● AI: 사랑의 은혜의 기계

AI가 세상을 더 나은 곳으로 바꿀 수 있는 방법

2024년 10월

다리오 아모데이

https://darioamodei.com/machines-of-loving-grace

요약

서론: AI의 위험과 가능성

다리오 아모데이(CEO, Anthropic)는 강력한 AI의 위험성과 가능성에 대해 깊이 고민하며, 그 위험을 줄이는 데 집중하고 있습니다. 하지만 그가 AI에 대해 비관적이거나 "파멸론자"라는 오해를 받는 경우가 있습니다. 아모데이는 오히려 AI가 세상을 긍정적으로 변화시킬 수 있는 가능성이 매우 크다고 믿고 있으며, 이러한 긍정적인 잠재력을 대부분의 사람들이 과소평가하고 있다고 주장합니다. 그는 AI의 위험을 강조하는 이유가 이러한 긍정적인 미래

에 방해가 될 수 있는 요소를 해결하기 위해서라고 설명합니다.

이 글에서 아모데이는 만약 AI의 위험을 성공적으로 관리하고 모든 것이 잘 풀린다면, 강력한 AI가 어떤 긍정적인 변화를 가져올 수 있을지에 대해 설명합니다. 물론 미래를 정확히 예측하는 것은 어렵지만, 가능한 시나리오를 통해 AI가 어떤 역할을 할지 가늠해보고자 합니다. 그는 구체적인 비전이 모호한 논의보다 더욱 실질적인 논의를 촉진할 수 있다고 믿으며, 따라서 세부적인 예측을 포함한 설명을 제시합니다.

왜 AI의 장점에 대해 더 많이 이야기하지 않는가?

아모데이는 그와 그의 회사가 AI의 장점보다는 위험성에 대해 더 많이 이야기하는 이유를 몇 가지로 설명합니다.

1. 레버리지를 극대화하기 위해: AI 기술의 개발은 시장의 힘에 의해 빠르게 진행될 것이므로, AI의 장점은 대부분 자연스럽게 따라올 것입니다. 반면, 위험 요소는 우리의 행동에 따라 달라질 수 있으며, 이를 관리하는 것이 더 큰 영향을 미칠 수 있습니다.

2. 선전의 위험을 피하기 위해: AI 회사가 AI의 장점만 강조하면, 그것이 선전처럼 보일 수 있고, AI의 위험성을 무시하는 인상을 줄 수 있습니다. 그는 AI의 장점에 대해 지나치게 홍보하는 것이 오히려 부정적인 영향을 미칠 수 있다고 생각합니다.

3. 과장의 위험을 피하기 위해: 아모데이는 AI의 장점을 과장하거나 AI가 모든 문제를 해결할 것이라는 식의 이야기를 경계합니다. 그는 현실적이고 실용적인 목표에 초점을 맞추는 것이 중요하다고 강조합니다.

4. "SF"적 분위기 피하기: AI에 대해 이야기할 때, 지나치게 공상 과학적인 요소를 강조하면 논의가 현실성에서 멀어질 수 있습니다. 그는 AI에 대해 비현실적인 기대를 가지는 것이 문제를 더욱 복잡하게 만들 수 있다고 우려합니다.

AI의 긍정적 비전: 좋은 세상의 모습

아모데이는 강력한 AI가 있는 세상이 긍정적일 수 있다는 비전을 제시합니

다. 그는 AI가 많은 문제를 해결할 수 있는 도구가 될 수 있으며, 인간이 직면한 도전 과제를 해결하는 데 큰 역할을 할 것이라고 믿습니다. AI는 단순히 문제를 해결하는 것이 아니라, 사람들이 더 나은 세상을 위해 협력하고, 도전을 극복하는 데 중요한 촉매제가 될 수 있다고 설명합니다. 두려움은 하나의 동기가 될 수 있지만, 그것만으로는 충분하지 않으며, 희망과 긍정적인 목표가 필요하다고 강조합니다.

AI가 가져올 긍정적 변화: 5가지 주요 분야

아모데이는 AI가 인간의 삶의 질을 직접적으로 개선할 수 있는 가장 중요한 5가지 분야를 제시하며, 그 변화를 기대하고 있습니다.

1. 생물학과 신체 건강

AI는 생물학 및 신체 건강 분야에서 획기적인 변화를 가져올 수 있습니다. 그는 다음과 같은 방식으로 AI가 생물학적 연구를 가속화할 것이라고 주장합니다.

데이터 분석 가속화: AI는 방대한 생물학적 데이터를 분석하고, 실험 결과를 더 빠르게 도출할 수 있습니다. 기존의 인간 연구자가 몇 년에 걸쳐 수행할

연구를 AI는 단기간에 수행할 수 있습니다.

새로운 실험 설계: AI는 인간처럼 실험을 설계하고 실행할 수 있으며, 연구 과정에서 더 나은 실험 방법을 발명할 수 있습니다. 이를 통해 연구 속도는 급격히 가속화될 수 있습니다.

의료 혁신: AI는 CRISPR와 같은 유전자 편집 기술, 세포 치료, 백신 개발 등 의료 혁신을 가속화하여, 질병을 퇴치하고 인간의 건강 수명을 연장하는 데 기여할 수 있습니다.

그는 AI가 향후 5100년에 걸쳐 달성할 연구 성과를 달성할 수 있을 것이라고 예측합니다. 이로 인해 암, 알츠하이머, 유전 질환 등 다양한 질병의 예방과 치료가 가능해질 것이며, 평균 수명을 두 배로 연장할 수 있는 가능성도 제기합니다.

2. 신경과학과 정신 건강

신경과학 분야에서도 AI는 매우 중요한 역할을 할 수 있습니다. AI는 다음과 같은 방식으로 정신 건강을 개선할 수 있습니다.

신경 측정 및 개입: AI는 신경 활동을 더 정밀하게 측정하고, 이를 바탕으로 정신 질환을 치료할 수 있습니다. 이를 통해 정신 질환의 원인을 밝혀내고 치료법을 개발할 수 있습니다.

행동 중재: AI는 행동 치료법을 개발하고, 이를 통해 정신 건강 문제를 겪는 사람들에게 도움을 줄 수 있습니다. 예를 들어, AI 코치가 사용자의 정신 상태를 모니터링하고, 최적의 정신 상태를 유지하도록 도울 수 있습니다.

유전적 예방: AI는 정신 질환의 유전적 요인을 파악하고, 이를 예방할 수 있는 방법을 제시할 수 있습니다.

아모데이는 AI가 PTSD, 우울증, 조현병 등 대부분의 정신 질환을 치료할 수 있을 것이며, 인간의 인지적 자유와 정신적 기능을 크게 확장할 수 있다고 주장합니다.

3. 경제 발전과 빈곤

AI는 경제 성장을 촉진하고, 특히 개발도상국의 빈곤 문제를 해결하는 데 기여할 수 있습니다. 그는 AI가 다음과 같은 방식으로 경제 발전에 기여할 수 있다고 설명합니다.

건강 중재의 확산: AI는 전 세계적으로 건강 중재를 빠르게 확산시키고, 질병을 퇴치하는 데 기여할 수 있습니다. 이는 개발도상국에서 질병을 효과적으로 퇴치하고, 경제 성장을 촉진하는 데 중요한 역할을 할 것입니다.

경제적 계획: AI는 경제적 의사결정을 돕고, 개발도상국이 선진국 수준으로 경제를 발전시키는 데 기여할 수 있습니다. AI는 중앙 계획과 시장 경제의 장점을 결합하여 경제 성장의 촉진을 도울 것입니다.

그는 AI가 전 세계 경제 성장에 기여할 것이며, 이를 통해 빈곤과 불평등 문제를 해결하는 데 중요한 역할을 할 것이라고 전망합니다.

4. 평화와 통치

AI는 국제 갈등을 줄이고, 민주주의를 강화하는 데 기여할 수 있습니다. 그는 AI가 다음과 같은 방식으로 국제적 평화와 통치를 도울 수 있다고 주장

합니다.

선진국 연합 전략: 민주주의 국가들이 AI를 활용하여 군사적 우위를 확보하고, 독재 국가에 대한 영향력을 강화할 수 있습니다. 이를 통해 세계적으로 민주주의가 확산될 수 있을 것입니다.

정보전에서의 우위 확보: AI는 독재 국가의 검열을 극복하고, 민주주의 국가들이 정보를 자유롭게 교류할 수 있도록 도울 수 있습니다. 이를 통해 독재에 맞서 싸우는 시민들이 더 큰 힘을 발휘할 수 있습니다.

아모데이는 AI가 민주주의를 지지하는 도구로 사용될 수 있으며, 이를 통해 전 세계적으로 자유와 평등이 증진될 수 있다고 강조합니다.

5. 일과 의미

AI가 많은 업무를 대신하게 되면, 인간은 새로운 방식으로 삶의 의미를 찾아야 할 것입니다. 그는 AI가 다음과 같은 방식으로 일과 의미에 대한 문제를 해결할 수 있다고 설명합니다.

비경제적 의미 추구: AI가 인간의 경제적 역할을 대신하게 되더라도, 인간은 여전히 경제적 가치를 창출하지 않는 활동에서 의미를 찾을 수 있습니다. 그는 사람들이 스포츠, 예술, 연구 등에서 자아실현을 이룰 수 있다고 주장합니다.

경제 구조의 변화: AI가 경제 구조를 근본적으로 바꾸게 되면, 인간은 새로운 방식으로 경제에 기여하거나, 기본소득과 같은 대안적인 경제 시스템을 통해 생계를 유지할 수 있을 것입니다.

그는 AI가 가져올 경제적 변화가 불가피하다고 보며, 이에 따라 인간이 새로운 방식으로 의미를 찾을 필요가 있음을 강조합니다.

결론: AI가 가져올 미래의 가치

다리오 아모데이는 AI가 가져올 긍정적인 미래가 현실화되기 위해서는 우리의 협력과 노력이 필요하다고 강조합니다. AI가 인간의 질병을 치료하고, 경제적 불평등을 줄이며, 민주주의를 강화하는 데 기여할 수 있지만, 이러한 변화는 저절로 일어나는 것이 아니라 많은 사람들의 노력이 필요하다는 점을 강조합니다. 그는 AI가 인간의 이상을 실현할 수 있는 도구가 될 수 있

다고 보며, 기술의 발전을 통해 더욱 나은 세상을 만들기 위한 싸움을 계속해야 한다고 주장합니다.

요약

1. 서론: 다리오 아모데이는 AI의 위험과 가능성에 대해 논의하며, 사람들이 AI의 긍정적인 잠재력을 과소평가하고 있다고 설명합니다. 그는 위험을 줄이는 것이 중요하지만, AI가 가져올 긍정적인 미래에 대한 논의도 필요하다고 강조합니다.

2. 왜 AI의 장점에 대해 더 많이 이야기하지 않는가: 그는 AI의 혜택이 시장에 의해 자연스럽게 발전할 것이라고 보며, 위험 요소에 더 집중해야 할 필요가 있다고 설명합니다. 또한, AI의 장점을 과장하거나 지나치게 "공상과학적"으로 접근하는 것을 경계합니다.

3. AI의 긍정적 비전: 아모데이는 AI가 인간에게 많은 문제를 해결할 수 있는 도구가 될 수 있으며, 더 나은 세상을 만들기 위한 협력의 촉매제가 될 수 있다고 주장합니다.

4. AI가 개선할 수 있는 5가지 주요 분야:

생물학과 신체 건강: AI는 생물학적 연구를 가속화하고, 암, 알츠하이머 등 다양한 질병을 치료하며, 인간의 평균 수명을 연장할 수 있습니다.

신경과학과 정신 건강: AI는 정신 질환을 치료하고, 인간의 인지적 자유와 정신적 기능을 확장할 수 있습니다.

경제 발전과 빈곤: AI는 전 세계적으로 경제 성장을 촉진하고, 개발도상국의 빈곤 문제를 해결할 수 있습니다.

평화와 통치: AI는 민주주의를 강화하고, 독재 국가의 검열을 극복하며, 국제적 평화를 증진할 수 있습니다.

일과 의미: AI가 경제 구조를 바꾸더라도, 인간은 새로운 방식으로 의미를 찾고 자아실현을 이룰 수 있을 것입니다.

5. 결론: 아모데이는 AI가 긍정적인 변화를 가져올 수 있는 도구가 될 것이라고 주장하며, 이를 위해 많은 사람들의 협력과 노력이 필요하다고 강조합니다. AI는 인간의 이상을 실현할 수 있는 강력한 도구이며, 이를 통해 더욱 나은 세상을 만들 수 있는 기회를 제공한다고 결론짓습니다.

다리오 아모데이

내용물

기본 가정 및 프레임워크

1. 생물학과 건강

2. 신경과학과 정신

3. 경제 발전과 빈곤

4. 평화와 통치

5. 일과 의미

재고 조사

사랑의 은혜 의 기계 1

AI가 세상을 더 나은 곳으로 바꿀 수 있는 방법

2024년 10월

저는 강력한 AI의 위험에 대해 많이 생각하고 이야기합니다. 제가 CEO로 있는 회사인 Anthropic은 이러한 위험을 줄이는 방법에 대한 많은 연구를 수행합니다. 이 때문에 사람들은 때때로 제가 AI가 대체로 나쁘거나 위험할 것이라고 생각하는 비관주의자 또는 "파멸론자"라고 결론을 내립니다. 저는 전혀 그렇게 생각하지 않습니다. 사실, 제가 위험에 집중하는 주된 이유 중 하나는 그것이 우리와 제가 근본적으로 긍정적인 미래로 보는 것 사이에 서 있는 유일한 것이기 때문입니다. 저는 대부분의 사람들이 AI의 이점이 얼마나 급진적일 수 있는지 과소평가하고 있다고 생각합니다 . 대부분의 사람들이 위험이 얼마나 심각할 수 있는지 과소평가하고 있다고 생각하는 것과 마

찬가지입니다.

이 글에서 저는 그 긍정적인 면이 어떨지, 모든 것이 잘 된다면 강력한 AI가 있는 세상이 어떨지 개략적으로 설명하려고 합니다 . 물론 아무도 미래를 확실성이나 정확성으로 알 수 없고, 강력한 AI의 영향은 과거의 기술적 변화보다 훨씬 더 예측 불가능할 가능성이 높으므로 이 모든 것은 불가피하게 추측으로 구성될 것입니다. 하지만 저는 적어도 교육적이고 유용한 추측을 목표로 하고 있으며, 대부분의 세부 사항이 틀렸더라도 무슨 일이 일어날지 그 맛을 포착합니다. 저는 구체적인 비전이 매우 모호하고 추상적인 비전보다 논의를 발전시키는 데 더 도움이 된다고 생각하기 때문에 많은 세부 사항을 포함시켰습니다.

하지만 먼저, 저는 저와 Anthropic이 강력한 AI의 장점에 대해 그렇게 많이 이야기하지 않은 이유와, 왜 우리가 전반적으로 위험에 대해 계속 많이 이야기할 것인지에 대해 간단히 설명하고 싶었습니다. 특히, 저는 다음과 같은 욕구에서 이 선택을 내렸습니다.

레버리지를 극대화합니다 . AI 기술의 기본적 개발과 그 혜택의 대부분(전부는 아님)은 불가피해 보이며(리스크가 모든 것을 탈선시키지 않는 한) 근본적으로 강력한 시장 힘에 의해 주도됩니다. 반면에 리스크는 미리 결정되지 않으며 우리의 행동은 그 가능성을 크게 바꿀 수 있습니다.

선전이라는 인식을 피하세요 . AI 회사가 AI의 놀라운 이점에 대해 이야기하는 것은 선전가처럼 보일 수 있고, 단점에서 주의를 돌리려는 것처럼 보일 수 있습니다. 또한 원칙적으로 "책에 대해 이야기"하는 데 너무 많은 시간을 보내는 것은 영혼에 좋지 않다고 생각합니다.

과장은 피하세요 . 저는 많은 AI 위험 대중 인물(AI 회사 리더는 말할 것도

없고)이 AGI 이후의 세계에 대해 이야기하는 방식에 종종 흥미를 잃습니다. 마치 선지자가 사람들을 구원으로 인도하듯이 혼자서 그것을 실현하는 것이 그들의 사명인 것처럼 말입니다. 저는 기업이 일방적으로 세상을 형성하는 것으로 보는 것은 위험하고, 실용적인 기술적 목표를 본질적으로 종교적 관점에서 보는 것은 위험하다고 생각합니다.

"SF"적 부담을 피하세요 . 대부분의 사람들이 강력한 AI의 장점을 과소평가한다고 생각하지만, 급진적인 AI 미래를 논의하는 소수의 커뮤니티는 종종 지나치게 "SF"적인 톤(예: 업로드된 마인드, 우주 탐사 또는 일반적인 사이버펑크 분위기)으로 논의합니다. 이로 인해 사람들이 주장을 덜 심각하게 받아들이고 일종의 비현실성을 부여한다고 생각합니다. 분명히 하자면, 문제는 설명된 기술이 가능하거나 가능성이 있는지 여부가 아닙니다(본문에서 이를 세부적으로 논의함). "분위기"가 많은 문화적 부담과 바람직한 미래 유형, 다양한 사회적 문제가 어떻게 전개될지 등에 대한 언급되지 않은 가정을 함축적으로 밀수한다는 것입니다. 그 결과 종종 좁은 하위 문화권의 판타지처럼 읽히지만 대부분의 사람들에게는 불쾌감을 줍니다.

그러나 위의 모든 우려에도 불구하고, 저는 강력한 AI가 있는 좋은 세상이 어떤 모습일지 논의하는 것이 중요하다고 생각하며, 위의 함정을 피하기 위해 최선을 다하고 있습니다. 사실, 저는 화재 진압 계획이 아니라 진정으로 고무적인 미래 비전을 갖는 것이 중요하다고 생각합니다 . 강력한 AI의 많은 의미는 적대적이거나 위험하지만, 결국에는 우리가 싸워야 할 무언가, 모든 사람이 더 나은 긍정적인 합계 결과, 사람들이 다툼을 극복하고 앞으로의 도전에 맞서도록 하는 무언가가 있어야 합니다 . 두려움은 한 가지 동기 부여 요인이지만, 그것만으로는 충분하지 않습니다. 우리에게는 희망도 필요합니다.

강력한 AI의 긍정적 응용 분야 목록은 매우 길지만(로봇공학, 제조, 에너지 등 포함), 저는 인간 삶의 질을 직접적으로 개선할 수 있는 가장 큰 잠재력

이 있는 것으로 보이는 소수의 분야에 집중할 것입니다. 제가 가장 기대하는 다섯 가지 범주는 다음과 같습니다.

생물학과 신체 건강

신경과학과 정신 건강

경제 발전과 빈곤

평화와 통치

일과 의미

제 예측은 대부분의 기준(SF "특이점" 비전 2 제외)으로 판단할 때 급진적일 것입니다. 하지만 저는 진지하고 성실하게 말하고 있습니다. 제가 말하는 모든 것은 매우 쉽게 틀릴 수 있습니다(위에서 제 요점을 반복하자면). 하지만 저는 적어도 다양한 분야에서 얼마나 많은 진전이 가속화될 수 있고 실제로 그것이 무엇을 의미할 수 있는지에 대한 반분석적 평가에 제 견해를 근거로 삼으려고 시도했습니다. 저는 생물학과 신경과학 모두 에서 전문적인 경험이 있어서 운이 좋았고 경제 개발 분야에서는 정보에 입각한 아마추어이지만, 제가 틀릴 부분이 많을 것이라고 확신합니다. 이 글을 쓰면서 깨달은 한 가지는 생물학, 경제학, 국제 관계 및 기타 분야의 전문가 그룹을 모아서 제가 여기서 작성한 내용을 훨씬 더 훌륭하고 정보에 입각한 버전으로 작성하는 것이 가치 있을 것이라는 것입니다. 아마도 제 노력을 그 그룹의 시작점으로 보는 것이 가장 좋을 것입니다.

기본 가정 및 프레임워크

이 글 전체를 더 정확하고 현실적으로 만들기 위해서는 강력한 AI가 무엇을 의미하는지(즉, 5~10년이 시작되는 임계치) 명확히 밝히고, 그러한 AI가 도입되면 그에 따른 영향에 대해 생각할 수 있는 프레임워크를 마련하는 것이 도움이 됩니다.

강력한 AI(저는 AGI라는 용어를 싫어합니다) 3가 어떤 모습일지, 그리고 언제(또는 도착할지)는 그 자체로 거대한 주제입니다. 저는 공개적으로 논의했고 완전히 별도의 에세이를 쓸 수 있는 주제입니다(아마 언젠가는 쓸 것입니다). 분명히 많은 사람들이 강력한 AI가 곧 만들어질 것이라고 회의적이며, 어떤 사람들은 그것이 전혀 만들어질 것이라고 회의적입니다. 저는 그것이 2026년에 나올 수 있다고 생각하지만, 훨씬 더 오래 걸릴 수 있는 방법도 있습니다. 하지만 이 에세이의 목적을 위해 저는 이러한 문제들을 제쳐두고, 그것이 상당히 빨리 나올 것이라고 가정하고, 그 후 5~10년 동안 무슨 일이 일어날지에 초점을 맞추고 싶습니다. 저는 또한 그러한 시스템이 어떤 모습일지 , 그 역량은 무엇이고, 어떻게 상호 작용할지에 대한 정의를 가정하고 싶습니다. 비록 이에 대해 의견이 다를 여지가 있지만요.

강력한 AI 라는 용어는 오늘날의 LLM과 형태가 비슷할 가능성이 크지만 다른 아키텍처를 기반으로 하고, 여러 상호 작용 모델을 포함하고, 다르게 훈련될 수 있는 다음과 같은 속성을 가진 AI 모델을 의미합니다.

순수한 지능 4 의 관점에서 보면 , 이 인공지능은 생물학, 프로그래밍, 수학, 공학, 글쓰기 등 대부분의 관련 분야에서 노벨상 수상자보다 더 똑똑합니다. 즉, 이 인공지능은 미해결 수학 정리를 증명하고, 매우 훌륭한 소설을 쓰고, 어려운 코드베이스를 처음부터 작성할 수 있습니다.

"대화하는 똑똑한 것"일 뿐만 아니라, 텍스트, 오디오, 비디오, 마우스 및 키보드 제어, 인터넷 접속을 포함하여 가상으로 작업하는 인간에게 사용 가능한 모든 "인터페이스"를 갖추고 있습니다. 인터넷에서 작업 수행, 인간에게 지시를 내리거나 지시하기, 재료 주문하기, 실험 지시하기, 비디오 시청하기, 비디오 만들기 등 이 인터페이스에서 지원하는 모든 작업, 커뮤니케이션

또는 원격 작업에 참여할 수 있습니다. 이 모든 작업을 세계에서 가장 유능한 인간을 능가하는 기술로 수행합니다.

이 앱은 질문에 수동적으로 대답하는 것이 아닙니다. 대신 완료하는 데 몇 시간, 며칠 또는 몇 주가 걸리는 작업을 맡고, 똑똑한 직원이 하듯 필요에 따라 설명을 요청하며 자율적으로 해당 작업을 수행합니다.

물리적인 실체는 없지만(컴퓨터 화면에 존재한다는 점 외에는) 컴퓨터를 통해 기존의 물리적 도구, 로봇, 또는 실험실 장비를 제어할 수 있다. 이론상으로는 그 자체가 사용할 로봇이나 장비를 설계할 수도 있다.

모델을 훈련하는 데 사용된 리소스는 수백만 개의 인스턴스를 실행하는 데 재활용될 수 있으며(이는 ~2027년까지 예상되는 클러스터 크기와 일치함), 모델은 대략 10배~100배의 인간 속도로 정보를 흡수하고 동작을 생성할 수 있습니다. 5 그러나 물리적 세계나 상호 작용하는 소프트웨어의 응답 시간에 의해 제한될 수 있습니다.

수백만 개의 사본은 각각 관련 없는 작업에 독립적으로 행동할 수도 있고, 필요한 경우 인간이 협업하는 것과 같은 방식으로 모두 함께 작업할 수도 있으며, 아마도 특정 작업에 특히 뛰어나도록 조정된 서로 다른 하위 집단이 있을 수도 있습니다.

이를 요약하면 "데이터 센터에 있는 천재들의 나라"라고 할 수 있습니다.

분명히 그러한 실체는 매우 어려운 문제를 매우 빠르게 해결할 수 있겠지만, 얼마나 빨리 해결할 수 있을지 파악하는 것은 쉬운 일이 아닙니다. 두 가지 "극단적인" 입장은 모두 저에게는 거짓으로 보입니다. 첫째, 초 또는 수 일 단위로 세상이 즉시 변형될 것이라고 생각할 수 있습니다(" 특이점 "). 뛰어난 지능이 스스로를 기반으로 구축되어 가능한 모든 과학, 엔지니어링 및 운영 작업을 거의 즉시 해결합니다. 이것의 문제점은 예를 들어 하드웨어를 구축하거나 생물학적 실험을 수행하는 것과 같은 실제적인 물리적 및

실제적 한계가 있다는 것입니다. 심지어 천재들의 새로운 나라조차도 이러한 한계에 부딪힐 것입니다. 지능은 매우 강력할 수 있지만 마법의 요정 가루는 아닙니다.

둘째, 반대로, 기술적 진보가 실제 세계 데이터나 사회적 요인에 의해 포화되거나 속도가 제한되고, 인간보다 뛰어난 지능이 거의 아무것도 더하지 않을 것이라고 믿을 수도 있습니다 . 6. 저에게는 이 역시 믿기 어렵습니다. 정말 똑똑한 사람들이 많이 모인다면 진보 속도가 엄청나게 빨라질 수 있는 과학적 또는 사회적 문제가 수백 가지나 생각납니다. 특히 그들이 분석에만 국한되지 않고 실제 세계에서 일을 해낼 수 있다면 말입니다(우리가 가정하는 천재 국가는 인간 팀을 지휘하거나 지원하는 것을 포함하여 할 수 있습니다).

저는 진실은 이 두 극단적인 그림의 지저분한 혼합물일 가능성이 높다고 생각합니다. 이는 작업과 분야에 따라 다르며 세부 사항이 매우 미묘합니다. 저는 이러한 세부 사항을 생산적인 방식으로 생각하기 위한 새로운 프레임워크가 필요하다고 생각합니다.

경제학자들은 종종 "생산 요소"에 대해 이야기합니다. 노동, 토지, 자본과 같은 것들입니다. "노동/토지/자본에 대한 한계 수익"이라는 문구는 주어진 상황에서 주어진 요소가 제한 요소가 될 수도 있고 아닐 수도 있다는 생각을 포착합니다. 예를 들어, 공군은 비행기와 조종사가 모두 필요하고 비행기가 없다면 조종사를 더 고용해도 별 도움이 되지 않습니다. 저는 AI 시대에 우리는 지능에 대한 한계 수익 7 에 대해 이야기하고 지능을 보완하고 지능이 매우 높을 때 제한 요소가 되는 다른 요소가 무엇인지 알아내려고 노력해야 한다고 생각합니다. 우리는 이런 방식으로 생각하는 데 익숙하지 않습니다.

"더 똑똑해지는 것이 이 작업에 얼마나 도움이 되며 어떤 시간 척도로 도움이 되나요?"라고 묻는 것은 아니지만 매우 강력한 AI가 있는 세상을 개념화하는 올바른 방법인 것 같습니다.

지능을 제한하거나 보완하는 요인 목록에 대한 내 추측은 다음과 같습니다.

외부 세계의 속도 . 지능형 에이전트는 일을 성취하고 또한 배우기 위해 세상에서 상호작용적으로 작동해야 합니다 .8 하지만 세상은 그렇게 빨리 움직일 뿐입니다. 세포와 동물은 고정된 속도로 움직이기 때문에 이에 대한 실험에는 일정 시간이 걸리며 이는 줄일 수 없는 시간일 수 있습니다. 하드웨어, 재료 과학, 사람과 소통하는 것과 관련된 모든 것, 심지어 기존 소프트웨어 인프라에도 마찬가지입니다. 더욱이 과학에서는 종종 많은 실험이 순서대로 필요하며, 각 실험은 마지막 실험에서 배우거나 이를 기반으로 구축됩니다. 이 모든 것은 주요 프로젝트(예: 암 치료법 개발)를 완료하는 속도가 지능이 계속 증가하더라도 더 이상 줄일 수 없는 최소값을 가질 수 있음을 의미합니다.

데이터 필요성 . 때로는 원시 데이터가 부족하고, 데이터가 없는 상황에서 더 많은 지능이 도움이 되지 않습니다. 오늘날의 입자 물리학자들은 매우 독창적이며 광범위한 이론을 개발했지만 입자 가속기 데이터가 너무 제한적이기 때문에 선택할 데이터가 부족합니다 . 그들이 초지능이라면 훨씬 더 나은 성과를 낼 수 있을지는 불분명합니다. 아마도 더 큰 가속기의 건설을 가속화하는 것 외에는 말입니다.

본질적인 복잡성 . 어떤 것들은 본질적으로 예측 불가능하거나 혼란스럽고, 가장 강력한 AI조차도 오늘날 인간이나 컴퓨터보다 상당히 더 잘 예측하거나 풀 수 없습니다. 예를 들어, 엄청나게 강력한 AI조차도 일반적으로 혼란스러운 시스템(예: 3체 문제)에서 약간 더 앞서 예측할 수 있을 뿐입니다 .

오늘날의 인간과 컴퓨터에 비하면 9

인간의 제약 . 법을 어기거나, 인간에게 해를 끼치거나, 사회를 엉망으로 만들지 않고는 할 수 없는 일이 많습니다. 정렬된 AI는 이런 일을 하기를 원하지 않을 것입니다(그리고 정렬되지 않은 AI가 있다면, 우리는 위험에 대해 다시 이야기하게 됩니다). 많은 인간 사회 구조는 비효율적이거나 심지어 적극적으로 해롭지만, 임상 시험에 대한 법적 요구 사항, 습관을 바꾸려는 사람들의 의지 또는 정부의 행동과 같은 제약을 존중하면서 변경하기 어렵습니다. 기술적 의미에서 잘 작동하지만 규제나 잘못된 두려움으로 인해 영향이 상당히 줄어든 발전의 예로는 원자력, 초음속 비행 , 심지어 엘리베이터가 있습니다 .

물리 법칙 . 이것은 첫 번째 요점의 더 강렬한 버전입니다. 깨질 수 없는 것처럼 보이는 특정 물리 법칙이 있습니다. 빛보다 빨리 여행하는 것은 불가능합니다. 푸딩은 풀리지 않습니다 . 칩은 신뢰할 수 없게 되기 전에 제곱센티미터당 트랜지스터 수에 제한이 있습니다 . 계산에는 지워진 비트당 특정 최소 에너지가 필요하므로 세계의 계산 밀도가 제한됩니다.

시간 척도 에 따른 추가적인 구분이 있습니다 . 단기적으로는 어려운 제약이 되는 것들이 장기적으로는 지능에 더 유연해질 수 있습니다. 예를 들어, 지능은 우리가 생체 동물 실험이 필요했던 것을 시험관 내에서 학습할 수 있게 해주는 새로운 실험 패러다임을 개발하거나, 새로운 데이터를 수집하는 데 필요한 도구(예: 더 큰 입자 가속기)를 구축하거나, (윤리적 한계 내에서) 인간 기반 제약을 해결할 방법을 찾는 데 사용될 수 있습니다(예: 임상 시험 시스템 개선 지원, 임상 시험에 관료주의가 덜한 새로운 관할권 생성 지원, 인간 임상 시험이 덜 필요하거나 더 저렴하도록 과학 자체 개선).

따라서 우리는 지능이 처음에는 다른 생산 요소에 의해 심하게 병목 현상을 보이나, 시간이 지나면서 지능 자체가 다른 요소를 점점 더 우회하는 그림을 상상해야 합니다. 다른 요소가 완전히 사라지지 않더라도(그리고 물리 법

칙과 같은 어떤 것들은 절대적입니다) 10. 핵심 질문은 그것이 얼마나 빨리, 어떤 순서로 일어나는가입니다.

위의 프레임워크를 염두에 두고, 서론에서 언급한 다섯 가지 영역에 대해 그 질문에 답해보도록 하겠습니다.

1. 생물학과 건강

생물학은 아마도 과학적 진보가 인간 삶의 질을 직접적이고 모호하지 않게 개선할 수 있는 가장 큰 잠재력을 가진 분야일 것입니다. 지난 세기에 가장 오래된 인간 질병 중 일부(천연두 등)가 마침내 정복되었지만, 여전히 많은 질병이 남아 있으며, 이를 물리치는 것은 엄청난 인도주의적 성과가 될 것입니다. 질병을 치료하는 것 외에도 생물학은 원칙적으로 건강한 인간의 수명을 연장하고, 우리 자신의 생물학적 과정에 대한 통제와 자유를 증가시키고, 현재 인간 상태의 불변의 일부라고 생각하는 일상적인 문제를 해결함으로써 인간 건강의 기본 품질을 개선할 수 있습니다.

이전 섹션의 "제한 요소" 언어에서 지능을 생물학에 직접 적용하는 데 있어 주요 과제는 데이터, 물리적 세계의 속도, 본질적인 복잡성입니다(사실, 이 세 가지는 모두 서로 관련이 있습니다). 인간의 제약은 임상 시험이 관련된 후기 단계에서도 역할을 합니다. 이것들을 하나씩 살펴보겠습니다.

세포, 동물, 심지어 화학적 과정에 대한 실험은 물리적 세계의 속도에 의해 제한됩니다. 많은 생물학적 프로토콜은 박테리아나 다른 세포를 배양하거나 단순히 화학 반응이 일어날 때까지 기다리는 것을 포함하며, 이는 때로는 며칠 또는 몇 주가 걸릴 수 있으며 속도를 높일 수 있는 명확한 방법이 없습니다. 동물 실험은 몇 달(또는 그 이상)이 걸릴 수 있으며 인간 실험은 종

종 몇 년(장기적 결과 연구의 경우 수십 년)이 걸립니다. 이와 다소 관련이 있어 데이터가 종종 부족합니다. 양이 아니라 질적으로 부족합니다. 관심 있는 생물학적 효과를 진행 중인 다른 10,000개의 혼란스러운 것에서 분리하거나 주어진 과정에 인과적으로 개입하거나 어떤 효과를 직접 측정하는(간접적이거나 시끄러운 방식으로 결과를 추론하는 것과 대조적으로) 명확하고 모호하지 않은 데이터가 항상 부족합니다. 제가 질량 분석 기술을 연구하는 동안 수집한 프로테오믹스 데이터와 같은 거대하고 정량적인 분자 데이터조차도 노이즈가 많고 많은 것을 놓치고 있습니다(이 단백질은 어떤 유형의 세포에 있었는가? 세포의 어느 부분에 있었는가? 세포 주기의 어느 단계에 있었는가?).

데이터와 관련된 이러한 문제에 대한 일부 책임은 내재적 복잡성입니다. 인간의 신진대사의 생화학을 보여주는 다이어그램을 본 적이 있다면 이 복잡한 시스템의 어떤 부분의 효과를 분리하는 것이 매우 어렵고, 정확하거나 예측 가능한 방식으로 시스템에 개입하는 것이 더 어렵다는 것을 알게 될 것입니다. 마지막으로, 인간에 대한 실험을 수행하는 데 걸리는 내재적 시간 외에도 실제 임상 시험에는 많은 관료주의와 규제 요구 사항이 포함되어 있어 (저를 포함한 많은 사람들의 의견에 따르면) 불필요한 추가 시간이 추가되고 진행이 지연됩니다.

이 모든 것을 감안할 때, 많은 생물학자들은 오랫동안 생물학에서 AI와 "빅 데이터"의 가치에 회의적 이었습니다. 역사적으로, 지난 30년 동안 생물학에 기술을 적용한 수학자, 컴퓨터 과학자, 물리학자들은 상당히 성공적이었지만, 처음에 기대했던 진정한 변혁적 영향은 없었습니다. 회의론 중 일부는 AlphaFold (그 개발자들에게 노벨 화학상을 수상한 바 있음)와 AlphaProteo 11 과 같은 주요하고 혁신적인 혁신으로 인해 줄어들었지만 , AI는 제한된 상황에서만 유용하고 앞으로도 계속 유용할 것이라는 인식이

여전히 있습니다. 일반적인 표현은 "AI는 데이터 분석을 더 잘할 수 있지만, 더 많은 데이터를 생성하거나 데이터의 품질을 개선할 수는 없습니다. 쓰레기가 들어가면 쓰레기가 나옵니다."

하지만 저는 비관적인 관점이 AI에 대해 잘못된 방식으로 생각하고 있다고 생각합니다. AI 발전에 대한 우리의 핵심 가설이 맞다면 AI를 데이터 분석 방법으로 생각하는 것이 아니라 실제 세계에서 실험을 설계하고 실행(실험실 로봇을 제어하거나 단순히 인간에게 어떤 실험을 실행할지 말해주는 것 - 수석 연구원이 대학원생에게 하는 것처럼), 새로운 생물학적 방법이나 측정 기법을 발명하는 것 등 생물학자가 하는 모든 작업을 수행하는 가상 생물학자로 생각하는 것이 옳습니다. AI가 진정으로 생물학을 가속화할 수 있는 것은 전체 연구 과정을 가속화하는 것입니다. 제가 AI가 생물학을 변화시킬 수 있는 능력에 대해 이야기할 때 가장 흔히 제기되는 오해이기 때문에 이 말을 반복하고 싶습니다. 저는 AI를 단순히 데이터를 분석하는 도구로 말하는 것이 아닙니다 . 이 글의 시작 부분에서 강력한 AI에 대한 정의에 따라 저는 AI를 사용하여 생물학자가 하는 거의 모든 것을 수행하고, 지시하고, 개선하는 것에 대해 이야기하고 있습니다.

가속이 어디에서 올 가능성이 있는지에 대해 더 구체적으로 말하자면, 생물학의 진보의 놀라울 정도로 큰 부분은 종종 광범위한 측정 도구나 기법과 관련된 매우 적은 수의 발견에서 비롯되었으며, 이는 생물학적 시스템 에서 정확하지만 일반화되거나 프로그래밍 가능한 개입을 허용합니다. 아마도 이러한 주요 발견은 1년에 약 1개 정도이며, 전체적으로는 생물학의 진보의 50% 이상을 주도한다고 주장할 수 있습니다. 이러한 발견은 본질적인 복잡성과 데이터 제한을 극복하고 생물학적 과정에 대한 이해와 제어를 직접적으로 증가시키기 때문에 매우 강력합니다. 10년에 몇 번의 발견으로 생물학에 대한 기본적인 과학적 이해의 대부분을 가능하게 했고, 가장 강력한 의

료 치료의 대부분을 주도했습니다.

다음은 몇 가지 예입니다.

CRISPR : 생물체의 모든 유전자를 라이브로 편집할 수 있는 기술(임의의 유전자 시퀀스를 다른 임의의 시퀀스로 대체). 원래 기술이 개발된 이래로 특정 세포 유형을 표적으로 삼고 정확도를 높이고 잘못된 유전자의 편집을 줄이기 위한 지속적인 개선이 있었으며 , 이 모든 것이 인간에게 안전하게 사용하기 위해 필요합니다.

정확한 수준에서 무슨 일이 일어나고 있는지 관찰하기 위한 다양한 종류의 현미경: 첨단 광학 현미경(다양한 종류의 형광 기술, 특수 광학 장치 등), 전자 현미경, 원자간력 현미경 등.

지난 수십 년 동안 게놈 시퀀싱 및 합성 비용이 몇 배나 낮아졌습니다.

광유전학 기술은 빛을 비춰 뉴런을 활성화시키는 기술입니다.

원칙적으로 무엇이든 백신을 설계한 다음 이를 빠르게 적용할 수 있는 mRNA 백신입니다 (물론 mRNA 백신은 COVID 기간 동안 유명해졌습니다).

CAR-T 와 같은 세포 치료법 은 면역 세포를 신체에서 꺼내 "재프로그램"하여 원칙적으로 모든 것을 공격할 수 있도록 합니다.

질병의 세균 이론이나 면역 체계와 암 사이의 연관성에 대한 인식과 같은 개념적 통찰력 13 .

저는 이 모든 기술을 나열하는 데 수고를 들이고 있습니다. 왜냐하면 저는 이 기술들에 대해 중요한 주장을 하고 싶기 때문입니다. 저는 재능 있고 창의적인 연구자들이 훨씬 더 많다면 이 기술의 발견 속도가 10배 이상 증가할 수 있다고 생각합니다 . 또는 다른 말로 하면, 저는 이러한 발견에 대한 지능에 대한 수익이 높고 생물학과 의학의 다른 모든 것이 대부분 이 기술

에서 비롯된다고 생각합니다.

왜 이렇게 생각할까요? "지능에 대한 수익"을 결정하려고 할 때 습관적으로 묻는 몇 가지 질문에 대한 답 때문입니다. 첫째, 이러한 발견은 일반적으로 소수의 연구자에 의해 이루어지며, 종종 동일한 사람들이 반복적으로 이루어지므로 무작위 검색이 아닌 기술을 시사합니다(후자는 장기 실험이 제한 요소임을 시사할 수 있음). 둘째, 이러한 발견은 종종 실제보다 수년 더 일찍 "만들어질 수 있었습니다." 예를 들어, CRISPR는 80년대 부터 알려진 박테리아의 면역 체계에서 자연적으로 발생하는 구성 요소였지만 , 사람들이 일반적인 유전자 편집에 재활용될 수 있다는 것을 깨닫는 데 25년이 더 걸렸습니다. 또한 유망한 방향에 대한 과학계의 지원이 부족하여 수년이 지연되는 경우가 많습니다(mRNA 백신 발명자에 대한 이 프로필 참조 . 비슷한 이야기가 많이 있습니다). 셋째, 성공적인 프로젝트는 종종 엉성하거나 사람들이 처음에 유망하지 않다고 생각했던 사후에 생각해 낸 것이지 막대한 자금이 투자된 노력이 아닙니다. 이는 발견을 이끄는 것이 막대한 자원 집중이 아니라 독창성임을 시사합니다.

마지막으로, 이러한 발견 중 일부는 "연속적 의존성"(발견 B를 만드는 도구나 지식을 갖기 위해 먼저 발견 A를 만들어야 함)이 있지만(다시 한번 실험적 지연을 초래할 수 있음) 많은 발견, 아마도 대부분이 독립적이어서 한 번에 많은 것을 병렬로 작업할 수 있습니다. 이러한 사실과 생물학자로서의 제 일반적인 경험은 과학자들이 더 똑똑하고 인류가 보유한 방대한 생물학적 지식 간의 연결을 만드는 데 더 능숙하다면 이러한 발견이 수백 개 더 이루어질 수 있다는 것을 강력히 시사합니다(다시 CRISPR 사례를 고려하세요). 수십 년 동안 신중하게 설계된 물리 모델링에도 불구하고 AlphaFold / AlphaProteo 가 인간보다 훨씬 효과적으로 중요한 문제를 해결하는 데 성공한 것은 앞으로 나아갈 길을 제시해야 하는 원리 증명(좁은 도메인에서

좁은 도구를 사용했지만)을 제공합니다.

따라서 강력한 AI가 이러한 발견의 속도를 최소한 10배로 높일 수 있을 것이라고 추측합니다. 그러면 5~10년 안에 50~100년 분의 생물학적 진보를 이룰 수 있을 것입니다. 14 왜 100배가 아닐까요? 아마도 가능할 수 있겠지만, 여기서는 직렬 의존성과 실험 시간이 모두 중요해집니다. 1년 안에 100년 분의 진보를 이루려면 동물 실험과 현미경 설계 또는 값비싼 실험실 시설과 같은 것들을 포함하여 처음에 제대로 진행되기 위해 많은 것이 필요합니다. 저는 실제로 5~10년 안에 1000 년 분의 진보를 이룰 수 있다는 (어쩌면 터무니없는 말처럼 들리는) 생각에는 열려 있지만, 1년 안에 100년 분의 진보를 이룰 수 있다는 것에는 매우 회의적입니다. 다른 말로 표현하자면, 피할 수 없는 지속적인 지연이 있다고 생각합니다. 실험과 하드웨어 설계에는 일정한 "지연"이 있으며 논리적으로 추론할 수 없는 것을 배우기 위해 일정한 "축소 불가능한" 횟수만큼 반복해야 합니다. 하지만 그 15개 위에 대규모 병렬 처리가 가능할 수도 있습니다 .

임상 시험은 어떨까요? 임상 시험에는 관료주의와 속도 저하가 많이 따르지만, 사실 임상 시험의 속도 저하의 상당 부분(전부는 아니지만!)은 거의 효과가 없거나 모호하게 효과가 있는 약물을 엄격하게 평가해야 하는 필요성에서 비롯됩니다. 오늘날 대부분의 치료법에서 유감스럽게도 이는 사실입니다. 평균적인 암 약물은 생존 기간을 몇 개월 늘리는 반면 신중하게 측정해야 하는 상당한 부작용이 있습니다(알츠하이머 약물의 경우도 비슷합니다). 이는 방대한 연구(통계적 힘을 얻기 위해)와 규제 기관이 일반적으로 잘 하지 못하는 어려운 상쇄로 이어집니다. 다시 말해 관료주의와 경쟁 이익의 복잡성 때문입니다.

무언가가 정말 잘 작동하면 훨씬 더 빨리 진행됩니다.승인 절차가 가속화되고 효과 크기가 클 때 승인이 훨씬 더 용이해집니다.코로나19에 대한 mRNA 백신은 일반적인 속도보다 훨씬 빠른 9개월 만에 승인되었습니다.그렇기는 하지만 이러한 조건에서도 임상 시험은 여전히 너무 느립니다.mRNA 백신은 아마도 약 2개월 만에 승인되어야 했습니다 . 하지만 이런 종류의 지연(약물의 경우 최종적으로 약 1년)과 대규모 병렬화, 그리고 반복이 필요하지만 너무 많지는 않은("몇 번의 시도") 것이 결합되면 5~10년 안에 급진적인 변화가 일어날 수 있습니다.훨씬 더 낙관적으로 말하면, AI 기반 생물학은 인간에게 일어날 일을 더 정확하게 예측하는 더 나은 동물 및 세포 실험 모델(또는 시뮬레이션)을 개발함으로써 임상 시험에서 반복의 필요성을 줄일 수 있습니다.이것은 수십 년에 걸쳐 진행되고 더 빠른 반복 루프가 필요한 노화 과정에 대항하는 약물을 개발하는 데 특히 중요할 것입니다.

마지막으로 임상 시험과 사회적 장벽이라는 주제에서, 어떤 면에서 생물의학 혁신은 다른 기술과는 대조적으로 성공적으로 배포된 이례적으로 강력한 실적을 가지고 있다는 점을 명확히 지적할 가치가 있습니다 .16 서론에서 언급했듯이 많은 기술은 기술적으로 잘 작동함에도 불구하고 사회적 요인으로 인해 방해를 받습니다. 이는 AI가 무엇을 성취할 수 있는지에 대한 비관적인 관점을 시사할 수 있습니다 . 그러나 생물의학은 약물 개발 과정이 지나치게 번거롭지만 일단 개발되면 일반적으로 성공적으로 배포되고 사용된다는 점에서 독특합니다.

위의 내용을 요약하자면, 저의 기본적인 예측은 AI 기반 생물학과 의학이 인간 생물학자들이 앞으로 50~100년 동안 이룰 수 있었던 진전을 5~10년으로 압축할 수 있게 해줄 것이라는 것입니다. 저는 이것을 "압축된 21세기"라고 부르겠습니다. 강력한 AI가 개발된 후, 몇 년 안에 21세기 전체에서 이

룰 수 있었던 생물학과 의학의 모든 진전을 이룰 것이라는 생각입니다.

강력한 AI가 몇 년 안에 무엇을 할 수 있을지 예측하는 것은 본질적으로 어렵고 추측적이지만, "인간이 다음 100년 안에 도움 없이 무엇을 할 수 있을까?"라는 질문에는 어느 정도 구체성이 있습니다. 20세기에 우리가 이룬 것을 살펴보거나, 21세기의 처음 20년을 외삽하거나, "10개의 CRISPR와 50개의 CAR-T"가 우리에게 무엇을 가져다줄지 묻는 것만으로도 강력한 AI에서 기대할 수 있는 일반적인 진보 수준을 추정할 수 있는 실용적이고 근거 있는 방법을 제공합니다.

아래에서 우리가 기대할 수 있는 것의 목록을 만들려고 합니다. 이는 엄격한 방법론에 근거하지 않으며 세부 사항에서 거의 확실히 틀릴 것이지만, 우리가 기대해야 할 일반적인 수준 의 급진주의를 전달하려고 합니다 .

거의 모든 17가지 자연 감염병의 신뢰할 수 있는 예방 및 치료. 20세기에 감염병에 대한 엄청난 진전을 감안할 때, 압축된 21세기에 우리가 어느 정도 "일을 끝낼 수 있을 것"이라고 상상하는 것은 급진적이지 않습니다. mRNA 백신과 유사한 기술은 이미 " 무엇이든 백신 "으로 가는 길을 가리킵니다. 감염병이 전 세계에서 완전히 근절될지 (일부 지역에서만 근절되는 것이 아니라)는 빈곤과 불평등에 대한 질문에 달려 있으며, 이는 섹션 3에서 논의됩니다.

대부분의 암 근절 . 암으로 인한 사망률은 지난 수십 년 동안 매년 약 2%씩 감소해 왔습니다 . 따라서 우리는 현재 인간 과학의 속도로 21세기에 대부분의 암을 근절할 궤도에 올랐습니다. 일부 아형은 이미 대부분 치료되었습니다(예: CAR-T 요법을 통한 일부 유형의 백혈병). 그리고 저는 아마도 암의 초기 단계에서 표적을 지정하고 암이 더 이상 자라지 않도록 막는 매우 선택적인 약물에 대해 더욱 기대하고 있습니다 . AI는 또한 암의 개별

유전체에 매우 정교하게 적용된 치료 요법을 가능하게 할 것입니다. 이는 오늘날 가능하지만 시간과 인간의 전문 지식 측면에서 엄청난 비용이 들며 AI는 이를 확장할 수 있도록 해야 합니다. 사망률과 발생률 모두에서 95% 이상 감소하는 것이 가능해 보입니다. 그렇긴 하지만 암은 매우 다양하고 적응력이 강하며 이러한 질병 중에서 완전히 파괴하기 가장 어려울 가능성이 큽니다. 희귀하고 어려운 악성 종양이 계속 존재한다면 놀라운 일이 아닐 것입니다.

매우 효과적인 예방 및 유전 질환의 효과적인 치료법 . 크게 개선된 배아 스크리닝은 대부분의 유전 질환을 예방할 수 있게 할 가능성이 높으며, CRISPR의 더 안전하고 신뢰할 수 있는 후손은 기존 사람들의 대부분의 유전 질환을 치료할 수 있습니다. 그러나 많은 세포에 영향을 미치는 전신 질환은 마지막 저항자일 수 있습니다.

알츠하이머 예방 . 우리는 알츠하이머의 원인을 알아내는 데 매우 어려움을 겪었습니다(베타 아밀로이드 단백질과 관련이 있는 듯하지만 실제 세부 사항은 매우 복잡해 보입니다). 생물학적 효과를 분리하는 더 나은 측정 도구로 해결할 수 있는 정확히 그런 유형의 문제인 듯합니다. 따라서 저는 AI가 이를 해결할 수 있는 능력에 대해 낙관적입니다. 실제로 무슨 일이 일어나고 있는지 이해하면 비교적 간단한 개입으로 결국 예방할 수 있는 가능성이 높습니다. 그렇긴 하지만 이미 존재하는 알츠하이머로 인한 손상은 역전하기가 매우 어려울 수 있습니다.

대부분의 다른 질병에 대한 치료 개선 . 이것은 당뇨병, 비만, 심장병, 자가면역 질환 등을 포함한 다른 질병에 대한 포괄적인 범주입니다. 이 중 대부분은 암과 알츠하이머병보다 해결하기 "쉬운" 것처럼 보이며 많은 경우 이미 급격히 감소하고 있습니다. 예를 들어, 심장병으로 인한 사망률은 이미 50% 이상 감소했으며 GLP-1 작용제 와 같은 간단한 개입은 이미 비만과 당뇨병에 대해 큰 진전을 이루었습니다.

생물학적 자유 . 지난 70년 동안 피임, 생식력, 체중 관리 등의 분야에서 발

전이 있었습니다. 하지만 저는 AI 가속 생물학이 가능한 것을 크게 확장할 것이라고 생각합니다. 체중, 외모, 생식 및 기타 생물학적 과정은 사람들이 완전히 통제할 수 있게 될 것입니다. 우리는 이를 생물학적 자유라는 제목으로 언급할 것입니다. 즉, 모든 사람이 자신이 되고 싶은 것을 선택하고 자신에게 가장 어필하는 방식으로 삶을 살 수 있는 권한을 가져야 한다는 생각입니다. 물론 글로벌 평등한 접근성에 대한 중요한 의문이 있을 것입니다. 이에 대해서는 섹션 3을 참조하세요.

인간 수명이 두 배로 늘어나는 것 18. 급진적으로 보일지 몰라도, 기대 수명은 20세기에 거의 2배로 늘어났습니다 (약 40년에서 약 75년). 따라서 "압축된 21세기"가 수명을 다시 두 배로 늘려 150년이 될 것이라는 것은 "추세"에 맞습니다. 분명히 실제 노화 과정을 늦추는 데 관련된 개입은 지난 세기에 질병으로 인한 (대부분 유아기) 조기 사망을 예방하기 위해 필요했던 개입과 다를 것이지만, 변화의 규모는 전례가 없는 것은 아닙니다 . 19. 구체적으로, 쥐의 최대 수명을 제한적인 부작용으로 25-50% 늘리는 약물이 이미 존재합니다 . 그리고 일부 동물(예: 거북이 일부 종류)은 이미 200년을 살기 때문에 인간은 분명히 이론적인 상한선에 도달하지 않았습니다. 추측컨대, 가장 중요한 것은 인간 노화의 신뢰할 수 있고 Goodhart-able이 아닌 바이오마커일 것입니다. 이를 통해 실험과 임상 시험에서 빠른 반복이 가능하기 때문입니다. 인간의 수명이 150세가 되면 우리는 "탈출 속도"에 도달할 수 있을 것입니다. 그러면 현재 살아있는 대부분의 사람들이 원하는 만큼 오래 살 수 있을 만큼 충분한 시간을 벌 수 있을 것입니다. 하지만 이것이 생물학적으로 가능하다는 보장은 전혀 없습니다.

이 목록을 보고 지금으로부터 7~12년 후에 이 모든 것이 달성된다면(공격적인 AI 타임라인과 일치할 것입니다) 세상이 얼마나 달라질지 생각해 보는 것은 가치가 있습니다. 말할 것도 없이, 그것은 상상할 수 없는 인도주의적 승리가 될 것이며, 수천 년 동안 인류를 괴롭혀 온 대부분의 재앙을 한꺼번에 없앨 것입니다. 저의 많은 친구와 동료가 자녀를 키우고 있으며, 그 아이

들이 자라면 질병에 대한 언급이 우리에게 괴혈병, 천연두 또는 흑사병이 들리는 것처럼 들리기를 바랍니다. 그 세대는 또한 생물학적 자유와 자기 표현의 증가로 이익을 얻을 것이고, 운이 좋으면 원하는 만큼 오래 살 수도 있을 것입니다.

강력한 AI를 기대했던 소수의 사람들을 제외한 모든 사람들에게 이러한 변화가 얼마나 놀라운지 과대평가하기는 어렵습니다. 예를 들어, 미국에서는 현재 수천 명의 경제학자와 정책 전문가가 사회 보장과 의료 보험을 어떻게 유지할 것인지 , 그리고 더 광범위하게는 의료비(대부분 70세 이상의 사람들, 특히 암과 같은 말기 질환을 앓고 있는 사람들이 소비함)를 어떻게 낮출 것인지에 대해 논의하고 있습니다. 이 모든 것이 실현된다면 이러한 프로그램의 상황은 근본적으로 개선될 가능성이 높습니다 .20 취업 연령과 은퇴 인구의 비율이 크게 변할 것이기 때문입니다. 의심할 여지 없이 이러한 과제는 새로운 기술에 대한 광범위한 접근성을 보장하는 방법과 같은 다른 과제로 대체될 것이지만, 생물학이 AI에 의해 성공적으로 가속화되는 유일한 분야 라 하더라도 세상이 얼마나 많이 변할지에 대해 생각해 볼 가치가 있습니다 .

2. 신경과학과 정신

이전 섹션에서는 신체 질환과 생물학 전반에 초점을 맞추었고 신경 과학이나 정신 건강에 대해서는 다루지 않았습니다. 하지만 신경 과학은 생물학의 하위 분야이며 정신 건강은 신체 건강만큼 중요합니다. 사실, 정신 건강은 신체 건강보다 인간의 웰빙에 더 직접적으로 영향을 미칩니다. 수억 명의 사람들이 중독, 우울증, 정신 분열증, 저기능 자폐증, PTSD, 정신병 21 또는 지적 장애와 같은 문제로 인해 삶의 질이 매우 낮습니다. 수십억 명 더 많은 사람들이 이러한 심각한 임상 장애 중 하나의 훨씬 더 가벼운 버전으로 해석될 수 있는 일상적인 문제로 어려움을 겪습니다. 그리고 일반 생물

학과 마찬가지로 문제를 해결하는 것을 넘어 인간 경험의 기본 품질을 개선하는 것이 가능할 수 있습니다.

제가 생물학에 대해 제시한 기본 프레임워크는 신경과학에도 동일하게 적용됩니다. 이 분야는 종종 측정이나 정확한 개입을 위한 도구와 관련된 소수의 발견에 의해 추진됩니다. 위의 목록에서 광유전학은 신경과학의 발견이었고, 최근에는 CLARITY 와 확대 현미경이 같은 맥락의 발전이며, 많은 일반적인 세포 생물학 방법이 신경과학으로 직접 이어졌습니다. 저는 이러한 발전 속도가 AI에 의해 비슷하게 가속화될 것이라고 생각하며, 따라서 "5~10년 안에 100년의 진보"라는 프레임워크는 생물학에 적용되는 것과 같은 방식으로 신경과학에도 같은 이유로 적용된다고 생각합니다. 생물학에서와 마찬가지로 20세기 신경과학의 진보는 엄청났습니다. 예를 들어, 우리는 1950년대까지 뉴런이 어떻게 또는 왜 발화하는지조차 이해하지 못했습니다 . 따라서 AI 가속화 신경과학이 몇 년 안에 빠른 진전을 이룰 것으로 기대하는 것이 타당해 보입니다.

이 기본적인 그림에 추가해야 할 것이 하나 있습니다. 최근 몇 년 동안 AI 자체에 대해 알게 된(또는 배우고 있는) 것 중 일부는 인간만이 계속해서 수행하더라도 신경 과학을 발전시키는 데 도움이 될 가능성이 있다는 것입니다. 해석 가능성 은 분명한 예입니다. 생물학적 뉴런은 표면적으로 인공 뉴런과 완전히 다른 방식으로 작동하지만(이들은 스파이크와 종종 스파이크 속도를 통해 통신하므로 인공 뉴런에는 존재하지 않는 시간적 요소가 있으며 세포 생리학 및 신경 전달 물질과 관련된 많은 세부 정보가 작동을 크게 수정함), "결합된 선형/비선형 작업을 수행하는 간단한 단위로 구성된 분산되고 훈련된 네트워크가 어떻게 함께 작동하여 중요한 계산을 수행하는가"라는 기본적인 질문은 동일하며, 저는 개별 뉴런 통신의 세부 사항이 계산 및 회로에 대한 대부분의 흥미로운 질문에서 추상화될 것이라고 강력히 의심합니

다 . 이에 대한 한 가지 예로, AI 시스템에서 해석 가능성 연구자들이 발견한 계산 메커니즘이 최근 쥐의 뇌에서 다시 발견 되었습니다.

인공 신경망에서 실험을 하는 것은 실제 신경망에서 하는 것보다 훨씬 쉽습니다(후자는 종종 동물의 뇌를 절단해야 합니다). 따라서 해석 가능성은 신경 과학에 대한 이해를 향상시키는 도구가 될 수 있습니다. 더욱이 강력한 AI는 아마도 인간보다 이 도구를 더 잘 개발하고 적용할 수 있을 것입니다.

그러나 해석 가능성 그 이상으로, AI로부터 지능형 시스템을 훈련하는 방법에 대해 배운 내용은 (아직은 그렇지 않다고 확신하지만) 신경 과학에 혁명을 일으킬 것입니다. 제가 신경 과학 분야에서 일할 때 많은 사람들이 제가 지금은 잘못된 학습 질문이라고 생각하는 것에 집중했습니다. 스케일링 가설 / 씁쓸한 교훈 이라는 개념 이 아직 존재하지 않았기 때문입니다. 간단한 목적 함수와 많은 데이터가 엄청나게 복잡한 행동을 유도할 수 있다는 생각은 목적 함수와 구조적 편향을 이해하는 것을 더 흥미롭게 만들고 새로운 계산의 세부 사항을 이해하는 것을 덜 흥미롭게 만듭니다. 저는 최근 몇 년 동안 이 분야를 면밀히 관찰하지 않았지만 계산 신경 과학자들이 여전히 그 교훈을 완전히 흡수하지 못했다는 막연한 느낌이 듭니다. 스케일링 가설에 대한 저의 태도는 항상 "아하 - 이것은 지능이 작동하는 방식과 그것이 얼마나 쉽게 진화하는지에 대한 높은 수준의 설명입니다"였지만, 저는 그것이 평균적인 신경 과학자의 견해라고 생각하지 않습니다. 부분적으로는 스케일링 가설이 "지능의 비결"이라는 것이 AI 내에서도 완전히 수용되지 않았기 때문입니다.

저는 신경과학자들이 이 기본적인 통찰력을 인간 뇌의 특성(생물학적 한계, 진화적 역사, 토폴로지, 운동 및 감각 입력/출력의 세부 사항)과 결합하여

신경과학의 핵심 퍼즐 중 일부를 파악하려고 해야 한다고 생각합니다. 일부는 그럴 가능성이 있지만 아직 충분하지 않을 것으로 생각하며, AI 신경과학자들은 이 각도를 더 효과적으로 활용하여 진전을 가속화할 수 있을 것입니다.

저는 AI가 네 가지 뚜렷한 경로를 따라 신경과학적 진보를 가속화할 것으로 기대합니다. 이 네 가지 경로가 모두 함께 작동하여 정신 질환을 치료하고 기능을 개선할 수 있기를 바랍니다.

전통적인 분자생물학, 화학, 유전학 . 이는 본질적으로 섹션 1의 일반생물학과 같은 이야기이며, AI는 동일한 메커니즘을 통해 이를 가속화할 수 있습니다. 뇌 기능을 변경하고, 각성이나 지각에 영향을 미치고, 기분을 바꾸는 등의 목적으로 신경전달물질을 조절하는 약물이 많이 있으며, AI는 우리가 더 많은 것을 발명하는 데 도움을 줄 수 있습니다. AI는 아마도 정신 질환의 유전적 기초에 대한 연구도 가속화할 수 있을 것입니다.

세밀한 신경 측정 및 개입 . 이는 많은 개별 뉴런 또는 신경 회로가 무엇을 하는지 측정하고 개입하여 행동을 변화시키는 능력입니다. 광유전학 및 신경 프로브는 살아있는 유기체에서 측정과 개입을 모두 수행할 수 있는 기술이며, 매우 진보된 방법(예: 많은 수의 개별 뉴런의 발사 패턴을 판독하는 분자 티커 테이프) 도 제안되었으며 원칙적으로 가능한 것으로 보입니다.

고급 계산 신경 과학 . 위에서 언급했듯이, 현대 AI의 구체적인 통찰력과 게슈탈트는 아마도 정신병이나 기분 장애와 같은 복잡한 질병의 실제 원인과 역학을 밝혀내는 것을 포함하여 시스템 신경 과학 의 질문에 유익하게 적용될 수 있을 것입니다 .

행동적 개입 . 신경과학의 생물학적 측면에 초점을 맞추었기 때문에 많이 언급하지는 않았지만, 정신과와 심리학은 물론 20세기에 걸쳐 광범위한 행

동적 개입 레퍼토리를 개발했습니다 . AI가 새로운 방법을 개발하고 환자가 기존 방법을 고수하도록 돕는 것 모두에서 이러한 개입을 가속화할 수 있다는 것은 당연한 이치입니다. 더 광범위하게, 항상 당신이 최고의 자신이 되도록 돕고, 당신의 상호작용을 연구하고, 더 효과적으로 되는 법을 배우도록 도와주는 "AI 코치"라는 아이디어는 매우 유망해 보입니다.

제 추측으로는, 이 네 가지 진행 경로가 함께 작용하면, 신체 질환과 마찬가지로, AI가 개입하지 않더라도 향후 100년 안에 대부분의 정신 질환을 치료하거나 예방할 수 있는 궤도에 오를 것이고, 따라서 합리적으로 AI 가속 5~10년 안에 완료될 수 있을 것입니다. 구체적으로 제가 추측한 바는 다음과 같습니다.

대부분의 정신 질환은 아마도 치료될 수 있습니다 . 저는 정신 질환의 전문가는 아니지만(신경 과학에서 보낸 시간은 작은 뉴런 그룹을 연구하기 위한 프로브를 만드는 데 보냈습니다) PTSD, 우울증, 정신 분열증, 중독 등의 질병은 위의 네 가지 방향을 조합하여 알아내고 매우 효과적으로 치료할 수 있다고 추측합니다. 답은 "생화학적으로 문제가 발생했습니다"(매우 복잡할 수 있지만)와 "높은 수준에서 신경망에 문제가 발생했습니다"의 조합일 가능성이 큽니다. 즉, 시스템 신경 과학 문제입니다. 하지만 위에서 논의한 행동 개입의 영향을 부정하는 것은 아닙니다. 특히 살아있는 인간에 대한 측정 및 개입 도구는 빠른 반복과 진전으로 이어질 가능성이 높아 보입니다.

매우 "구조적"인 조건은 더 어려울 수 있지만 불가능한 것은 아닙니다 . 정신병이 명백한 신경 해부학적 차이와 관련이 있다는 증거가 있습니다 . 즉, 일부 뇌 영역이 정신병자에게서 단순히 더 작거나 덜 발달되었다는 것입니다. 정신병자는 또한 어린 나이부터 공감 능력이 부족한 것으로 여겨집니다. 그들의 뇌가 다른 점이 무엇이든, 아마도 항상 그랬을 것입니다. 일부 지적 장애와 아마도 다른 조건에도 마찬가지일 수 있습니다. 뇌를 재구조화하는 것은 어렵게 들리지만 지능에 대한 높은 보상이 있는 작업처럼 보입니다.

아마도 성인 뇌를 더 일찍 또는 더 가소성 있는 상태로 유도하여 재형성할 수 있는 방법이 있을 것입니다. 이것이 얼마나 가능한지는 잘 모르겠지만, 제 본능은 AI가 여기서 무엇을 발명할 수 있는지에 대해 낙관적입니다.

정신 질환의 효과적인 유전적 예방은 가능해 보입니다 . 대부분의 정신 질환은 부분적으로 유전 되며 , 게놈 전체 연관 연구는 종종 수가 많은 관련 요인을 식별하는 데 있어 주목을 받기 시작했습니다 . 신체 질환의 경우와 마찬가지로 배아 검사를 통해 이러한 질병의 대부분을 예방하는 것이 가능할 것입니다. 한 가지 차이점은 정신 질환은 다유전자적일 가능성이 더 높다는 것입니다(많은 유전자가 기여함). 따라서 복잡성으로 인해 질병과 관련된 긍정적인 특성을 모르게 선택할 위험이 증가합니다. 그러나 이상하게도 최근 몇 년 동안 GWAS 연구는 이러한 상관 관계가 과장되었을 수 있음 을 시사하는 것 같습니다 . 어떤 경우든 AI 가속 신경 과학은 이러한 사항을 파악하는 데 도움이 될 수 있습니다. 물론 복잡한 특성에 대한 배아 검사는 여러 사회적 문제를 제기하고 논란의 여지가 있지만 대부분의 사람들이 심각하거나 쇠약해지는 정신 질환에 대한 검사를 지지할 것이라고 추측합니다.

우리가 임상적 질병이라고 생각하지 않는 일상적인 문제도 해결될 것입니다 . 우리 대부분은 일반적으로 임상적 질병 수준으로 생각하지 않는 일상적인 심리적 문제를 가지고 있습니다. 어떤 사람들은 화를 잘 내고, 다른 사람들은 집중하는 데 어려움을 겪거나 종종 졸리고, 어떤 사람들은 두렵거나 불안해하거나 변화에 나쁘게 반응합니다. 오늘날에는 경계심이나 집중력(카페인, 모다피닐, 리탈린)을 돕는 약물이 이미 존재하지만 이전의 많은 다른 분야와 마찬가지로 훨씬 더 많은 것이 가능할 것입니다. 아마도 그러한 약물이 훨씬 더 많이 존재하고 발견되지 않았을 것이고, 표적 광 자극(위의 광유전학 참조)이나 자기장과 같은 완전히 새로운 개입 양식이 있을 수도 있습니다. 20세기에 인지 기능과 감정 상태를 조절하는 약물이 얼마나 많이 개발되었는지 감안할 때, 저는 모든 사람이 뇌를 조금 더 잘 작동시키고 일상

생활을 더욱 만족스럽게 경험할 수 있는 "압축된 21세기"에 대해 매우 낙관적입니다.

인간의 기본 경험은 훨씬 더 나을 수 있습니다 . 한 걸음 더 나아가 많은 사람들이 놀라운 계시의 순간, 창조적 영감, 연민, 성취, 초월, 사랑, 아름다움 또는 명상적 평화를 경험했습니다. 이러한 경험의 특성과 빈도는 사람마다 크게 다르고 같은 사람 내에서도 다른 시간에 다르며 때로는 다양한 약물에 의해 유발될 수도 있습니다(비록 종종 부작용이 있지만). 이 모든 것은 "경험할 수 있는 것의 공간"이 매우 광범위하고 사람들의 삶의 더 많은 부분이 이러한 놀라운 순간으로 구성될 수 있음을 시사합니다. 아마도 전반적으로 다양한 인지 기능을 개선하는 것도 가능할 것입니다. 이것은 아마도 "생물학적 자유" 또는 "연장된 수명"의 신경과학적 버전일 것입니다.

AI에 대한 공상과학 묘사에서 자주 등장하지만, 여기서는 의도적으로 논의하지 않은 주제 중 하나는 "마인드 업로딩"으로, 인간 뇌의 패턴과 역학을 포착하여 소프트웨어로 구현하는 아이디어입니다. 이 주제는 그 자체로 에세이의 주제가 될 수 있지만, 업로딩이 이론적으로는 거의 확실히 가능하다고 생각하지만 , 실제로는 강력한 AI를 사용하더라도 상당한 기술적, 사회적 과제에 직면해 있으며, 우리가 논의하는 5~10년 창구를 벗어나게 할 가능성이 크다고만 말씀드리겠습니다.

요약하자면, AI 가속 신경과학은 대부분의 정신 질환에 대한 치료법을 크게 개선하거나 심지어 치유할 가능성이 높으며, "인지 및 정신적 자유"와 인간의 인지 및 감정 능력을 크게 확장할 것입니다. 이는 이전 섹션에서 설명한 신체 건강의 개선만큼이나 급진적일 것입니다. 세상은 겉보기에 눈에 띄게 다르지 않을지 몰라도, 인간이 경험하는 세상은 훨씬 더 나은, 더 인간적인 곳이 될 것이며, 자아 실현을 위한 더 큰 기회를 제공하는 곳이 될 것입니다. 저는 또한 정신 건강이 개선되면 정치적 또는 경제적으로 보이는 문제를 포함하여 많은 다른 사회적 문제가 완화될 것이라고 생각합니다.

3. 경제 발전과 빈곤

이전 두 섹션은 질병을 치료하고 인간 삶의 질을 개선하는 새로운 기술을 개발하는 것에 관한 것입니다 . 그러나 인도주의적 관점에서 분명한 질문은 "모든 사람이 이러한 기술에 접근할 수 있을까?"입니다.

질병에 대한 치료법을 개발하는 것은 한 가지이고, 질병을 세상에서 근절하는 것은 또 다른 것입니다. 더 광범위하게, 많은 기존의 건강 개입이 아직 전 세계에 적용되지 않았으며, (비건강) 기술 개선에도 일반적으로 마찬가지입니다. 이를 다른 방식으로 표현하면, 세계 여러 지역의 생활 수준은 여전히 절망적으로 열악합니다. 사하라 이남 아프리카의 1인당 GDP는 약 2,000달러인 반면, 미국의 1인당 GDP는 약 75,000달러입니다. AI가 개발도상국을 돕는 데 거의 도움이 되지 않으면서 선진국의 경제 성장과 삶의 질을 더욱 증가시킨다면, 우리는 그것을 끔찍한 도덕적 실패이자 이전 두 섹션에서의 진정한 인도주의적 승리에 대한 오점으로 보아야 합니다. 이상적으로는 강력한 AI가 개발도상국이 선진국 을 따라잡는 데 도움이 되어야 하며 , 동시에 후자를 혁신해야 합니다.

저는 AI가 불평등과 경제 성장에 대처할 수 있다는 것에 대해 기본적인 기술을 발명할 수 있다는 것에 대해 확신하는 만큼 확신하지 못합니다. 기술은 지능에 대한 명백히 높은 수익(복잡성과 데이터 부족을 우회하는 능력 포함)이 있는 반면, 경제는 인간의 많은 제약과 많은 양의 내재적 복잡성을 수반하기 때문입니다. 저는 AI가 유명한 " 사회주의 계산 문제 " 23를 해결할 수 있다는 것에 다소 회의적 이며, 정부가 그러한 기관에 경제 정책을 넘겨줄 것이라고 생각하지 않습니다(또는 넘겨줘야 한다고 생각하지 않습니다). 효과가 있지만 사람들이 의심할 수 있는 치료법을 받아들이도록 사람들

을 설득하는 방법과 같은 문제도 있습니다.

개발도상국이 직면한 과제는 민간 및 공공 부문 모두에서 만연한 부패 로 인해 더욱 복잡해졌습니다 . 부패는 악순환을 만듭니다. 부패는 빈곤을 악화 시키고 , 빈곤은 다시 더 많은 부패를 낳습니다. AI 주도 경제 개발 계획은 부패, 취약한 제도 및 기타 매우 인간적인 과제를 고려해야 합니다.

그럼에도 불구하고, 저는 낙관할 만한 상당한 이유가 있다고 봅니다. 질병은 근절 되었고 많은 국가가 가난에서 부유로 바뀌 었으며 , 이러한 작업에 관련된 결정은 지능에 대한 높은 수익을 보인다는 것이 분명합니다(인간의 제약과 복잡성에도 불구하고). 따라서 AI는 현재 수행되고 있는 것보다 더 잘 수행할 가능성이 높습니다. 또한 인간의 제약을 우회하고 AI가 집중할 수 있는 표적 개입이 있을 수 있습니다. 그러나 더 중요한 것은 우리가 시도해야 한다는 것입니다. AI 회사와 선진국 정책 입안자 모두 개발도상국이 소외되지 않도록 각자의 역할을 해야 합니다. 도덕적 명령이 너무 큽니다. 따라서 이 섹션에서는 낙관적인 주장을 계속할 것이지만, 성공이 보장되지는 않으며 우리의 집단적 노력에 달려 있다는 점을 명심하십시오.

아래에서 저는 강력한 AI가 개발된 후 5~10년 동안 개발도상국에서 상황이 어떻게 전개될지에 대한 몇 가지 추측을 내립니다.

건강 개입의 분배 . 제가 가장 낙관적인 분야는 아마도 전 세계에 건강 개입을 분배하는 것입니다. 질병은 실제로 상향식 캠페인을 통해 근절되었습니다. 천연두는 1970년대에 완전히 근절 되었고 , 소아마비와 기니벌레는 연간 100건 미만으로 거의 근절되었습니다. 수학적으로 정교한 역학 모델링은 질병 근절 캠페인에서 적극적인 역할을 하며, 인간보다 더 똑똑한 AI 시스

템이 인간보다 더 잘 해낼 여지가 매우 큰 것으로 보입니다. 분배의 물류도 아마도 크게 최적화될 수 있을 것입니다. GiveWell 의 초기 기부자로서 제가 배운 한 가지는 일부 건강 자선 단체가 다른 자선 단체보다 훨씬 더 효과적이라는 것입니다. AI 가속화 노력이 훨씬 더 효과적일 것이라는 희망이 있습니다. 또한 일부 생물학적 진보는 실제로 분배의 물류를 훨씬 더 쉽게 만듭니다. 예를 들어, 말라리아는 질병에 걸릴 때마다 치료가 필요하기 때문에 근절하기 어려웠습니다. 한 번만 투여하면 되는 백신은 물류를 훨씬 더 단순하게 만들어줍니다(그리고 말라리아에 대한 그러한 백신은 실제로 현재 개발 중입니다). 더 간단한 유통 메커니즘도 가능합니다. 어떤 질병은 원칙적으로 동물 보균자를 표적으로 삼아 근절될 수 있습니다. 예를 들어, 질병을 운반하는 능력을 차단하는 박테리아에 감염된 모기를 풀어놓거나 (그러면 다른 모든 모기를 감염시킵니다) 단순히 유전자 드라이브를 사용하는 것입니다.모기를 없애는 것입니다. 이를 위해서는 개별적으로 수백만 마리를 치료해야 하는 조정된 캠페인보다는 하나 또는 몇 가지 중앙 집중화된 조치가 필요합니다. 전반적으로, 저는 AI 기반 건강 혜택의 상당 부분(아마도 50%)이 세계에서 가장 가난한 나라에도 전파되기에 5~10년이 적당한 시간이라고 생각합니다. 강력한 AI가 도입된 지 5~10년 후에 개발도상국이 적어도 오늘날 선진국보다 상당히 더 건강해지는 것이 좋은 목표가 될 수 있습니다. 선진국보다 계속 뒤처지더라도 말입니다. 이를 달성하려면 물론 글로벌 건강, 자선 활동, 정치적 옹호 및 기타 많은 노력에 막대한 노력이 필요하며, AI 개발자와 정책 입안자 모두가 이를 도와야 합니다.

경제 성장 . 개발도상국이 건강뿐만 아니라 경제적으로 전반적으로 선진국을 빠르게 따라잡을 수 있을까요? 이에 대한 선례가 있습니다. 20세기 마지막 수십 년 동안 동아시아 여러 경제권이지속적인 ~10%의 연간 실질 GDP 성장률을 달성하여 선진국을 따라잡을 수 있었습니다. 인간 경제 기획자는 전체 경제를 직접 통제하지 않고 몇 가지 핵심 레버(수출 주도 성장의 산업 정책, 천연 자원 부에 의존하려는 유혹을 저항하는 것 등)를 당겨 이러한 성

공으로 이어진 결정을 내렸습니다. "AI 재무 장관과 중앙 은행가"가 이 10% 성과를 복제하거나 초과할 수 있을 가능성이 있습니다. 중요한 질문은 개발도상국 정부가 자기 결정의 원칙을 존중하면서 이를 채택하도록 하는 방법입니다. 일부는 이에 열광할 수 있지만 다른 일부는 회의적일 가능성이 높습니다. 낙관적인 측면에서, 이전 요점의 많은 건강 개입은 경제 성장을 유기적으로 증가시킬 가능성이 높습니다. AIDS/말라리아/기생충을 근절하면 생산성에 혁신적 효과가 있을 뿐만 아니라 일부 신경과학적 개입(기분과 집중력 향상 등)이 선진국과 개발도상국 모두에 미칠 경제적 이점은 말할 것도 없습니다. 마지막으로, 비건강 AI 가속 기술(예: 에너지 기술, 운송 드론, 개선된 건축 자재, 더 나은 물류 및 유통 등)은 단순히 자연스럽게 세상에 스며들 수 있습니다. 예를 들어, 휴대전화조차도 자선 활동 없이도 시장 메커니즘을 통해 사하라 이남 아프리카에 빠르게 스며들었습니다. 더 부정적인 측면에서 AI와 자동화는 많은 잠재적 이점이 있지만, 특히 아직 산업화되지 않은 국가의 경제 개발에 과제를 안겨줍니다. 자동화가 증가하는 시대에도 이러한 국가가 여전히 경제를 개발하고 개선할 수 있는 방법을 찾는 것은 경제학자와 정책 입안자가 해결해야 할 중요한 과제입니다. 전반적으로, 꿈의 시나리오(아마도 목표로 삼을 만한 목표)는 개발도상국의 연간 GDP 성장률이 20%이고, 각각 10%가 AI 기반 경제적 결정과 AI 가속 기술의 자연스러운 확산(건강을 포함하되 이에 국한되지 않음)에서 발생하는 것입니다. 이것이 달성된다면, 사하라 이남 아프리카는 5~10년 안에 중국의 현재 1인당 GDP 수준으로 올라갈 것이고, 나머지 개발도상국의 많은 부분은 현재 미국 GDP보다 더 높은 수준으로 올라갈 것입니다. 다시 말해서,이건 꿈의 시나리오지, 기본적으로 일어나는 일이 아닙니다. 우리 모두가 힘을 합쳐서 실현 가능성을 높여야 하는 일이죠.

식량 안보 24. 더 나은 비료와 살충제, 더 많은 자동화, 더 효율적인 토지 이용과 같은 작물 기술의 발전은 20세기에 걸쳐 작물 수확량을 크게 증가시켜 수백만 명의 사람들을 굶주림으로부터 구했습니다. 유전자 조작은 현재

많은 작물을 더욱 개량하고 있습니다 . 이를 수행하는 더 많은 방법을 찾고 농업 공급망을 더욱 효율적으로 만드는 것은 AI 주도의 두 번째 녹색 혁명을 가져올 수 있으며 , 개발도상국과 선진국 간의 격차를 줄이는 데 도움이 될 수 있습니다.

기후 변화 완화 . 기후 변화는 개발도상국에서 훨씬 더 강하게 느껴질 것이며, 이는 개발을 방해할 것입니다. AI가 대기 탄소 제거 및 청정 에너지 기술 부터 탄소 집약적 공장식 농장에 대한 의존도를 줄이는 실험실에서 재배한 고기 에 이르기까지 기후 변화를 늦추거나 방지하는 기술의 개선으로 이어질 것으로 예상할 수 있습니다 . 물론, 위에서 논의했듯이 기술이 기후 변화의 진전을 제한하는 유일한 요소는 아닙니다. 이 글에서 논의한 다른 모든 문제와 마찬가지로 인간 사회적 요인도 중요합니다. 그러나 AI 강화 연구를 통해 기후 변화 완화를 훨씬 덜 비용이 많이 들고 파괴적으로 만들어 많은 반대 의견을 무의미하게 만들고 개발도상국이 더 많은 경제적 진전을 이룰 수 있는 수단을 제공할 것이라고 생각할 만한 충분한 이유가 있습니다.

국가 내 불평등 . 저는 불평등을 글로벌 현상(저는 이것이 가장 중요한 표현이라고 생각합니다)으로 주로 이야기했지만, 물론 불평등은 국가 내에서도 존재합니다 . 진보된 건강 개입과 특히 수명이나 인지 향상 약물의 급격한 증가로 인해 이러한 기술이 "부자만을 위한 것"이라는 타당한 우려가 확실히 있을 것입니다. 저는 두 가지 이유에서 특히 선진국의 국가 내 불평등에 대해 더 낙관적입니다. 첫째, 시장은 선진국에서 더 잘 기능하고, 시장은 일반적으로 시간이 지남에 따라 고가 기술의 비용을 낮추는 데 능숙합니다 . 25 둘째, 선진국 정치 기관은 시민들에게 더 잘 반응하고 보편적 접근 프로그램을 실행할 수 있는 국가 역량이 더 큽니다. 그리고 저는 시민들이 삶의 질을 근본적으로 개선하는 기술에 대한 접근을 요구할 것으로 기대합니다. 물론 그러한 요구가 성공할 것이라고 미리 결정된 것은 아닙니다. 그리고 여기 우리가 집단적으로 공정한 사회를 보장하기 위해 할 수 있는 모든 것

을 해야 하는 또 다른 곳이 있습니다. 부의 불평등 (생명을 구하고 삶을 향상시키는 기술에 대한 접근성 불평등과는 대조적으로) 이라는 별도의 문제가 있는데 , 이 문제는 더 어려운 것으로 보이며, 5장에서 이에 대해 논의하겠습니다.

옵트아웃 문제 . 선진국과 개발도상국 모두에서 우려되는 문제 중 하나는 사람들이 AI 지원 혜택을 거부하는 것 입니다 (반백신 운동이나 더 일반적으로 러다이트 운동과 유사). 예를 들어, 가장 좋은 결정을 내릴 수 없는 사람들이 의사 결정 능력을 향상시키는 기술을 거부하여 격차가 계속 커지고 심지어 디스토피아적 하위 계층이 생성되는 등 나쁜 피드백 주기가 발생할 수 있습니다(일부 연구자들은 이것이 민주주의를 훼손 할 것이라고 주장했습니다 . 이 주제는 다음 섹션에서 자세히 논의합니다). 이는 다시 한 번 AI의 긍정적인 발전에 도덕적 오점을 남길 것입니다. 이는 사람들에게 강요하는 것이 윤리적으로 괜찮다고 생각하지 않기 때문에 해결하기 어려운 문제이지만, 적어도 사람들의 과학적 이해를 높이기 위해 노력할 수 있으며 아마도 AI 자체가 이를 도울 수 있을 것입니다. 희망적인 신호 중 하나는 역사적으로 반기술 운동이 공격보다는 짖는 소리가 더 컸다는 것입니다. 현대 기술에 대한 비난은 인기가 있지만, 대부분의 사람들은 결국 그것을 채택합니다. 적어도 개인의 선택 문제일 때는 말입니다. 개인은 대부분의 건강 및 소비자 기술을 채택하는 경향이 있는 반면, 핵 에너지와 같이 진정으로 방해받는 기술은 집단적인 정치적 결정이 되는 경향이 있습니다.

전반적으로 저는 AI의 생물학적 진보를 개발도상국의 사람들에게 빠르게 제공하는 데 대해 낙관적입니다. 저는 AI가 전례 없는 경제 성장률을 가능하게 하고 개발도상국이 적어도 선진국이 지금 있는 수준을 뛰어넘을 수 있도록 할 수 있기를 바라지만, 확신하지는 않습니다. 저는 선진국과 개발도상국 모두에서 "옵트아웃" 문제가 우려되지만, 시간이 지남에 따라 사라질 것이고 AI가 이 과정을 가속화하는 데 도움이 될 것이라고 생각합니다. 완벽한 세상은 아닐 것이고, 뒤처진 사람들은 적어도 처음 몇 년 동안은 따라잡지 못

할 것입니다. 하지만 우리가 강력한 노력을 기울이면, 우리는 상황을 올바른 방향으로, 그리고 빠르게 움직일 수 있을 것입니다. 그렇게 한다면, 우리는 지구상의 모든 인간에게 빚진 존엄성과 평등에 대한 약속에 대한 선급금을 최소한 지불할 수 있을 것입니다.

4. 평화와 통치

처음 세 섹션의 모든 것이 잘 진행된다고 가정해 보자. 질병, 빈곤, 불평등이 상당히 줄어들고 인간 경험의 기준이 상당히 높아진다. 그렇다고 해서 인간 고통의 모든 주요 원인이 해결되는 것은 아니다. 인간은 여전히 서로에게 위협이 된다 . 민주주의와 평화로 이어지는 기술적 개선과 경제 발전 추세가 있지만 , 그것은 매우 느슨한 추세이며 빈번하고(최근 에) 퇴보한다. 20세기 초, 사람들은 전쟁을 뒤로 하고 있다고 생각했다 . 그런 다음 두 차례의 세계 대전이 일어났다. 30년 전 프랜시스 후쿠야마는 " 역사의 종말 "과 자유민주주의의 최종 승리에 대해 썼다. 아직은 그런 일이 일어나지 않았다. 20년 전 미국 정책 입안자들은 중국과의 자유 무역이 중국이 부유해짐에 따라 자유화되도록 만들 것이라고 믿었다. 그런 일은 거의 일어나지 않았고, 우리는 이제 다시 부상하는 권위주의 블록과의 두 번째 냉전으로 향하는 듯하다 . 그리고 그럴듯한 이론에 따르면 인터넷 기술은 실제로 권위주의에 유리할 수 있으며 , 처음 믿었던 것처럼 민주주의에 유리하지 않을 수 있습니다(예: "아랍의 봄" 시기). AI가 평화, 민주주의, 자유의 문제와 얼마나 교차하는지 이해하는 것이 중요해 보입니다.

안타깝게도 저는 AI가 민주주의와 평화를 우선적으로 또는 구조적으로 증진시킬 것이라고 믿을 만한 강력한 이유가 보이지 않습니다. 제가 AI가 구조적으로 인간의 건강을 증진시키고 빈곤을 완화시킬 것이라고 생각하는 것과 같은 방식입니다. 인간의 갈등은 적대적이며 AI는 원칙적으로 "좋은 사람"과 "나쁜 사람" 모두를 도울 수 있습니다. 오히려 일부 구조적 요인이 걱정스럽

습니다. AI는 독재자의 도구 키트에서 주요 도구인 훨씬 더 나은 선전과 감시를 가능하게 할 가능성이 높아 보입니다. 따라서 올바른 방향으로 사물을 기울이는 것은 개별 행위자로서 우리에게 달려 있습니다. AI가 민주주의와 개인의 권리를 선호하기를 원한다면 그 결과를 위해 싸워야 할 것입니다. 저는 국제적 불평등보다 이 문제에 대해 더 강하게 느낍니다. 자유주의 민주주의와 정치적 안정의 승리는 보장되지 않으며 , 아마도 가능성이 없을 수도 있으며, 과거에 종종 그랬듯이 우리 모두의 큰 희생과 헌신이 필요할 것입니다.

저는 이 문제가 두 가지 부분으로 구성되어 있다고 생각합니다. 국제 갈등과 국가의 내부 구조입니다. 국제적인 측면에서 강력한 AI가 만들어질 때 민주주의가 세계 무대에서 우위를 차지하는 것이 매우 중요한 것 같습니다. AI 기반 권위주의는 생각하기에는 너무 끔찍해 보이므로 민주주의는 권위주의에 압도당하는 것을 피하고 권위주의 국가 내에서 인권 침해를 방지하기 위해 강력한 AI가 세상에 출시되는 조건을 정할 수 있어야 합니다.

현재 제가 추측하기로는 이를 수행하는 가장 좋은 방법은 "협상 전략" 26을 통한 것입니다 . 여기서 민주주의 연합은 강력한 AI의 공급망을 확보하고, 빠르게 확장하며, 칩과 반도체 장비와 같은 핵심 리소스에 대한 적의 접근을 차단하거나 지연시킴으로써 강력한 AI에 대한 명확한 이점(일시적일지라도)을 얻으려 합니다. 이 연합은 한편으로는 AI를 사용하여 강력한 군사적 우월성(막대기)을 달성하는 동시에 강력한 AI의 이점(당근)을 점점 더 많은 국가에 분배하여 연합의 민주주의 증진 전략을 지원하는 대가로 제공할 것입니다(이는 " 평화를 위한 원자력 "과 약간 유사할 것입니다). 연합은 점점 더 많은 세계의 지지를 얻고, 최악의 적을 고립시키고, 결국 그들이 나머지 세계와 같은 거래를 하는 것이 더 나은 위치에 놓이게 하는 것을 목표로 할 것입니다. 모든 이점을 받기 위해 민주주의와의 경쟁을 포기하고 더 뛰어난

적과 싸우지 않는 것입니다.

우리가 이 모든 것을 할 수 있다면, 민주주의가 세계 무대를 선도하고 독재 정권에 의해 훼손되거나 정복되거나 방해받지 않을 경제적, 군사적 힘을 갖추고 AI 우월성을 지속 가능한 이점으로 활용할 수 있는 세상이 될 것입니다. 이는 낙관적으로 "영원한 1991"로 이어질 수 있습니다. 민주주의가 우위를 차지하고 후쿠야마의 꿈이 실현되는 세상입니다. 다시 말하지만, 이는 달성하기 매우 어려울 것이며, 특히 사적 AI 기업과 민주 정부 간의 긴밀한 협력과 당근과 채찍의 균형에 대한 매우 현명한 결정이 필요할 것입니다.

모든 것이 잘 되더라도, 각 국가 내에서 민주주의와 독재 사이의 싸움이라는 문제가 남습니다 . 여기서 무슨 일이 일어날지 예측하기는 분명 어렵지만, 민주주의가 가장 강력한 AI를 통제하는 글로벌 환경에서 AI 가 실제로 구조적으로 모든 곳에서 민주주의를 선호할 수 있다는 낙관론이 있습니다. 특히, 이 환경에서 민주 정부는 우수한 AI를 사용하여 정보 전쟁에서 승리할 수 있습니다. 독재 정권의 영향력과 선전 활동에 대항할 수 있으며, 독재 정권이 차단하거나 모니터링할 수 있는 기술적 능력이 없는 방식으로 정보와 AI 서비스 채널을 제공함으로써 전 세계적으로 자유로운 정보 환경을 조성할 수도 있습니다. 선전을 전달할 필요는 없고, 악의적인 공격에 대응하고 정보의 자유로운 흐름을 차단 해제하기만 하면 됩니다. 당장은 아니지만, 이와 같은 공평한 경쟁 환경은 여러 가지 이유로 글로벌 거버넌스를 점차적으로 민주주의로 기울일 수 있는 좋은 기회가 있습니다.

첫째, 1-3절의 삶의 질 향상은 모든 조건이 동등하다면 민주주의를 증진시킬 것입니다. 역사적으로 적어도 어느 정도는 그렇습니다. 특히 정신 건강, 웰빙, 교육의 향상이 민주주의를 증진시킬 것으로 기대합니다. 왜냐하면 이

세 가지 모두 권위주의적 지도자에 대한 지지 와 부정적으로 상관관계 가 있기 때문입니다 . 일반적으로 사람들은 다른 요구가 충족될 때 더 많은 자기 표현을 원하며, 민주주의는 다른 것들 중에서도 자기 표현의 한 형태입니다. 반대로 권위주의는 두려움과 원망에 의해 번창합니다.

두 번째, 권위주의자들이 검열할 수 없는 한, 무료 정보가 실제로 권위주의를 약화시킬 가능성이 높습니다. 그리고 검열되지 않은 AI는 개인에게 억압적인 정부를 약화시킬 수 있는 강력한 도구를 제공할 수도 있습니다. 억압적인 정부는 사람들에게 특정한 종류의 상식을 거부하고 "황제는 옷을 입지 않았다"는 것을 깨닫지 못하게 함으로써 살아남습니다. 예를 들어 세르비아에서 밀로셰비치 정부를 전복하는 데 도움을 준 스르자 포포비치는 권위주의자들의 권력을 심리적으로 강탈하고, 주문을 깨고, 독재자에 대한 지지를 모으는 기술에 대해 광범위하게 기술했습니다. 모든 사람의 주머니에 있는 초인적 효율성을 가진 AI 버전의 포포비치(그 기술은 지능에 대한 높은 수익으로 보임)는 독재자가 차단하거나 검열할 힘이 없는 것으로, 전 세계의 반체제 인사와 개혁가들의 등에 바람을 일으킬 수 있습니다. 다시 말씀드리자면, 이는 길고 지루한 싸움이 될 것이며 승리가 보장된 싸움은 아닙니다. 하지만 우리가 AI를 올바른 방법으로 설계하고 구축한다면 적어도 전 세계 자유를 옹호하는 사람들이 유리한 싸움이 될 수 있을 것입니다.

신경과학과 생물학에서와 마찬가지로, 우리는 또한 사물이 "정상보다 더 나아질 수 있는" 방법을 물을 수 있습니다. 독재를 피하는 방법뿐만 아니라 민주주의를 오늘날보다 더 나아지게 하는 방법입니다. 민주주의 내에서도 불의는 항상 일어납니다. 법치주의 사회는 시민들에게 모든 사람이 법 앞에서 평등할 것이고 모든 사람이 기본 인권을 누릴 자격이 있다고 약속하지만, 사람들이 항상 실제로 그 권리를 누리는 것은 아닙니다. 이 약속이 부분적으로라도 이행된다는 것은 자랑스러운 일이지만, AI가 우리가 더 잘하는 데

도움이 될 수 있을까요?

예를 들어, AI가 의사 결정과 프로세스를 보다 공평하게 만들어 우리의 법률 및 사법 시스템을 개선할 수 있을까요? 오늘날 사람들은 대부분 법률 또는 사법적 맥락에서 AI 시스템이 차별의 원인이 될 것이라고 걱정하고 있으며 , 이러한 걱정은 중요하며 방어해야 합니다. 동시에 민주주의의 활력은 위험에 대응하는 것이 아니라 새로운 기술을 활용하여 민주적 기관을 개선하는 데 달려 있습니다. AI의 진정으로 성숙하고 성공적인 구현은 편견을 줄이고 모든 사람에게 더 공평할 수 있는 잠재력이 있습니다.

수세기 동안 법 체계는 법이 공평해야 하지만 본질적으로 주관적이어서 편파적인 인간이 해석해야 한다는 딜레마에 직면해 왔습니다. 법을 완전히 기계적으로 만들려는 시도는 현실 세계가 지저분하고 항상 수학적 공식으로 포착할 수 없기 때문에 효과가 없었습니다. 대신 법 체계는 " 잔혹하고 비정상적인 처벌 "이나 " 사회적 중요성을 전혀 회복하지 못하는 "과 같은 악명 높게 부정확한 기준에 의존하는데, 그런 다음 인간이 이를 해석하고 종종 편파성, 편애 또는 자의성을 보이는 방식으로 해석합니다. 암호화폐의 " 스마트 계약 "은 일반 코드가 그렇게 많은 관심사를 판단할 만큼 똑똑하지 않기 때문에 법을 혁신하지 못했습니다. 하지만 AI는 이를 위해 충분히 똑똑할 수 있습니다. 반복 가능하고 기계적인 방식으로 광범위하고 모호한 판단을 내릴 수 있는 최초의 기술입니다.

저는 판사를 문자 그대로 AI 시스템으로 대체하자고 제안하는 것이 아니지만, 공정성과 지저분하고 현실적인 상황을 이해하고 처리하는 능력의 조합은 법과 정의에 심각한 긍정적 적용이 있어야 한다고 생각합니다 . 최소한 그러한 시스템은 의사 결정을 위한 보조 도구로 인간과 함께 작동할 수 있습

니다. 그러한 시스템에서 투명성은 중요할 것이고, 성숙한 AI 과학은 그것을 제공할 수 있을 것입니다. 그러한 시스템의 훈련 과정을 광범위하게 연구할 수 있고, 고급 해석 기술을 사용하여 최종 모델 내부를 보고 숨겨진 편견을 평가할 수 있는데, 이는 인간으로는 단순히 불가능한 방식입니다. 그러한 AI 도구는 또한 사법 또는 경찰 맥락에서 기본권 침해를 모니터링하는 데 사용될 수 있으며, 헌법을 보다 자체적으로 시행할 수 있습니다.

비슷한 맥락에서 AI는 의견을 모으고 시민들의 합의를 이끌어 갈등을 해결하고 공통점을 찾고 타협을 모색하는 데 사용될 수 있습니다. 이 방향의 초기 아이디어 중 일부는 Anthropic과의 협업을 포함하여 계산 민주주의 프로젝트 에서 수행되었습니다 . 더 많은 정보를 갖추고 사려 깊은 시민은 분명히 민주주의 기관을 강화할 것입니다.

또한 AI를 사용하여 원칙적으로 모든 사람이 이용할 수 있지만 실제로는 종종 심각하게 부족하고 어떤 지역에서는 다른 지역보다 더 심각한 건강 혜택이나 사회 서비스와 같은 정부 서비스를 제공하는 데 사용할 수 있는 명확한 기회가 있습니다.여기에는 건강 서비스, DMV, 세금, 사회 보장, 건축법 집행 등이 포함됩니다.정부가 합법적으로 부여한 모든 것을 이해할 수 있는 방식으로 제공하는 것이 임무인 매우 사려 깊고 정보에 입각한 AI가 있고 종종 혼란스러운 정부 규칙을 준수하도록 돕는다면 큰 의미가 있을 것입니다.국가 역량을 늘리는 것은 법 앞에서의 평등이라는 약속을 이행하는 데 도움이 되고 민주적 거버넌스에 대한 존중을 강화합니다.현재 제대로 구현되지 않은 서비스는 정부에 대한 냉소주의의 주요 원인 입니다 .27

이 모든 것은 다소 모호한 아이디어이며, 이 섹션의 시작 부분에서 말했듯이, 저는 생물학, 신경 과학, 빈곤 완화 분야의 발전만큼 그 실현 가능성에

대해 확신하지 못합니다. 비현실적으로 유토피아적일 수 있습니다. 하지만 중요한 것은 야심 찬 비전을 갖고 큰 꿈을 꾸고 시도할 의지가 있는 것입니다. 자유, 개인의 권리, 법적 평등을 보장하는 AI의 비전은 싸우지 않을 수 없을 만큼 강력한 비전입니다. 21세기 AI 기반 정치는 개인의 자유를 더 강력하게 보호할 수 있고, 자유주의 민주주의를 전 세계가 채택하고자 하는 정부 형태로 만드는 데 도움이 되는 희망의 등대가 될 수 있습니다.

5. 일과 의미

앞의 네 섹션의 모든 것이 잘 되더라도—질병, 빈곤, 불평등을 완화할 뿐만 아니라 자유주의 민주주의가 지배적인 정부 형태가 되고, 기존의 자유주의 민주주의가 더 나은 버전이 되더라도—적어도 하나의 중요한 의문은 여전히 남습니다. "우리가 기술적으로 진보된 세상에 살고 공정하고 괜찮은 세상에 사는 건 좋은 일이죠"라고 누군가는 반대할 수 있습니다. "하지만 AI가 모든 것을 한다면 인간은 어떻게 의미를 가질 수 있을까요? 사실, 그들은 어떻게 경제적으로 살아남을 수 있을까요?"

저는 이 질문이 다른 질문들보다 더 어렵다고 생각합니다. 다른 질문들보다 반드시 더 비관적이라는 뜻은 아닙니다(도전은 보지만요). 저는 이 질문이 사회가 어떻게 조직되는지에 대한 거시적인 질문과 관련이 있고, 시간이 지나고 분산된 방식으로만 해결되는 경향이 있기 때문에 더 모호하고 미리 예측하기 어렵다는 뜻입니다. 예를 들어, 역사적 수렵채집 사회는 사냥과 다양한 종류의 사냥 관련 종교 의식 없이는 삶이 무의미하다고 생각했을 수도 있고, 우리의 잘 먹고 사는 기술 사회는 목적이 없다고 생각했을 수도 있습니다. 또한 우리 경제가 모든 사람을 부양할 수 있는 방법이나 기계화된 사회에서 사람들이 어떤 기능을 유용하게 수행할 수 있는지 이해하지 못했을 수도 있습니다.

그럼에도 불구하고, 이 섹션이 간결하다는 것이 제가 이 문제를 심각하게 받아들이지 않는다는 신호로 받아들여져서는 안 된다는 점을 명심하면서, 적어도 몇 마디라도 말씀드리는 것이 좋습니다. 오히려, 이는 명확한 답변이 부족하다는 신호입니다.

의미에 대한 질문에 대해, AI가 더 잘할 수 있다는 이유만으로 당신이 수행하는 작업이 무의미하다고 믿는 것은 매우 실수일 가능성이 높다고 생각합니다. 대부분의 사람들은 어떤 일에서든 세계 최고가 아니며, 그들은 그것이 특별히 신경 쓰이지 않는 것 같습니다. 물론 오늘날에도 그들은 여전히 비교 우위를 통해 기여할 수 있으며, 그들이 생산하는 경제적 가치에서 의미를 얻을 수 있지만, 사람들은 또한 경제적 가치를 생산하지 않는 활동을 크게 즐깁니다. 저는 비디오 게임을 하고, 수영하고, 밖을 걷고, 친구들과 이야기하는 데 많은 시간을 보내는데, 이 모든 것은 경제적 가치를 전혀 생성하지 않습니다. 저는 비디오 게임을 더 잘하거나 산을 오르는 데 더 빨리 노력하기 위해 하루를 보낼 수도 있고, 어딘가에 누군가가 그런 일에 훨씬 더 뛰어나다는 것은 저에게는 별로 중요하지 않습니다. 어쨌든 저는 의미는 주로 경제적 노동이 아니라 인간 관계와 연결에서 온다고 생각합니다. 사람들은 성취감, 심지어 경쟁감을 원하며, AI 이후의 세계에서는 오늘날 사람들이 연구 프로젝트에 착수하거나 할리우드 배우가 되려고 하거나 회사를 설립할 때 하는 것과 비슷하게 복잡한 전략을 가지고 매우 어려운 작업을 시도하는 데 몇 년을 보내는 것이 완벽하게 가능할 것입니다 . 28. (a) 어딘가의 AI가 원칙적으로 이 작업을 더 잘 수행할 수 있다는 사실, (b) 이 작업이 더 이상 세계 경제의 경제적으로 보상되는 요소가 아니라는 사실은 나에게 그다지 중요하지 않은 것 같습니다.

경제적 부분은 실제로 의미적 부분보다 더 어려워 보입니다. 이 섹션에서 "

경제적"이라는 말은 대부분 또는 모든 인간이 충분히 발전된 AI 주도 경제에 의미 있게 기여할 수 없을 수 있는 문제를 의미합니다. 이는 3절에서 논의한 새로운 기술에 대한 접근성 불평등과 같은 별도의 불평등 문제보다 더 거시적인 문제입니다.

우선, 단기적으로 비교 우위가 인간을 계속해서 관련성 있게 유지 하고 실제로 생산성을 증가시킬 것이며, 어떤 면에서는 인간 간의 경쟁 환경을 평준화 할 수도 있다는 주장에 동의합니다 . AI가 주어진 작업의 90%에서만 더 나은 한, 나머지 10%는 인간이 고도로 레버리지되어 보상을 늘리고 실제로 AI가 잘하는 것을 보완하고 증폭하는 많은 새로운 인간 일자리를 창출하여 "10%"가 확장되어 거의 모든 사람을 계속 고용하게 될 것입니다. 사실, AI가 인간보다 100%의 일을 더 잘할 수 있더라도 일부 작업에서는 비효율적이거나 비용이 많이 들거나 인간과 AI에 대한 리소스 입력이 의미 있게 다르다면 비교 우위의 논리가 계속 적용됩니다. 인간이 상당한 기간 동안 상대적(또는 절대적) 이점을 유지할 가능성이 있는 한 영역은 물리적 세계입니다. 따라서 인간 경제는 "데이터 센터의 천재 국가"에 도달하는 지점을 약간 지나서도 계속 의미가 있을 것이라고 생각합니다.

하지만 저는 장기적으로 AI가 매우 광범위하게 효과적이고 저렴해져서 더 이상 적용되지 않을 것이라고 생각합니다. 그 시점에서 우리의 현재 경제 구조는 더 이상 의미가 없게 될 것이고, 경제를 어떻게 조직해야 하는지에 대한 보다 광범위한 사회적 대화가 필요할 것입니다.

미친 짓처럼 들릴지 몰라도, 사실 문명은 과거에 수렵채집에서 농사로, 농사에서 봉건제로, 봉건제에서 산업주의로의 주요 경제적 변화를 성공적으로 헤쳐 나갔습니다. 저는 새롭고 이상한 것이 필요할 것이라고 생각하며, 오늘날

아무도 제대로 상상하지 못한 것이라고 생각합니다. 모든 사람을 위한 대규모 보편적 기본 소득처럼 간단할 수도 있지만, 저는 그것이 해결책의 작은 부분일 뿐이라고 생각합니다. AI 시스템의 자본주의 경제가 될 수 있으며, 그런 다음 AI 시스템이 인간에게 보상하는 것이 합리적이라고 생각하는 것(궁극적으로 인간의 가치에서 파생된 어떤 판단에 기반)에 따라 인간에게 자원(전체 경제적 파이가 거대해질 것이므로 엄청난 양)을 제공할 수 있습니다. 아마도 경제는 Whuffie 포인트 로 운영될 것 입니다. 아니면 인간은 결국 일반적인 경제 모델에서 예상하지 못한 방식으로 경제적으로 계속 가치가 있을 것입니다. 이러한 모든 해결책에는 수많은 문제가 있을 수 있으며, 많은 반복과 실험 없이는 의미가 있는지 알 수 없습니다. 그리고 다른 도전과 마찬가지로, 우리는 여기서 좋은 결과를 얻기 위해 싸워야 할 것입니다. 착취적이거나 디스토피아적인 방향도 분명히 가능하며 예방해야 합니다. 이러한 질문에 대해 훨씬 더 많은 글을 쓸 수 있으며, 나중에 그렇게 하기를 바랍니다.

재고 조사

위의 다양한 주제를 통해 저는 AI가 모든 것을 잘 한다면 실현 가능하고 오늘날의 세상보다 훨씬 나은 세상에 대한 비전을 제시하려고 노력했습니다 . 이 세상이 현실적인지 저는 모릅니다. 그리고 설령 현실적이라 하더라도 많은 용감하고 헌신적인 사람들의 엄청난 노력과 투쟁 없이는 이룰 수 없을 것입니다. 모든 사람(AI 회사 포함!)이 위험을 예방하고 혜택을 충분히 실현하기 위해 각자의 역할을 다해야 할 것입니다.

하지만 싸울 만한 세상입니다. 이 모든 것이 실제로 5~10년 안에 일어난다면 – 대부분 질병의 패배, 생물학적 및 인지적 자유의 성장, 수십억 명의 사람들을 빈곤에서 벗어나 새로운 기술을 공유하고, 자유 민주주의와 인권의 르네상스 – 저는 그것을 보는 모든 사람이 그것이 자신에게 미치는 영향에

놀랄 것이라고 생각합니다. 저는 모든 새로운 기술의 혜택을 개인적으로 얻는 경험을 말하는 것이 아니지만, 그것은 확실히 놀랍습니다. 저는 오랫동안 품어온 이상들이 한꺼번에 우리 앞에 실현되는 것을 보는 경험을 말합니다. 저는 많은 사람이 말 그대로 감동하여 눈물을 흘릴 것이라고 생각합니다.

이 글을 쓰는 내내 저는 흥미로운 긴장감을 느꼈습니다. 어떤 면에서 여기에 제시된 비전은 극도로 급진적입니다. 거의 모든 사람이 다음 10년 안에 일어날 것이라고 기대하는 것이 아니며, 많은 사람에게 터무니없는 환상으로 여겨질 것입니다. 어떤 사람들은 그것이 바람직하다고 생각하지 않을 수도 있습니다. 그것은 모든 사람이 동의하지 않을 가치와 정치적 선택을 구체화합니다. 하지만 동시에 그것에 대해 눈부시게 명백한 무언가, 지나치게 결정된 무언가가 있습니다. 마치 좋은 세상을 상상하려는 많은 다른 시도가 필연적으로 대략 여기로 이어지는 것처럼 말입니다.

Iain M. Banks의 The Player of Games 29 에서 주인공은 문화라는 사회의 구성원으로, 여기서 제가 설명한 원칙과 크게 다르지 않은 원칙에 기반을 두고 있습니다. 복잡한 전투 게임에서 경쟁을 통해 리더십이 결정되는 억압적이고 군국주의적인 제국으로 여행을 떠납니다. 그러나 게임은 너무 복잡해서 플레이어의 전략이 자신의 정치적, 철학적 관점을 반영하는 경향이 있습니다. 주인공은 게임에서 황제를 물리치는 데 성공하여 그의 가치(문화의 가치)가 무자비한 경쟁과 적자생존에 기반한 사회가 설계한 게임에서도 승리 전략을 나타낸다는 것을 보여줍니다. Scott Alexander의 잘 알려진 게시물에는 경쟁이 스스로를 패배시키고 연민과 협동에 기반한 사회로 이어지는 경향이 있다는 동일한 주장이 있습니다. " 도덕적 우주의 호 "는 또 다른 유사한 개념입니다.

저는 문화의 가치가 승리 전략이라고 생각합니다. 왜냐하면 그것은 명확한 도덕적 힘을 지닌 수백만 개의 작은 결정의 합계이며 모든 사람을 같은 편으로 끌어들이는 경향이 있기 때문입니다. 공정성, 협동, 호기심, 자율성에 대한 기본적인 인간적 직관은 반박하기 어렵고, 우리의 더 파괴적인 충동이 종종 그렇지 않은 방식으로 누적됩니다. 우리가 예방할 수 있다면 아이들이 질병으로 죽지 않아야 한다고 주장하는 것은 쉽고, 거기에서 모든 사람의 아이들이 동등하게 그 권리를 누릴 자격이 있다고 주장하는 것은 쉽습니다. 거기에서 우리 모두가 뭉쳐서 이 결과를 달성하기 위해 지성을 적용해야 한다고 주장하는 것은 어렵지 않습니다. 사람들이 불필요하게 다른 사람을 공격하거나 해친 것에 대해 처벌을 받아야 한다는 데 동의하지 않는 사람은 거의 없으며, 거기에서 처벌은 사람들 사이에서 일관되고 체계적이어야 한다는 생각으로의 도약은 그리 어렵지 않습니다. 사람들이 자신의 삶과 선택에 대해 자율성과 책임을 가져야 한다는 것도 마찬가지로 직관적입니다. 이러한 간단한 직관은 논리적 결론으로 이어지면 결국 법치주의, 민주주의, 계몽주의 가치로 이어집니다. 불가피하지는 않더라도 적어도 통계적 경향으로 볼 때, 이는 인류가 이미 향하고 있던 방향입니다. AI는 단순히 우리를 더 빨리 목적지에 도달하게 할 수 있는 기회를 제공합니다. 논리를 더 뚜렷하게 하고 목적지를 더 명확하게 하기 위해서입니다.

그럼에도 불구하고, 그것은 초월적인 아름다움의 대상입니다. 우리는 그것을 실현하는 데 작은 역할을 할 기회가 있습니다.

이 글의 초안을 검토해 주신 Kevin Esvelt, Parag Mallick, Stuart Ritchie, Matt Yglesias, Erik Brynjolfsson, Jim McClave, Allan Dafoe, 그리고 Anthropic의 많은 분들께 감사드립니다.

2024년 노벨 화학상 수상자에게 , 우리에게 모든 것을 보여주셔서 감사합니다.