Department for
Science, Innovation
& Technology

Viscount Camrose
Parliamentary Under
Secretary of State
Department for Science,
Innovation & Technology
100 Parliament Street
London SW1A 2BQ

www.gov.uk/dsit

Baroness Stowell
House of Lords
London
SW1A 0PW

Tuesday 16th January 2024

Dear Baroness Stowell,

Following on from providing evidence to the Committee on Communications and Digital on Large Language Models, I shared further detail on the Frontier AI Task Force's spend to date; the AI Safety Institute's budget, and existing and planned spend. We also previously provided information on headcount for specific roles within the Department for Science, Innovation and Technology (DSIT), including the number of full-time equivalent (FTE) staff employed directly within the central AI risk function.

In response to your inquiries, I'm pleased to share an outline of cross-government efforts to mitigate emerging AI security risks.

Cyber Security

Regarding the Committee's inquiry on cyber threats and generative AI, the Government maintains a proactive approach to monitoring potential risks associated with advancements in generative AI. To counter these challenges, we established the AI Safety Institute (AISI) and the central AI risk function (CAIRF) within DSIT. These entities collaborate closely to identify, track, and mitigate emerging threats.

The AI Safety Institute (AISI) offers insights into the advanced capabilities of frontier AI and foundation models by evaluating model capabilities, assessing their impacts, and developing the sociotechnical infrastructure needed to understand the risks.

CAIRF subsequently helps translate and explain what AISI's evaluations of model capabilities mean for policymakers across government, helping them respond to risks posed by developing specific solutions. CAIRF also evaluates the effectiveness of risk reduction strategies aimed at addressing AI associated risks and identifies gaps in government's risk mitigation efforts. Furthermore, CAIRF will work collaboratively with a partnership of AI experts on the risk domain and risk assessment to ensure a consistent approach to understanding impacts and likelihoods, including risks from generative AI.

The cyber security team within DSIT is undertaking work as part of the National Cyber Strategy to help address a range of threats to AI models and systems, such as data poisoning and model manipulation. DSIT are working closely with the National Cyber Security Centre to create a robust evidence base and develop interventions that will ensure users can securely benefit from AI. Cabinet Office leads a program of work focused on addressing cyber security risks from AI models.

While details on resource uplifts for specific AI initiatives reside with individual government departments and intelligence agencies, DSIT maintains close collaboration to ensure effective utilisation of resources in addressing evolving AI security threats.

Child Sexual Abuse Material

The Home Office is identifying and tracking emerging technologies, and developing a robust understanding of the risks and opportunities these technologies pose with regards to tackling child sexual abuse. The Government recognises the many benefits AI can provide towards efforts to ensure public safety. However, capabilities such as generative AI, also pose significant risks to our efforts to tackle the prevalence and proliferation of child sexual abuse. A key area of concern lies in the growing exploitation of generative AI for the creation of hyper-realistic child sexual abuse materials (CSAM).

Possessing, sharing, or searching for any CSAM, regardless of its origin, is already a criminal offense. The Online Safety Act plays a crucial role in curbing CSAM and protecting children by mandating stringent safeguards from all stakeholders. AI applications must not be exempt from this responsibility, and robust measures to prevent their misuse for child sexual abuse are paramount. Child safety cannot be an afterthought in the face of technological advancements, and AI companies must be held accountable for the risks posed by their services. The Government's 2021 Child Sexual Abuse Strategy remains a strong foundation for tackling all forms of abuse, and continued investment in its implementation is essential. Only through robust legislation, industry accountability, and ongoing commitment can we truly protect children from online and offline exploitation.

The United Kingdom plays a leading role on the international stage in combating child sexual abuse (CSAM). Leveraging its domestic expertise, the UK fosters collaborative action with like-minded governments, technology companies, and civil society partners. In September 2023, the UK Home Secretary and US Homeland Security Secretary pledged to explore further efforts against the surge in AI-generated CSAM, building on the Government's November

2023 AI Safety Summit. The Summit, which convened charities, tech firms, academics, and international representatives, culminated in the release of a Joint Statement: Tackling Child Sexual Abuse in the Age of Artificial Intelligence. This statement, co-signed by over 33 entities including TikTok, Snap, Stability AI, the US Department of Justice, and the Australian Government, established key principles for collectively addressing the risks posed to child safety by Generative AI. These principles serve as a foundation for continued discussions with the global community to identify and explore new mitigation measures.

The Home Office has strong relationships with external stakeholders across civil society, industry, and law enforcement. The department has dedicated resources focussed on engaging with the technology sector, identifying emerging technologies, risks posed to children, and developing potential mitigation measures to these.

The Home Office has strengthened the National Crime Agency by increasing its budget by 21% in the past two years from £711m in 2021-22 to £860m in 2023-24. This reflects the importance the Government attaches to the Agency's role in tackling serious and organised crime, and will help it continue to develop critical data, intelligence, and investigative capabilities. With regards to tackling child sexual abuse specifically, the Home Office has invested in high harm disruption across the NCA, GCHQ and policing which is aimed at bringing the highest harm, most technologically sophisticated CSA offenders to justice.

Additionally, the Home Office has supported the development of the Online Safety Act, which places new legal duties on companies to tackle child sexual abuse as a top priority. The Home Office are also engaging with Ofcom to focus on future threats, with international partners such as the G7, EU and Five Countries, as well as direct engagements with AI companies. The Home Office is in the process of building a new central strategic function, to drive a holistic departmental approach for emerging technology, including Generative AI, across several public safety and national security threats. Child sexual abuse remains a priority threat area for the department.

The Home Office will continue to work across government to ensure additional opportunities to increase investment in innovation for new mitigation measures in partnership with industry, and new AI capabilities to enhance the policing response.

Misinformation and Disinformation

Through the central AI risk function, government is identifying, tracking, and responding to AI-related risks including misinformation, disinformation and deepfakes. DSIT actively works across departments, social media platforms, civil society groups, academia, and international partners to tackle emerging threats in this area. In 2024, over 50 countries worldwide will hold elections that will cover 4 billion people. As highlighted at the November 2023 AI Safety Summit active international collaboration and industry partnership is key to tackling this shared challenge. DSIT is actively supporting wider cross-government efforts to protect UK democratic processes, including through the Defending Democracy Taskforce and Election Cell.

To mitigate AI-driven misinformation and disinformation, government are encouraging social media platforms to implement effective policies specifically addressing the threat posed by generative AI, ensure consistent enforcement of these policies while upholding the principle of freedom of expression. Implementation of the Online Safety Act and the Foreign Interference Offence will make a significant difference here. As well as ensuring companies take action to keep users safe, government are taking steps to educate and empower users with the skills and knowledge they need to keep themselves safe online through work on media literacy.

Government takes the issue of information threats to national security very seriously, DSIT continues to work closely with a range of cross-government teams bringing different types of expertise to bear. DSIT's National Security Online Information Team (NSOIT, formerly known as the Counter Disinformation Unit) is focused on threats from foreign states, as well as risks emerging from the use of AI and deepfakes. This includes the imitation of elected leaders, election interference, and the use of bots to drive disinformation surrounding conflict in Israel-OPT.

Addressing the challenges of disinformation and wider AI-enabled threats is a whole government effort, led by DSIT. DSIT will continually review the requirement and work flexibly across government to ensure this work is sufficiently resourced, allowing us to surge capacity where it's needed.

Terrorism

Government acknowledges the need to stay ahead of the curve when it comes to generative AI's potential for misuse in the context of terrorism, as outlined in the counterterrorism strategy CONTEST 2023. HMG will continue to invest in the identification of future threats and opportunities that derive from technology. We will adapt policies to mitigate emerging technology enabled threats and seize new opportunities to ensure that counter-terrorism responses are enabled by technology at every stage. This is reflective of governments approach towards understanding the threats and opportunities arising from Artificial Intelligence.

Government is also committed to understanding how emerging technologies may be exploited for terrorist purposes and are ensuring that effective mitigations are in place. The impact of Generative AI on terrorist activity online is yet to be fully established. While there is evidence of early experimentation with generative AI for terrorist purposes, including to facilitate the production of synthetic terrorist propaganda, government have not seen evidence of widespread adoption of generative AI. Steps are being taken to build knowledge of the risks relating to generative AI and to consider appropriate mitigations, including leveraging potential opportunities associated with the technology.

The Counter-terrorism strategy CONTEST 2023 implements a robust cross-government framework, ensuring accountability and facilitating effective interdepartmental coordination on all matters related to terrorism.

On behalf of the cross-government teams involved in responding to the questions I would like to thank the Committee for the opportunity to address their questions. We will be happy to provide further detail, should the Committee have any further queries.

Yours sincerely,

Viscount Camrose
Parliamentary Under Secretary of State at the
Department for Science, Innovation & Technology