**To:**

**From:**

AGI(인공 일반 지능) 개발을 늦추기 위한 즉각적인 규제 조치와 조약의 필요성에 대해

친애하는,

최근 저는 인공지능(AI)의 위험성, 인공지능이 사회에 미칠 심대한 영향, 그리고 이 문제에 대한 공론화(정보에 입각한 공론화는 말할 것도 없고)가 부족하다는 사실에 경각심을 갖게 되었습니다.

Chat-GPT나 Midjourney와 같이 이미 존재하는 AI 시스템으로 인해 다양한 비즈니스에서 잘못된 정보나 부적절한 구현과 같은 많은 즉각적인 위험이 존재합니다. 멀지 않은 미래에 로봇과 AI 시스템으로 일자리가 대체되어 대량 실업이 발생할 경우 보편적 기본소득 정책과 같은 특정 안전망이 활성화되어야 할 것입니다.

그러나 가장 우려되는 것은 인간 수준 이상의 인지 능력을 갖춘 시스템인 AGI(인공 일반 지능)**를 향해 나아가고 있는 속도입니다**. 인공지능과 인간의 차이점은 인공지능이 스스로를 복제하고 산불처럼 퍼져나가는 데 몇 분밖에 걸리지 않는다는 점입니다. AGI는 컴퓨터 속도로 사고하며 인간의 법의 적용을 받지 않을 것입니다. 그것은 우리의 환경과 종으로서 우리를 지배할 수 있을 것입니다.

지난 몇 달 동안 AGI로 가는 길에 많은 진전이 있었습니다. 구글이나 마이크로소프트와 같은 기업들은 지금까지 진정한 스마트 시스템을 만드는 데 큰 성과를 거둔 거대한 실험을 진행하고 있습니다. 또한, 많은 진전을 이끈 OpenAI와 기타 연구소는 AGI 개발이라는 분명한 목표를 가지고 있습니다.

최근의 성공은 특히 미국을 중심으로 AI 역량에 대한 투자 물결을 일으켰습니다. 기업들은 안전성, '정렬' 등에 대해 주장하지만, 이러한 주장 중 상당수는 AI 정렬에 대한 연구자들의 견해와 일치하지 않습니다. 기업은 일반적으로 민주적 수단을 통해 대중의 동의 없이 일방적으로 결정을 내립니다.

지난 몇 달 동안 인공지능 업계에 쏟아진 막대한 자금은 로비스트와 대중의 대변자를 매수할 것입니다. **따라서 이제 우리가 선출한 대표자들이 공공의 이익을 위해 직접 나서서 문제를 해결하는 것이 중요합니다**.

트위터에서 이 주제에 대해 많은 논쟁이 벌어지고 있습니다. Eliezer Yudkowsky의 **"Pausing AI Developments Isn't Enough. We Need to Shut it All Down**". 유드코프스키의 의견에 동의하지 않더라도, 새로운 AI 기능의 개발과 확장을 중단하고 공개 토론과 위험에 대한 일반적인 이해를 위한 시간을 갖는 것이 현명한 접근 방식일 것입니다.

우리나라 대부분의 사람들은 AGI가 개인적으로나 전 세계에 어떤 의미가 있는지 이해하지 못합니다. 그들은 그것에 대해 생각하지 않았고, 뉴스에서 그것에 대해 읽지 않았으며,이 방향의 진행 상황이나 가능한 시나리오를 평가할 수있는 기술적 지식이 없습니다.

또한 이러한 기술을 개발한 사람들의 입장도 걱정스러운 부분입니다. OpenAI는 우리보다 훨씬 더 똑똑한 모델을 조정할 수 있는 완벽한 방법이 없다는 것을 인정합니다. **Microsoft는 GPT-4 내부에서 무슨 일이 일어나고 있는지 이해하지 못한다고 인정합니다**.

인공지능 연구소의 잘못된 결정, 자금의 홍수, 규제의 부재 등 모든 것이 인공지능을 향한 기술 발전 속도를 가속화하는 데 끊임없이 작용하고 있습니다. 우리 정부는 거대한 AI 실험에 제동을 걸 수 있는 조약을 체결하기 위해 매우 빠르게 행동해야 합니다. **유드코프스키가 제안하는 조약 중 하나는 AI의 목표와 가치가 우리의 목표와 일치한다는 확신을 가질 수 있을 때까지 전 세계적으로 대규모 모델 훈련을 중단하는 것입니다.**

현재 전 세계적으로 이를 위해 합리적인 정책을 장려하는 많은 단체가 생겨나고 있습니다. 예를 들어, AI 안전을 위한 캠페인은 권고안을 작성했습니다. 생명의 미래 연구소는 자체 문서를 가지고 있습니다.

인간은 우리보다 수천 배 더 빠르고 예리한 '디지털 마인드'와 함께 살아온 적이 없습니다. 소수의 사람들이 인류의 운명을 결정하도록 내버려둘 수는 없습니다.

안부 전해주세요,