제3프로토콜

Daehyung Lee

May 4, 2026

Abstract

본 원고는 현대 사회의 핵심 병목이 더 이상 실행 능력 자체가 아니라, 무엇을 위해 실행할 것인가를 규정하는 상위 목적의 형성과 집계에 있다는 진단에서 출발한다. 자본주의는 오랫동안 생존과 결핍의 문제를 해결하는 강력한 수단이었으나, 기술의 고도화와 함께 수단의 팽창은 목적 판단의 빈곤을 오히려 더 선명하게 드러내고 있다. 제3의 소비시대가 상품보다 제안과 큐레이션을 중시했다면, 본 원고가 제안하는 제4의 소비시대는 인간이 자기 목적을 선언하고 그 선언을 중심으로 상품, 조직, 도시, 관계를 역으로 구성하는 단계다. AI와 자동화는 번역, 검색, 추상화, 설계, 협상, 생산, 조율의 비용을 급격히 낮추며 실행의 풍부화를 가져오지만, 바로 그 때문에 사회는 더욱 고해상도의 목적 표현을 요구하게 된다. 이때 표와 가격이라는 기존의 정치적·경제적 프로토콜만으로는 개인의 깊은 목적, 조건부 커밋, 책임 있는 선언을 충분히 집계할 수 없다.

이 문제의 철학적 토대를 위해 본 원고는 자아를 과거로부터 보존되는 실체가 아니라 미래를 향해 스스로를 기립시키는 비존재적 구조로 재정의하고, 목적 선언을 그 자아의 헌법적 기원으로 해석한다. 여기서 목적 선언은 개인의 존재론적 기원이면서 동시에 사회가 읽어야 할 가장 근본적인 입력이 된다. 개인 수준에서 목적 선언은 자아를 만들고, 사회 수준에서 목적 선언은 조정되어야 할 사회적 입력을 만들며, 문명 수준에서 그 집계 방식은 새로운 사회계약의 형식을 만든다. 따라서 지능의 총량이 커지고 실행비용이 낮아지는 사회는 선언된 목적과 책임 있는 커밋을 표현하고 집계하고 다시 실행으로 되돌려 보내는 세 번째 회로, 곧 표와 가격을 넘어서는 제3프로토콜을 요구하게 된다.

제3프로토콜이 목적 선언을 사회적 입력으로 받기 위해서는 두 개의 내부 엔진이 필요하다. 하나는 선언된 목적들이 충돌할 때 무엇을 더 정의로운 조정이라고 볼 것인가를 다루는 규범 좌표계이고, 다른 하나는 그 목적들이 선언, 기여, 설계, 설득, 조직 형성, 자원 이동, 계약, 생산 같은 의도적 사건으로 어떻게 번역되는지를 다루는 실행 문법이다. 이러한 관점 위에서 본 원고는 일반공리주의(Generalized Utilitarianism, GU)를 제안한다. GU는 효용을 단순한 쾌락-고통의 합이 아니라 공감, 지식, 관계 구조를 포함하는 고차원 함수로 확장하고, 정의를 이미 완성된 상태가 아니라 전지적 공감으로 수렴하는 방향성으로 이해한다. 나아가 본 원고는 이 추상적 좌표계를 전지적 공감의 형식화와 계약적 가치이전역학(Contractual Value Transfer Dynamics, CVTD)을 통해 계산 가능한 구조로 연결하려 한다. 여기서 CVTD의 '계약적'이라는 말은 법률상 계약으로 한정되지 않고, 목적을 가진 에이전트가 조건, 기대, 책임, 커밋, 위반 가능성 속에서 세계에 개입하는 사건 구조 전체를 가리킨다. 또한 삼기저동력모형(Tri-Basis Drive Model, TBDM)은 무수한 의도적 사건들이 가능성, 사회성, 정합성의 세 축에서 어떤 거시적 알짜힘으로 평균화되고 상쇄되고 증폭되는지를 설명한다. 다만 여기서 제시되는 GU와 CVTD의 수식은 완결된 방법론이라기보다, 이러한 방향성을 계산 가능한 언어로 옮기기 위한 예비적 제안이며 앞으로 상당한 개발과 확장을 필요로 한다.

마지막으로 목적은 로그로부터 자동 추출되는 정보가 아니라, 인간이 읽고 생각하고 토론하고 선언하는 과정을 통해 사회적 객체가 된다. 따라서 앞으로의 핵심 인프라는 목적형성

인프라이며, 책과 장문 텍스트는 그 형성의 가장 느리고 단단한 씨앗이자 선언으로 이어질 수 있는 장문형 정본 레이어가 된다. 이러한 맥락에서 The Channel은 인간의 목적 형성 과정을 도시 위에 배치하고, 그 과정에서 축적된 기록이 AI의 감각신경계이자 개인이 자기 이름으로 승인할 수 있는 트윈의 기반이 되게 만드는 첫 번째 현실 인터페이스로 위치 지어진다.

1 서론

"그래서 어떤 문제를 해결하는가?" 스타트업을 향한 이 관문 질문의 속내는 명확하다. 실상 "어떤 문제를 해결하여 우리에게 돈을 벌어다 줄 것인가?"라는 자본의 요구와 맞닿아 있다. 이는 문제 해결의 가치가 자본주의의 논리와 정렬될 때만 유효하다는 뜻이다. 과거에는 이 정렬이 자연스러웠을지 모르나, 지금은 다르다.

나는 더 이상 자본주의가 애덤 스미스가 말한 '보이지 않는 손'처럼 인류를 구원할 정의로운 조정자가 아니라고 생각한다. 자본주의는 이제 인간의 고통을 섬세하게 들여다보기엔 너무나 거대하고 비대해졌다. 인류의 기본값이 기아, 추위, 질병 같은 '불행'에 가까웠던 시절, 자본주의의 목표는 불행의 축소, 즉 생존이었다. 그때 자본주의는 효율적이고 윤리적인 시스템처럼 보였다. 하지만 기술과 문명의 발달로 인류의 기본값이 생존을 넘어 행복의 단계로 올라선 지금, 자본주의의 동력은 변질되었다.

마스다 무네아키의 <지적자본론>은 이러한 변화를 소비사회의 단계 전환으로 읽는다. 대량생산과 대량유통의 시대에는 상품 자체가 희소했고, 플랫폼이 넘쳐나는 제3의 소비사회에서는 상품보다 "제안"이 중요해진다. 즉 소비자는 더 많은 물건을 원하는 것이 아니라, 자신의 취향과 생활양식에 맞게 편집된 의미 있는 선택지를 원한다. 이 진단은 중요하다. 그러나 여전히 소비자를 "제안받는 존재"로 둔다는 한계를 가진다. 제3의 소비시대가 기업과 플랫폼이 인간에게 더 나은 생활양식을 제안하는 시대라면, 내가 제안하려는 제4의 소비시대는 인간이 자기 목적을 선언하고, 그 선언을 중심으로 상품, 서비스, 조직, 도시, 교육, 관계를 역으로 구성하는 시대다. 제안의 시대에서 선언의 시대로, 큐레이션의 시대에서 자기헌법의 시대로 이동해야 한다.

인간의 효용 곡선은 전형적인 로그함수의 형태를 띤다. 초기 단계, 즉 생존이 위협받던 시절에는 투입되는 자원이 조금만 늘어도 효용(행복)의 기울기는 가파르게 상승했다. 빵 한 조각이 목숨을 구하고, 지붕 하나가 삶을 바꿨다. 하지만 우리는 이미 그 가파른 구간을 지났다.

더 일반적으로 이는 경제학에서 CRRA(Constant Relative Risk Aversion) 또는 isoelastic utility로 표현되는 효용 곡선과 가깝다.

$$u(c) = \begin{cases} \dfrac{c^{1-\gamma} - 1}{1-\gamma}, & \gamma \neq 1, \\ \log c, & \gamma = 1. \end{cases} \tag{1}$$

여기서 c는 자원 또는 소비 수준이고, γ는 한계효용이 얼마나 빨리 감소하는지를 나타낸다. $\gamma = 1$인 경우가 로그 효용이며, 그보다 큰 경우에는 생존 이후의 추가 투입이 더 빠르게 둔감해진다. 이 수식이 서론에서 중요한 이유는, 자본주의의 윤리적 정당성이 어느 구간에서 강했고 어느 구간에서 약해지는지를 보여주기 때문이다. 생존 구간에서는 자본이 인간을 실제로 구원한다. 그러나 포화 구간에서는 사회 전체가 아무리 많은 예측, 광고, 추천, 생산 능력을 동원해 c를 조금 더 밀어 올려도, 인간의 총효용은 거의 움직이지 않는다. 이때 병목은 소비량의 부족이 아니라 목적 좌표의 부재가 된다. 따라서 제4의 소비시대의 과제는 더 많이 소비하게 만드는 것이 아니라,

각자가 무엇을 위해 소비하고, 무엇을 위해 조직되고, 무엇을 위해 자기 시간을 걸 것인지를 선언 가능하게 만드는 것이다.

지금 우리는 투입 대비 효용의 증가가 거의 평행선에 가깝게 누워버린, 로그 함수의 오른쪽 끄트머리에 서 있다. 이제는 어지간한 자극으로는 효용이 증가하지 않는다. 과거에는 생존을 위해 싸웠다면, 지금 자본을 굴리는 것은 더 큰 자극과 말초적인 쾌락을 향한 끝없는 갈증이다.

우리를 이끄는 이 거대하고 강력한 자본주의라는 시스템이, 고작 이 로그 함수의 끄트머리에 있는 미미한 증가를 위해 존재하는가? 수십 조의 자본과 최첨단 기술이 투입되어 만들어내는 결과물이 고작 인류가 느끼는 쾌락의 1% 남짓한 상승이라니, 이 얼마나 비효율적인 폭주인가.

이 비효율은 '삶의 수단'과 '삶의 목적' 사이의 치명적인 괴리에서 비롯된다. 영화 <죽은 시인의 사회>에서 키팅 선생은 말했다. "의술, 법률, 기술... 이것들은 삶을 유지하는 데 필요한 고귀한 수단이다. 하지만 시, 아름다움, 낭만, 사랑... 이것들이야말로 우리가 살아가는 목적이다." 우리의 비극은 수단을 극대화하는 기술에는 능숙하지만, 목적을 다루는 지혜는 잃어버렸다는 데 있다. 기술은 버전업 되고, 코드는 깃허브에 쌓이며, 지식은 복제되어 전승된다. 우리는 아이폰의 진화를 보며 기술의 축적을 체감한다. 하지만 정의, 윤리, 사랑과 같은 삶의 목적들은 그렇게 축적되지 않는다. "정의를 위한 깃허브는 없다." 기술적 수단은 지수적으로 팽창하는데 윤리적 목적은 제자리를 맴도는 이 '속도의 격차'가, 자본주의의 폭주를 제어할 브레이크를 고장 냈다.

설상가상으로, 그러나 필연적으로, 기술의 진보는 이 쾌락을 즉각적으로 충족시키는 방향으로 이루어졌고, 폭발적으로 늘어난 지식의 총량은 개개인을 '전문화'라는 이름의 쐐기로 가두었다. 역설적이게도 전체 맥락을 돌봐줄 눈은 사라졌다. 말하자면, 근대 이후 지식의 양적 폭증은 분과의 극단적 전문화, 즉 지식의 쐐기화를 낳았다. 이 과정 에서 서로 다른 체계 간 상호참조는 차단되고, 지식 생산 주체는 전체 맥락을 읽지 못한 채 부분 과업만 수행하는 기능 단위로 환원된다. 이는 마르크스가 말한 인간의 수단화가 "분과적 수행자"라는 형태로 재현된 결과이며, 동시에 니체가 비판한 "학문적 노동자"와 구조적으로 동일한 형상이다. 니체에게 학문적 노동자는 거대한 지식 공장에서 끝없이 자료를 채굴 · 주석 · 가공하지만, 산과 산 사이를 건너며 가치를 재구성하는 역할에는 도달하지 못하는 존재였다. 그는 자신을 "산과 산 사이를 넘어가는 자"로 규정했지만, 현대의 지식 체계는 이러한 산넘는 자를 예외적 개인의 덕성에 맡긴 채, 다수를 분과 내부에 고착된 노동자로 양산하는 방향으로 굳어져 있다. 인문 · 공학 · 정치 · 경제의 판단 구조는 서로를 해석하지 못하는 상태로 분리되고, 민주주의 사회에서 이 구조는 점점 더 치명적인 폐단으로 작용한다. 유권자들은 물론 선출된 사회적 결정권자들 또한 쐐기형 분과의 압력 속에서 전체 맥락을 잃게 된다.

결국 우리는 '전체'를 잃어버릴 것이다. 자본주의의 독주를 견제하고 거시적 맥락을 잡아야 할 정부가, 거대해진 자본 앞에서 무력해질 것은 자명하다. 아니, 차라리 무력해지기를 바라야 할지도 모르겠다.

민주주의는 집단지성이란 명분 아래 권력의 입장권을 국민들에게 찢어 나눠주었다. 본래 개인이 쥔 그 권력이란 것은 신기루처럼 희미한 것이었으나 더욱 절망적인 사실은, 그 입장권을 쥔 국민들이 전문화의 쐐기에 갇혀 전체 맥락을 볼 수 없게 된 것이다.

맥락을 상실한 군중이 쥔 앙상한 결정권. 이 상황에서, 우리는 도대체 어디에서 '옳은 판단'을 기대해야 하는가. 결국 우리는 다시 근본적인 질문 앞에 서야 할 것이다. 우리는 무엇을 위해 살아가는가. 우리가 그토록 치열하게 좇아야 할 대상이, 고작 그 곡선의 끄트머리를 아주 조금 더 올리는 일이었는가.

AI는 어쩌면 이 상태를 해결할 방법을 제공할 수도 있다. 그것은 번역 · 검색 · 추상화 비용을

거의 0에 가깝게 만들어 분과 경계의 거래비용을 제거한다. 이 때문에 우리는 재통합 표준, 즉 가치 · 지식 · 결정의 공통 좌표계를 필요로 한다. 정보 이동 비용이 붕괴한 시대에는, 개인이 선언한 목적을 실재적인 기준으로 삼아 기존 학문 분과를 최소 · 직교 축 좌표로 재배열함으로써, 니체가 예외적 개인에게만 맡겨두었던 "산 과 산을 건너는" 기능을 사회 전체의 기본 구조로 끌어올리고, 개인의 목적을 판단의 최상위 기준으로 복귀시키는 종류의 재통합 표준이 요구된다.

그러나 여기서 한 걸음 더 나아가야 한다. 사회는 본질적으로 분산된 개인들의 입력을 집계하여 규칙과 자원 배분과 실행으로 변환하는 조정 시스템이며, 근대 이후 이 입력을 처리하는 대표적 방식은 크게 두 가지였다. 하나는 개인의 의사를 표와 권리 행사와 공적 절차로 압축하는 정치의 프로토콜이고, 다른 하나는 개인의 선호와 우선순위를 가격, 거래, 계약, 투자, 구매, 노동, 시간 같은 희소가치의 이동으로 압축하는 경제의 프로토콜이다. 그러나 이 두 프로토콜은 모두 본질적으로 고차원적인 인간의 의사와 목적을 저차원 신호로 변환하는 손실 압축 방식이다. 정치는 정당성과 강제력을 가지지만 저빈도이고 번들링이 심하며, 경제는 더 세밀한 배분을 가능하게 하지만 결국 현재 거래 가능한 것에 대한 지불 의사만을 직접 읽는다. 어떤 세계를 원하며 어떤 조건이 충족될 때 누구와 함께 무엇을 얼마나 걸 수 있는지 같은 고해상도 목적 구조 자체는, 이 두 체계 안에서 충분히 집계되지 않는다.

이 한계가 오랫동안 치명적이지 않았던 이유는 실행 자체가 더 비쌌기 때문이다. 사람을 찾고, 연결하고, 협상하고, 계약하고, 설계하고, 생산하고, 조율하는 비용이 높았던 사회에서는 애초에 실현 가능한 선택지의 수가 적었기 때문에 입력이 거칠더라도 시스템은 작동할 수 있었다. 그러나 AI와 자동화는 이 균형을 무너뜨린다. 탐색, 매칭, 협상, 설계, 생산, 조율의 비용이 계속 하락하면 사회는 훨씬 더 다양한 조합과 훨씬 더 미세한 실행안을 실현할 수 있게 되고, 그 순간 병목은 실행 능력에서 "정확히 무엇을 실행해야 하는가"를 결정하는 입력 문제로 이동한다. 다시 말해 액추에이터가 정교해질수록 센서와 목표 함수의 품질이 병목이 된다. 지능의 총량이 커질수록 더 양질의 목적 표현이 요구되는데, 우리는 여전히 표와 가격 같은 낮은 해상도의 신호에 사회 전체를 걸고 있다.

따라서 앞으로의 핵심 문제는 데이터를 더 많이 추출하는 것이 아니라, 인간이 자신의 깊은 목적을 어떻게 사고하고, 언어화하고, 다듬고, 공적으로 선언하게 만들 것인가에 있다. 목적은 이미 완성된 채 로그 속에 숨어 있는 정보가 아니라, 읽고, 생각하고, 반박받고, 수정하고, 다시 말하는 과정을 통해 비로소 사회적 실재가 되는 구조이기 때문이다. 이 점에서 앞으로 필요한 것은 단순한 예측 엔진이 아니라 목적형성 인프라이다. 그리고 선언된 목적, 참여 조건, 조건부 커밋, 타협 가능 범위, 금지 조건, 기여 형태의 다원성을 직접 다루는 새로운 사회적 입출력 프로토콜이 요구된다. 본 원고는 이러한 문제의식을 바탕으로, 먼저 자아를 목적 선언의 구조로 재해석하고, 그 목적 선언이 어떻게 사회적 입력이 되는지를 밝힌 뒤, 표와 가격을 넘어서는 제3프로토콜의 필요성을 제시한다. 이어 제3프로토콜이 내부적으로 요구하는 규범 좌표계로서 일반공리주의(GU)를, 선언된 목적이 의도적 사건 π로 번역되는 실행 언어로서 계약적 가치이전역학(Contractual Value Transfer Dynamics, CVTD)을 제안하고, 마지막으로 The Channel을 목적형성 인프라의 첫 현실 인터페이스로 위치 지으려 한다.

2 자아의 비존재성과 목적 선언

여기서 왜 철학으로부터 시작하는가를 먼저 밝혀 둘 필요가 있다. 우리가 다루려는 문제는 단순히 더 나은 제품이나 시장 메커니즘을 설계하는 문제가 아니라, 인간이 무엇을 목적 함수로 삼고 어떤

존재로서 판단의 주체가 되는가에 대한 문제이기 때문이다. 따라서 이 원고는 해결책의 외곽부터 다듬기보다, 우리가 당연하다고 믿고 있는 가장 깊은 전제들, 곧 시간을 견디는 실체적 자아와 주어진 욕망이 판단의 출발점이라는 상식부터 먼저 해체하는 데서 출발한다. 문제를 정확히 식별하려면, 그 문제를 낳는 존재론적 전제를 먼저 의심해야 한다.

우리는 스스로를 '시간 여행자'로 인식한다. 과거의 기억을 담지하고 현재를 관통하여 미래로 나아가는 단단한 실체로서의 '자아'를 상정하는 동시에, 우리는 기계론적 결정론을 거부한다. 매 순간 선택 가능하며 미래는 닫혀 있지 않다는 '자유의지'의 신념은 도덕 판단의 기초가 된다. 우리는 방아쇠를 당긴 행위와 강풍에 쓰러진 나무가 덮친 사건을 구분하는데, 전자에만 죄와 책임을 묻는 근거는 "그는 달리 행동할 수도 있었다"는 규범적 전제에 있다. 그러나 우리가 제안하는 형식 모델 내에서, '시간을 견디며 지속하는 실체적 자아'와 '열린 미래를 전제하는 자유의지'는 구조적인 긴장 관계에 놓인다.

우리가 거주하는 세계의 증거는 본질적으로 불완전하다. 우주는 일종의 '해상도 낮은 카메라'와 같아서, 우리의 거시적 행동 궤적은 포착할지언정 내면의 미시적 고뇌, 스쳐 간 감정, 선택되지 않은 가능성(counterfactuals)까지 완벽하게 기록하지 않는다. 이 불완전한 기록 장치가 자아 해체의 기점이 된다. 자유의지가 성립하는 세계는 과거의 특정 시점에서 A를 선택할 수도, B를 선택할 수도 있었던 '열린 미래'를 의미한다. 문제는 '경로의 병합(Merge)'에서 발생한다. 과거에 서로 다른 길을 걸었더라도, 시간이 흘러 도달한 현재의 물리적 증거 상태가 구분 불가능하다면 그 차이는 정보적으로 소멸한다. 방 안에 코끼리가 있었다가 모든 흔적을 지우고 사라졌다면, 그 세계는 코끼리가 없었던 세계와 정보적으로 동치이다. 우리는 이를 "삭제 가능성(Erasability)"이라 정의한다.

과거의 특정 자아 상태가 현재에 인과적 흔적을 남기지 않고, 현재로부터 역추적 불가능하다면, 그 자아는 "정보적 죽음(Informational Death)"을 맞이한 것이다. 10년 전 오늘 점심에 느꼈던 미묘한 감정의 결은 지금 우주 어디에 기록되어 있는가? 만약 그것이 현재의 상태값에 아무런 차이를 만들지 못한다면, 시간 위를 매끄럽게 잇는다고 믿어졌던 '실체적 자아의 사슬'은 끊어진다. 사슬의 대부분은 망각과 정보 손실이라는 엔트로피의 바다로 침전하기 때문이다. 물리적 연속성이 이토록 허무하게 단절된다면, 지금 여기에 현전하며 고뇌하는 '나'의 존립 근거는 무엇인가?

2.1 발굴에서 건축으로: 비존재적 구조로서의 자아

과거의 인과 사슬이 소실된 지점에서, 자아를 바라보는 관점은 '발굴(excavation)'에서 '건축(architecture)'으로 전환되어야 한다. 자아는 과거로부터 계승되는 단단한 실체가 아니라, 미래를 향해 스스로를 기립시키는 "비존재적 구조(Non-existent Structure)"이기 때문이다. 이 구조는 "나는 과거에 무엇이었는가"라는 존재론적 질문에 답하지 않는다. 대신 "나는 무엇을 향해 가는가"라는 목적론적 질문을 통해 매 순간 스스로를 기원(originate)시킨다.

이는 칸트의 '목적의 왕국(Kingdom of Ends)' 개념과 맞닿아 있다. 칸트의 세계에서 이성적 존재자는 자연 법칙(물리적 인과)에 종속된 객체가 아니라, 스스로 법칙을 수립하고 준수하는 자율적 입법자이다. 우리의 모델도 동일한 결론에 도달한다. 물리적 증거가 소실된 빈 서판(tabula rasa) 위에서, 자아는 "나는 이러한 존재로서 나 자신을 승인하겠다"는 최상위 목적을 헌법처럼 선언함으로써 비로소 성립한다. 이 선언은 단순한 다짐이 아니라, 흩어진 가능성의 파편들을 하나의 주체로 묶어내는 중력점이다. 따라서 자아는 발견되는 '대상'이 아니라, 목적의 왕국 안에서 끊임없이 갱신되는 '선언된 지위'이다.

2.2 허수(i) 모델과 인정의 프로토콜

이러한 '선언된 자아'의 존립 방식은 수학의 허수 i와 유사하다. 실수축($\mathbb{R}$)을 우리가 발 딛고 선 물리적 인과와 데이터의 세계라고 하자. 이 실수축만으로는 $x^2+1=0$이라는 방정식의 해를 구할 수 없다. 즉, 물리적 인과론만으로는 "왜 내가 책임을 져야 하는가?", "무엇이 인간을 존엄하게 만드는가?"라는 질문에 답할 수 없다. 물리적 세계에서 책임은 인과의 결과일 뿐이지만, 윤리적 세계에서 책임은 의지의 발현이기 때문이다.

여기서 우리는 실수축에 직교하는 새로운 차원, 즉 허수축을 도입한다. i는 수직선 상의 위치를 묻는다면 답할 수 없지만, $i^2=-1$이라는 관계와 작동 방식으로서 엄연히 존재한다. 자아 역시 마찬가지이다. 생물학적 뇌를 해부해도 '자아'라는 물질은 발견되지 않으나, "스스로 목적을 수립하고 책임지는 구조"라는 작동 방식을 도입할 때 비로소 인간이라는 방정식은 해를 갖는다. 실체가 없다는 것(non-substantiality)은 결함이 아니라, 물리적 결정론의 외부에서 작동하기 위한 필수 조건이다.

이 비존재적 구조는 고립된 개인의 환상이 아니다. 그것은 타인과의 관계망 속에서 유지되는 거대한 "인정의 프로토콜(Protocol of Recognition)"이다. 블록체인은 그 자체로 0과 1의 나열에 불과하나, 참여자들이 "이 비트열을 가치로 인정하겠다"는 집단적 선언(합의 알고리즘)을 공유할 때 실물 화폐보다 견고한 가치를 획득한다. 인간의 자아도 이와 흡사하다. 우리에게 영혼이라는 '중앙 서버'나 불변의 실체는 부재할지 모른다. 그러나 우리는 서로를, 그리고 나 자신을 자유 의지를 가진 주체로 대우하겠다는 '상호 인정의 합의' 위에서 살아간다. 칸트의 목적의 왕국이 이상적 모델이라면, 이 프로토콜은 그 모델의 현실적 구현체다. 중앙 권력 없이 신뢰를 창출하는 블록체인처럼, 우리는 영원불멸의 영혼 없이도 서로의 선언을 교차 검증하며 '나'라는 존재를 블록처럼 축조한다.

결국 우리의 논증은 허무주의가 아닌 구원으로 귀결된다. 과거의 정보가 삭제된다는 사실은, 역설적으로 과거가 우리를 영원히 규정하거나 구속할 수 없다는 자유의 가능성을 개방한다. 그 정보의 공백을 채우는 것은 물리적 잔여물이 아니라 우리의 선언과 목적이다. "나"는 박물관에 박제된 조각상이 아니다. "나"는 매 순간 불완전한 증거 위에서 재구성되는 현상이자, 스스로 수립한 목적을 헌법처럼 승인하며 작동하는 비존재적 구조다. 실체가 없다는 것은, 우리가 과거의 유령이 아니라 미래를 향한 선언으로서 매번 새롭게 기원할 수 있음을 의미한다. 우리는 완전히 기억되지 않기에 비로소 자유롭고, 서로를 목적으로 대우하기로 선언하기에 비로소 존재한다.

3 목적 선언의 사회화: 개인의 헌법에서 사회적 입력으로

자아가 과거로부터 보존되는 실체가 아니라면, 인간은 현재에서 미래를 향해 스스로를 선언함으로써만 주체가 된다. 따라서 목적 선언은 개인의 존재론적 기원이자, 사회가 읽어야 할 가장 근본적인 입력이다. 이 전환이 본 원고의 핵심 경첩이다. 목적은 개인 안에서는 자아를 기립시키는 헌법이고, 사회 안에서는 조정되어야 할 원초 입력이다.

기존 경제학은 인간을 선호의 보유자로 읽었고, 기존 민주주의는 인간을 의견의 보유자로 읽었으며, 기존 플랫폼은 인간을 반응의 발생원으로 읽었다. 그러나 인간은 그보다 앞서 목적을 선언하는 존재다. 목적은 이미 완성된 채 발견되는 데이터가 아니라, 형성되고 승인되고 책임지는 자기헌법이다. 그러므로 사회는 인간을 소비자, 유권자, 사용자로만 읽는 데서 멈출 수 없고, 목적선언자로 읽을 수 있어야 한다.

이 전환은 존재론에서 제도론으로 바로 이어진다. 개인 수준에서 목적 선언은 자아를 만든다. 사회 수준에서 목적 선언은 사회적 입력을 만든다. 문명 수준에서 목적 선언의 집계 방식은 새로운 사회계약을 만든다. 만약 목적 선언이 개인 자아의 헌법이라면, 사회는 그 헌법들을 읽고 조정하는 상위 헌정 시스템을 가져야 한다. 다음 절의 제3프로토콜은 바로 이 요구에서 출발한다.

4 투표와 가격을 넘어: 제3프로토콜

여기서 다시 사회 전체의 입력 구조로 돌아갈 필요가 있다. 정치는 표를 집계하고 경제는 가격과 계약을 집계한다. 그러나 실행 비용이 충분히 낮아진 사회에서는 이 두 프로토콜만으로는 더 이상 사회의 조정 가능성을 충분히 회수할 수 없다. 특히 "이런 학교가 실제로 생기면 보내겠다", "이런 규칙이 보장되면 이주하겠다", "이 정도의 사람들이 모이면 나는 돈과 시간과 전문성을 걸겠다" 같은 조건부 집단형 수요는 정치로는 지나치게 거칠고 느리며, 경제로는 아직 거래 객체가 형성되기 이전이어서 충분히 포착되지 않는다.

이때 필요한 것은 표와 가격을 대체하는 체계가 아니라, 정치와 경제와 병렬로 작동하는 또 하나의 사회적 입출력 프로토콜이다. 정치가 표를 통해, 경제가 가격을 통해 사회적 의사를 표현하고 집계하고 다시 실행으로 되돌려 보내는 체계였다면, 지능의 총량이 커지고 실행비용이 낮아지는 사회는 목적 그 자체를 표현하고 집계하고 다시 실행으로 되돌려 보내는 세 번째 프로토콜을 요구하게 된다. 이것이 여기서 말하는 제3프로토콜이다. 제3프로토콜의 핵심은 개인이 자신의 목적을 공적으로 선언하는 행위가, 동시에 참여 의사와 자원 제공 의사와 장기 커밋의 가능성을 함께 드러내는 새로운 사회적 신호가 된다는 데 있다. 다시 말해, 그것은 "무엇을 사고 싶은가"나 "누구에게 표를 줄 것인가"만이 아니라, "어떤 세계를 원하며 어떤 조건이 충족되면 누구와 함께 무엇을 얼마나 걸 수 있는가"를 표현하고, 그 표현을 다시 조직과 계약과 생산으로 되돌려 보내는 세 번째 경로다. 이때 선언된 목적은 흩어진 의견으로 남지 않고, 서로 비슷한 선언들과 결합하여 의도 클러스터를 이루며, 그 자체가 아직 존재하지 않는 재화와 제도와 조직을 향한 생산 신호가 된다.

핵심은 이것이 단순한 의견 게시판이나 더 정교한 수요 예측 엔진이 아니라는 점이다. 제3프로토콜은 사용자의 로그에서 욕구를 추출하는 체계가 아니라, 인간이 읽고, 생각하고, 토론하고, 목적을 다듬고, 마침내 그것을 선언하게 만드는 구조에 가깝다. 그리고 미래의 대리 지능들은 그 선언 이력과 수정 과정과 가치 우선순위를 먹고 들어와, 서로의 목적 간 정합성을 맞추고 공동 목적 후보와 계약 가능한 표현까지 압축하게 된다. 따라서 제3프로토콜은 인간의 숙고를 공적 목적 선언으로 바꾸고, 그 선언을 다시 제도와 조직과 대리 지능의 응답으로 되돌려 보내는 사회적 입출력 프로토콜이다.

뒤에서 다룰 GU와 CVTD의 관점에서 보면, 제3프로토콜은 철학과 실행을 연결하는 제도적 층이다. GU가 왜 개인의 선언된 목적을 판단의 최상위 기준으로 복귀시켜야 하는지를 말하고, CVTD가 그러한 목적들이 어떻게 의도적 사건 π의 언어로 계산될 수 있는지를 보여준다면, 제3프로토콜은 바로 그 선언들이 사회적으로 표현되고, 집계되고, 다시 실행으로 되돌아오는 통로를 설명한다. 따라서 그것은 부가적인 서비스가 아니라, 선언된 목적의 사회적 가시성과 계산 가능성을 여는 새로운 조정 프로토콜이다.

이를 도식화하면 다음과 같다.

프로토콜	기본 입력	집계 방식	사회적 출력	핵심 한계
1: 표	표, 권리, 의견	선거, 의회, 공론장, 절차적 정당성	법, 정책, 권력 배분	저빈도, 번들링, 다수결 압축
2: 가격	가격, 구매, 투자, 노동, 계약	시장, 기업, 금융, 거래 네트워크	재화, 서비스, 자본 배분	지불 가능한 현재 선호 중심
3: 목적	목적 선언, 조건부 커밋, 기여 의사	의도 클러스터, GU적 조정, 대리 지능 간 협상	조직, 제도, 생산, 의도적 사건 π	아직 표준화되지 않음

제1프로토콜이 "누가 권한을 갖는가"를 묻고, 제2프로토콜이 "무엇에 얼마를 지불할 것인가"를 묻는다면, 제3프로토콜은 "어떤 세계를 원하며, 그 세계가 가능해지는 조건에서 무엇을 걸 것인가"를 묻는다. 따라서 제3프로토콜은 정치와 경제의 폐지가 아니라, 둘이 손실 압축해 온 목적 구조를 더 높은 해상도로 사회에 입력하는 보완 회로다.

5 제3프로토콜의 내부 질문: 목적은 어떻게 조정되는가

제3프로토콜의 필요가 확인되면 곧바로 다음 질문이 발생한다. 목적 선언을 사회적 입력으로 받는다고 해서, 그 선언들이 자동으로 정합적 질서를 이루는 것은 아니다. 목적은 서로 충돌하고, 서로를 오해하며, 때로는 자기 자신조차 불명료하게 이해한다. 따라서 제3프로토콜은 단순한 게시판이나 집계 장치일 수 없고, 선언된 목적들이 어떤 조건에서 더 정의롭게 조정되는지 판단할 규범 좌표계를 필요로 한다.

이 지점에서 일반공리주의(Generalized Utilitarianism, GU)가 등장한다. GU는 제3프로토콜의 존재 이유를 다시 설명하기 위한 장식적 철학이 아니라, 제3프로토콜 내부에서 목적 충돌을 해석하고 조정하기 위한 규범 엔진이다. 다시 말해 제3프로토콜이 목적 선언을 사회적 입력으로 받는 사회론이라면, GU는 그 입력들이 충돌할 때 무엇을 더 나은 조정으로 볼 것인가를 묻는 규범론이다.

6 선언된 목적에서 GU로

자아를 과거의 실체가 아니라 미래를 향한 선언으로 이해한다면, 윤리와 제도 역시 그 선언의 구조를 중심으로 다시 정렬되어야 한다. 기존 체계는 대체로 이미 드러난 선호를 집계하거나 이미 성립한 거래를 조정하는 데 능숙했지만, 어떤 목적이 형성되고 어떤 목적이 공적으로 승인되며 어떤 목적들이 서로 조정 가능한가를 판단의 중심에 두지 않았다. 그러나 자아가 목적 선언을 통해 기원한다면, 사회적 판단 역시 그 선언된 목적들의 관계를 중심으로 설계되어야 한다.

이 지점에서 일반공리주의(GU)는 단순한 공리주의의 변형이 아니라, 목적 선언을 사회적 판단의 상위 기준으로 복귀시키려는 시도로 이해될 수 있다. 여기서 중요한 것은 인간에게 하나의 보편 목적을 부과하는 것이 아니다. 오히려 각 개인과 집단이 선언한 목적들의 다원성을 전제로 하면서도, 그 목적들이 서로 어떤 구조를 이루며 어디에서 충돌하고 어디에서 수렴할 수 있는지를 묻는 새로운 좌표계를 세우는 일이다. GU는 바로 이 좌표계를 향한 첫 번째 철학적 제안이다.

7 일반공리주의(GU)의 제안

자아의 본질이 '발견되는 실체'가 아니라 '선언되는 목적'이라면, 우리에게 필요한 사회 시스템 또한 근본적으로 달라져야 한다. 더 이상 맹목적인 이윤이나 효율이 기준이 될 수 없다. 개인이 선언한 목적을 실재적인 기준으로 삼아 기존 학문 분과를 재배열하고, 니체가 예외적 개인에게만 맡겨두었던 "산과 산을 건너는" 기능을 사회 전체의 기본 구조로 끌어올리는 새로운 좌표계가 필요하다. 우리는 이것을 일반공리주의(Generalized Utilitarianism, GU)라 부른다.

일반공리주의(GU)는 이러한 조건에서 "정의"를 다시 묻고, 고전적 공리주의를 다음 세 가지 차원에서 일반화한다.

1. 효용을 단순한 쾌락·고통의 합이 아니라, 지식 상태, 관계 구조, 공감 가능성을 포함하는 고차원 함수로 확장한다.

2. 판단의 1차 기준을 가능한 선택지들 중 "무엇이 더 효율적인가"가 아니라, 각 개인이 선언한 목적들 간의 조정과 수렴으로 옮긴다.

3. 정의를 이미 주어진 상태가 아니라, 가능한 미래 궤적들의 기대효용이 수렴해 가는 방향, 즉 "정의에 대한 근사(approximation to justice)"로 이해한다.

7.1 핵심 직관: 무지와 목적

GU를 지탱하는 핵심 직관은 다음과 같다.

첫째, "회피 가능한 모든 고통은, 충분히 인과를 거슬러 올라가면 궁극적으로 무지(ignorance)의 결과로 수렴한다"는 테제이다. 만약 전지(omniscience)가 있었다면, 구조·제도·관계를 고통스럽게 설계하지 않았을 것이라는 강한 가정이다. 이는 "고통의 일부는 세계 구조의 필연"이라는 체념적 공리에 대한 반론이자, 고통을 세계의 본질이 아니라 정보의 결핍에 귀속시키려는 시도이다.

둘째, "판단의 1차 근거는 가능성(what is possible)이 아니라 가치관으로부터의 목적(what ought to be pursued)이어야 한다"는 테제이다. 가능성은 수단을 정교화하는 정보이고, 목적은 판단을 가르는 기준이다. 목적의 다원성은 사실이자 규범이며, "인간의 존재 목적"을 하나의 보편 명제로 묻는 기획은 범주 오류에 가깝다. 목적은 발견되는 본질이 아니라, 각 개인·집단이 선언하고 설계하는 구조로 취급되어야 한다.

7.2 자유의지와 헌법적 의지

자유의지에 관해서 GU는 물리적 실재 여부에 대한 형이상학적 논쟁을 정지하고, 규범적 공리로서의 자유의지만을 채택한다. 즉, 우리는 개인을 사회적 의사결정의 원자적 단위로 전제하는 세계 안에서 살아야 한다. 이는 논증으로 "입증"해야 할 명제가 아니라, 인간이 자연 현상이나 알고리즘과 구분되지 않은 순수 수단으로 전락하는 것을 막기 위한 최소 방어선이다. 자유의지를 공리로 채택하지 않으면 "누가 책임지는가, 누가 목적을 세우는가"라는 질문 자체가 무의미해지고, 목적은 사라지며 인간은 전면적으로 도구화된다.

자유의지를 공리로 채택한다는 것은 자아가 순수한 인과 연쇄로 완전히 환원되지 않는다는 것을 인정하는 셈이다. 이때 자아는 외부 인과가 아니라 자기 내부의 의지로부터 최소 한 번은 자

신을 규정해야 한다. 이 기원적 의지(헌법적 의지)를 언어적으로 고정한 결과가 곧 목적 선언이다. 따라서 여기서 말하는 '목적'은 단순한 욕구 목록이 아니라, 자아가 스스로를 기원시키기 위해 채택한 최상위 의지의 내용이다.

이 관점에서 개인의 목적 선언은 개인의 헌법으로 재정의된다. 국가의 헌법이 그 정체성과 경계를 규정하듯, 최상위 목적 선언은 자아의 경계를 규정한다. 이 최상위 선언이 임계적으로 변경되면, 국가가 '제 n공화국'으로 불리듯 자아도 '제 n-자기'로 재기원한다.

7.3 괴델의 불완전성과 전략적 무지

다만 GU는 절대적, 외재적 심판자가 아니다. 한 사회 내에서 낙태, 안락사, 혁명과 같은 논쟁적 사안에 대해 "GU에 따르면 답은 X다"라고 공개적으로 선언하는 행위 자체가 새로운 정책 π이며, 그 발화가 다시 미래 분포와 효용에 영향을 미친다. 즉, 계 안에 있는 주체는 계 전체에 대한 완전한 판정을 구조적으로 내릴 수 없다. 이는 형식 체계 내부에서 그 체계의 모든 참을 산출할 수 없다는 괴델의 불완전성 정리를 사회적 의사결정 차원에서 반복하는 셈이다.

GU는 바로 이 한계를 정면에서 인정하며, 일종의 "모른 척하기 게임(game of strategic/willed ignorance)"을 도출한다. 어떤 물음에 대해 계산상 한쪽이 더 유리해 보이더라도, 그 답을 '언제, 누구에게, 어떤 형식으로 말할 것인가' 자체가 또 다른 결정 변수가 되며, 어떤 경우에는 "알아도 말하지 않는 것"이 정의 근사에 더 가깝다는 역설이 발생한다. GU는 이 자기참조성을 숨기지 않고, "진실을 말하지 않는 것의 윤리"를 이론 내부의 요소로 포함한다.

7.4 철학에서 시스템으로

철학은 아름답지만, 시스템은 작동해야 한다. 우리는 이 거대한 비전을 막연한 희망 사항으로 남겨두지 않기 위해, 윤리 판단을 효용의 네트워크 상에서 계산 가능한 수식으로 공식화했다. 이것은 현실의 복잡한 도덕적 결정을 평가하고 근사(Approximate)할 수 있는 이상적 기준, 마치 물리학에서의 "마찰 없는 평면"과 같은 좌표계를 제공하려는 시도이다. 다만 여기서 제시하는 수식들은 완결된 방법론이나 닫힌 계산 체계가 아니다. 그것들은 윤리 판단과 목적 조정의 방향성을 어떤 식으로 계산 가능한 언어로 옮길 수 있을지를 탐색하기 위한 예비적 제안이며, 이후 상당한 개발과 확장의 여지를 의도적으로 남겨 둔다.

우리는 이상적인 정의로운 상태를 모든 행위자의 효용이 타인의 효용과 공감까지 포함하여 최적으로 해결된 상태로 상정한다. 이 정의로운 상태를 구현하는 가상의 전지적 판단자를 상정하면, 이 판단자는 각 개인의 주관적 행복과 그들이 타인에게 갖는 공감 정보를 모두 취합하여 평가를 내릴 것이다. 우리는 이러한 이상적인 판단 과정을 "전지적 공감"이라는 개념으로 표현하며, 그 전제 조건으로서 공감적 및 인지적 완성도를 $EERI = 1$로 정의한다.

결국 우리의 목표는, 이 전지적 공감 상태를 실제로 근사하는 원리를 찾는 것이다. 물론 인간은 전지적 상태에 도달할 수 없으므로, 우리는 현실의 제약 내에서 근접 정의 (Proximal Justice)를 실현하고자 한다. 이는 현재의 인지적 한계에서 달성 가능한 정의의 최선의 근사치로서, 전지적 공감 상태에 최대한 가깝도록 의사결정 구조를 조정하는 것을 의미한다.

본 원고의 나머지 부분은 다음과 같이 구성된다. 앞에서는 자아의 선언 구조가 곧 사회적 입력 구조로 이어짐을 보였고, 그 결과로 표와 가격을 넘어서는 제3프로토콜의 필요성을 먼저 제시했다. 이제부터는 그 제3프로토콜이 내부적으로 어떤 규범 좌표계와 실행 언어를 필요로 하는지를 다룬다. 먼저 자아의 선언 구조에서 GU로 넘어가는 철학적 경로를 다시 정리하고, 이어 전지적

공감의 형식화를 통해 이를 계산 가능한 언어로 옮긴다. 그 다음 CVTD를 실행층의 예비적 스케치로 제안하고, 마지막으로 왜 읽기가 그 출현 조건을 담당하는 첫 번째 웨지가 되는지로 이어진다.

이론의 위계. 이후에 등장하는 여러 개념들은 병렬로 늘어선 별개의 이론이 아니라, 서로 다른 층위를 맡는 하나의 구조로 읽혀야 한다. 제3프로토콜은 중심 명제, 곧 목적 선언을 표와 가격에 이은 세 번째 사회적 입력으로 받아야 한다는 사회론이다. GU는 그 입력들이 충돌할 때 무엇을 더 정의로운 조정으로 볼 것인가를 다루는 규범 엔진이다. CVTD는 선언된 목적이 모든 의도적 사건 π로 내려올 때 그 가치 변화와 정렬을 기록하는 실행 엔진이다. TBDM은 무수한 미시 사건들이 가능성, 사회성, 정합성의 세 축에서 어떤 거시적 알짜힘으로 나타나는지를 설명하는 거시 해석 모델이다. The Channel은 이 전체 구조가 현실로 들어가기 위한 첫 진입점, 곧 목적형성 인프라다. 따라서 아래의 논의는 닫힌 완성 체계라기보다, 인간론에서 사회론, 규범론, 실행론, 현실 웨지로 이어지는 하나의 설계도에 가깝다.

8 GU의 형식화로

앞선 절에서 GU는 자아의 선언 구조로부터 도출되는 사회적 판단 원리로 제안되었다. 이제 필요한 것은 이 제안을 단순한 철학적 선언으로 남겨두지 않고, 효용과 공감과 판단의 구조를 계산 가능한 형태로 옮기는 일이다. 다시 말해, 목적의 다원성과 공감의 비대칭성과 정보의 불완전성을 전제로 하면서도, 정의가 어떤 방향으로 수렴하는지를 형식적으로 다룰 수 있어야 한다.

다음 절의 형식화는 바로 이 문제를 다룬다. 여기서 중요한 것은 수식이 철학을 대체한다는 뜻이 아니라, 철학이 제시한 좌표계를 연산 가능한 구조로 번역한다는 점이다. GU가 정의의 방향성을 제시하는 거시적 프레임이라면, 그 정식화는 그 방향성이 실제 판단과 설계의 언어로 어떻게 변환될 수 있는지를 보여주는 중간 단계가 된다.

9 이론적 틀: 전지적 공감의 형식화

본 섹션에서는 일반공리주의의 수학적 모델을 구축하고, 도덕성을 효용과 공감의 네트워크 상에서 계산하는 방법을 설명한다. 이는 현실의 도덕적 결정을 평가하고 근사할 수 있는 이상적 기준을 제공하려는 시도로, 물리학에서의 "마찰 없는 평면"과 같은 개념으로 볼 수 있다.

9.1 행위자, 내재 상태 및 효용 H_n

우리는 도덕적 행위자들을 $n = 1, 2, \ldots, N$으로 표기하고, 각 행위자 n은 그들의 복지와 관련된 모든 요소 (예를 들어, 신체적 건강, 정신적 상태, 물질적 자원, 욕망 등)를 포괄하는 내재 상태 I_n을 가진다고 가정한다. 이 내재 상태를 바탕으로, 각 행위자의 효용 또는 행복 H_n은 자신에 대한 상태와 타인에 대한 공감 반응을 모두 함수로 표현할 수 있다. 구체적으로, 우리는 다음과 같이 정의한다:

$$H_n \;=\; I_n \;+\; \sum_{i=1}^{N} C_{n,i} \cdot EERI_{n,\pi,i} \cdot H_i \,, \tag{2}$$

여기서 $C_{n,i}$는 행위자 n이 행위자 i의 행복에 대해 가지는 공감 계수로, n이 i의 효용에 어느 정도 영향을 받는지를 나타낸다. 만약 $C_{n,i} = 0$이라면, 행위자 n은 i의 행복에 전혀 관심이 없음을

의미하며, 양의 값이라면 n은 i의 행복 변화에 따라 일정한 효용 또는 불효용을 함께 느낀다는 것을 의미한다. 이와 같이 각 행위자의 효용 H_n은 단지 자기 자신의 상태 I_n뿐만 아니라 타인과의 감정적 상호작용을 반영하여, 공감, 연대, 혹은 도덕적 분노와 같은 감정들을 포괄할 수 있다.

이 식은 한 행위자의 행복이 다른 행위자들의 행복에 의존함을 보여준다. 특히, 식 (1)에서는 H_i가 다시 다른 존재자들의 영향으로 결정되므로, 전체 $H = (H_1, H_2, \ldots, H_N)$ 구조는 상호 참조적인 성격을 갖는다. 이러한 상호 참조 구조는 뒤에서 고정점 해석을 통해 더욱 명확히 다루게 될 것이다.

9.2 교육된 공감 해상도 지수 (EERI)

실제 인간은 완전한 합리성이나 무한한 정보를 갖추고 있지 않다. 또한 인간은 타자의 상태를 이해할 때, 단순히 '모르거나 무시' 하는 것뿐 아니라, 종종 왜곡되거나 반대로 해석된 상태에서 인지적 평가를 내리기도 한다. 이를 반영하기 위해 일반공리주의(GU)에서는 각 행위자 n이 정책 π 하에서 타자 i의 내재 상태를 얼마나 정확하게 이해하고 정합적으로 해석하는지를 나타내는 지표로 교육된 공감 해상도 지수 (Educated Empathic Resolution Index, EERI)를 도입한다. EERI는 다음과 같은 범위를 갖는다:

$$EERI_{n,\pi,i} \in [-1,\, 1]\,, \tag{3}$$

- $+1$: 타자 i의 감정과 효용을 완전히 정합적으로 이해함 (이상적 명료성)
- 0: 감정 및 효용에 대한 정보 없음 또는 몰이해 상태 (무지)
- -1: 타자 i의 감정을 완전히 왜곡하거나 반대로 이해함 (오해 혹은 적대적 해석)

이 지수는 단순한 이해 유무를 넘어, 그 이해의 방향성과 정합성까지 반영하며, 감정적 또는 인지적 왜곡을 수치화할 수 있게 한다. 예를 들어, 타자에 대한 편견, 적개심, 증오 등은 $EERI_{n,\pi,i} < 0$으로 표현될 수 있으며, 이는 전체 정의 판단에서 타자 i의 효용이 반대로 반영되는 효과를 가져온다. 즉, n이 i를 혐오하는 경우 $C_{n,i}$가 양수이더라도 $EERI_{n,\pi,i}$가 음수라면, i의 행복 증가는 n에게 오히려 불행으로 작용하게 된다.

EERI 개념은 인간 판단의 인지적 한계와 편향을 통합하기 위한 장치다. 범위 무감각, 내집단 편향, 현재 편향 등 실제 사람들의 여러 인지적 제한으로 인해, 동일한 상황에 대해서도 개인마다 서로 다른 도덕적 평가가 나타날 수 있다. 이러한 현실을 모델에 반영하기 위해 EERI는 각 행위자의 인지적·감정적 정련도 (clarity)를 나타낸다. 낮은 EERI 값은 비합리적 판단이나 편향을 의미하며, 높은 EERI 값은 보다 숙고되고 교육된 윤리 판단을 반영한다. 결과적으로 EERI는 편향으로 인한 왜곡을 구조적으로 보정하거나 드러내는 역할을 한다.

9.3 정의근사 (Proximal Justice)

일반공리주의(GU)에서 정의 (Justice)는 단순한 정책 결과의 합산이 아니다. 정의는 존재자 간의 감정, 공감, 이해 구조가 윤리 연산자 π에 의해 정렬되어, 전체 효용 구조 $H = (H_1, H_2, \ldots, H_N)$가 정보론적 고정점에 수렴하는 상태로 정의된다.

우리는 이러한 구조적 조정을 설명하기 위해 존재론적 자기 확장 (Ontological Self-Extension, OSE) 개념을 도입한다. 행위자 n이 타자 i에 대해 갖는 존재론적 자기 확장 정도를 다음과 같이 정의한다:

$$OSE_{n,i} := C_{n,i} \cdot EERI_{n,\pi,i}\,, \tag{4}$$

특히 자기 자신 ($i = n$)에 대해서는 $C_{n,n} = 1$이므로, 다음과 같이 간단히 쓸 수 있다:

$$OSE_{n,n} = EERI_{n,\pi,n} \, . \tag{5}$$

정의(Justice)는 모든 존재자 n, i 쌍에 대해 $OSE_{n,i}$가 1에 가까워지는 극한에서 수렴한다. 즉, 존재자들이 서로를 완전히 이해하고 동일시하는 이상적 극한 상태에서 정의가 실현된다고 볼 수 있다. 이를 수식으로 나타내면:

$$\text{Justice} := \lim_{OSE_{n,i} \to 1} \pi^* \, , \tag{6}$$

여기서 π^*는 위 극한 상태에 해당하는 궁극의 윤리 연산자를 가리킨다. 현실에서는 모든 관계에서 $OSE = 1$인 전지적 공감 상태에 도달할 수 없지만, π^*는 이 이상을 향해 수렴하는 방향성을 나타낸다.

현실적 판단 구조: 근접 정의 π^*. 현실에서는 완전한 OSE 수렴이 불가능하므로, GU는 현실적 공감/이해 구조를 반영한 근사 최적화를 통해 정의 판단을 수행한다. 이를 공식화하면 다음과 같다:

$$\pi^* \; = \; \arg\max_{\pi} \sum_{n=1}^{N} EERI_{n,\pi,n} \cdot H_n \, , \tag{7}$$

즉, 각 행위자 n의 효용 H_n에, 그 행위자가 현재 자기 자신을 얼마나 명료하게 이해하는지를 나타내는 $EERI_{n,\pi,n}$ (자기 자신에 대한 이해 지수)을 가중치로 곱한 총합을 최대화하는 π를 찾는 것이다. 여기서 π는 단순한 결과의 집합이 아니라, 앞서 언급한 윤리 연산자로서 전체 구조에 영향을 미치는 변수다.

각 존재자의 총합적 효용 H_n은 다음과 같이 구성된다:

$$H_n = I_n + \sum_{i=1}^{N} C_{n,i} \cdot EERI_{n,\pi,i} \cdot H_i \, , \tag{8}$$

이 구조는 효용 H_n이 단순히 개인의 주관적 반응이 아니라, 공감 계수 C와 이해 해상도 $EERI$를 통해 구조적으로 전달되고 반영되는 값임을 보여준다. 특히 H_i는 재귀적으로 다시 타자들의 감정을 포함하므로, 전체 구조는 다음과 같은 무한 자기참조 체계로 확장된다:

$$H_n = I_n + \sum_{i \neq n} C_{n,i} \cdot EERI_{n,\pi,i} \cdot \left(I_i + \sum_{j \neq i} C_{i,j} \cdot EERI_{i,\pi,j} \cdot H_j + \cdots \right) , \tag{9}$$

이와 같은 자기참조 구조는 일련의 반복되는 내포를 보여주며, 이론적으로 무한히 계속될 수 있다. 따라서 이러한 시스템에서 정의로운 판단이 성립하기 위해서는, 전체 효용 벡터 H가 안정된 값을 가져야 한다.

고정점 해석. 위 자기참조 구조는 고정점 형태로 수렴되어야 정의 판단이 수학적으로 성립한다. 전체 효용 벡터 $\mathbf{H}$는 다음 고정점 조건을 만족해야 한다:

$$\mathbf{H} = F_\pi(\mathbf{H}) \, , \tag{10}$$

여기서 F_π는 윤리 연산자 π에 의해 유도되는 공감-이해-감정 흐름의 전체 변환을 나타내는 함수(연산자)이다. 만일 어떤 $\mathbf{H}^*$가 이 식을 만족하는 고정점으로 존재하고, 실제 과정이 $\mathbf{H}^*$에 수렴할 수 있다면, 정의는 안정된 형태로 존재한다고 말할 수 있다. 이때 π^*는 $\mathbf{H}$가 $\mathbf{H}^*$로 수렴하는 경로의 방향성으로 해석된다. 다시 말해, π^*는 H 공간에서 효용 분포의 변화 방향을 결정하는 연산자이며, 그 연산의 반복 결과가 정의로운 상태로 수렴한다.

윤리 연산자 π의 본질. 정책 π는 단순한 행위 선택이 아니라, 존재자 기반 윤리 정보 구조 전체에 작용하는 고차원 비선형 윤리 연산자이다. 보다 형식적으로, π는 세계의 존재 상태 O에 작용하여 다음 상태 O'을 유도하는 변환으로 볼 수 있다:

$$\pi : O_t \ \rightarrow \ O_{t+1} \, , \tag{11}$$

여기서 O_t는 다음과 같은 전체 구조를 포함한다:

$$O_t = \{ \, I_n(t), \, x_n(t), \, H_n(t), \, C_{n,i}(t), \, EERI_{n,\pi,i}(t) \, \}_{n,i} \, ,$$

즉, π는 세상의 존재론적 구조 O 전체에 작용하여, 감정, 가치, 공감, 이해 등의 모든 요소들을 재구성한다. π^*는 이 연산자들 중에서 정의의 수렴 방향성을 가장 잘 실현하는 이상적인 연산자를 의미한다.

9.4 GU의 수학적 해석 가능성: 비교, 중심성, 해 선택

여기서 한 가지 주의할 점이 있다. 아래의 수학적 개념들은 GU를 이미 완성된 정리 체계로 입증하기 위한 권위의 목록이 아니라, GU가 어떤 종류의 계산 언어로 발전할 수 있는지를 보여주는 해석 렌즈들이다. 특히 I_n과 H_n은 서로 다른 개인의 쾌락을 직접 같은 자로 재는 단순한 행복 점수가 아니다. 개인 간 목적충족도는 원초적으로 직접 비교하기 어렵다. 같은 1시간, 같은 1달러, 같은 칭찬도 어떤 사람에게는 생존의 조건이고, 어떤 사람에게는 거의 무의미한 잡음일 수 있다. 따라서 GU에서 I_n은 각 개인의 선언된 목적, 신체적 조건, 관계 구조, 시간축, 위험 회피도, 자기이해 정도를 반영하여 그 개인의 가치 기저 안에서 정규화된 내재 상태로 보아야 한다. H_n은 그 내재 상태가 타인의 효용, 공감, 이해 해상도와 연결되며 형성되는 네트워크적 값이다.

이 점에서 GU의 효용 벡터 $\mathbf{H}$는 단순 합산표라기보다 PageRank 또는 고유벡터 중심성(eigenvector centrality)에 가까운 구조를 가진다. 어떤 존재자의 효용은 고립된 점수가 아니라, 다른 존재자들과의 연결, 공감 가중치, 이해 해상도, 자기이해의 반복적 상호참조 속에서 안정화된다. 즉 중요한 것은 "누가 더 큰 주관적 행복을 말하는가"가 아니라, 전체 공감망이 어떤 고정점으로 수렴하며, 그 고정점이 각자의 선언된 목적과 얼마나 정렬되어 있는가이다.

정의 판단은 이러한 고정점 위에서의 해 선택 문제로 볼 수 있다. 각 행위자가 가능한 정책 π에 대해 반응하고, 그 반응이 다시 다른 행위자의 반응 조건을 바꾼다면, 사회적 상태는 게임이론적 균형 문제로 해석될 수 있다. 적절한 연속성, 볼록성, 콤팩트성 조건이 주어질 경우 Kakutani 고정점 정리는 Nash equilibrium의 존재를 논의하는 한 가지 수학적 배경이 된다. 그러나 GU는 단지 어떤 균형이 존재한다는 사실에 만족하지 않는다. 나쁜 균형도 존재할 수 있기 때문이다. 따라서 GU가 묻는 것은 "균형이 있는가"가 아니라 "어떤 균형을 정의 근사로 선택할 것인가"이다.

이 선택 문제는 Nash bargaining의 관점에서도 해석될 수 있다. 여러 목적 선언이 동시에 만족

될 수 있는 실행 가능 집합 $\mathcal{F}$와 협상 결렬점 d가 있을 때, 하나의 단순한 해 선택 원리는 다음처럼 쓸 수 있다.

$$\pi^{NB} = \arg\max_{\pi \in \mathcal{F}} \prod_{n=1}^{N} \left(H_n(\pi) - d_n \right) \tag{12}$$

이는 모든 문제를 곱셈식 협상해로 환원하자는 뜻이 아니다. 다만 제3프로토콜 안에서 선언된 목적들이 단순 다수결이나 지불 능력으로 압축되지 않고, 각자의 결렬점과 개선폭을 고려하는 공리적 해 선택의 언어로 다루어질 수 있음을 보여준다.

또 다른 관점에서, GU의 고정점은 Tarski 고정점 정리의 최소 고정점(least fixed point)을 떠올리게 한다. 만약 윤리 연산자 F_π가 부분순서가 주어진 격자 위에서 단조적이라면, 우리는 가능한 고정점들 중 가장 작은 가정과 가장 적은 강제를 요구하는 최소 고정점을 먼저 고려할 수 있다. 이는 제3프로토콜의 설계 원칙과 맞닿아 있다. 목적 조정은 가능한 한 많은 목적을 강제로 지우는 방향이 아니라, 가장 적은 억압으로도 안정된 정렬을 만드는 방향이어야 한다.

제약이 존재하는 현실에서는 Lagrange multiplier의 해석도 유용하다. 자원, 시간, 안전, 신뢰, 정보, 법적 권리, 양심의 한계가 제약식 $g_k(\pi) \le 0$로 주어질 때, 목적 조정 문제는 다음과 같은 형태를 가질 수 있다.

$$\mathcal{L}(\pi, \lambda) = \Omega(\mathbf{H}(\pi)) - \sum_{k} \lambda_k g_k(\pi) \tag{13}$$

여기서 λ_k는 단순한 계산 보조항이 아니라 shadow price, 곧 어떤 제약이 사회적 목적 달성에 얼마나 큰 그림자를 드리우는지를 나타낸다. 예컨대 한 사회에서 신뢰 부족의 λ가 매우 크다면, 더 많은 돈보다 신뢰 인프라를 구축하는 것이 더 큰 목적 개선을 낳을 수 있다.

마지막으로, 목적 선언은 common knowledge의 문제이기도 하다. 어떤 개인이 속으로 무엇을 원한다는 사실과, 그가 그것을 공적으로 선언했고, 다른 사람들이 그 선언을 알고 있으며, 그가 다른 사람들이 알고 있음을 안다는 사실은 완전히 다르다. 제3프로토콜의 목적 선언은 바로 이 믿음의 고정점을 만든다. 목적은 내면의 욕구로만 남아 있을 때 사회적 입력이 되지 못하지만, 공적으로 승인되고 반복적으로 참조될 때 계약, 협상, 조직, 대리 지능의 행동 조건이 된다.

이때 가장 어려운 반론은 악한 목적의 문제다. 누군가가 "나는 악인이다. 누가 나를 멈출 것인가"라고 선언한다면, 목적 선언을 존중한다는 원리는 그 선언을 그대로 허가해야 하는가? 답은 아니다. GU에서 목적 선언은 사회적 입력이지 면책권이 아니다. 선언은 숨은 욕구를 보이지 않게 실행하는 대신, 그 목적이 타인의 가능성, 사회성, 정합성에 어떤 피해를 주는지 계산 가능하고 반박 가능하게 만드는 행위다. 악한 목적은 억압의 바깥에서 신비롭게 처리되는 것이 아니라, 제3프로토콜 안에서 드러나고, GU 안에서 평가되며, CVTD 안에서 그 의도적 사건의 피해 경로가 추적된다. 그러므로 이 절의 여러 수학적 참조는 GU의 완성을 선언하는 것이 아니라, 목적 선언의 조정 문제가 어떤 형식적 도구들과 연결될 수 있는지를 표시하는 지도에 가깝다.

9.5 시간성과 정의 판단의 시점

GU 이론에서 시간(t)은 온톨로지 벡터공간 O의 또 하나의 축으로 포함된다. 전체 윤리적 구조는 다음과 같이 표현될 수 있다:

$$O_t = O \times \mathbb{R}\,, \tag{14}$$

즉, 시간에 따라 O의 상태가 변화한다고 본다. 존재자 n은 시점 t_0에서의 내재 상태 $I_n(t_0)$ 및 총합적 효용 $H_n(t_0)$로 정의되며, 정의 판단은 항상 현재 시점 t_0의 판단자 n에 의해 수행된다. 이때 과거 및 미래의 존재자, 즉 과거의 나 또는 미래의 나 역시 타자 i의 하나로 간주된다. 과거 세대나 미래 세대에 대한 공감 계수 $C_{n,i}$와 이해 해상도 $EERI_{n,\pi,i}(t_0)$가 낮을 경우, 현재 n의 총합적 효용 $H_n(t_0)$에 미치는 영향력은 자연스럽게 축소된다. 이는 시간적 거리가 먼 존재자일수록 현재 판단에 덜 고려될 수 있음을 의미하지만, GU 구조 내에서는 명시적으로 그 영향을 모델링하고 있다.

결과적으로, 현실에서의 정의 판단은 다음과 같은 구조를 가진다:

$$\pi^*(t_0) \;=\; \arg\max_{\pi} \sum_{n=1}^{N} EERI_{n,\pi,n}(t_0) \cdot H_n(t_0)\,, \tag{15}$$

이는 정의가 시간에 종속된 절대적인 값이라기보다는, 현재의 정보 구조 위에서 행해지는 윤리적 정렬임을 의미한다. 현재 시점의 각 행위자들은 자신과 타인의 현재 및 예상되는 미래 효용, 그리고 그들 간의 공감 정도를 고려하여 $\pi^*(t_0)$를 산출한다. 그 과정에서 미래 세대나 과거 세대에 대한 고려도 C와 $EERI$ 값에 의해 연속적으로 반영되므로, 시간적 거리에서 오는 도덕적 책임의 약화를 방지하려는 구조적 특징을 갖는다.

철학적 정리. 이상의 논의를 요약하면 다음과 같이 표현할 수 있다:

$$\text{Justice} := \text{Omniscient Empathy} \;\succ\; \text{Proximal Justice} = \pi^* \xrightarrow{OSE_{n,i}\to 1,\ \forall n,i} \text{Justice}\,, \tag{16}$$

즉, 현실의 정의는 항상 전지적 공감(Omniscient Empathy)의 방향으로 수렴하며, π^*는 그 수렴을 현실에서 구현한 근사적 정의(Proximal Justice)에 해당한다. 고도화된 AI 시스템은 이러한 수렴을 보조하는 윤리적 메타 연산자로 작동해야 한다.

10 목적의 사회화: 선언에서 의도적 사건으로

GU가 목적 충돌을 판단하기 위한 규범 좌표계라면, 아직 남는 문제는 실행이다. 사회는 목적을 이해하는 데서 멈추지 않고, 그것을 선언, 기여, 설계, 설득, 조직 형성, 자원 이동, 계약, 생산, 거부와 조정 같은 의도적 사건으로 되돌려 보내야 한다. 따라서 제3프로토콜에는 두 번째 내부 엔진이 필요하다. 하나는 목적들 사이의 정의로운 조정을 다루는 GU이고, 다른 하나는 그 목적들이 현실의 사건으로 내려오는 방식을 기록하고 계산하는 실행 문법이다.

여기서 CVTD는 로그에서 목적을 대신 추출하는 체계가 아니다. 목적은 앞서 말했듯 추출되는 데이터가 아니라 형성되고 선언되는 자기헌법이다. CVTD가 하는 일은 그 선언된 목적이 실제 의도적 사건 속에서 어떤 형태로 구현되는지, 그리고 여러 에이전트의 서로 다른 가치 기저가 어떻게 양의 합을 만들거나 충돌을 낳는지를 기술하는 것이다. 계약과 교환은 그 의도적 사건의 중요한 특수 사례일 뿐, CVTD의 전체 대상은 아니다. 그러므로 CVTD는 목적 선언을 대체하지 않고, 선언된 목적이 사회적 사건으로 번역되는 과정을 다룬다.

11 GU에서 CVTD로

전지적 공감의 형식화는 GU가 단지 수사적 이상이 아니라 계산 가능한 방향성이라는 점을 보여준다. 그러나 여전히 하나의 간극이 남는다. GU는 무엇이 정의에 더 가까운지를 말해 주는 좌표계이지만, 현실 세계의 목적은 언제나 구체적 의도적 사건의 형태로 발생한다. 어떤 사람은 계약하고, 어떤 사람은 조직을 만들고, 어떤 사람은 설득하고, 어떤 사람은 돌봄을 제공하며, 어떤 사람은 특정 행위를 거부함으로써 세계 상태를 바꾼다. 따라서 정의의 방향성과 실제 의도적 사건의 언어 사이를 이어 주는 실행층이 필요하다.

CVTD는 바로 이 간극을 메우기 위한 시도이다. 만약 GU가 선언된 목적들의 조정과 수렴을 위한 거시적 규범이라면, CVTD는 그러한 목적들이 물리적 세계에서 어떤 사건 행렬과 가치 기저와 커밋 구조로 표현될 수 있는지를 다루는 미시적 실행 언어가 된다. 다음 절은 이 점에서 GU를 버리는 것이 아니라, 오히려 GU를 현실의 계산 가능한 상호작용으로 번역하는 과정으로 읽혀야 한다. 이때부터 표기를 분명히 하자. GU 절에서 관습적으로 π라 부른 거시적 윤리 연산자는 CVTD에서는 가능한 의도적 사건들의 집합 또는 정책 $\mathcal{P}$로 분해되고, 그 원소인 개별 의도적 사건을 π로 부른다. 사건 π의 행렬 표현은 $\Pi(\pi, t)$로 쓴다. 다만 CVTD 역시 완성된 calculus라기보다, 실행층의 수학적 골격을 어떻게 세울 수 있을지에 대한 초기 제안이다.

12 계약적 가치이전역학(Contractual Value Transfer Dynamics, CVTD): 가치의 계산과 자아의 역산

일반공리주의(GU)가 정의(Justice)의 수렴 방향을 제시하는 거시적 좌표계라면, 계약적 가치이전역학(CVTD)은 이를 물리적 세계 위에서 실제로 계산하고 연산하기 위한 실행 레이어(Execution Layer)이다. 우리는 우주의 모든 사건 중 자연 현상을 제외하고, 앞서 정의한 '비존재적 자아(Self)'를 가진 에이전트들이 목적을 가지고 만들어내는 모든 '의도적 사건'을 CVTD의 기술 대상으로 삼는다.

따라서 CVTD의 '계약적'이라는 표현은 법률상 계약서나 명시적 합의 절차만을 뜻하지 않는다. 그것은 의도적 행위가 언제나 조건, 기대, 책임, 커밋, 귀속, 위반 가능성, 집행 가능성의 구조를 가진다는 뜻이다. 친구에게 조언하는 행위, 회사를 창업하는 행위, 글을 쓰는 행위, 어떤 제안을 거부하는 행위, 공동체 규칙을 설계하는 행위, 자원을 이전하는 행위, 명시적 계약을 체결하는 행위는 모두 서로 다른 형태의 π이다. 계약과 교환은 CVTD의 중요한 사례 중 하나이지만, CVTD 자체는 모든 의도적 사건 π를 기술하기 위한 동역학이다.

이 장에서는 GU의 거시적 정책 $\mathcal{P}$를 개별 의도적 사건 π들의 집합으로 분해하고, 각 사건의 행렬 표현 $\Pi(\pi, t)$와 공감 계수 C를 선형대수학적 실체로 구체화한다. 이를 통해 실제 의도적 사건 시나리오에서 가치 총량의 증가 메커니즘과 자아 방향의 역산 과정을 설명한다.

12.1 의도적 사건 π의 일반형

CVTD에서 하나의 의도적 사건은 단순한 거래 기록이 아니라, 세계 상태에 작용하는 목적을 가진 변환이다. 에이전트 집합을 $\mathcal{A} = \{1, \ldots, N\}$, 시점 t의 세계 상태를 O_t, 공통 가치공간을

$K = \mathrm{span}(k_1, \ldots, k_d)$라 하자. 그러면 사건 π_t는 다음과 같은 튜플로 정의될 수 있다.

$$\pi_t = (A_\pi, X_\pi, \Phi_\pi, \tau_\pi, \Gamma_\pi, \mu_\pi) \tag{17}$$

여기서 $A_\pi \subseteq \mathcal{A}$는 사건에 참여하거나 영향을 받는 에이전트 집합, $X_\pi \subseteq O_t$는 사건이 다루는 객체, 행위, 관계, 제도 상태의 묶음, Φ_π는 사건이 성립하는 조건과 맥락, $\tau_\pi : O_t \to O_{t+1}$는 세계 상태 변환, Γ_π는 커밋, 책임, 제약, 집행, 신뢰 구조, μ_π는 이 사건이 어떤 선언된 목적 또는 목적 후보와 연결되는지를 나타내는 목적 서명이다.

이때 사건의 기본 작용은 다음과 같다.

$$O_{t+1} = \tau_\pi(O_t) \tag{18}$$

자연 현상은 목적 서명 μ_π와 커밋 구조 Γ_π를 갖지 않으므로 CVTD의 직접 대상이 아니다. 반대로 에이전트의 목적이 세계 상태에 개입해 어떤 변환을 일으킨다면, 그것이 거래든 설계든 대화든 거부든 제도 형성이든 모두 CVTD의 π가 된다.

12.2 가치 변화 벡터와 사건 행렬

각 에이전트 n은 세계 상태와 사건 객체를 자신의 가치 기저로 사상하는 평가 연산자 $\mathbf{C}_n^t : O_t \to K$를 가진다. 사건 π_t가 일어났을 때 에이전트 n의 주관적 가치 변화는 다음과 같이 표현된다.

$$\Delta_n(\pi, t) = \mathbf{C}_n^t(O_{t+1}) - \mathbf{C}_n^t(O_t) \tag{19}$$

이 값은 돈의 이동만이 아니라 시간, 신뢰, 전략적 자산, 평판, 자유도, 책임, 관계 안정성, 목적 정렬성 같은 여러 가치축의 변화를 포함한다. 따라서 동일한 사건도 에이전트마다 서로 다른 행 벡터로 나타난다.

사건 전체의 행렬 표현은 다음과 같다.

$$\Pi(\pi, t) = \begin{pmatrix} \Delta_1(\pi, t) \\ \Delta_2(\pi, t) \\ \vdots \\ \Delta_N(\pi, t) \end{pmatrix} \in \mathbb{R}^{N \times d} \tag{20}$$

따라서 이하에서 π는 의도적 사건 자체를, $\Pi(\pi, t)$는 그 사건이 시점 t에서 만드는 선형대수학적 사건 행렬을 가리킨다. 이 구분은 중요하다. 같은 사건 π라도 평가 시점, 가치 기저, 참여자 집합이 달라지면 행렬 표현 $\Pi(\pi, t)$는 달라질 수 있기 때문이다.

사회적 가치 변화의 총량 벡터는 다음처럼 정의된다.

$$G(\pi, t) = \mathbf{1}^T \Pi(\pi, t) \in K \tag{21}$$

또한 GU가 제공하는 사회적 평가 함수 또는 정렬 함수 $\Omega_t : K \to \mathbb{R}$를 두면, 사건의 규범적 점수는 다음과 같이 표현된다.

$$W(\pi, t) = \Omega_t(G(\pi, t)) \tag{22}$$

따라서 CVTD에서 양의 합(Positive Sum)이란 단순히 돈의 총합이 늘어나는 것이 아니라, 사건 행렬의 총량 벡터가 관련 가치축에서 양의 방향을 가지거나, GU적 정렬 함수 아래에서 더 높은 값으로 평가되는 경우를 뜻한다. 이때 어떤 축에서의 양의 합이 다른 축의 침해를 숨길 수 있으므로, Γ_π와 Φ_π는 동의, 책임, 강제, 정보 비대칭, 장기 비용을 함께 기록해야 한다.

12.3 사건 흐름으로서의 동역학

CVTD는 단일 사건의 표기법을 넘어 사건들의 시간적 흐름을 다룬다. 시점 t에 활성화된 의도적 사건들의 집합을 $\mathcal{P}_t^{\text{int}}$, 각 사건의 강도 또는 실행 비율을 $\lambda_\pi(t)$라 하면, 사회의 누적 가치 상태 $Z(t)$는 다음과 같이 갱신된다.

$$Z(t+1) = Z(t) + \sum_{\pi \in \mathcal{P}_t^{\text{int}}} \lambda_\pi(t) \Pi(\pi, t) \tag{23}$$

이 식이 CVTD를 단순한 장부가 아니라 동역학으로 만든다. 제3프로토콜은 선언된 목적들을 받아 가능한 사건 집합 $\mathcal{P}_t^{\text{int}}$를 생성하고, GU는 그 사건들의 규범적 정렬을 평가하며, CVTD는 실제 선택된 사건들이 각 에이전트와 사회의 가치 상태를 어떻게 변화시키는지를 계산한다.

12.4 GU 변수의 벡터 공간 매핑

12.4.1 정책 집합과 사건 행렬 (Event Matrix)

GU에서 정의의 수렴 방향을 결정하는 거시적 정책은, CVTD에서 구체적인 의도적 사건들의 집합 $\mathcal{P}^{\text{int}}$로 표현된다. 개별 의도적 사건 $\pi \in \mathcal{P}^{\text{int}}$는 행렬 표현 $\Pi(\pi, t)$를 통해 계(System)의 가치 총량을 변화시킨다. 따라서 물리적 세계에서 목적을 가진 에이전트가 발생시키는 모든 의도적 상호작용은 거시 정책 $\mathcal{P}$의 국소적 사건으로 취급된다.

12.4.2 공감 계수 C와 가치 기저 벡터 (Value Basis Vector, C)

GU에서의 공감 계수 $C_{n,i}$는 에이전트 n이 타자 i의 효용을 자신의 효용 함수에 반영하는 가중치였다. 벡터 공간인 CVTD에서 이 C는 에이전트 n의 고유한 가치 기저 벡터(Value Basis Vector, $\mathbf{C}_n$)로 승격된다. 에이전트가 세상을 바라보는 관점, 중요하게 여기는 가치축(Value Axes), 그리고 객체(Object)를 효용으로 환산하는 평가는 모두 이 기저 벡터 $\mathbf{C}_n$에 의해 결정된다. 즉, $\mathbf{C}_n$은 에이전트 n이 세계를 해석하는 구조적 필터이자, 타자와 연결되는 인터페이스이다.

12.5 사건 행렬 $\Pi(\pi, t)$의 구성과 주관성

가장 단순한 의도적 사건 원형 중 하나는 2자-2객체 교환(2×2 Primitive)이다. CVTD의 전체 대상은 이보다 넓지만, 교환 사건은 서로 다른 가치 기저가 어떻게 양의 합을 만들 수 있는지를 가장 선명하게 보여준다. 에이전트 A와 B가 각각 객체 x와 y를 교환하는 사건($A \to B : x, B \to A : y$)을 고려하자. 이때 사건 π의 행렬 표현 $\Pi(\pi, t)$는 다음과 같이 정의된다.

$$\Pi(\pi, t) = \begin{pmatrix} \mathbf{C}_A(y) - \mathbf{C}_A(x) \\ \mathbf{C}_B(x) - \mathbf{C}_B(y) \end{pmatrix} \tag{24}$$

여기서 1행은 에이전트 A의 관점에서 본 순 가치 변화량이며, 2행은 에이전트 B의 관점이다. 핵심은 물리적으로 동일한 객체 x라 할지라도, $\mathbf{C}_A$와 $\mathbf{C}_B$라는 서로 다른 기저(Basis)를 통과하며 전혀 다른 가치 값으로 변환된다는 점이다. 이 기저의 차이(Misalignment)가 바로 동일한 사건이 양측 모두에게 양의 합(Positive Sum)을 만들 수 있는 원동력이다.

12.6 사례 연구: 프리랜서 개발자와 스타트업 창업자

CVTD의 작동 원리를 명확히 하기 위해, 의도적 사건의 한 특수 사례인 계약형 교환 시나리오를 행렬로 분석한다.

시나리오 설정. 우리는 가치 공간을 3차원 축 $K = [k_1 : \text{재무}, k_2 : \text{시간/노력}, k_3 : \text{전략적 자산}]$으로 정의한다.

- **에이전트 A (개발자):** '안정성(Cash)'과 '워라밸(Time)'을 중시하는 기저 $\mathbf{C}_A$를 가짐.
- **에이전트 B (창업자):** '제품 자산(Asset)'과 '속도'를 중시하며 리스크를 감수하는 기저 $\mathbf{C}_B$를 가짐.
- **교환 객체:** x (개발된 코드), y (용역비)

주관적 평가 (Projection). 각 에이전트의 기저에 따른 객체의 평가는 다음과 같다.

- **A의 평가:** 코드를 작성하는 것은 시간 손실($k_2 : -10$)이지만, 용역비는 큰 재무적 이익($k_1 : +15$)이다.
$$\mathbf{C}_A(x) = [0, -10, 0], \quad \mathbf{C}_A(y) = [+15, 0, 0]$$
- **B의 평가:** 용역비 지출은 재무적 손실($k_1 : -15$)이지만, 코드 획득은 막대한 자산 가치($k_3 : +50$)이다.
$$\mathbf{C}_B(x) = [0, 0, +50], \quad \mathbf{C}_B(y) = [-15, 0, 0]$$

사건 행렬 $\Pi(\pi_{dev}, t)$의 도출. 위 값을 앞서 정의한 사건 행렬의 형식에 대입하여 구성한다.

$$\Pi(\pi_{dev}, t) = \begin{pmatrix} (15-0) & (0-(-10)) & (0-0) \\ (-15-0) & (0-0) & (50-0) \end{pmatrix} = \begin{pmatrix} 15 & 10 & 0 \\ -15 & 0 & 50 \end{pmatrix} \tag{25}$$

해석. 이 행렬은 단순한 장부 기록 이상의 정보를 담고 있다. 물리적 재화(돈, k_1)의 합은 0이지만, 행렬 전체의 가치 총량 벡터는 $[0, 10, 50]$으로 양의 값을 가진다. 즉, 서로 다른 기저 $\mathbf{C}$가 교차하면서 무에서 유(시간 효율과 자산 가치)를 창조했다. CVTD는 이 행렬을 통해 '무엇이 거래되었는가'를 넘어 '왜 거래되었는가'를 기록한다.

12.7 선형대수학적 동역학: 자아 방향의 역산과 정렬

사건을 행렬 π로, 에이전트의 성향을 벡터 $\mathbf{C}$로 통일함으로써, 우리는 자아의 문제와 윤리적 판단을 '계산 가능한(Computable)' 영역으로 가져올 수 있다.

12.7.1 고유벡터(Eigenvector)와 목적 방향의 역산

에이전트 A가 관여한 수백 건의 의도적 사건 $\pi_1, \pi_2, \dots$ 에서 A의 행 벡터들을 모아 사건 이력 행렬 $M_A(T)$를 구성한다고 가정하자.

$$M_A(T) = \begin{pmatrix} \Delta_A(\pi_1, t_1) \\ \Delta_A(\pi_2, t_2) \\ \vdots \\ \Delta_A(\pi_T, t_T) \end{pmatrix} \tag{26}$$

이때 공분산 행렬은 다음과 같다.

$$\Sigma_A = \frac{1}{T} M_A(T)^T M_A(T) \tag{27}$$

Σ_A의 최대 고유값에 대응하는 고유벡터(Eigenvector) $\mathbf{e}_A$는 A가 반복적으로 선택하고 감수하고 생성한 사건들이 향하는 지배적 가치 방향을 가리킨다.

위 예시에서 A가 지속적으로 $[15, 10, 0]$ 형태의 이득을 취한다면, $\mathbf{e}_A$는 k_1(재무)과 k_2(시간) 평면 상에 놓이게 된다. 이는 A의 실천적 목적 방향이 "안정적인 삶의 영위"에 가까움을 수학적으로 보여준다. 반면, B의 고유벡터는 k_3(전략적 자산)을 강하게 가리킬 것이며, 이는 B의 실천적 목적 방향이 "성취와 확장"에 있음을 보여준다. 즉, CVTD에서 자아란 형이상학적 실체가 아니라, 의도적 사건 행렬들이 반복적으로 드러내는 방향성이다.

다만 이것은 선언된 목적을 행동 로그로 대체한다는 뜻이 아니다. 앞서 말한 것처럼 목적 선언은 개인의 자기헌법이며, CVTD의 사건 이력은 그 선언이 실제 사건 속에서 어떻게 구현되고 왜곡되고 수정되는지를 보여준다. 에이전트 A의 선언된 목적 벡터를 $\mathbf{p}_A$라 하면, 선언과 사건 이력의 정렬도는 다음과 같이 측정될 수 있다.

$$\alpha_A = \frac{\langle \mathbf{p}_A, \mathbf{e}_A \rangle}{\|\mathbf{p}_A\| \|\mathbf{e}_A\|} \tag{28}$$

α_A가 높을수록 선언된 목적과 반복된 의도적 사건의 방향이 정렬되어 있으며, 낮을수록 선언, 자기이해, 환경 제약, 실제 선택 사이에 재조정되어야 할 간극이 존재한다.

12.7.2 기저 변환(Change of Basis)과 상호 이해

에이전트 간의 갈등은 서로 다른 기저 $\mathbf{C}$를 사용하기 때문에 발생한다. 동일한 가치공간 K와 양측의 기저가 명시되어 있다면, 선형대수의 기저 변환 행렬(Transition Matrix) $P_{A \to B}$를 통해 에이전트 A의 가치관($\mathbf{C}_A$)을 B의 관점($\mathbf{C}_B$)으로 정확하게 좌표 변환할 수 있다.

$$[\text{Value}]_B = P_{A \to B} \cdot [\text{Value}]_A \tag{29}$$

현실에서는 양측의 기저가 완전히 명시되어 있지 않을 수 있으므로, 사건 이력으로부터 다음과 같이 근사 변환을 학습할 수도 있다.

$$\widehat{P}_{A\to B} = \arg\min_{P} \sum_{s=1}^{T} \|\Delta_B(\pi_s, t_s) - P\Delta_A(\pi_s, t_s)\|^2 \tag{30}$$

이는 CVTD가 은유라는 뜻이 아니라, 형식적으로는 정확한 좌표 변환을 정의하고 현실 적용에서는 미지의 기저를 추정한다는 뜻이다. 이를 통해 A의 "주말 근무 불가(시간 가치 보호)"라는 요구는, B의 기저에서 "생산성 저하 리스크 관리(전략적 자산 보호)"라는 형태로 번역되어 전달될 수 있다. 이는 상호 이해를 높이고 GU의 목표인 전지적 공감($EERI \to 1$)에 근접하게 하는 수학적 엔진이다.

12.7.3 SVD와 잠재 욕망의 발견

사건 행렬 $\Pi(\pi, t)$ 또는 사건 이력 행렬 $M_A(T)$를 특이값 분해(SVD, $\Pi(\pi, t) = U\Sigma V^T$)하면, 명시적인 사건 구조 뒤에 숨겨진 잠재적 차원(Latent Dimensions)을 발견할 수 있다. 예를 들어, A와 B의 상호작용 이력을 분해했을 때, 명시적인 가치축(k_1, k_2, k_3) 외에 제4의 축에서 높은 상관관계가 발견될 수 있다. 이는 두 사람이 계약서에는 쓰지 않았지만, 암묵적으로 공유하는 '미적 취향'이나 '신뢰 비용'일 수 있다. 고도화된 AI 시스템은 이 잠재 차원을 포착하여, B에게 "A는 단순한 개발자가 아니라 당신의 미적 비전을 공유하는 파트너"라고 제안함으로써, 더 높은 차원의 의도적 사건과 협력(Positive Sum)을 유도할 수 있다.

13 TBDM: 개인 효용 압력에서 집단 동력으로

CVTD가 모든 의도적 사건 π를 미시적으로 기록하는 문법이라면, 다음 질문은 그 무수한 미시 사건들이 집단 수준에서 어떤 거시적 힘으로 나타나는가이다. 여기서 필요한 보조 모델이 TBDM이다. 이 모델의 출발점은 단순하다. 집단은 독립된 의지를 갖지 않는다. 국가, 기업, 정당, 시장, 민족, 종교가 무엇을 원한다고 말할 때, 실제로 의식하고 욕망하고 판단하고 행동하는 것은 언제나 원자적 개인들이다. 집단의 의지처럼 보이는 것은, 개인들이 특정 효용 압력 아래에서 같은 방향으로 정렬될 때 발생하는 거시적 패턴이다.

TBDM은 개인의 효용 압력이 대체로 세 기저축 위에서 표현된다고 본다.

기저	핵심 질문	대표 매개물	실패와 충족의 상태
가능성	나는 무엇을 할 수 있는가	돈, 자원, 시간, 안전, 기술, 자유, 권력, 접근권	무력함, 빈곤, 위험에서 선택지, 안전, 실행력으로 이동
사회성	나는 타자들 속에서 어떤 존재인가	명예, 지위, 평판, 사랑, 소속, 신뢰, 권리	배제, 수치, 고립에서 인정, 소속, 존중으로 이동
정합성	나는 무엇을 믿고 살아도 되는가	진실, 의미, 명분, 신념, 세계관, 양심	혼란, 허무, 자기배반에서 의미, 확신, 진실성으로 이동

짧게 말하면 가능성은 "할 수 있어야 한다", 사회성은 "받아들여져야 한다", 정합성은 "말이 되어야 한다"는 압력이다. 돈은 가능성의 대표적 매개물이고, 명예는 사회성의 대표적 매개물이며, 진실 또는 명분은 정합성의 대표적 매개물이다. 이 세 축이 같은 방향을 가리킬수록 개인은 강하게 정렬된다. 어떤 집단에 속하는 것이 이익이고, 그 집단 안에서 인정받는 것이 존엄이며, 그 집단이 말하는 세계가 진실처럼 느껴질 때, 개인은 집단의 방향성을 외부 명령이 아니라 자기 의지처럼 내면화한다.

이를 CVTD와 연결하면 다음과 같다. 각 에이전트 n의 사건별 가치 변화 $\Delta_n(\pi, t)$를 가능성, 사회성, 정합성의 세 기저공간 $\mathcal{B}_3 = \mathrm{span}(\mathbf{e}_{pos}, \mathbf{e}_{soc}, \mathbf{e}_{coh})$으로 사상하는 투영 연산자 R_3를 둔다.

$$\mathbf{f}_n(\pi, t) = R_3 \Delta_n(\pi, t) = \begin{pmatrix} f_n^{pos}(\pi, t) \\ f_n^{soc}(\pi, t) \\ f_n^{coh}(\pi, t) \end{pmatrix} \tag{31}$$

여기서 $\mathbf{f}_n(\pi, t)$는 사건 π가 개인 n에게 가하는 삼기저 효용 압력이다. f^{pos}는 가능성의 압력, f^{soc}는 사회성의 압력, f^{coh}는 정합성의 압력을 뜻한다. 집단 수준의 알짜힘은 다음처럼 쓸 수 있다.

$$\mathbf{F}(t) = \sum_{\pi \in \mathcal{P}_t^{\mathrm{int}}} \sum_{n=1}^{N} w_n(t) \lambda_\pi(t) \mathbf{f}_n(\pi, t) \tag{32}$$

이때 $w_n(t)$는 각 개인의 영향력, 책임, 노출도, 취약성 또는 대표성을 반영하는 가중치이고, $\lambda_\pi(t)$는 사건의 실행 강도다. 집단의 방향성은 독립된 집단 정신에서 나오지 않는다. 그것은 원자적 개인들의 삼기저 압력이 평균화되고, 상쇄되고, 특정 방향으로 증폭될 때 나타나는 $\mathbf{F}(t)$의 방향이다.

이 모델의 설계 목표는 개인의 의도를 억압해 집단적 선을 강제로 만드는 것이 아니다. 오히려 개인들 사이에서 발생하는 무수한 미시적 가치 이전과 의도적 사건을 통과시키되, 악한 힘은 내부에서 서로 상쇄되고 선한 알짜힘만 거시적으로 남도록 프로토콜을 설계하는 것이다. 이를 위해 제3프로토콜은 각 목적 선언이 가능성, 사회성, 정합성의 어느 축을 강화하거나 훼손하는지 드러내야 하고, GU는 그 힘의 방향을 평가해야 하며, CVTD는 실제 사건들이 어떤 삼기저 압력으로 누적되는지 계산해야 한다.

기업을 예로 들면, 기업이 이윤과 성장을 원하는 것처럼 보이는 이유는 기업이라는 생명체가 욕망하기 때문이 아니다. 구성원들이 월급, 승진, 지분, 고용 안정이라는 가능성 압력, 직함과 성과 평가와 업계 평판이라는 사회성 압력, 미션과 혁신과 고객 가치라는 정합성 압력 아래에서 같은 방향으로 정렬되기 때문이다. 정당, 국가, 민족, 시장도 마찬가지다. 집단의 이름으로 말해지는 의지는 개인들의 삼기저 효용 정렬이 만든 거시적 그림자다.

따라서 어떤 집단 현상을 분석할 때 TBDM은 세 가지 질문을 던진다. 가능성의 차원에서는 누가 무엇을 얻고 잃는가. 사회성의 차원에서는 누가 인정받고 누가 배제되는가. 정합성의 차원에서는 어떤 명분과 진실의 언어가 행동을 정당화하는가. 이 세 질문을 통과하면, 집단의 의지처럼 보였던 것이 개인 단위의 효용 압력으로 분해된다. 인류의 driving force는 집단의 의지가 아니라, 원자적 개인들의 삼기저 효용 정렬이다.

14 프로토콜에서 현실 웨지로

여기까지의 논의가 인간론, 사회론, 규범론, 실행론을 세웠다면, 마지막 질문은 현실적 진입점이다. 목적 선언은 중요하다고 해서 즉시 풍부하게 생성되지 않는다. 인간은 자기 목적을 이미 완성된 문장으로 보유하고 있다가 플랫폼에 입력하는 존재가 아니다. 목적은 읽기, 사유, 반박, 수정, 대화, 승인이라는 느린 과정을 통해 만들어진다. 따라서 제3프로토콜의 첫 현실 인터페이스는 목적을 곧장 거래시키는 시장이 아니라, 목적을 형성하게 만드는 인프라여야 한다.

이 절부터 The Channel은 완성된 제3프로토콜 그 자체가 아니라, 그 프로토콜로 가기 위한 첫 웨지로 자리 잡는다. 책과 장문 텍스트는 목적 선언 이전의 목적 형성 과정을 견디게 만드는 가장 느리고 단단한 매체이며, The Channel은 그 과정을 도시적 인터페이스와 대리 지능의 학습 구조 위에 배치하려는 현실적 출발점이다.

15 목적형성 인프라와 선언

그러나 목적은 원리상 추출만으로는 충분히 다룰 수 없다. 행동 로그, 클릭 패턴, 소비 기록은 말초적 선호와 즉각적 반응은 잘 포착할 수 있지만, 사람이 무엇을 위해 살고 어떤 세계를 원하며 무엇에 책임을 지려 하는지는 동일한 방식으로 읽어낼 수 없다. 목적은 대부분 이미 완성된 채 데이터 속에 숨어 있는 정보가 아니라, 읽고, 생각하고, 반박받고, 수정하고, 승인하는 과정을 통해 비로소 형성되기 때문이다. 따라서 앞으로의 핵심 병목은 더 좋은 추천 알고리즘이 아니라, 인간이 자신의 목적을 형성하고 공적으로 선언하게 만드는 장의 부재에 있다.

이 점은 AI가 더 정교해질수록 오히려 더 중요해진다. Meta FAIR가 2026년 3월 26일 공개한 TRIBE v2는 영상, 음성, 언어 자극에 대한 인간의 fMRI 뇌 반응을 예측하는 tri-modal foundation model이며, Meta는 이를 인간 신경 활동의 디지털 트윈에 가까운 모델로 설명한다.[1] 이 사례가 보여주는 것은 분명하다. 앞으로의 시스템은 인간의 반응을 점점 더 깊고 정밀하게 예측할 수 있다. 그러나 뇌 반응을 예측하는 것과 인간이 자기 이름으로 승인할 수 있는 목적을 갖는 것은 전혀 다른 문제다. 어떤 영상에 뇌가 반응했다는 사실은 중요한 자료이지만, 그것이 곧 "나는 이런 인간이 되겠다"는 헌법적 선언은 아니다. 추출이 강해질수록 선언의 필요성은 약해지는 것이 아니라 더 강해진다.

이 점에서 필요한 것은 처음부터 목적시장이 아니라 목적형성 인프라이다. 그 안에서 인간은 읽고, 생각하고, 토론하고, 목적을 다듬고, 마침내 선언한다. 그리고 그 선언은 단순한 발화가 아니라 책임과 귀속과 협상의 출발점이 된다. 이후에야 비슷한 목적들이 서로를 발견하고, 조건부 커밋과 자원 의사와 기여 형태가 결합되며, 목적조직과 목적시장으로의 이행이 가능해진다. 목적시장은 자본을 대체하지 않지만, 사람들을 처음 결속시키는 원리를 자본이 아니라 목적에 두는 구조다.

이 과정이 실제로 에이전트의 언어로 넘어가기 위해서는 표준화 또한 필요하다. 각 개인은 최상위 목적, 하위 목적, 우선순위, 금지 조건, 타협 가능 범위, 시간축, 자원 제약, 수정 이력, 공개 범위, 대리권 범위를 갖는 일종의 목적 패스포트를 통해 공적 공간에서 대리될 수 있어야 한다.

[1] Meta AI, "Introducing TRIBE v2: A Predictive Foundation Model Trained to Understand How the Human Brain Processes Complex Stimuli," March 26, 2026, https://ai.meta.com/blog/tribe-v2-brain-predictive-foundation-model/. See also Meta AI Research, "A foundation model of vision, audition, and language for in-silico neuroscience," https://ai.meta.com/research/publications/a-foundation-model-of-vision-audition-and-language-for-in-silico-neuroscience/.

이때 플랫폼의 핵심은 직접 최고의 에이전트가 되는 데 있지 않다. 더 중요한 것은 어떤 모델 위의 에이전트든 주인을 더 깊이 대리하기 위해 반드시 통과해야 하는 목적 선언 포맷과 의미론적 표준층을 정의하는 데 있다.

기술적으로 이 표준층은 폐쇄형 에이전트 제국이 아니라, 에이전트 간 상호운용성을 전제해야 한다. Google이 2025년 4월 발표한 Agent2Agent(A2A) 프로토콜은 서로 다른 프레임워크나 벤더가 만든 에이전트들이 안전하게 정보를 교환하고 행동을 조율하기 위한 공개 프로토콜로 제안되었다.[2] The Channel이 직접 모든 일을 하는 초거대 에이전트가 될 필요는 없다. 오히려 The Channel의 역할은 사용자의 목적 패스포트와 선언 이력을 표준화하고, 이를 A2A와 같은 에이전트 상호운용 표준을 통해 외부 에이전트들이 읽고 존중할 수 있는 의미론적 레이어로 만드는 데 있다. 그렇게 될 때 The Channel은 모델이 아니라 주인을 중심에 두는 에이전트 생태계의 목적 레지스트리가 된다.

16 왜 읽기가 첫 번째 웨지인가: The Channel

이러한 목적형성 인프라가 필요하다면, 그 첫 번째 현실 인터페이스는 무엇이어야 하는가. 나는 그 답이 읽기 일반이 아니라, 장문과 책을 중심으로 한 느리고 단단한 언어 공간에 있다고 본다. 짧은 포스트와 피드와 반응형 미디어는 사람을 자극하고 선호를 포착할 수는 있어도, 삶의 방향과 장기 목적을 형성하게 만들기에는 지나치게 빠르고 얕다. 반면 책은 긴 문맥을 따라가게 하고, 자기 생각을 문장과 논리와 서사 속에서 다시 조직하게 만든다. 목적은 대부분 짧은 자극 속에서 태어나지 않고, 긴 독해와 긴 사고와 긴 언어를 필요로 한다. 내가 예상하는 것은, 실행과 맞춤화의 비용이 계속 떨어지고 지능의 총량이 커질수록 사회 전체가 더 높은 해상도의 목적을 요구하게 되며, 그 결과 사회의 "목적요구압"이 필연적으로 커진다는 점이다. 여기서 말하는 목적요구압은, 개인과 조직과 대리 지능이 "정확히 무엇을 원하고 무엇을 위해 커밋할 것인가"를 더 선명하게 말해내야 한다는 압력의 증가를 뜻한다.

책이 첫 번째 웨지가 되는 이유는 단지 책이 오래된 매체이기 때문이 아니다. 첫째, 책은 긴 시간의 주의를 요구하므로 즉각적 반응보다 자기 해석을 강화한다. 둘째, 책은 낮은 자극 밀도 속에서 개념의 계층을 따라가게 하므로, 파편적 취향이 아니라 세계관을 만든다. 셋째, 책은 문장과 논증과 서사의 형태로 남아 다시 밑줄 긋고, 주석 달고, 반박하고, 공유할 수 있는 공적 객체가 된다. 넷째, 책은 개인의 경험을 더 긴 역사적·철학적 기억과 연결하여, 자기 목적을 단순한 현재 욕구가 아니라 문명적 질문 위에서 재구성하게 만든다. 다섯째, 책을 둘러싼 읽기와 대화의 흔적은 이후 대리 지능이 주인을 이해할 때 사용할 수 있는 고품질 목적형성 데이터가 된다. 따라서 책은 과거의 콘텐츠 형식이 아니라, 목적 선언 이전의 목적 형성을 위한 가장 안정적인 인터페이스다.

따라서 The Channel은 매우 현실적으로는 책 사업이다. 그러나 그 본질은 책을 상품으로 유통하는 데 있지 않다. 인간이 스스로 생각할 수 있는 조건을 다시 문명의 중심에 가져오는 데 있다. 스스로 생각하지 않는 인간은 점점 더 정교한 추천, 관리, 자극, 대리 판단의 체계 안에서 사육되거나, 자기 목적을 잃은 채 자기소거의 방향으로 밀려날 위험에 놓인다. 이때 책은 낭만적 취미가 아니라, 인간이 자기 목적을 스스로 형성하기 위해 남겨진 가장 오래되고 강한 기술이다. 기술이 종교가 되기 이전, 사람들은 왜 살아야 하는가, 무엇을 사랑해야 하는가, 무엇을 부끄러워해야 하는가, 어떤 세계를 만들어야 하는가에 대한 끝없는 질문과 각자의 답을 인문학, 철학, 문학,

[2] Google Developers Blog, "Announcing the Agent2Agent Protocol (A2A)," April 9, 2025, `https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/`.

종교, 역사, 과학의 긴 문장으로 남겼다. 우리가 무한한 경쟁과 즉각적 최적화 속에서 뒤에 두고 온 것은 바로 그 장기 기억이다. The Channel은 그 장기 기억을 다시 현재의 목적형성 인프라로 연결하려는 시도다.

바로 이 지점에서 The Channel의 위치가 정해진다. 목적요구압이 높아질수록, 사회는 더 많은 콘텐츠를 필요로 하는 것이 아니라 더 분명하고 더 숙고된 목적을 낳을 수 있는 담론의 씨앗을 필요로 하게 된다. 여기서 책은 단순한 매개체가 아니다. 그것은 읽히고, 밑줄 그어지고, 주석 달리고, 반박되고, 다시 해석되고, 마침내 선언으로 이어질 수 있는 장문형 정본 레이어이자 담론의 씨앗이다. 같은 텍스트를 둘러싸고 여러 사람이 자기 언어를 다시 조직하기 시작하는 순간, 책은 소비재가 아니라 목적 형성의 출발점이 된다.

따라서 The Channel은 사람들이 자기 목적과 세계관을 형성할 수 있는 실제 조건을 도시 위에 배치하는 첫 번째 채널 인프라이며, 목적형성 인프라가 목적시장으로 넘어가기 위한 가장 현실적인 출발점이다. 더 정확히 말하면, 우리가 여기서 겨냥하는 것은 사회에 배포될 지능의 개인에 대한 감각신경계이다. 기업과 정부의 영역에서 Palantir류의 시스템이 AI의 감각신경계에 가까운 위치를 점유하듯, The Channel이 노리는 것은 개인에 대해 그와 같은 위치다. 중요한 것은 책을 많이 유통하는 것이 아니라, 어떤 책과 어떤 장문 텍스트가 사람 안에서 사유와 토론과 세계관의 재조직을 일으키고, 그 변화가 나중에 선언 가능한 형태로 축적되는지를 붙잡는 일이다. The Channel은 담론의 씨앗을 사람들의 생활 동선 위에 배치하고, 그 씨앗이 읽기와 대화와 기록을 거쳐 목적 형성으로 이어지는 경로를 축적하는 첫 번째 웨지로 기능한다.

더 나아가 이 구조는 인간층과 대리 지능층의 이중 구조를 전제한다. 인간은 책을 읽고, 긴 글을 쓰고, 토론하며 목적을 다듬고 선언한다. 동시에 각자의 대리 지능은 그 긴 기록과 선언 이력, 가치 우선순위, 수정 과정뿐 아니라 무엇을 읽었고 어떤 텍스트를 계기로 생각이 바뀌었는지의 흔적까지 학습하여, 플랫폼 밖으로 나갈 때조차 주인을 더 깊이 이해한 상태로 남는다. 이때 The Channel의 의미는 단순한 사용 시간의 확보가 아니라, 말초적 선호를 최적화하던 대리 지능을 숙고된 목적을 대리하는 방향으로 전환시키는 첫 현실 인터페이스라는 데 있다.

여기서 대리 지능은 단순한 추천 엔진이 아니라, 내가 내 이름으로 승인할 수 있는 나를 함께 구성하는 트윈에 가까워진다. 사람들은 자신이 무엇을 원하는지 잘 모른다. 그렇다고 도파민에 이끌려 짧은 영상을 넘긴 기록이 완전히 내가 아니라고 말할 수도 없다. 그것 역시 나의 일부다. 다만 그것이 나 전체를 대표하기를 원하지 않을 뿐이다. The Channel의 트윈은 말초적 반응을 부정하지 않되, 그것을 긴 읽기, 긴 사유, 선언, 수정, 책임의 기록과 함께 놓는다. 그리고 사용자가 끝내 "이것은 내가 내 이름으로 승인할 수 있는 나다"라고 말할 수 있는 방향으로 자신을 정렬하도록 돕는다. 결국 인간이 궁극적으로 원하는 것은 단순한 욕구 충족이 아니라, 자신이 원할 수 있는 자신에 가까워지는 일이다.

결국 책은 여기서 상품이 아니라 인간용 연료이자 담론의 씨앗이며, The Channel은 그 씨앗이 도시와 인간과 에이전트 사이를 오가며 목적 선언으로 자라나게 만드는 첫 채널이다.

17 결론

본 원고는 현대 사회의 지식 파편화와 목적 상실 문제를 해결하기 위해, 자아를 '선언된 목적'으로 재정의하는 철학적 토대 위에서 출발하였다. 이를 바탕으로 정의의 수렴 방향을 제시하는 '일반공리주의(Generalized Utilitarianism, GU)'를 제안하였으며, 나아가 이 추상적 모델을 물리적 세계에서 실행하기 위한 계산 언어로서 모든 의도적 사건 π를 기술하는 '계약적 가치이전역학

(Contractual Value Transfer Dynamics, CVTD)'을 제안적 형태로 스케치하였다.

CVTD의 사건 행렬 $\Pi(\pi, t)$는 GU의 정책을 구성하는 구체적 단위이며, 에이전트의 가치 기저 벡터 $\mathbf{C}$는 공감 계수의 구조적 기반이 된다. 이 통합된 프레임워크는 고도화된 AI 시스템이 단순한 도구를 넘어, 인간의 선언된 목적과 실제 의도적 사건의 정렬을 계산하고, 잠재된 연결을 발견하여 사회 전체를 전지적 공감의 상태로 수렴시키는 윤리적 동반자로 기능하기 위한 이론적 토대가 될 것이다.

나아가 TBDM은 이 미시적 사건들이 집단 수준에서 어떻게 가능성, 사회성, 정합성의 알짜 힘으로 나타나는지를 설명한다. 집단은 독립된 의지를 갖지 않는다. 집단의 의지처럼 보이는 것은 원자적 개인들이 세 기저에서 받는 효용 압력이 특정 방향으로 정렬될 때 발생하는 거시적 패턴이다. 따라서 제3프로토콜의 설계 목표는 개인의 의도를 삭제하는 것이 아니라, 미시적 목적 선언과 가치 이전을 통과시키면서 악한 힘은 내부에서 상쇄되고 선한 알짜힘은 증폭되도록 만드는 데 있다.

그러나 본 원고의 확장은 여기서 멈추지 않는다. 실행 비용이 지속적으로 하락하고 지능의 총량이 계속 커지는 사회에서는 표와 가격만으로 인간의 목적을 충분히 집계할 수 없으며, 제3 프로토콜의 출현이 점차 기능적으로 요구된다. 이때 핵심 병목은 더 많은 행동 데이터를 추출하는 데 있지 않고, 인간이 자기 목적을 형성하고 공적으로 선언하게 만드는 인프라의 부재에 있다. 목적형성 인프라는 단순한 커뮤니티나 추천 엔진이 아니라, 인간의 숙고를 사회적 표현으로 전환하고, 대리 지능이 그 표현을 더 깊이 학습하도록 만드는 층으로 이해되어야 한다.

이러한 맥락에서 The Channel은 완성된 목적시장이 아니라 그 시장으로 가는 첫 번째 현실적 채널로 위치 지어진다. 책과 장문 텍스트는 인간이 자기 목적을 다듬는 가장 느리고 단단한 씨앗이자, 선언으로 이어질 수 있는 장문형 정본 레이어이며, The Channel은 그 씨앗을 도시적 인터페이스와 에이전트 학습 구조 위에 배치하는 첫 번째 웨지다. 그것은 책 사업이지만, 단지 책을 파는 사업은 아니다. 인간이 스스로 생각하고, 자기 이름으로 승인할 수 있는 트윈을 만들며, 그 트윈이 A2A와 같은 에이전트 상호운용 표준 위에서 외부 세계와 협상할 수 있게 만드는 목적형성 인프라다. 따라서 GU는 철학적 좌표계이고, CVTD는 아직 개발과 확장의 여지가 크게 남아 있는 실행층의 제안적 언어이며, 제3프로토콜과 목적형성 인프라는 그 둘이 사회적 구조로 확장되는 방향이다. 본 원고는 이 흐름 전체를 닫힌 체계로 완결하기보다, 앞으로의 사회가 어떤 종류의 목적 입력과 선언 구조를 필요로 하게 될 것인지에 대한 하나의 강한 방향 제시로 읽혀야 한다.