THE PEOPLE'S AI SUMMIT

CRONYISM: AI SUMMIT EDITION

How the UK's newly formed Frontier AI Taskforce, driving the AI Summit, is entrenched in Big Tech money and connections

Lead researcher and author:

Florence Wildblood
Journalist, Byline Investigates

Co-author:

Manasa Narayanan
Journalist, The Citizens

INTRODUCTION

Investigation by the Citizens of the members making up the Frontier AI Taskforce — the group driving the summit agenda as well as responsible for the larger AI framework in the UK — reveals a number of troubling links. Our analysis shows the group is riddled with conflicts of interest to Big Tech and government, and includes several adherents to the controversial 'effective altruism' philosophy. Further questions must be asked about why and how public money is being funnelled into private enterprise.

The new Frontier AI Taskforce, set up by PM Rishi Sunak, a body that effectively sits within the Department of Science, Innovation and Technology (DSIT), is at the core of the upcoming AI Safety Summit; which is to see global leaders join mostly tech industry representatives in a closed doors affair to discuss existential risks posed by Frontier AI.

In choosing to majorly focus efforts on a far-off apocalypse rather than the real and present risks posed by AI's current capabilities, the summit has attracted criticism. But an analysis of those driving the summit makes clear why far-off existential risks from Frontier AI are on the agenda, instead of grave immediate harms: a combined influence of Big Tech and Effective Altruism philosophy.

Taking a 'doomist' outlook has become en vogue amongst big tech heavyweights, who stand to gain by shifting attention away from their present transgressions and failings. They are pulling up the drawbridge to other innovations now that their monopolies have been established. [It is the few Big Tech companies that are already dominating the tech space that are at the helm of AI innovations too.]

Added to this mix is the Effective Altruism (EA) movement, which is supposedly a movement about doing good in the most logical manner possible, in a bid to maximise benefit to all. Or so its proponents claim. But driving this rather vague movement are Big Tech bigwigs who quite ironically are the ones responsible for the immediate and urgent problems we are facing today - from encoded algorithmic biases reinforcing oppressive systems, massive monopolies that are even strongly consolidating around the AI market, to dangerous election (or otherwise) disinformation that is compromising democratic systems.

These current risks from social media and AI systems need urgent addressing, But these do not really figure in the summit or in the minds of those driving UK AI policy. How will they, when those in charge are so cosy with Big Tech?

An 18 October government press release states that 'Ian Hogarth [the Chair of the taskforce] and DSIT have responsibility to identify and address any actual, potential or perceived personal or business interests which may conflict, or may be perceived to conflict, with the Chair's public duties.'

But that must also apply to the rest of the members of the taskforce? Given this is a government taskforce, there should be more transparency around how these members were chosen? If we were to treat AI as a public good, something that everyone should benefit from, should it be private interests that drive the policies?

FINDINGS ON THE TASKFORCE

1/ Paul Christiano and Taskforce cosy up to OpenAI

On the taskforce is Paul Christiano, a leading researcher in the field of AI Alignment, co-founder of the Alignment Research Centre (ARC), who previously ran the language model alignment team at OpenAI.

Christiano worked at OpenAI until 2021. Recently, the government announced that the Frontier AI Taskforce (of which Christiano is an integral part of) would now be working closely and directly with companies like OpenAI, Anthropic and Deepmind, who will get "early or priority access to models for research and safety purposes."

OpenAI, which created ChatGPT, is a black box when it comes to its operations. While initially created as open-source (meaning that anyone could access their code), in 2019 OpenAI shifted to a 'capped-profit' model, closing access and partnering with Microsoft. It is now receiving billions in investment from Microsoft. Whether Christiano retains any financial interest in OpenAI is unclear.

Christiano is also deeply embedded in the effective altruism movement. He has been vocal in his support for the philosophy across its content platforms and his own blogs. He was a technical advisor for Open Philanthropy, the movement's grantmaker, at the time it gave $30 million to OpenAI – his then employer – in 2017. He was also at that point living with its executive director.

This same executive director, Holden Karnofsky, now advises his Alignment Research Centre (ARC). ARC has received extensive funding from EA-affiliated bodies and is one of the 'leading technical organisations' that the UK government is partnering with on the AI summit.

> Why was a researcher who is a previous employee of a major private AI lab brought into a government taskforce that is responsible for AI oversight? How did his links to OpenAI and other Big Tech money and links via the EA movement help him?

2/ Investor on sabbatical, now PM representative for AI Summit, has deep links to DeepMind

Matt Clifford is the Prime Minister's joint Representative for the AI Safety Summit. Before taking on this role, Clifford was working with Entrepreneur First, an investment company that he co-founded that puts funding behind tech startups. He has taken a sabbatical from Entrepreneur First to be in this role under Rishi Sunak.

His company, Entrepreneur First, raised £130 million in a funding round last year. Investors including Demis Hassabis and Mustafa Suleyman, co-founders of DeepMind, one of the world's foremost AI research organisations, now owned by Google. Hassabis and Suleyman also invested in Entrepreneur First in 2017.

As well as the questions that are raised by the financial relationship between Hassabis and Clifford's company, Hassabis's prior work with the UK government on AI and healthcare has been controversial. A [misuse of private information claim](#) was brought on behalf of 1.6 million people whose data was used to help DeepMind train an app, after the ICO ruled that the data protection act had been breached. It has since been struck out. There was also pushback when it emerged Hassabis [had attended](#) a SAGE meeting during the COVID-19 pandemic, having been [brought into the fold](#) by Dominic Cummings.

Also to note, Matt Clifford is [linked to Dominic Cummings](#) as the chair of ARIA. Based on the US's Defense Advanced Research Projects Agency (DARPA), it was the brainchild of Boris Johnson's former right-hand-man, which they claim will fund 'high risk, high reward' research and development.

> Why was a private investor, deep in Google money and links with the co-founders of one of the foremost AI labs (Deepmind), brought into a taskforce that is responsible for AI oversight? What role did his existent connections with Big Tech and the government — through ARIA and Dominic Cummings — play in this selection?

· ·

3/ Bengio, yet more Google money and Element AI

Yoshua Bengio, Turing Award winner, is a Professor at Université de Montréal, and the Founder and Scientific Director of Mila – Quebec AI Institute.

He's known as one of the three 'godfathers' of AI, and founder of Mila, the world's largest academic research centre for deep learning. Mila [received](#) $4.5 million from Google over a three-year period starting in 2016; $4 million (also over a three-year period) in 2020; and an additional [$1.5 million](#) in 2022.

'Google Canada's generous support, and our pledge toward Mila's mission, further solidifies Mila and Google's longstanding mutual commitment to continue to develop AI for the benefit of everyone,' Bengio [wrote on the corporation's blog](#) in 2020.

In 2016, Bengio co-founded Element AI, an AI incubator designed around business applications of AI research. In 2020, it was sold to ServiceNow, with the *Globe and Mail* [revealing](#) that it was 'running out of money and options when it inked a deal ... to sell itself for $230 million to Silicon Valley software company ServiceNow Inc.' ServiceNow agreed to pay $10M to key employees including Bengio, who remains involved as a [research advisor](#).

ServiceNow has worked across the UK government, [including](#) with the NHS, DEFRA and DWP. It signed a Memorandum of Understanding with the Crown Commercial Service this year 'to support consistent innovation across government and the wider public sector'. In October, the Crown Prosecution Service invested £9M in DXC, ServiceNow's [leading ecosystem partner](#), to operate and transform its IT infrastructure (run by ServiceNow), making use of AI.

3/ Bengio continued

> While Bengio's credentials as an AI researcher are unparalleled, why was his lab's deep entrenchment in Google money not a factor of contention? Did his links to Service Now (via Element AI) and its work with the UK government over various projects have any role to play?

. . . .

4/ David Krueger and EA bankrolling

The other expert researcher on the taskforce is David Krueger, who, like Paul Christiano, is a very prominent figure in the effective altruism movement.

He studied under Yoshua Bengio at Mila and has received significant funding from EA bodies: $1 million from Open Philanthropy, $880,000 from the Survival and Flourishing Fund, another $250,000 from Open Philanthropy, $200,000 from the Long-Term Future Fund, and $10,000 from the EA Foundation Fund.

He also assisted in sourcing the Open Philanthropy project grant to Mila to the tune of $2.4 million, and worked as a freelance career mentor for EA-affiliated 80,000 hours, providing career advice for those interested in AI alignment.

> Why was Krueger and others, who are so deep in Effective Altruism money, made core to the taskforce and UK policymaking? Why not other researchers and experts with a different outlook towards Artificial Intelligence?