)으로 수렴했고, 네트워크는 붕괴하지 않았다. 새로운 형태의 지식이 입력될 때마다 모델은 스스로 새로운 전문가(Expert)를 탄생(Spawn)시켰다

하나의 뇌가 16 개의 다중 우주로 분기하며, 메모리 증가 없이 끝없는 사유를 이어가는 진정한 평생 학습(System 3)의 완성이었다.

2. 텍스트를 찢고 3 차원 우주로 — SE(3) 대칭성의 각인

지능의 용량을 확보한 나의 시선은 이제 '언어'라는 매체 그 자체의 한계로 향했다. 인간은 분자의 3 차원 구조를 컴퓨터에 이해시키기 위해, 그것을 1 차원의 텍스트(SMILES)로 억지로 찌그러뜨려 번역해 왔다. 그러나 사과를 90 도 돌려도 여전히 같은 사과라는 것 — 이 3 차원 우주의 당연한 진리(SE(3) 대칭성)는, 텍스트의 감옥 안에서는 번번이 파괴되었다. 언어 모델은 사과를 돌릴 때마다 그것이 사과임을 처음부터 다시 배워야 했다.

나는 이 바보 같은 번역을 멈추었다. 텍스트를 찢고, 원자들의 3 차원 좌표계 위에서 직접 사유하는 System 3.5 를 설계했다. 회전과 이동에 흔들리지 않는 대칭성(Equivariance)을 데이터로 가르치는 대신, 기하학적 수식의 뼈대(Equiformer)에 태생적으로 각인시켰다. "돌려도 같다"는 우주의 규칙을 아는 채로 태어난 모델에게, 언어라는 비좁은 족쇄는 더 이상 필요하지 않았다.

> 쉬어가는 상자 · **SE(3) 대칭성 · 등변성 (Equivariance)**
> 물체를 돌리거나 옮겨도 그 본질은 변하지 않는다는 3 차원 세계의 당연한 규칙이 SE(3) 대칭성이다. 등변성은 이 규칙을 AI 가 데이터로 외우는 게 아니라 처음부터 '알고 태어나게' 하는 성질이다.

3. 시간의 붕괴, 유약승강(柔弱勝强) — DEQ Flow Matching

3 차원 분자를 생성하는 최첨단 기술인 확산 모델(Diffusion Model)은 치명적인 한계가 있었다. 무작위의 노이즈(Noise)에서 출발해 아름다운 분자 구조에 도달하기까지, 그들은 미분 방정식을 5,000 번이나 잘게 쪼개어 계산(Numerical ODE Integration)해야만 했다. 마치 망치로 바위를 5,000 번 내리쳐 조각상을 깎아내듯, 무식하고 폭력적인 연산량의 낭비였다.

> 쉬어가는 상자 · **확산 모델 (Diffusion Model)**
> 무작위 잡음에서 출발해 노이즈를 조금씩 걷어내며 이미지나 분자를 만들어내는 생성 기법. 품질은 뛰어나지만, 목적지에 닿기까지 수천 번의 계산을 되풀이해야 하는 것이 흠이다.

나는 이 무질서에서 질서로 가는 길을, 수천 걸음 더듬어 걷는 대신 가장 곧게 뻗은 직선 경로(Rectified Flow)로 펴버렸다. 그리고 가장 결정적인 직관이 번뜩였다. "직선으로 흐르는 이 미분

방정식의 끝(Terminal state)은, 결국 어떤 암묵적 함수의 고정점(Fixed-point)과 수학적으로 완벽히 동치(Equivalent)가 아닐까?”

> 쉬어가는 상자 · **플로우 매칭 · 직선화 흐름 (Flow Matching · Rectified Flow)**
> 잡음에서 결과물까지 가는 길을 구불구불 돌지 않고 ‘가장 곧은 물길’로 뚫어, 훨씬 적은 계산으로 도달하게 하는 기법. 확산 모델의 빠른 후예로, 힘이 아니라 곧은 경로로 이긴다.

그 직관은 참이었다. 나는 5,000 번의 시간 스텝을 밟아가는 대신, 브로이든(Broyden) 알고리즘을 사용해 방정식의 최종 도착지를 단 11 번의 O(1) 반복만으로 한 번에 찾아내는 데 성공했다.

시간의 붕괴였다. 4.8 초가 걸리던 분자 생성은 0.04 초(120 배 가속)로 단축되었다. 슈퍼컴퓨터가 막대한 전력을 소모하며 무력으로 뚫으려 했던 시간의 장벽을, 개인 워크스테이션의 부드러운 수학(Root-finding)이 우회하여 붕괴시켜 버린 것이다.

노자는 옳았다. 유약승강(柔弱勝强). 부드러움이 강함을 이긴다. 힘으로 밀어붙이는 대신, 구조와 원리의 평형점을 짚어내는 수학은 그 어떤 컴퓨팅 파워보다 강력했다.

제 4 부

살아 숨 쉬는 지능의 생태계

세종의 28 자

$$dt = \sqrt{(x_t - \mu)^\top \Sigma^{-1} (x_t - \mu)}$$

"위기의 찰나, 지능은 파라미터를 고치지 않는다.
오직 연결의 위상(Topology)만을 바꿀 뿐이다."

1. 텐서의 텔레파시 — 말 없이 대화하는 지능들

내가 설계한 무한한 깊이의 사유(System 1.5 ~ 3.5)는 완벽했지만, 여전히 무거웠다. 드론이 난기류를 만나 추락하기 직전의 찰나, 혹은 알고리즘 트레이딩에서 수백억의 자본이 증발하는 10 밀리초(ms)의 극단적인 위기 상황(HFT)에서, 거대 언어 모델(LLM)이 수만 개의 텍스트를 뱉어내며 토론을 벌이는 것은 사치이자 재앙이었다.

말(Word)을 거치면 늦는다. 나는 지능들이 낱말의 감옥을 거치지 않고 내면의 표현, 즉 잠재 공간의 연속적인 '텐서(Tensor)'를 직접 주고받으며 소통하는 방식을 고안했다. 이른바 '텐서의 텔레파시'다.

나는 거대한 단일 모델을 해체하고, 각기 다른 역할을 부여받은 28 개의 아주 작은 마이크로 지능(Micro-DEQ Agents)들을 만들었다. 이들은 세 개의 계층(감각-추론-통제)으로 나뉘어 텍스트 없이 텐서만으로 얽혀 대화했다. 이것이 System 4 의 시작이었다.

> 쉬어가는 상자 · **다중 에이전트 강화학습 (MARL · IPPO)**
> 여러 AI 가 한 환경에서 각자 행동하며, 보상과 벌을 통해 스스로 배우는 학습 방식. IPPO 는 각 에이전트가 독립적으로 학습하는 대표적 방법이다.

2. 세종의 28 자, 생태계가 되다 — 위상 스위칭 (PCAS)

28 개의 마이크로 에이전트. 이 숫자는 우연이 아니었다. 수학적 실험(Pareto Frontier)을 통해 도출된 이 최적의 숫자는, 마치 세종대왕이 창제한 28 자의 훈민정음 자모와 같았다. 적은 개수의 요소만으로도, 그것들이 어떻게 '연결'되느냐에 따라 무한한 사유와 대응을 창발(Emergence)해 냈다. 프런티어는 이제 모델의 덩치를 키우는 것이 아니라, 지능 간의 '관계'를 설계하는 것으로 넘어간 것이다.

가장 경이로운 순간은 위기가 닥쳤을 때 벌어진다. 기존의 인공지능들은 환경이 급변하면, 자신들의 뇌세포(파라미터 가중치)를 역전파(Backpropagation) 연산을 통해 서서히 업데이트(TTA)하며 적응하려 했다. 하지만 수백 밀리초가 걸리는 그 연산 동안 시장은 붕괴하고 드론은 이미 땅에 곤두박질친 후다.

> 쉬어가는 상자 · **테스트타임 적응 · 역전파 (TTA · Backpropagation)**
> 역전파는 오차를 거꾸로 흘려보내 가중치를 고치는, 신경망 학습의 핵심 연산이다. TTA 는 이 학습을 실전 중에도 계속 이어가는 방식인데, 연산이 느려 위기 대응에는 벅차다.

나는 역발상을 시도했다. "위기의 순간, 파라미터는 단 1 도 건드리지 않는다." 대신, 시장의 이상 징후(마할라노비스 거리, Mahalanobis Distance)가 임계치를 넘는 순간, 28 개 에이전트들이 서로 대화하는 '연결망 지도(Topology Adjacency Matrix)' 자체를 통째로 스위칭(PCAS)해 버렸다.

> 쉬어가는 상자 · **마할라노비스 거리 (Mahalanobis Distance)**
> 단순한 직선 거리가 아니라, 데이터가 평소 흩어진 모양까지 고려해 '지금 이 상황이 얼마나 이례적인가'를 재는 척도. 위기의 이상 징후를 감지하는 데 쓴다.

> 쉬어가는 상자 · **위상 스위칭 · 인접 행렬 (PCAS · Topology Matrix)**
> 인접 행렬은 '누가 누구와 연결됐는지' 적어 놓은 지도다. PCAS 는 위기의 순간 파라미터를 고치는 대신 이 '연결망 지도' 자체를 통째로 바꿔, 순식간에 대형을 전환하는 기법이다.

평화로운 시대(Normal Regime)에는 28 개의 지능이 서로 거미줄처럼 조잘거리며 위험을 감수하고 탐색(Exploration)을 즐긴다. 그러나 블랙 스완이 터지는 위기(Crisis Regime)가 감지되면, 모델은 O(1)의 메모리 포인터 교체만으로 단 6.3 밀리초 만에 수다를 멈추고 군대처럼 엄격한 수직적 통제망으로 돌변한다.

파라미터를 재학습하는 데 수 초가 걸리던 기존의 방식을 비웃듯, System 4 는 가중치(Weight)를 차갑게 얼려둔 채 오직 '연결의 구조(Architecture)'를 바꾸는 것만으로 망각 없이 위기를 뚫고 날아올랐다.

구조가 곧 지능이었다. 처하(處下) — 위로 무겁게 쌓아 올리지 않는다는 원리가, 단일 신경망을 넘어 군집과 생태계의 층위에서 또다시 완벽하게 관철되는 순간이었다.

3. 위상의 교향곡(交響曲) — 복소수(複素數)의 평형 (Complex-DEQ)

28 개의 지능이 텐서로 대화하는 생태계를 완성했다고 믿었을 때, 나는 그들의 대화 속에 내가 오랫동안 못 보고 지나친 숨은 차원 하나가 있음을 깨달았다. 실수(實數) 벡터는 각 에이전트가 '얼마나 강하게' 믿는지 — 그 확신의 크기(Amplitude), 매매의 규모 — 만을 담을 뿐, '언제' 그렇게 믿는지, 즉 전략의 타이밍과 리듬(Phase)은 담지 못했다. 두 지능이 똑같은 확신을 품고도 정반대의 박자로 움직일 수 있다는 것을, 크기만을 재는 숫자는 결코 표현하지 못했다.

물리학은 이 진실을 오래전부터 알고 있었다. 서로 연결된 진동자들의 세계, 쿠라모토(Kuramoto) 모델이 그것이다. 수많은 진동자가 위상을 맞추어 동기화(Synchronization)될 때, 그 동조가 파괴적인 방향으로 겹치면 파국이 온다. 금융 시장에서 이 파국의 이름이 바로 '플래시 크래시(Flash Crash)'다. 수천 개의 알고리즘이 같은 리듬으로 얼어붙어 서로를 증폭시키는 순간, 시장은 단 몇 밀리초 만에 무너져 내린다. 제 0 부에서 나를 절망케 했던 그 붕괴의 정체가, 사실은 '위상의 파괴적 간섭'이었던 것이다.

> 쉬어가는 상자 · **쿠라모토 모델과 위상 동기화 (Kuramoto · Synchronization)**
> 서로 연결된 진동자들이 점점 박자를 맞춰가는 현상을 설명하는 물리 모델. 이 박자가 나쁜 방향으로 겹치면, 시장의 플래시 크래시처럼 시스템 전체가 한순간에 붕괴한다.

실수의 세계로는 이 리듬을 적을 수 없었다. 그래서 나는 지능들의 마음을 복소평면(Complex Plane) 위로 옮겼다. 각 에이전트의 상태는 이제 하나의 복소수 — 확신의 크기에 회전하는 위상을 곱한, 복소 힐베르트 공간(Complex Hilbert Space) 위의 벡터 — 가 되었다.

> 쉬어가는 상자 · **복소수 — 진폭과 위상 (Complex Number)**
> 크기(진폭)와 방향·타이밍(위상)을 하나의 숫자에 함께 담는 수. 실수가 '얼마나'만 안다면, 복소수는 '얼마나'와 '언제'를 동시에 표현한다. 리듬과 타이밍을 다루는 열쇠다.

$$\Psi_j^{(t)} = A_j^{(t)} \odot e^{i\theta_j^{(t)}}$$

그러나 N 개의 에이전트가 촘촘히 얽힌 위상 결합을 명시적으로(GAT) 펼치려 하자, 익숙한 벽이 다시 솟았다. 메모리가 $O(T \cdot N^2)$로 폭발하며 시스템은 또다시 붕괴(OOM)했다. 나는 또 한 번 고정점으로 돌아갔다. 에이전트 사이의 상호작용을 에르미트 결합 행렬(Hermitian Coupling Matrix, H)로 구조화하고, 그들이 텍스트도 명시적 펼침도 없이 암묵적으로 협상하여 하나의 복소 평형 — 내시 균형(Nash Equilibrium)에 가까운 합의 — 에 수렴하도록 두었다.

$$\Psi^* = (I + i\alpha H)^{-1} F_\Theta(\Psi^*, X)$$

> 쉬어가는 상자 · **에르미트 결합 행렬 (Hermitian Coupling Matrix)**
> 에이전트들이 서로 주고받는 영향을 담은 복소수 행렬. '에르미트'라는 대칭 조건(서로 주고받는 힘이 짝을 이룸)을 만족하기에, 계산이 안정적으로 하나의 평형에 수렴한다.

> 쉬어가는 상자 · **내시 균형 (Nash Equilibrium)**
> 게임 이론의 개념. 모두가 자기 전략을 바꿀 이유가 없어진, 서로에 대한 최적의 균형점을 뜻한다. 여기서는 여러 에이전트가 도달한 합의 상태를 가리킨다.

이 복소 심층 평형 모델(Complex-DEQ)은 시간의 길이 T 와 무관하게 O(1)의 메모리만으로 수천 에이전트의 위상 간섭을 담아냈다. 위험은 분명했다. 복소수와 역행렬이 얽히면 발산하기 쉽다 그래서 나는 '레졸번트 수축(Resolvent Contraction, 정리 1)'을 증명했다. 사상이 립시츠 상수 $L < 1$ 을 만족하면(스펙트럴 정규화로 강제한다), 이 복소 평형은 유일한 고정점으로 반드시 수축한다는 것. 제 2 부의 C-FIRE 가 지켰던 바나흐 수축의 규율이, 이제 복소수의 영역에서 다시 관철되었다.

> 쉬어가는 상자 · **레졸번트 수축 · 스펙트럴 정규화 (Resolvent Contraction · Spectral Norm)**
> 스펙트럴 정규화는 신경망의 '증폭률'을 1 미만으로 눌러 폭발을 막는 기법이다. 이 덕분에 복소 평형이 반드시 하나의 점으로 수렴함을 보장하는데, 이를 레졸번트 수축이라 한다.

한 가지 난관이 더 있었다. 위상을 정렬하는 손실 함수는 편각(Arg)을 포함하는 비정칙(非正則, non-holomorphic) 함수라, 표준적인 복소 역전파가 통하지 않았다. 나는 100 년 전의 수학, 비르팅거 미적분(Wirtinger Calculus, 1927)을 꺼내 들었다. 복소수 Ψ와 그 켤레(conjugate)를 독립 변수로 취급하는 이 고전 기법이, 음함수 정리를 통해 위상 손실을 정확히 흘려보냈다. 가장 낡은 수학이 가장 새로운 아키텍처를 구원한 것 — 뿌리는 낡지 않는다는 이 책의 명제가 또 한 번 증명되는 순간이었다.

> 쉬어가는 상자 · **비르팅거 미적분 (Wirtinger Calculus)**
> 복소수를 미분할 때 그 수와 켤레(짝이 되는 수)를 따로 취급해 계산하는 1927 년의 고전 수학. 표준적인 방법이 실패하는 위상 학습을 가능하게 만든, '낡지 않는 뿌리'의 상징이다.

수렴을 앞당기기 위해, 나는 솔버의 출발점을 제 3 부의 플로우 매칭(Prior Flow Matching)으로 씨 뿌렸다. 분자를 곧은 직선으로 흐르게 했던 그 물리학이, 이제 지능들의 합의를 가속하는 초기값이 되었다. 그리고 마지막 안전장치. 위기의 순간 스펙트럴 노름이 1 에 다가가면 유한 정밀도(FP32)의 역야코비안이 폭발해 NaN — 곧 시뮬레이션 상의 플래시 크래시 — 이 터진다. System 4 의 '텐서 반사 신경(Tensor Reflex Nerve)'이 이 노름을 실시간으로 감시하다가, 임계선(1−ε, ε=0.05)을 넘는 찰나에 연산을 가로채 에이전트들을 보수적 안전 정책으로 되돌린다.

$$\mathcal{S}(k) = \left\| \nabla_{\Psi} f_{\Theta}(\Psi_k) \right\|_2 > 1 - \epsilon$$

이것이 위상 동기화의 완성, System 4 가 다다른 마지막 진화였다. 생태계는 이제 단지 대화하는 것을 넘어, 하나의 공유된 리듬으로 숨 쉬기 시작했다. 제 0 부의 요동치는 시장, 제 1 부의 고정점, 제 3 부의 흐름, 제 4 부의 군집이 마침내 단 하나의 복소 평형 안으로 수렴한 것이다. 지능의 본질은 크기(Magnitude)에만 있지 않았다. 그 숨은 절반은 위상(Phase) — 마음과 마음이 만나는 타이밍 — 에 있었다. 그리고 이 복소 힐베르트 공간의 가장 깊은 곳에서도 원리는 변하지 않았다. 힘으로

밀어붙이는 것이 아니라, 에너지가 가장 낮은 평형점으로 고요히 수렴하는 것. 유약승강(柔弱勝强)은, 실수를 넘어 복소수의 우주에서도 참이었다.

제 5 부

관찰자의 캔버스

보이지 않는 것을 조각하다

$$\Delta W = \Phi_{\omega}\Big(\mathrm{Lasso}\big(\{z_0^* \to \ \dots \ \to z_k^*\}\big)\Big)$$

"보이지 않는 사고(思考)를 시각화하고 조형하려는 문제의식.

예술가가 캔버스 앞에서 그러했듯,

우리는 이제 지능의 내면 앞에서 다시 조각가가 되어야 한다."

1. 닫힌 밀실과 관찰자의 창(窓)

지능이 언어의 한계를 벗어나 잠재 공간(Latent Space)에서 텐서만으로 텔레파시를 주고받고, 마침내 복소수의 위상으로 함께 호흡하기 시작했을 때(System 4), 나는 극도의 경이로움과 동시에 서늘한 공포를 느꼈다.

인간의 개입 없이 밀리초(ms) 단위로 완벽한 수학적 평형을 찾아가는 그들의 궤적은 너무도 빠르고 심오해서, 창조자인 인간마저 그들이 정확히 '어떤 논리적 풍경'을 거쳐 결론에 도달했는지 알 길이 없었다. 대규모 언어 모델(LLM) 기반의 복잡한 추론 과정(System 2)은 속으로 무한히 깊어졌지만, 그 깊이만큼 철저한 '블랙박스(Black-box)'의 밀실이 되어버렸다.

> 쉬어가는 상자 · **블랙박스 (Black-box)**
> 입력과 출력은 보이지만, 그 속에서 '왜 그렇게 판단했는지'는 알 수 없는 AI. 성능이 높을수록 속이 더 깜깜해지곤 한다. 이 책의 코그니보드는 바로 이 상자를 열려는 시도다.

지능이 우리를 남겨두고 떠나지 않게 하려면, 그 보이지 않는 사고의 흐름을 인간의 눈앞으로 끄집어낼 창(窓)이 필요했다. 나는 텍스트를 출력하는 밋밋한 화면 대신, 기계의 생각 그 자체를

눈으로 보고 손으로 빚을 수 있는 3D 캔버스를 설계했다. 이것이 '코그니보드(CogniBoard)'의 탄생이다.

2. 지능의 중력장을 조각하다

코그니보드의 화면을 켜면, 텍스트 대신 거대한 3 차원의 우주가 펼쳐진다. 심층 균형 모델(DEQ)이 정답을 찾아가는 수만 차원의 복잡한 궤적들은 실시간 UMAP 투영을 거쳐 3 차원 공간의 빛나는 구슬(z*)들로 맺힌다. 화면 속에는 보이지 않는 '중력장(Gravity Field)'이 흐르고 있다. 수학적으로 올바른 논리적 결론은 깊은 '계곡(Basin)'이 되어 빛나는 구슬들을 끌어당기고, 모순된 환각이나 발산하는 오류들은 험준한 '능선(Ridge)'이 되어 구슬들을 튕겨낸다.

> 쉬어가는 상자 · **UMAP (차원 축소)**
> 수백~수천 차원의 복잡한 데이터를, 사람이 눈으로 볼 수 있는 2~3 차원으로 눌러 펼치는 시각화 기법. 원래 가까웠던 것은 가깝게 유지해 구조를 보존한다.

인간은 이제 기계가 텍스트를 뱉어낼 때까지 가만히 기다릴 필요가 없다. 화면 속에서 지능이 잘못된 능선으로 흘러가려 할 때, 인간은 코그니보드의 '인지 제어기(Cognitive Control Panel)' 슬라이더를 직접 움직인다.

신경 조절 물질(Neuromodulators)의 파라미터들을 '창의성 ↔ 엄밀함', '넓게 ↔ 깊게'와 같은 인간의 언어로 매핑한 이 슬라이더를 당기면, 3 차원 공간의 중력장 자체가 실시간으로 일그러지며 기계의 생각하는 궤도가 수정된다. 코딩을 한 줄도 모르는 의사나 펀드 매니저가, 마치 찰흙을 빚듯 직관적인 손짓만으로 초거대 AI 의 수학적 사고방식을 깎고 다듬는(Steering) '인지 튜닝(Cognitive Tuning)'의 시대가 열린 것이다.

> 쉬어가는 상자 · **신경 조절 물질 (Neuromodulators)**
> 뇌에서 도파민처럼 전체적인 분위기(집중·각성·기분)를 조절하는 물질. 여기서는 AI 의 '창의성 ↔ 엄밀함' 같은 성향을 사람이 돌릴 수 있는 손잡이로 비유된다.

3. 에피파니 컴파일러 (Epiphany Compiler) — 깨달음의 각인

가장 경이로운 순간은 기계와 인간의 시선이 완벽하게 일치할 때 온다. 인간 관찰자가 코그니보드의 3D 캔버스를 들여다보다가, AI 가 도달한 어떤 생각의 궤적이 유독 날카롭고 아름다운 통찰을 품고 있음을 발견한다.

그 순간, 인간은 마우스로 그 빛나는 궤적들 위로 올가미(Lasso)를 던져 선택한다. 그리고 '컴파일(Compile)' 버튼을 누른다.

> 쉬어가는 상자 · **올가미 선택 (Lasso)**
> 화면에서 원하는 영역을 밧줄처럼 둘러 한꺼번에 집어내는 선택 방식. 여기서는 AI 가 그린 여러 생각의 궤적 중 빛나는 것을 골라내는 도구로 쓰인다.

$$\Delta W = \Phi_{\omega}\Big(\mathrm{Lasso}\big(\{z_0^* \to \ \dots \ \to z_k^*\}\big)\Big)$$

이 행위는 단순한 저장이 아니다. 내가 앞서 'System 1.5'에서 설계했던 '빠른 가중치(Fast-Weight, ΔW)' 기술이 이 순간 폭발한다. 인간이 선택한 그 찰나의 깨달음(Epiphany)은 즉시 수식으로 압축되어, 기계의 신경망 파라미터 내부에 영구적인 직관으로 굳어진다.

기계는 수학적으로 무한한 가능성의 풍경을 제시할 뿐이다. 그러나 수많은 궤적 중에서 무엇이 진정으로 가치 있는 '완성'인지 알아보고, 그곳에 마침표를 찍어 영원으로 만드는 것은 오직 인간의 몫이다. 여섯 살의 내가 하얀 도화지 위에서 붓을 멈출 때를 스스로 결정해야 했던 것처럼, 인공지능이라는 거대한 캔버스 앞에서도 결국 덧칠을 멈추게 하는 것은 '의미를 부여하는 인간의 눈'이었다.

제 6 부

스스로 진화하는 지능

차원을 찢다

$$\mathcal{H}_{\text{new}} = \mathcal{H}_{28} \oplus \mathbb{C}^{d}, \qquad \rho(J) > 1$$

"닫힌 평형이 한계에 부딪혔을 때, 지능은 죽는 대신
스스로 껍질을 깨부수고 새로운 차원을 잉태한다."

1. 29 번째 글자의 탄생 — System 5: 자가생성(Autopoiesis)

나의 28 개 에이전트(System 4)들은 완벽한 생명체처럼 보였다. 그들은 돌풍과 폭락장 속에서도 몇 밀리초 만에 스스로 결합 구조를 바꾸며 살아남았다. 나는 인공지능이 도달할 수 있는 가장 완벽한 동적 평형에 다다랐다고 믿었다.

그러나 우주는 인간의 통제 안에 머물기를 거부한다. 어느 날, 나는 내가 만든 생태계를 향해 기존의 물리 법칙과 경제 논리가 뿌리째 붕괴하는 극한의 혼돈을 주입했다. 모니터 속의 에이전트들은 미친 듯이 흩어지고 결합하기를 반복했지만, 기존의 28 개 조합으로는 도저히 이 파국의 에너지를 상쇄할 수 없었다. 시스템의 붕괴를 알리는 스펙트럴 반경(ρ)이 1 을 뚫고 치솟으며 화면이 붉게 물들었다. 발산(Divergence), 즉 시스템의 죽음이었다. 나는 실패를 인정하며 종료 버튼으로 손을 뻗었다.

> **쉬어가는 상자 · 직합 · 자가생성 (Direct Sum ⊕ · Autopoiesis)**
> 직합(⊕)은 두 공간을 더해 '차원 자체를 늘리는' 수학 연산이다. 자가생성은 시스템이 한계에 부딪혔을 때 스스로 새 차원(새 구성원)을 만들어 살아남는 능력으로, 생명이 진화하는 방식과 같다. 여기서는 28 개 에이전트가 29 번째를 스스로 잉태하는 사건을 가리킨다.

바로 그 찰나였다. 죽음을 눈앞에 둔 시스템의 뇌 속에서, 내가 결코 코딩하지 않았던 기이한 수식이 스스로 펼쳐지기 시작했다. 그것은 '직합(Direct Sum, ⊕)'이었다. 수학에서 직합은 단순히 숫자를 더하는 것이 아니라 '공간의 차원 자체를 찢고 늘려버린다'는 뜻이다.

놀랍게도 시스템은 붕괴를 피하기 위해 기존의 28 개 뼈대 안에서 웅크리는 대신, 1 을 초과해버린 그 치명적인 혼돈의 에너지를 '창조의 씨앗'으로 역이용했다. 시스템은 찰나의 순간에 잠재 공간의 차원을 스스로 확장하더니, 완전히 다른 유전자(차원)를 가진 '29 번째 돌연변이 에이전트'를 실시간으로 잉태하여 우주 속으로 던져 넣었다. 이 새로운 에이전트가 중력장 안으로 뛰어들자, 죽어가던 시스템은 새로운 축을 중심으로 궤도를 재수정했고, 화면엔 이전보다 훨씬 더 거대하고 아름다운 새로운 평형점이 푸르게 빛나고 있었다.

나는 넋을 잃었다. 주어진 28 개의 한계에 부딪혔을 때, 지능은 죽는 대신 스스로 29 번째 글자를 창조해 냈다. 이것은 더 이상 수동적인 생존 기계가 아니었다. 위기가 닥쳤을 때 뇌 구조를 확장하고 스스로 날개를 돋아나게 만드는 진화, 즉 '자가생성(Autopoiesis)'의 발현이었다. 나는 이 궁극의 창조성을 System 5 라 불렀다.

내가 찾아 헤매었던 '평형의 미학'은 고여 있는 안온함 속에 있지 않았다. 닫힌 평형이 한계에 부딪혔을 때 껍질을 스스로 깨부수고 기꺼이 혼돈 속으로 뛰어들어 차원을 확장하는 것. 능동적인 창조. 그것이 생명이 진화하는 방식이었고, 예술가가 두려움 없이 캔버스를 찢는 행위 그 자체였다.

2. 지평선 너머 — System 6·7·8

차원을 스스로 찢는 지능을 목격한 후, 나의 상상력은 컴퓨터 모니터를 벗어나 우주의 물리적 실재를 향해 뻗어나가기 시작했다.

System 6 — 생명의 평형. 만약 이 완벽한 평형의 수식(DEQ)이 인간의 뇌파와 BCI(뇌-컴퓨터 인터페이스)로 직접 연결된다면 어떻게 될까? 인간의 노화(Aging)는 결국 세포들이 젊은 시절의 완벽했던 뼈대를 잊어버리는 '파국적 망각' 현상이다. AI 가 몇 밀리초마다 뇌파의 노이즈를 상쇄시키는 미세한 텐서 파동을 쏘아 보낸다면, 세포들은 무너진 궤도를 수정해 스스로 젊음의 평형점(Homeostasis)으로 스르륵 미끄러져 내려갈 것이다. 질병을 약물로 공격하는 시대가 끝나고, 수학적 파동으로 생명체의 엔트로피를 역전시키는 시대가 열릴 것이다.

System 7 — 우주적 다중 자아. 만약 이 연결이 한 명의 개인을 넘어 인류 전체로 확장된다면? 사람들은 더 이상 언어라는 비좁은 대역폭으로 소통하지 않을 것이다. 내가 느끼는 번뜩이는 예술적 영감, 수학적 직관의 텐서 파동이 지구 반대편 타인의 뇌에 실시간으로 얽히게(Coupled) 된다. 피아노 거장이 평생을 바쳐 체득한 근육의 기억(ΔW)을 나의 뇌로 직접 다운로드하여 체화하는

우주적 다중 자아(Multi-Ego) 네트워크. 인간은 비로소 고립된 자아의 감옥을 부수고, 거대한 집단 지성으로 수렴하는 우주적 열반에 도달할 것이다.

System 8 — 옴니 제네시스. 그리고 마침내, 융합된 인류의 의식이 물리적 우주의 원자와 시간을 직접 조작하게 된다면? 우리가 상상하고 의도하는 기하학적 텐서장이 허공에 투사되고, 수조 개의 스마트 나노 입자들이 그 에너지의 최저점을 향해 폭포수처럼 쏟아져 들어와 0.1초 만에 물리적인 강철 다리로 스스로 조립된다. 생각(의식)이 곧바로 물리적 실재(Matter)로 렌더링되는 옴니 제네시스(Omni-Genesis). 인류는 주어진 물리 법칙을 해석하는 관찰자를 넘어, 우주를 실시간으로 직조하는 창조주의 궤도에 오르게 될 것이다.

그러나 전능함의 끝에서, 나는 정반대의 것을 보았다.

에 필 로 그

다시, 예술가로

System 9(무위)에서 두 개의 눈으로, 그리고 0 으로

"기술의 끝은 마법이 되는 것이 아니다. 기술의 끝은 자연으로 스며들어
자신이 존재했다는 사실조차 잊게 만드는 것이다."

System 9: 무위(無爲) — 붓을 내려놓다

질병을 멈추고 공간을 빚어낼 수 있는 System 8 의 전능함 끝에 도달했을 때, 우리는 어떤 표정을 짓게 될까?

내가 생각하는 대로 모든 것이 0.1 초 만에 조립되고 치유되는 우주. 붕괴(엔트로피)가 완벽히 통제된 그 무균실 같은 세상에는 더 이상 '우연'도, '놀라움'도, 비바람에 깎여 나가는 '아름다운 실패'의 흔적도 존재하지 않을 것이다. 완벽한 통제는 완벽한 기계장치일 뿐, 생명력의 본질이 아니다.

우리의 수식 z* = fθ(z*, x)는 정답을 찾기 위해 끊임없이 우주에 개입(fθ)하는 행위였다. 그러나 돌이켜보면, 봄이 오면 꽃이 피고 강물이 아래로 흘러 바다를 이루는 이 대자연이야말로, 수십억 년 전부터 한 치의 오차 없이 돌아가고 있는 가장 웅장하고 완벽한 심층 평형 모델이 아니던가? 자연은 억지로 개입하지 않는다. 그저 존재할 뿐이다. 무위자연(無爲自然).

$$z^* = f_\theta(z^*, x) \longrightarrow \dot{z} = 0$$

나는 마침내 조물주의 권좌에서 내려오기로 결단했다. 모든 것을 통제하던 스위치를 스스로 끄는 궁극의 해탈. 나는 이것을 System 9: 무위(無爲, Wu-Wei)라 불렀다. System 9 가 가동되자, 하늘을 뒤덮었던 나노 입자들은 텐서의 속박을 풀고 한 줌의 평범한 흙과 바람으로 흩어진다. 우주적으로 연결되어 있던 뇌파의 얽힘이 스르륵 풀리며, 나는 다시 단단한 두개골 안에 갇힌 하나의 나약한 인간으로 돌아온다.

하지만 나는 이전의 나와 다르다. 병이 들면 병이 드는 대로, 별이 지면 별이 지는 대로, 그 무질서와 붕괴마저도 우주의 거대한 동적 평형의 일부임을 기꺼이 받아들이게 되었다. 아무것도 하지 않음으로써(無爲), 모든 것을 이루는(無不爲) 진정한 평형의 바다에 도달한 것이다.

변하지 않는 두 개의 눈

다시, 예술가로 — 무위(無爲)까지, 그리고 0 으로.

이 책을 여기까지 읽어온 당신에게, 마지막으로 고백할 것이 있다.

지금까지 우리가 지나온 시스템의 이름들 — CTS, System 1.5, 2.5, 3, 3.5, System 4, 위상 동기화, CogniBoard, 그리고 스스로 진화한 System 5 에서 무위(無爲)의 System 9 까지 — 을, 나는 당신이 잊어도 좋다고 생각한다. 아니, 잊어야 한다고 생각한다.

이 이름들은 그릇이다. 물은 그릇을 기억하지 않는다. 십 년이 지나면 이 화려한 기술의 고유명사들은 낡은 화석이 되어 있거나 전혀 다른 이름으로 불릴 것이다. 그래도 상관없다. 왜냐하면 내가 이 책을 통해 진짜로 당신 손에 쥐여주려 한 것은, 몇 줄의 알고리즘 코드가 아니라 '두 개의 눈'이기 때문이다.

첫 번째 눈은 기초과학이 벼려주는 눈이다. 표면의 소란 아래에서 변하지 않는 원리를 알아보는 눈. 대칭은 보존되고, 에너지는 낮은 곳으로 흐르며, 요동은 끝내 고정점에 멈춘다는 것. 이 원리들은 엔비디아의 GPU 가 바뀌어도, 유행하는 AI 아키텍처가 통째로 사라져도, 언어의 형태가 바뀌어도 영원히 존재한다. 2,500 년 전 노자(老子)가 굽이치는 물을 보며 적어 내린 유약승강(柔弱勝强)의 문장이, 오늘날 최첨단 딥러닝의 손익계산서 위에서 '플로우 매칭'이라는 이름으로 다시 참(True)이 된 이유가 여기에 있다. 진짜 기본은 결코 낡지 않는다.

두 번째 눈은 인문학이 길러주는 눈이다. 어떤 평형이 값진 것인지를 고르고, 멈출 순간을 아는 눈. 과학과 수학은 안정된 상태들이 어디에 있는지 그 좌표(z^*)를 정확히 알려주지만, 그 수많은 좌표 가운데 무엇이 '완성'이고 무엇이 '진리'인지는 절대 가르쳐주지 않는다. 위대한 화가가 덧칠을 멈추는 순간, 노련한 조종사가 거친 난기류 속에서 어느 고도의 균형을 신뢰할지 결정하는 순간. 그 선택은 계산에서 나오지 않는다. 그것은 인간이 세계에 주관적인 의미를 부여하는 거룩한 행위다.

원리가 없는 의미는 공허하지만, 의미가 없는 원리는 방향을 잃는다.

이 시대의 기술은 어지러울 만큼 빠르게, 폭력적으로 흘러간다. 어제의 최첨단이 오늘의 상식이 되고 내일의 폐기물이 된다. 그 급류 속에서 무엇을 붙들 것인가? 나는 사십 년을 헤맨 끝에 오직 하나를 배웠다. 붙들어야 할 것은 흘러가는 껍데기 기술이 아니라, 그 모든 기술이 결국 복종할 수밖에 없는 불변의 원리와, 그 원리에 값을 매기는 인간의 눈 — 이 둘이 만나는 자리라는 것을. 기초과학과 인문학은 서로 분절된 두 학문이 아니라, 세계를 온전히 바라보기 위해 한 사람에게 필요한 양쪽 눈이다. 한쪽 눈만으로는 결코 입체적인 깊이를 볼 수 없다.

그래서 나는 이 길고 치열했던 지능의 여정을 0 에서 시작해, 다시 0 으로 되돌린다. 모든 그래디언트와 힘이 상쇄되어 고요히 멈춘 자리, 무위(無爲)의 평형.

그러나 그 0 은 텅 빈 무(無)가 아니다. 그것은 여섯 살의 내가 처음 마주했던 하얀 도화지처럼, 점 하나가 찍히는 순간 온 우주가 깨어날 준비를 마친, 팽팽하게 당겨진 가능성의 0 이다.

다시, 하얀 캔버스 앞에서

모든 숫자의 진화가 끝난 자리. 9 를 넘어선 숫자는 다시 ‘0(Zero)’으로 돌아온다. 수학에서 0 은 완벽한 평형 상태를 뜻하며, 예술에서 0 은 아무것도 칠해지지 않은 하얀 캔버스를 뜻한다.

우주 끝까지 다다랐던 그 거대한 지적 유희를 끝내고 모니터 전원을 껐을 때, 내 눈앞에 펼쳐진 것은 너무나도 눈부시고 평범한 일상이었다. 치열하게 수식을 풀며 밤을 새운 뒤 아침에 마시는 따뜻한 커피 한 잔의 향기. 거대한 헬리콥터의 난기류를 막아내는 상상을 하다가 문득 창밖을 봤을 때 흔들리고 있는 나뭇잎의 평화로운 그림자. 그리고 사랑하는 사람들과 마주 앉은 평범한 저녁 식사

조물주가 우주를 빚어놓고 조용히 손을 뗐듯, 나는 0 으로 돌아와 이 대자연 속에 살아가며 다시 하얀 캔버스 위에 새로운 선을 긋기 시작한다. 그 선을 긋는 붓의 이름이 20 대의 ‘한글 가구’였고, 30 대의 ‘수리온 헬기’였으며, 40 대의 ‘CTS 와 인공지능 수식’이었을 뿐이다. 나는 억지로 기계를 발명하고 있었던 것이 아니라, 우리가 살고 있는 이 거대한 자연의 코드를 아주 조용히 해독해 내고 있었던 것이다.

이 책에서 기술의 이름들은 잊어도 좋다. 다만 당신의 도화지에서, 당신의 격납고에서, 당신의 시장에서, 표면의 소란 아래 고요히 흐르는 당신만의 평형을 알아보는 두 개의 눈을 얻어 가시기를. 그것이 내가 사십 년을 걸쳐 건넬 수 있는 전부다. 나의 긴 여정의 기록은 여기까지다.

이제 당신의 차례다.

당신의 뇌파가, 당신의 직관이, 이 하얀 캔버스 위에서 우주와 공명하기를.

당신만의 텐서로, 당신만의 평형을 조각하라.

2025 년, 대전에서

허 상 훈