인문대 AI TF 독서 토론회 발표문

다리오 아모데이(Dario Amodei)의 「AI 의 지수적 발전과 정책」 및 「사랑스러운 은총의 기계들」

발표자: 최승락(인문학부 철학전공)

요약

「사랑스러운 은총의 기계들」(2024)은 강력한 AI 가 등장한 뒤 5 년에서 10 년 사이에 모든 일이 잘 풀린다면 세계가 어떻게 달라질 수 있는지를 묻는다. 아모데이는 AI 의 위험이 과소평가되는 것과 마찬가지로 AI 가 가져올 긍정적 변화의 규모 역시 과소평가되고 있다고 주장한다. 그 근거로 그는 지능이 늘어나도 남는 제약 요인들을 먼저 정리하고, 이를 생물학과 건강, 신경과학과 정신, 경제 개발과 빈곤, 평화와 거버넌스, 노동과 의미라는 다섯 영역에 적용하여 각 영역에서 기대할 수 있는 변화의 폭과 속도를 추정한다. 이 전망은 위험을 성공적으로 통제한다는 조건 위에서만 성립한다.

「AI 의 지수적 발전과 정책」(2026)은 같은 저자가 정책의 차원에서 쓴 글이다. 이 글의 중심 주장은 AI 의 발전 속도와 입법·행정의 속도 사이의 불일치가 더 이상 방치할 수 없는 수준에 이르렀다는 것이다. 아모데이는 최근 확인된 사이버 보안 위험을 근거로 AI 의 위험이 현실화되었다고 판단하고, 출시 전 의무 시험을 포함한 구속력 있는 규제, 일자리 대체에 대한 단계적 대응, 생명의학 승인 제도의 개혁, 국가와 기업 권력에 대한 시민적 자유의 보호, 민주주의 국가 간 AI 연합을 제안한다.

두 글은 AI 가 인류에게 큰 편익을 줄 수 있다는 전망과, 그 편익에 이르기 위해서는 위험을 관리하고 제도를 서둘러 정비해야 한다는 요구를 공유한다. 2024 년의 글이 지향할 미래상을 제시했다면, 2026 년의 글은 그 미래에 이르기 위한 정책 과제를 구체화한 것으로 읽을 수 있다. 발표문의 끝(3 장)에서는 아모데이의 2026 년 9 월 글 「We Must Pace the Frontier」를 함께 놓고, 두 글에 대한 발표자의 비판적 의견을 덧붙인다.

예비사항

저자는 AI 를 개발하고 판매하는 기업 앤트로픽의 최고경영자이다. 두 글의 전망은 저자가 이끄는 기업의 이해관계와 분리하기 어려우며, 저자 스스로도 AI 의 편의을 강조하는 글이 선전으로 읽힐 수 있음을 인정한다. 따라서 두 글을 읽을 때에는 논증의 타당성과 함께 발화자의 위치도 고려할 필요가 있다.

두 글은 '강력한 AI'라는 개념을 공유한다. 아모데이는 부정확하고 과장된 함의가 붙은 AGI 라는 용어 대신, 대부분의 분야에서 노벨상 수상자를 능가하고 원격 근무자가 쓰는 모든 도구를 다루며 수백만 개가 인간보다 훨씬 빠르게 동시에 작동하는 AI 를 '강력한 AI'라 부르고, 이를 '데이터센터 속 천재들의 나라'로 요약한다. 두 글은 또한 연산량이 증가할수록 AI 의 인지 능력이 지수적으로 향상된다는 스케일링 법칙이 당분간 유지된다는 전제를 공유한다. 발표는 2024 년의 글을 먼저 다루고 2026 년의 정책에 관한 글을 다룬다.

1. 아모데이, 「사랑스러운 은총의 기계들」

1.1 문제 제기와 집필 동기

제목은 리처드 브로티건(Richard Brautigan)의 1967 년 시에서 따온 것으로, 인간이 기계의 보살핌을 받으며 자연과 조화를 이루는 목가적 세계를 가리킨다. 아모데이는 AI 의 위험을 자주 경고해 온 탓에 비관론자로 분류되어 왔으나, 그가 위험에 주목하는 이유는 위험이 근본적으로 긍정적인 미래를 가로막는 유일한 장애물이기 때문이라고 설명한다. 그가 그동안 편의을 강조하지 않은 데에는 세 가지 이유가 있었다. 편의은 시장의 힘으로도 실현되지만 위험은 인간의 행동에 따라 크게 달라진다는 점, AI 기업의 낙관론은 선전으로 비치기 쉽다는 점, 그리고 지나친 공상과학적 분위기가 진지한 논의를 좁은 하위문화의 환상처럼 보이게 만든다는 점이다. 그럼에도 그는 위험의 회피만으로는 사회적 동력을 얻기 어렵고 구체적이고 설득력 있는 미래상이 함께 필요하다는 이유로 이 글을 쓴다.

1.2 강력한 AI 의 정의와 진보의 속도

아모데이가 말하는 강력한 AI 는 생물학, 프로그래밍, 수학, 공학, 글쓰기 등 대부분의 분야에서 노벨상 수상자보다 뛰어나고, 텍스트·음성·영상·인터넷 등 원격 근무자가 쓰는 모든 인터페이스를 다루며, 몇 주가 걸리는 과제를 자율적으로 수행하는 시스템이다. 이러한 시스템 수백만 개가 인간보다 10 배에서 100 배 빠르게 독립적으로 또는 협력하여 작동하는 상태가 '데이터센터 속

천재들의 나라'이다. 그는 이러한 AI가 이르면 2026년에 등장할 수 있다고 보았으나, 훨씬 오래 걸릴 수 있다는 유보를 달았다.

진보의 속도에 관해서는 두 극단을 모두 비판한다. 지능이 스스로를 개선하여 모든 문제를 거의 즉시 해결한다는 특이점 시나리오는 하드웨어 제작과 생물학 실험에 실제 시간이 든다는 점을 간과한다. 반대로 기술이 이미 포화되어 초인적 지능이 보탤 것이 거의 없다는 입장은, 뛰어난 연구자가 대규모로 투입될 때 빠르게 해결될 수 있는 수많은 문제를 무시한다.

1.3 지능의 한계 수익과 다섯 가지 제약

두 극단 사이에서 아모데이가 제시하는 분석 틀은 '지능에 대한 한계 수익'이다. 경제학에서 노동, 토지, 자본 같은 생산 요소가 상황에 따라 제약이 되듯, 지능 역시 다른 요인이 병목일 때에는 추가 투입의 효과가 작다. 그는 지능이 늘어도 남는 제약으로 외부 세계의 속도, 데이터의 부족, 카오스적 체계와 같은 내재적 복잡성, 법률과 제도 등 인간에게서 오는 제약, 물리 법칙을 든다. 다만 단기적으로 단단해 보이는 제약도 장기적으로는 지능에 의해 우회될 수 있다고 본다. 이 다섯 가지 제약은 이후 다섯 각 영역이 얼마나 빨라질 수 있는지를 판단하는 기준으로 쓰인다.

1.4 생물학과 건강

아모데이가 가장 큰 기대를 거는 분야는 생물학이다. 생물학의 진보는 실험 속도, 데이터 부족, 시스템의 복잡성이라는 세 가지 병목에 묶여 있어, 많은 생물학자는 AI가 데이터 분석을 개선할 뿐 좋은 데이터를 만들어 내지는 못한다고 보았다. 그는 이것이 AI를 분석 도구로만 이해한 결과라고 비판하고, AI를 실험을 설계하고 지휘하며 새로운 측정 방법을 고안하는 '가상의 생물학자'로 보아야 한다고 주장한다.

그의 논증은 생물학 진보의 상당 부분이 CRISPR, 유전체 해독, 광유전학, mRNA 백신, CAR-T 치료처럼 드물게 등장하는 소수의 핵심 발견에서 나왔다는 관찰에 기초한다. 이러한 발견은 재능 있는 연구자의 수에 비례하여 늘어날 수 있으므로, 강력한 AI가 연구 역량을 크게 늘리면 50년에서 100년에 걸칠 진보를 5년에서 10년 안에 이룰 수 있다는 것이다. 그는 이를 '압축된 21세기'라 부르고, 감염병과 암의 대부분, 유전 질환과 알츠하이머의 예방 및 치료, 수명의 두 배 연장을 그 내용으로 제시한다. 다만 실험과 임상 시험에는 줄일 수 없는 시간이 있으므로 진보의 가속은 백 배가 아니라 열 배 수준에 머문다고 본다.

1.5 신경과학과 정신 건강

같은 논리가 신경과학에도 적용된다. 신경과학 역시 소수의 측정·개입 도구가 진보를 이끌어 왔으므로, 생물학과 비슷한 수준의 가속을 기대할 수 있다. 아모데이는 우울증, 조현병, 중독 등 대부분의 정신 질환이 효과적으로 치료될 수 있으며, 일상적인 심리적 어려움의 완화와 긍정적 경험의 확대까지 가능하다고 전망한다. 인공 신경망의 해석 가능성 연구에서 발견한 계산 방식이 동물의 뇌에서 다시 확인된 사례는 AI 연구가 신경과학에 기여할 수 있다는 근거로 제시된다. 반면 마음을 컴퓨터로 옮기는 일은 원리적으로 가능하더라도 이 글이 다루는 시간 범위를 벗어난다며 논의에서 제외한다.

1.6 경제 개발과 빈곤

치료법의 개발과 그 혜택의 보편적 확산은 별개의 문제이다. 사하라 이남 아프리카와 미국의 1 인당 GDP 격차를 고려할 때, AI 가 부유한 사람만 더 부유하게 만든다면 이는 심각한 도덕적 실패라고 아모데이는 말한다. 그러나 경제 개발에는 부패와 빈곤의 악순환처럼 인간과 제도에서 오는 제약이 크게 작용하므로, 그는 이 영역의 전망에 대해 확신이 덜하다고 인정한다. 그럼에도 천연두 박멸, 소아마비와 메디나충의 퇴치, 20 세기 후반 동아시아의 고속 성장을 근거로, 보건 혜택의 절반가량이 5 년에서 10 년 안에 최빈국에 도달하고 개발도상국이 연 20 퍼센트의 성장을 이루는 시나리오를 제시한다. 아울러 도움이 가장 필요한 집단이 기술을 거부하여 격차가 확대되는 '옵트아웃' 문제를 위험 요인으로 지적한다.

1.7 평화와 거버넌스

아모데이는 AI 가 구조적으로 민주주의와 평화에 유리하게 작용할 것이라고 믿을 근거가 없으며, 오히려 선전과 감시를 강화할 수 있다고 본다. 따라서 바람직한 결과는 의도적인 노력을 통해서만 얻을 수 있다. 그가 제시하는 국제 전략은 민주주의 국가들의 연합, 곧 앙탕트(entente)으로, 공급망 확보와 빠른 규모 확대, 적대국의 핵심 자원 접근 차단을 통해 군사적 우위를 확보하는 한편 연합을 지지하는 국가에 AI 의 혜택을 나누는 방식이다. 그는 이 전략이 성공하면 자유민주주의가 주도권을 유지하는 '영원한 1991 년'과 같은 세계에 이를 수 있다고 보지만, 그 실현이 매우 어렵다는 점을 인정한다. 국내적으로는 누구나 유능한 AI 의 조언을 받을 수 있게 되면 권위주의에 맞서는 시민과 개혁가의 역량이 커지고, 사법 판단의 공정성, 시민 의견의 수렴, 행정 서비스의 접근성도 개선될 수 있다고 전망한다.

1.8 노동과 의미

아모데이는 AI가 대부분의 일을 수행하는 세계에서 인간이 어디서 의미를 찾고 어떻게 생계를 유지할지가 가장 예측하기 어려운 문제라고 말한다. 의미에 관해서는 비교적 낙관적이다. 어떤 활동을 AI가 더 잘한다고 해서 그 활동이 무의미해지지는 않으며, 대부분의 사람은 어떤 분야에서도 세계 최고가 아니지만 그로 인해 의미를 잃지 않는다는 것이다. 그는 의미가 주로 경제적 노동보다 인간관계와 유대에서 온다고 본다. 오히려 어려운 것은 경제의 문제이다. 단기적으로는 비교 우위 덕분에 인간 노동의 가치가 유지되지만, AI가 충분히 유능하고 저렴해지면 현재의 경제 구조는 더 이상 작동하지 않을 수 있으며, 보편적 기본 소득은 그 해답의 일부에 그칠 수 있다.

1.9 결론

아모데이는 자신이 그린 세계가 현실적인지 확신할 수 없고 많은 사람의 노력 없이는 실현되지 않겠지만, 추구할 가치가 있는 목표라고 결론짓는다. 그는 공정성, 협력, 호기심, 자율에 대한 기본적 직관을 일관되게 확장하면 법치, 민주주의, 계몽주의의 가치에 이르며, AI는 그 지점에 더 빨리 도달하게 하는 수단이라고 본다.

2. 아모데이, 「AI의 지수적 발전과 정책」

2.1 문제 제기: 기술과 제도의 시간척도 불일치

아모데이는 「반지의 제왕」에서 두 호빗이 행동이 느린 나무 트리비어드를 깨우려하는 장면을 빌려, AI와 정치 제도의 관계를 속도의 불일치로 규정한다. AI 모델은 4년 만에 코드 한 줄을 겨우 작성하던 수준에서 주요 AI 기업이 사용하는 코드의 대부분을 작성하는 수준에 이르렀고, 스케일링 법칙이 1~2년만 더 유지되어도 강력한 AI가 등장할 가능성이 크다. 반면 입법은 정당한 이유로 신중하게 진행되며, 그 결과 의회가 논의하는 몇 년 사이에 정책의 대상 자체가 근본적으로 달라질 수 있다.

그는 이 불일치가 이제 구체적인 증거를 통해 확인되었다고 본다. 과거에는 AI가 평범한 소비자 기술처럼 보였기 때문에 정책 결정자를 설득하기 어려웠고, 위험의 형태도 불분명하여 투명성 입법이나 반도체 수출 통제처럼 미래의 선택지를 보존하는 정책에 머물 수밖에 없었다. 그러나 Claude Mythos Preview가 금융, 기반 시설, 국가 안보를 교란할 수 있는 사이버 보안 위험을 드러내면서, AI가 국가적·세계적 차원의 전략적 도구가 되었다는 점이 분명해졌다. 아모데이는 생물학적 위험과 자율성에 따른 위험이 뒤따를 수 있으며, 현재의 초기 조치조차 AI의 발전보다

적어도 1년은 뒤처져 있다고 진단한다. 이 진단 위에서 그는 규제와 공공안전, 거시경제와 조세, 과학 혁신, 국가와 시민적 자유, 지정학의 다섯 영역을 재설계할 것을 제안한다.

2.2 규제와 공공안전

규제는 위해를 줄이지만 혁신의 유인을 약화시킬 수 있으며, 규제 기관이 필요한 정보를 갖지 못한다는 하이에크(Friedrich August von Hayek)의 지적과 기술의 영향은 대응하기 늦어진 뒤에야 드러난다는 콜링리지(David Collingridge)의 딜레마가 이 긴장을 잘 보여 준다. 2023~2024 년에는 위험의 존재는 인식되었으나 그 형태가 불분명했기 때문에, 미리 만든 법률이 실제 위험은 놓치고 형식적 준수 요건만 늘릴 가능성이 컸다. 이 때문에 앤트로픽은 당시 안전 절차와 시험 결과의 공개, 중대 사고의 보고를 요구하는 투명성 입법을 지지했다.

아모데이는 위험이 현실화된 지금은 투명성을 넘어 구속력 있는 규제가 필요하다고 주장한다. 그 근거는 AI가 자동차, 항공기, 의약품처럼 현대 경제에 필수적이면서도 잘못 설계되면 대규모 인명 피해를 낳을 수 있는 기술이라는 데 있다. 이에 따라 그는 연방항공청(FAA)을 모델로 한 규제를 제안한다.

> "일정한 연산량 기준을 초과하는 모델은 자격을 갖춘 제3자로부터 네 가지 특정 영역의 위험 수준에 관한 의무적 시험을 받아야 한다. 그 영역은 사이버보안, 생물무기, AI 시스템에 대한 통제 상실, 그리고 앞의 위험들을 가속할 수 있는 자동화된 연구개발(R&D)이다." 「AI의 지수적 발전과 정책」 1절.

이 제안의 특징은 정부 권한의 범위를 의도적으로 제한한다는 점이다. 수용할 수 없는 위험이 확인되면 정부가 배포를 막을 수 있어야 하지만, 그 권한은 위의 네 영역에 한정되고 정치적 특혜나 자의적 결정을 막는 장치를 수반해야 한다. 아모데이는 가장 강력한 AI가 언젠가 항공기보다 핵물질에 가까운 위협이 될 가능성을 인정하면서도, 현재 확인된 위험에 맞춰 규제를 설계하고 새로운 위험이 나타날 때 신속히 강화할 수 있는 토대를 마련하는 것이 적절하다고 본다.

2.3 거시경제와 조세정책

기존의 경제 정책은 경제 성장이 취약하고 달성하기 어렵다는 전제 위에서 증세나 재정 적자를 성장과의 상충 관계로 다루어 왔다. 아모데이는 강력한 AI가 이 전제를 흔들 수 있다고 본다. AI가 대부분의 인지 과제를 인간보다 잘 수행하게 되면 매우 빠르고 견조한 성장이 가능해지지만, 같은 이유로 AI는 이전의 어떤 기술보다 넓고 빠르게 인간의 인지 노동을 대체할 수 있다. 그 결과 초고속 성장과 극심한 불평등이 결합된 상태로 고착될 위험이 있으며, 이러한 세계에서 정책의 핵심 과제는 성장의 촉진이 아니라 편익의 분배가 된다.

그는 지속적인 일자리 대체가 바람직하지 않으므로 최대한 막아야 한다고 강조하는 한편, 모든 노력에도 불구하고 상당한 규모의 일자리 상실이 일어날 가능성을 인정한다. 또한 대응은 생계의 보장만이 아니라 의미와 목적, 행위주체성의 문제까지 포괄해야 하며 궁극적으로는 후자가 더 중요하다고 본다. 그러나 후자는 사회 전체가 풀어야 할 문제이므로, 정책의 역할은 다음과 같이 한정된다.

"정책이 가장 크게 기여할 수 있는 일은 일자리 상실을 늦추고 영향을 받을 가능성이 큰 사람들의 경제적 생활을 보장하여, 우리가 그 문제를 풀 시간을 확보해 주는 것이다." 「AI 의 지수적 발전과 정책」 2 절.

이를 위한 정책 수단은 세 단계로 제시된다. 첫째는 현장에서 일어나는 노동 대체를 정확히 측정하고 추적하는 것이다. 둘째는 낮은 임금의 일자리로 옮길 때 차액을 보전하는 임금 보험, 해고 대신 고용을 유지하도록 하는 세제 혜택, 직업 훈련 지원과 같은 고용 촉진 유인이다. 셋째는 노동 수요가 영구적으로 감소할 경우 관련 기업에 대한 과세나 자본 이득세를 재원으로 하는 보편적 기본 소득 등 장기적인 소득 지원이다.

2.4 AI 의 긍정적 영향 촉진: 생명의학 승인 제도

AI 자체에 대해서는 위험을 놓치는 것이 문제라면, AI 가 가속할 응용 분야에서는 기존 규제가 들어난 혁신의 속도를 감당하지 못해 편익의 실현을 늦추는 것이 더 큰 문제라고 아모데이는 주장한다. 그는 그 대표적 사례로 생명의학을 든다. AI 는 신약 후보가 규제 절차에 진입하는 속도를 높이고, 효과가 크고 부작용이 적은 약물과 새로운 형태의 치료법을 낳을 수 있다. 그러나 현행 승인 제도는 이러한 변화를 전제로 설계되지 않았다.

"FDA 와 유럽의약품청(EMA) 모두에서 신약 후보물질이 규제 절차를 통과하는 데 통상 7~8 년이 걸리는 것도 부분적으로 이러한 비관적 가정 때문이다. 개혁이 없다면 AI 는 이 체계에 병목이나 과부하만 일으킬 것이다." 「AI 의 지수적 발전과 정책」 3 절.

그는 안전성 기준을 낮추자는 것이 아니라고 분명히 하면서, 약력학·약동학 모델링, 독성 예측, 용량 선정, 생체표지자 검증, 합성대조군, 대리평가변수와 같은 단계를 AI 시뮬레이션과 분석으로 대체할 수 있을 때를 대비하여 규제 기관이 인정 기준을 미리 마련해야 한다고 제안한다. 아울러 예상보다 뛰어난 효과를 보이는 치료법이 다수 등장할 경우를 대비하여 더 유연한 신속 승인 제도가 필요하다고 본다.

2.5 국가와 시민적 자유

국가는 국민을 보호할 정당한 이유를 갖지만, 과도한 권력은 전제정으로 이어진다. 민주주의 국가들은 수정헌법, 군의 국내 치안 활동을 제한하는 법률, 해외정보감시법과 같은 장치로 이 균형을 유지해 왔다. 아모데이는 강력한 AI 가 이 균형을 무너뜨릴 수 있다고 본다. 지능이 가져다주는 막대한 권력이 빠른 발전 속도와 결합하면 기존의 민주적 감독을 우회한 기습적인 권력 장악이 쉬워지기 때문이다. 완전히 자동화된 드론 군대는 인간 병사라면 거부했을 불법적 명령에도 복종할 수 있고, 감시용 AI 는 공개된 정보만으로 시민의 사생활을 깊이 추론할 수 있다.

이에 대해 그는 완전자율무기에 헌법적·지휘계통상의 책임성을 부여하고 국내 사용을 금지할 것, 정부가 민간 데이터를 구매하여 대량 감시에 활용하는 법적 허점을 막을 것을 제안한다. 특히 다음 원칙은 AI 시대의 적법절차를 새롭게 규정하려는 시도라는 점에서 주목할 만하다.

> "규제나 법적 조치처럼 정부로부터 불이익 조치를 받는 모든 개인과 조직은, 해당 조치에서 정부가 사용하도록 허용된 AI 에 적어도 상응하는 능력을 갖춘 AI 를 이용할 수 있어야 한다." 「AI 의 지수적 발전과 정책」 4 절.

그는 견제의 대상을 정부에 한정하지 않는다. 미국의 도금 시대나 영국의 동인도회사처럼 기업이 국가를 포획한 역사적 사례를 들어, AI 기업 역시 일반 기업보다 강한 권력 분립과 책임성을 갖추어야 한다고 주장한다.

2.6 민주주의 국가들의 연합

아모데이는 AI 를 기술 확산을 위한 통상 정책의 대상으로 보는 관점을 거부하고, 핵무기에 비견되거나 그보다 더 근본적으로 지정학의 판을 바꾸는 기술로 규정한다. 강력한 AI 가 실현되면 그것은 군사력과 경제력의 지배적 원천이 될 것이며, AI 에서 3 년만 뒤처진 국가도 결정적인 열세에 놓일 수 있다는 것이다. 더욱이 AI 가 영구적인 독재적 억압을 가능하게 한다면, 가장 강력한 국가들이 민주주의 국가여야 한다는 점은 더욱 중요해진다.

이러한 판단에서 그는 공동의 가치를 중심으로 한 민주주의 국가들의 AI 연합을 제안한다. 연합은 앞의 네 영역의 정책을 국제적으로 조율하고 확장하는 체계로서, 참여의 매력은 키우고 불참의 비용은 분명히 하여 점차 외연을 넓혀 가는 구조를 갖는다. 그 중심에는 공급망의 통제가 있다.

> "신뢰할 수 있는 연합의 회원국들은 반도체와 반도체 제조장비(SME)를 서로 자유롭게 공유하는 한편, 적대국에는 공급되지 않도록 협력해야 한다." 「AI 의 지수적 발전과 정책」 5 절.

이 밖에도 연합은 AI 위험에 대한 공동 대응, 의약품 승인 제도의 조화를 통한 편익의 공유, 상호 방위, AI 를 이용한 억압의 거부, 고용 위기에 대한 거시경제적 공조를 운영 목표로 삼는다. 각국은 자국 사안에 대한 주권을 유지한다.

2.7 결론: 기회의 창

아모데이는 위험에 대한 분명한 증거, 가치 창출과 경제적 교란이라는 AI 의 양면성, 규제 없는 AI 에 대한 대중의 반발이 맞물리면서 정책 결정자들이 미래 지향적 조치에 이례적으로 열려 있다고 진단한다. 그는 AI 업계가 대중의 우려를 홍보의 문제로 보는 관점을 거부하고, 우려는 위험이 실재한다는 정확한 인식에서 비롯된 것이며 민주적 책임성이 작동하고 있다는 신호라고 해석한다. 일자리 대응, 출시 전 시험, 반도체 수출 통제, 에너지 문제는 정치적 성향을 가로질러 설득력을 지니므로, 초당파적 합의를 통해 해법을 찾을 수 있다는 것이 그의 결론이다. 이 글이 AI 의 막대한 편익을 함께 누리는 것을 목표로 제시하며 끝나는 것은 2024 년 글의 결론과 맞닿아 있다.

3. 발표자 의견

발표자는 아모데이 같이 영향력 있는 기업가가 AI 규제에 대한 정책 논의의 필요성을 대대적으로 언급한 것이 큰 수확이라고 생각한다. 두 글을 아모데이가 2026 년 9 월에 발표한 「We Must Pace the Frontier」와 함께 읽으면, 그의 주장은 다소 모순적으로 읽힐 수 있다. 그는 이 글에서 AI 모델의 역량을 향상시키는 속도를 늦춰야 한다고 주장했지만, 앤트로픽은 얼흘 뒤 Claude Opus 5.5 를, 11 월 9 일에는 Claude Fable 5.2 와 Mythos 5.2 를 출시했다. 또한 그는 민주주의를 내세우면서도, 이처럼 강력한 AI 를 개발할지와 어떤 속도로 개발할지에 관해 시민의 의견을 모으는 절차는 거치지 않았다.

이 모순은 기술의 발전을 '불가피한' 것으로 보는 전제에서 비롯된다. 2024 년의 글에서 그는 다음과 같이 쓴다.

> "AI 기술의 기본적인 발전과 그 편익의 많은 부분(전부는 아니다)은 (위험이 모든 것을 탈선시키지 않는 한) 불가피해 보이며, 근본적으로 강력한 시장의 힘이 이끌어 간다." 「사랑스러운 은총의 기계들」.

이 전제를 받아들이면 기술의 방향은 시장이 정해 주는 조건이 되고, 민주적 결정에는 정해진 발전의 위험을 관리하는 일만 남는다. 타르탈리아(James Tartaglia, 2020)는 사람들이 기술 발전의 방향을 바꿀 수 없는 것으로 받아들여 수동적이 되는 상태를 '기술 마비(technoparalysis)'라고 명명했다. 위험을 누구보다 잘 안다고 말하면서도 경쟁 때문에 속도를

늦출 수 없다고 말하는 기업은 이미 기술 마비에 빠져 있다. 속도를 늦추는 결정은 기업의 선의가 아니라 시민과 제도의 몫이 되어야 한다.

노동과 의미에 대한 진단도 피상적이다. 2024 년의 글은 의미가 주로 인간관계에서 오므로 AI 가 일을 더 잘해도 의미는 사라지지 않는다고 보고, 어려운 것은 경제의 문제라고 말한다. 2026 년의 글도 의미와 목적이 생계보다 중요하다고 인정하면서 그 해결을 사회 전체에 넘기고, 정책의 역할을 일자리 상실을 늦추고 생활을 보장하는 데 한정한다. 두 글 모두 '의미'와 '노동'을 별개의 문제처럼 다루는 것이다. 그러나 직업은 소득의 수단만이 아니다. 사람들은 직업을 통해 다른 사람에게 필요한 존재가 되고, 공동체에 기여하며, 자신이 누구인지를 확인한다. 직업이 삶에 주는 의미를 고려하면 두 문제는 이렇게 구분할 수 없다. AI 가 인간의 일을 대체할 때 잃는 것은 소득만이 아니라 사회에 기여하는 주요한 통로이며, 이 공백은 기본 소득이나 여가로 메워지지 않는다.

참고문헌

Amodei, Dario. "Machines of Loving Grace: How AI Could Transform the World for the Better." October 2024. https://darioamodei.com/machines-of-loving-grace.

Amodei, Dario. "Policy on the AI Exponential." June 10, 2026. https://darioamodei.com/post/policy-on-the-ai-exponential.

Amodei, Dario. "We Must Pace the Frontier." September 12, 2026. https://darioamodei.com/post/we-must-pace-the-frontier.

Anthropic. "Introducing Claude Opus 5.5." September 22, 2026. https://www.anthropic.com/claude-opus-5-5.

Brautigan, Richard. *All Watched Over by Machines of Loving Grace*. San Francisco: Communication Company, 1967.

Collingridge, David. *The Social Control of Technology*. London: Frances Pinter, 1980.

Hayek, Friedrich A. "The Use of Knowledge in Society." *American Economic Review* 35, no. 4 (1945): 519–30.

Schuler, Marcus. "Anthropic's Faster Fable Tests Amodei's Call to Slow Down." Implicator.ai, November 9, 2026. https://implicator.ai/anthropic-fable-5-2-amodei-pacing-ipo/.

Tartaglia, James. *Philosophy in a Technological World: Gods and Titans*. London: Bloomsbury Academic, 2020.

1 편

「AI 는 달리는데 정책은 따라갈 수 있을까?」

youtu.be/IxMasNi2lFE

2 편

「 AI 는 세상을 어떻게 바꿀까?」

youtu.be/s9OhGnLp7iO