**From:**

**To:**

Over de noodzaak van onmiddellijke regelgeving en verdragen om de ontwikkeling van kunstmatige algemene intelligentie (AGI) te beteugelen

Geachte ,

Onlangs werd ik gealarmeerd door de risico's van kunstmatige intelligentie (AI), de ingrijpende effecten die het zal hebben op de samenleving en het gebrek aan publiek debat (laat staan geïnformeerd publiek debat) over deze kwestie.

Er zijn veel directe risico's van reeds bestaande AI-systemen zoals Chat-GPT of Midjourney, zoals desinformatie of onjuiste implementatie in verschillende bedrijven. In de niet zo verre toekomst zullen bepaalde vangnetten moeten worden geactiveerd (zoals een universeel basisinkomen) in het geval van massale werkloosheid door de verdringing van banen door robots en AI-systemen.

Maar het meest zorgwekkend is de snelheid waarmee we oprukken in de richting van AGI (kunstmatige algemene intelligentie) - **systemen die cognitieve vermogens zullen hebben op of boven het menselijke niveau.** Het verschil tussen een AGI en een mens is dat een AGI maar een paar minuten nodig heeft om zichzelf te kopiëren en zich als een lopend vuurtje te verspreiden. AGI zal op computersnelheid denken en zal niet onderworpen zijn aan de handhaving van menselijke wetten. Het zal in staat zijn om onze omgeving en ons als soort te domineren.

De afgelopen maanden is er veel vooruitgang geboekt op het pad naar AGI. Er zijn gigantische experimenten ondernomen door bedrijven als Google en Microsoft die tot nu toe vruchtbaar zijn geweest in het creëren van echt slimme systemen. Bovendien hebben OpenAI en andere onderzoekslaboratoria achter veel van de vooruitgang een expliciet doel om AGI te ontwikkelen.

Hun recente successen hebben een golf van investeringen in AI-capaciteiten teweeggebracht, vooral in de VS. Hoewel bedrijven beweringen doen over veiligheid, "afstemming", enzovoort, komen veel van die beweringen niet overeen met de standpunten van onderzoekers over het onderwerp van AI-afstemming. Bedrijven nemen over het algemeen eenzijdig beslissingen, zonder toestemming van het publiek via democratische middelen.

De stortvloed aan geld die de AI-industrie de afgelopen maanden heeft ontvangen, zal lobbyisten en openbare pleitbezorgers opleveren. **Het is daarom cruciaal dat onze gekozen vertegenwoordigers nu het heft in eigen handen nemen in het belang van het publiek**.

Op Twitter wordt veel gediscussieerd over dit onderwerp. Ik hoop dat je het artikel van Eliezer Yudkowsky met de titel **"Pausing AI Developments Isn't Enough. We Need to Shut it All Down".** Zelfs als je het niet eens bent met de heer Yudkowsky, zou de verstandige aanpak zijn om te stoppen met het ontwikkelen en opschalen van nieuwe AI-capaciteiten, zodat er tijd is voor een publiek debat en algemeen begrip van de risico's.

De meeste mensen in ons land begrijpen niet wat AGI voor hen persoonlijk en voor de wereld in het algemeen zou betekenen. Ze hebben er niet over nagedacht, ze hebben er niet over gelezen op het nieuws, ze hebben geen technische kennis om de vooruitgang in deze richting of de mogelijke scenario's te beoordelen.

Wat ook zorgwekkend is, zijn de bekentenissen van de makers van dergelijke technologie. OpenAI geeft toe dat het geen waterdichte manier heeft om een model af te stemmen dat significant slimmer is dan wij. **Microsoft geeft toe niet te begrijpen wat er gebeurt in GPT-4**.

De misplaatste vastberadenheid van AI-laboratoria, de stortvloed aan geld en het gebrek aan regelgeving versnellen de technische vooruitgang in de richting van AGI. Onze regering moet extreem snel handelen om aan verdragen te werken die reusachtige AI-experimenten een halt toeroepen. De heer Yudkowsky stelt zo'n verdrag voor: **wereldwijd stoppen met het trainen van grote modellen totdat we met vertrouwen de doelen en waarden van AI kunnen afstemmen op de onze**.

Op dit moment duiken er wereldwijd veel groepen op die hiervoor verstandig beleid promoten. De Campaign for AI Safety heeft bijvoorbeeld zijn aanbevelingen opgeschreven. Het Future of Life Institute heeft zijn eigen document.

Mensen hebben nooit geleefd met "digitale geesten" die duizend keer sneller en scherper waren dan wij. We kunnen kleine groepen mensen niet over het lot van de mensheid laten beslissen.

Hoogachtend,