**To:**

**From:**

AGI（人工知能）の発展を遅らせるための早急な規制措置と条約の必要性について

最近、私は人工知能（AI）の危険性、AIが社会に及ぼすであろう深刻な影響、そしてこの問題に対する一般市民の議論の欠如（ましてや情報に基づいた一般市民の議論）に警鐘を鳴らすようになりました。

Chat-GPTやMidjourneyのようなすでに存在するAIシステムからは、偽情報や様々なビジネスへの不適切な導入など、多くの直接的なリスクがある。そう遠くない将来、ロボットやAIシステムによる雇用の置き換えによって大量の失業者が発生した場合、一定のセーフティネット（ユニバーサル・ベーシック・インカム政策など）を作動させる必要があるだろう。

**しかし、最も懸念されるのは、AGI（人工知能）、つまり人間並みかそれ以上の認知能力を持つシステムに向かうスピードである**。AGIと人間の違いは、AGIが自分自身をコピーして山火事のように広がるのに数分しかかからないことだ。AGIはコンピューターのスピードで思考し、人間の法律の適用を受けることはないだろう。AGIは我々の環境と我々という種を支配することができるだろう。

AGIへの道では、ここ数カ月で多くの進展があった。グーグルやマイクロソフトのような巨大な実験が行われ、純粋にスマートなシステムを作ることに成功している。さらに、OpenAIやその他の研究所は、AGIの開発を明確な目標として掲げている。

彼らの最近の成功は、特にアメリカにおいてAI能力への投資の波を扇動した。企業は安全性や「整合性」などについて主張するが、それらの主張の多くは、AIの整合性に関する研究者の見解と一致していない。企業は一般的に、民主的な手段を通じて国民の同意を得ることなく、一方的に決定を下す。

ここ数カ月でAI業界に放たれた資金の洪水は、彼らにロビイストや公的擁護者を買わせるだろう。**したがって、選挙で選ばれた議員たちが、国民の利益のために自らの手で問題を解決することが極めて重要なのだ。**

ツイッター上では、この話題について多くの議論が交わされている。エリエゼル・ユドコフスキーの「**Pausing AI Developments Isn't Enough. We Need to Shut it All Down**」と題されたエリエゼル・ユドコフスキーの記事を読んでほしい。たとえユドコフスキー氏の意見に反対であったとしても、賢明なアプローチは、新しいAI能力の開発とスケーリングを停止し、一般市民の議論とリスクに対する一般的な理解の時間を確保することだろう。

私たちの国のほとんどの人々は、AGIが個人的に、そして世界全体にとって何を意味するのか理解していない。AGIについて考えたこともなく、ニュースで読んだこともなく、この方向への進展や起こりうるシナリオを評価する技術的知識もない。

さらに心配なのは、このような技術を開発したクリエイターたちが認めていることだ。OpenAIは、我々よりはるかに賢いモデルを調整する確実な方法がないことを認めている。**マイクロソフトは、GPT-4の内部で何が起きているのか理解していないことを認めている**。

AI研究所の誤った決意、氾濫する資金、規制の欠如はすべて、AGIに向けた技術進歩の速度を加速させるために容赦なく働く。私たちの政府は、巨大なAI実験に歯止めをかける条約に取り組むため、極めて迅速に行動する必要がある。ユドコフスキー氏は、そのような条約のひとつを提案している。**それは、AIの目標と価値観が我々と確実に一致するまで、世界中で大規模なモデルの訓練を停止するというものだ**。

現在、この目的のために賢明な政策を推進するグループが世界中に数多く誕生している。例えば、「AI安全キャンペーン」はその提言を記している。The Future of Life Instituteには独自の文書がある。

人類は、私たちの何千倍も速く鋭い「デジタル・マインド」とともに生きてきたわけではない。小さな集団に人類の運命を決めさせるわけにはいかないのだ。