Luminate

MEDIA BRIEFING

Addressing AI Safety immediately
Democracy, human rights and the UK's AI Safety Summit

November 1, 2023

The United Kingdom hosts the AI Safety Summit on 1 and 2 November at Bletchley Park.

This first-of-its-kind global AI gathering will focus on the existential threats of "frontier AI" — defined as models more powerful than those available today.

The summit will steer the global AI conversation, but its "frontier" focus and limited attendance are flawed choices: they run a high risk of limiting the safety design conversation to Big Tech corporations and governments, side-stepping critical conversations about how to address the immediate harms AI creates around the world.

Luminate: working for safer AI

At Luminate, we're working to address global digital threats to democracy. We fund policy change, campaigns, partnerships, and litigation that seek to challenge the business model of the Big Tech companies and social media platforms, ensuring that we have oversight and regulations required to move us from today's polarising digital sphere – where hate speech and disinformation are rewarded and monetized – to a healthier digital environment that works for democracy and social cohesion.

As we approach the Summit, there are several critical issues we believe should be front of mind for those attending, or following:

1. Frontline, not frontier: reframing the AI safety conversation

What's needed is to put the full diversity of civil society and **citizens at the heart of the AI conversation** and agenda. New technology must be built and tested incorporating all our hard-won rights, not dismantling them.

This means securing **data rights**, enforcing **transparency around healthy algorithms**, and building **accountability and international human rights standards** into how AI systems are designed and deployed today. Democratic oversight of AI is essential to ensure technology serves humanity rather than concentrate power in the hands of the few.

Evidence further shows how narrow the AI safety discourse currently is. New research by Rootcause looked at 30,000 YouTube videos and 1000 UK and EU traditional media articles published in May 2023, finding that:

• Corporate actors are by far the most frequently quoted by journalists covering AI, with no civil society voices featuring amongst the top-20 most-quoted people or organisations;

• Only two of the top-20 most-quoted people are women.

2. Start with enforcing data protection

Tech corporations like Meta continue to champion the "Move fast and break things" philosophy.

The company's current AI trajectory doubles down on the two broken systems of this business model: first, the **surveillance model of data extraction from users**, and second, the building of **algorithms trained on this data to maximize user engagement**, at the expense of people's well-being and of societies' shared reality all over the world.

That's why any discussion on AI needs to start first and foremost with the first guardrail: data protection.

Our first line of defense against the unchecked rise of AI risk is **enforcing existing laws**, including the global standard set by the General Data Protection Regulation (GDPR) in the EU.

While staking a claim to leadership in AI safety at this summit, experts warn that the UK's new Data Protection and Digital Information Bill (DPDI) risks compromising international standards.

Calling for the DPDI Bill to be rethought or withdrawn, the Open Rights Group (ORG) have produced a comprehensive analysis of how it will impact citizens negatively. Watering down the UK's data protection standards – and risking the EU data-sharing agreement that relies on having similar rules – could cost British business more than £1 billion and make the UK a "leaky valve" of sensitive personal data worldwide.

It's bad for business — and citizens do not want weaker protection of data either: a recent Luminate/YouGov poll in the UK showed that 88% of people felt it was important for social media companies to let users exercise their right to object to data collection and processing to target them.

Meta's President of Global Affairs, Nick Clegg, has acknowledged that there are real concerns regarding the large-scale training of Meta's AI tools on users' Instagram and Facebook public posts, stating, "I strongly suspect that's going to play out in litigation." Luminate is currently funding litigation challenging Meta's data practices.

3. Detox addiction algorithms to protect children, social cohesion and democracy

Democracy is already in steep decline around the world, and Gen AI is a major new threat vector.

At least 16 countries have deployed this technology in the last year to "sow doubt, smear opponents, or influence public debate." The engagement-based addictive algorithms of social media are being further supercharged with the creation of AI tools such as using voice Gen AI mimicry to sow confusion.

Despite the risks, this technology is being integrated into search and social media. Following the introduction of generative AI to its recommender systems, the company announced that time spent on Instagram's reels increased by 24% with a corresponding jump in revenue.

Gen AI opens up a whole new frontier to increase engagement in the attention economy – with serious consequences for democracy. In a recent experiment, Luminate's partner organisation Eko created advertising using Gen AI tools with extreme hate speech, which Meta accepted for publication. Eko withdrew the ads before they could go live, but the experiment showed the ease with which democratic elections could be derailed on an industrial scale, in breach of social media platform's obligations under the EU's new Digital Services Act.

Protecting democracy, in fact, is the first test case for AI safety. Next year, nearly 65 major elections are scheduled, with more than 1 in 4 people globally living in a country going to the polls, including the European Parliament, India, Indonesia, Mexico, South Africa, South Korea, Taiwan, the UK and the US. In the category of immediate risk, the unregulated rollout of "generative AI" (Gen AI) raises urgent questions for the integrity of elections on hyper-personalised social media platforms that amplify bias and disinformation on unprecedented scales, at a fraction of the previous effort and cost.

Over 50 civil society groups under the banner of the People vs Big Tech network in Europe are demanding enforcement of these rules to tackle election-related threats and a global coalition is putting direct pressure on the companies to release their election Action Plans.

Quotes from Luminate and our partners

"The development of generative AI represents a significant new acceleration in the informational ecosystem. Generative AI enhances impersonation and refines the social engineering that serves as the hook for different kinds of attacks in the online space, all while reducing costs and automating production and distribution.

The result is a hyper-polarised environment where the audience is increasingly vulnerable to deceptions that make it harder to tell the truth from lies. This paradigm calls for the development of new responses, not only from platforms or governments but also carried by an independent civil society."

— **Alexandre Alaphilippe**, Executive Director, EU Disinfo Lab

"To confront the unfolding climate crisis we need strong democracies, bold political leadership, and a compassionate, truth-based public square — all things imperilled by AI-fuelled disinformation.

Attempts at AI governance must not become dazzled by potential 'helpful' applications of the technology, but instead recognise AI as an accelerant of the attention economy and the grave societal impacts it has created – both for climate and more widely."

— **Oliver Hayes**, Policy and Campaigns Lead, Global Action Plan

"Year after year, Big Tech corporations have failed to tackle hate speech, disinformation and other harms facilitated by their platforms with devastating consequences. We have no faith that these corporations will voluntarily adopt a responsible approach to AI that respects everyone's rights. The AI Safety Summit can only be taken seriously if it paves the way for AI regulation that addresses existing dangers and takes into account the views of all stakeholders, not just the billionaire tech CEOs who have an interest in using this technology to further enrich themselves."

— **Naomi Hirst**, Campaign Strategy Lead, Global Witness

"The AI Safety Summit looks years into the future, while ignoring the harms and damages of AI happening today. We need a real conversation about the harms of AI on democracy, privacy and the economy. And we reject the idea that Big Tech should be given a forum to report progress to world leaders against voluntary commitments. This is not how regulation works – and it's not how we'll achieve 'AI Safety.'"

— **Clara Maguire**, Executive Director, The Citizens

"Social media companies with a shameful reputation of disregarding human rights and allowing divisive rhetoric to fester across the world are now leading the rollout of generative AI and large language models (LLMs).

In the UK, we've witnessed how this divisive, algorithmically amplified rhetoric transcends borders and influences community cohesion in the diaspora: in Leicester, this was a clear factor in seeing Indian Hindus and Indian Muslims suddenly grapple with polarization across religious and political lines, after decades of peaceful co-existence.

This is why we need to be overtly critical and excessively careful of the big promises of generative AI. We can't allow Big Tech to further distort our information environments."

— **Dr. Ritumbra Manuvie**, Executive Director, The London Story

"Facing up to the dangers of AI is urgent. But by focusing on theoretical future risks at the 'frontier' of AI, the AI Safety Summit skirts the grave harms that are already here.

As at least 65 countries prepare to go to the polls next year, disinformation campaigns that are turbocharged by AI pose a serious threat to free and fair elections around the world. At this critical juncture for global democracy, it has never been more important to hear the voices of civil society and those most affected by AI harms, yet they have been excluded from this Summit.

Generative AI systems are trained on vast amounts of personal data, collected from consumers without their knowledge or consent. This can be used to manipulate, scam, or otherwise cause harms to people. It is essential that these concerns are addressed as the technology is becoming normalized and embedded into consumers' everyday lives."

— **Finn Myrstad**, Director of Digital Policy, Norwegian Consumer Council

"The public is being led to believe that AI is some kind of runaway train that is careering towards vague and faraway future risks. This is deeply disingenuous, given that the same handful of corporations that are stoking such fears are driving the train themselves.

The reality is that the harms are not vague and faraway, they are concrete and they are already happening. What we need is less fantasy and rhetoric from the UK government, and more regulation to bring these corporations under control. A good start would be to enforce the laws we already have rather than dismantling them as the government is doing with GDPR."

— **Tanya O'Carroll**, Senior Fellow, Foxglove

"In Ireland, platforms like X, Facebook and TikTok have become a dangerous vector for far-right hate, disinformation and manipulative content, a trend Gen AI looks set to worsen.

Many of these social media platforms profiting from growing these AI-driven disruptions are based in Dublin. This presents a unique opportunity for the Irish government and for the Data Protection Commission to demonstrate real leadership -- not to mention, do their job -- by properly enforcing GDPR and holding corporate tech leaders accountable. Without regulation, our democracies face escalating threats."

— **Siobhan O'Donoghu**e, Executive Director, Uplift

"Is AI safety possible if you give more power to giant profiteering tech corporations over citizens? You don't need to be an AI expert to know the right answer. Yet the British government's new data law (DPDI Bill) will steamroll our hard-won rights and that's why it must be withdrawn."

— **Alaphia Zoyab**, Director for Campaigns and Media, Luminate

"If governments meeting at the AI Safety Summit want these technologies to be safe, they can't overlook the real and present dangers they pose to us all.

We are heading into a uniquely significant year for democracy around the world. This is not a time to focus solely on the distant horizon. Unless we foster an open conversation about these tools, AI-fuelled disinformation could decide the future for us by dividing our societies like never before."

— **Stephen King**, CEO, Luminate

Questions for reporters to consider around the Summit

- Why is the government watering down basic data protection rules by introducing a new Data Protection and Digital Information Bill?

- Why isn't the Summit focusing on safety rules applicable to both current and (hypothetical) frontier AI models?

- How can we ensure community involvement in shaping AI rules, preventing companies and governments – including authoritarian ones – from banding together to decide the future of AI?

- What measures are we considering to counteract AI outputs that are discriminatory or propagate disinformation? How can we use existing laws to mitigate AI harms?

- What measures did the Summit discuss to protect the safety of democratic elections?

- Is the government creating any guidance for platforms on the use of synthetic media/ generative AI in electioneering?

Further reading

SURVEY | **Luminate: Bots versus ballots: Europeans fear AI threat to elections and lack of control of personal data**

Survey results reveal a deep unease among European respondents regarding AI's threat to elections and the unchecked use of their personal data.

REPORT | **Norwegian Consumer Council: Ghost in the Machine: Addressing the consumer harms of generative AI**

Examines generative AI, summarizes various challenges, risks, and harms of this emerging technology, and presents human-centered principles that can help shape and use AI systems.

STUDY | **AlgorithmWatch and AI Forensics: ChatGPT and Co: Are AI-driven search engines a threat to democratic elections?**

Explores ways AI can be dangerous to the formation of public opinion in a democracy.

REPORT | **AI Now 2023 Landscape: Confronting Tech Power**

Investigates the concentration of power in the tech industry and identifies strategic priorities that are critical to this period of development of artificial intelligence.

OPINION | **Financial Times: When it comes to AI and democracy, we cannot be careful enough**

International Cyber Policy expert Marietje Schaake argues that next year's elections could descend into chaos without adequate precautions.