



How AI and digital technologies enable gendered harms

Written evidence to the UN Working Group on discrimination against women and girls

November 2025

Dr Stefanie Felsberger, Dr Stefanie Ullmann, Dr Clementine Collett, Professor Gina Neff, and Thomas Lacy



**MINDEROO
CENTRE FOR
TECHNOLOGY
& DEMOCRACY**

How AI and digital technologies enable gendered harms

Written evidence to the UN Working Group on discrimination against women and girls for the 2026 Thematic Report on Gender Equality, the Digital Space and the Age of Artificial Intelligence.

Submission made by the Minderoo Centre for Technology and Democracy (MCTD) at the University of Cambridge.

Authored by Dr Stefanie Felsberger, Dr Stefanie Ullmann, Dr Clementine Collett, Professor Gina Neff, and Thomas Lacy.

13 November 2025.

DOI: <https://doi.org/10.17863/CAM.123589>

About this submission

In Autumn 2025, the UN Working Group on discrimination against women and girls [called for inputs](#) from States, NGOs, the private sector, and other stakeholders to inform the production of the 2026 Thematic Report on *Gender Equality, the Digital Space and the Age of Artificial Intelligence*. The final report is expected in June 2026.

We submitted our evidence in November 2025. This document reproduces our submission.

About MCTD

The Minderoo Centre for Technology and Democracy (MCTD) is an independent team of academic researchers at the University of Cambridge who are radically rethinking the power relationships between digital technologies, society, and the planet.

Through our ambitious research agenda, we are enhancing public understanding of digital technologies and delivering positive changes to society's relationship with these technologies.

Find more at mctd.ac.uk

Introduction

Our submission focuses on current harms that digital technologies and Artificial Intelligence (AI) pose to women and girls. We note that the same technologies also present positive impacts in many contexts, but we wish to use this submission to highlight the areas where better technology governance is required to protect women and girls. In many of the areas we explore, gendered harm is preventable and being actively made worse by the way digital technologies are (mis)governed.

Throughout our submission, we signpost specific questions from the call, but we group our answers thematically. Our submission touches on direct harms enabled by AI, on harms resulting from AI supply chains and infrastructure and, mainly, on harms that result from how AI tools are designed, deployed, and governed. These harms include gendered bias and algorithmic discrimination, demeaning representations of women and girls perpetuated with technology, enabling abuse and the prevalence of misogynistic content and abuse on social media platforms. We stress that online and offline harms are not easily separated, with the impacts of emerging technology being felt by women and girls in every aspect of life.

We also call for better access to data for researchers seeking to study the impacts of AI and digital technologies on women and girls, and briefly reflect on some of MCTD's recent relevant submissions and consultations within the UK jurisdiction.

Impact of AI on the rights of women & girls

In this section, we draw attention to the harms to women and girls that AI tools can enable.

First, digital technologies and AI facilitate direct harm and abuse. Examples include non-consensual intimate images, (including so-called 'deepfakes'),¹ abuse of Internet of Things devices in gendered intimate partner violence,² and cyber-stalking and other digital surveillance.³ These harms disproportionately impact women and girls.

Second, on a material level, serious harms to women and girls emerge from the extraordinary demand for resources and infrastructure to build AI technologies. The high demand for water to cool data centres—often built in arid climates—curtails the access of women (and all others) to water rights.⁴ The extraction of materials like copper has led to increased health risks and poverty in local communities.⁵ The high demand for cheap labour to facilitate content moderation undermines women-worker's rights and can expose them to psychological trauma,⁶ and the demand for data labelling work has locked many men and women into very precarious employment situations.⁷

¹ See for example: Karolina Mania, 'Legal Protection of Revenge and Deepfake Porn Victims in the European Union: Findings From a Comparative Legal Study', *Trauma Violence and Abuse*, 25.1 (24 December 2022). <https://doi.org/10.1177/15248380221143772>.

² Andi Brown, Diarmaid Harkin, & Leonie Maria Tanczer . 'Safeguarding the "Internet of Things" for Victim-Survivors of Domestic and Family Violence: Anticipating Exploitative Use and Encouraging Safety-by-Design' *Violence Against Women*, 31(5), 1039-1062. (2024) <https://doi.org/10.1177/10778012231222486>

³ 'Review of IOPC cases involving stalking', *Independent Office for Police Conduct* (September 2024). <https://www.policeconduct.gov.uk/complaints/super-complaints/police-response-stalking/review-of-IOPC-cases-involving-stalking#cyber-stalking> [accessed 11 November 2025].

⁴ 'Where Cloud Meets Cement: A Case Study Analysis of Data Center Development', *The Maybe* (2025), https://cdn.prod.website-files.com/680949ac5f1b50bd51e4a964/682e72957dd411556eae230_The%20Maybe_Where%20cloud%20meets%20cement-%20a%20case%20study%20analysis%20of%20data%20center%20development.pdf. [accessed 5 November 2025]

⁵ Karen Hao. *Empire of AI: Inside the Reckless Race for Total Domination*. London: Allen Lane, 2025.

⁶ Claire Wilmot and Rachel Hall, "Suicide Attempts, Sackings and a Vow of Silence: Meta's New Moderators Face Worst Conditions Yet," *The Bureau of Investigative Journalism* (27 April 2025) <https://www.thebureauinvestigates.com/stories/2025-04-27/suicide-attempts-sackings-and-a-vow-of-silence-metas-new-moderators-face-worst-conditions-yet>. [accessed 12 November 2025].

⁷ Hao, *Empire of AI*. (*Supra*, Note 5).

We focus most of this submission on a third category of harm, namely the negative impact of digital technologies on the rights of women and girls. How digital technologies are designed and governed are important socio-technical considerations, because they shape technologies' use in society. In many cases, gendered harms are preventable or mitigable through better design or governance choices.

Many of these problems at the intersection of gender, technology and rights are transnational and cross borders. Decisions in one country by the owners of technology companies can affect millions of women and girls as users, or as workers in global supply chains.

For one, bias in machine learning and algorithms, including gendered discrimination in search,⁸ content moderation,⁹ and hiring algorithms,¹⁰ are well documented, and will require active work to mitigate.

For example, a recent study found that setting a profile to male or female on Google's Ad Settings page resulted in the female profiles getting fewer instances of an ad related to high paying jobs than the male profiles.¹¹ In another study, researchers at the London School of Economics found that the large language models used by over half of England's local councils to assist social workers in their decision making downplayed women's physical and mental health issues in comparison to men's.¹²

How women and girls are presented in technology design and by technological outputs impacts how women are treated in society. There is evidence that the design of virtual personal assistants such as Siri and Alexa capture and perpetuate unhelpful

⁸ Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018).

⁹ Carolina Are, 'How Instagram's Algorithm Is Censoring Women and Vulnerable Users but Helping Online Abusers', *Feminist Media Studies* 20 (5): 741–44. (2020) <https://doi.org/10.1080/14680777.2020.1783805>

¹⁰ Clementine Collett, Livia Gouvea Gomes, and Gina Neff. *The effects of AI on the working lives of women*. UNESCO Publishing, 2022. <https://unesdoc.unesco.org/ark:/48223/pf0000380861>

¹¹ Amit Datta, Michael Carl Tschantz, and Anupam Datta (2015) 'Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination', *Proceedings on Privacy Enhancing Technologies*, pp. 92–112 (2025). <https://doi.org/10.48550/arXiv.1408.6491>

¹² 'AI Tools Risk Downplaying Women's Health Needs in Social Care', *London School of Economics* (11 August 2025), <https://www.lse.ac.uk/news/latest-news-from-lse/ai-tools-risk-downplaying-womens-health-needs-in-social-care> [accessed 12 November 2025]; Sam Rickman, 'Evaluating gender bias in large language models in long-term care' *BMC Medical Informatics and Decision Making* 25, 274 (2025). <https://doi.org/10.1186/s12911-025-03118-0>.

stereotypes of women as domestic, overly feminised, and submissive, and this has been seen to carry impacts in wider society.¹³

Even tools such as machine translation perpetuate biases and sexist stereotypes, for example by struggling to provide balanced forms for words when translating into a gendered language, especially for minority and low-resource languages (e.g., doctor being translated into the male form only).¹⁴ Despite more than a decade passing since the problem of gender bias in machine translation was first stressed,¹⁵ the problem still remains and it turns out to be much more complex than anticipated.¹⁶

AI and digital systems do not operate in a cultural or social vacuum. There is an ongoing problem of women and minoritised people being systematically underrepresented in training data that is used in AI and machine learning.¹⁷ Bias therefore is a multifaceted and deeply contextual sociotechnical issue, emerging both from data and algorithms, and from the social norms, power structures, and linguistic practices that inform their design and use. Addressing these challenges requires only technical refinement and participatory methods to bring diverse communities into system development. Users should not merely be subjects of data collection but active contributors shaping how their languages and identities are represented.

As AI applications increasingly integrate multimodal inputs and outputs (for example not only text but also images, voice, gesture, or cultural context) the complexity of bias grows. 'Intersectionality' adds yet another layer, as individuals embody multiple identities that interact in ways that are not easily captured through simplified model assumptions. We argue for a rights-based, trans-national and cross-sectoral approach to counter the harmful impact of AI on women and girls and to ensure that women and girls can fully participate in society.

¹³ 'The Effects of AI on the Working Lives of Women', *UNESCO, OECD, and IDB* (2022) <https://unesdoc.unesco.org/ark:/48223/pf0000380861> [accessed 12 November 2025].

¹⁴ Stefanie Ullmann, 'Gender Bias in Machine Translation Systems', in: Ariane Hanemaayer (ed) *Artificial Intelligence and Its Discontents. Social and Cultural Studies of Robots and AI*. Palgrave Macmillan, Cham (2022) https://doi.org/10.1007/978-3-030-88615-8_7

¹⁵ Londa Schiebinger, 'Scientific research must take gender into account', *Nature* 507, 9 (2014). <https://doi.org/10.1038/507009a>

¹⁶ Beatrice Savoldi, Jasmijn Bastings, Luisa Bentivogli, and Eva Vanmassenhove, 'A decade of gender bias in machine translation', *Patterns* 2025, 6, 101257. <https://doi.org/10.1016/j.patter.2025.101257>

¹⁷ J. Q. H. Ho, A. Hartanto, A. Koh, and N. M. Majeed. 'Gender Biases within Artificial Intelligence and ChatGPT: Evidence, Sources of Biases and Solutions', *Computers in Human Behaviour: Artificial Humans*, forthcoming 2025.

Algorithmic harm and content policies on Social Media Platforms

Online algorithms are a particular locus of AI-driven gendered harms, particularly on social media platforms.

We argue that business models that are tied to monetising data extraction mean companies prioritise maximum user engagement over user safety or public wellbeing, increasing the risk of misogynist or gendered harms.¹⁸ Often these harms come from design choices about the 'gendered affordances' that afford users of platforms and systems different capacities for participation based on gender, meaning things are that seemingly neutral like privacy or reach play out differently in the lives of men and women.

With such engagement models, harmful and inaccurate¹⁹ content disseminates faster than other forms of content.²⁰ This means misogynistic content is allowed to exist online, and algorithms that reward virality actively incentivise it.²¹

Recent research demonstrates the increasing scale of misogynist content that is algorithmically amplified online. One study from 2024 about TikTok found that as young men spent more time on the app, they were increasingly recommended content focused on anger and blame, with a four-fold increase in the level of misogynistic content, up to 56% of content.²² Other researchers found that YouTube

¹⁸ We explored these themes in recent written evidence, see for example: Hugo Leal, Stefanie Felsberger, and Gina Neff, 'Harmful by Design: Current Approaches to Misinformation and How to Improve Harm Mitigation', Written evidence submitted to the House of Commons Science, Innovation and Technology Committee (18 December 2024) <https://www.mctd.ac.uk/wp-content/uploads/2025/01/Written-Evidence-Harmful-by-Design-current-approaches-to-misinformation-and-how-to-improve-harm-mitigation.pdf> [accessed 12 November 2025]; and 'Written evidence from Minderoo Centre for Technology and Democracy, University of Cambridge [MIS0042]', *House of Commons Women and Equalities Select Committee* (6 June 2025), <https://committees.parliament.uk/writtenevidence/142618/pdf/> [accessed 12 November 2025].

¹⁹ Becca Schwartz and Gina Neff (2019). The gendered affordances of Craigslist "new-in-town girls wanted" ads. *New Media & Society*, 21(11-12), 2404-2421. <https://doi.org/10.1177/1461444819849897>

²⁰ Soroush Vosoughi, Deb Roy, and Sinan Aral, "The Spread of True and False News Online," *Science* 359.6380 (09 March 2018): 1146–51, <https://doi.org/10.1126/science.aap9559>

²¹ See for example Zizi Papacharissi, *Affective Publics: Sentiment, Technology, Politics*, Oxford, Oxford University Press, 2015; and Centre for Countering Digital Hate, *Deadly by Design* [report], 15 December 2022, https://counterhate.com/wp-content/uploads/2022/12/CCDH-Deadly-by-Design_120922.pdf [accessed 11 November 2025].

²² Kaitlyn Regehr, Caitlin Shaughnessy, Minzhu Zhao, and Nicola Shaughnessy, 'Safer Scrolling: How algorithms popularise and gamify online hate and misogyny for young people' (5 February 2024)

algorithms promoted misogynistic content and that the prevalence of this content is increasing on the platform.²³

Misogynist online content has chilling effects on women's participation online and in public life, causing harm to reputations, careers, and mental wellbeing. Online misogyny can also lead to 'real-world' violence and political radicalisation. A 2024 study argued that online misogynist content can function as a gateway to the alt-right, far right, and potentially violent extremism.²⁴ In written evidence to the House of Commons Science, Innovation and Technology Select Committee, the UK Home Office cited the online environment as a significant factor in inciting violence surrounding the Southport attacks in the UK 2024.²⁵

We argue tackling such harms from online content requires a 'network' approach. Currently, most content moderation strategies take an individual user or content-focused approach: examining individual posts or content on a case-by-case basis. Alternative models of 'context' moderation, which tackle networks and clusters of harmful content, represent more effective models for harm mitigation.²⁶

Content moderation dynamics are complicated on a global scale due to the topographies of cultural and political difference that affect decisions on appropriateness. These can impact what content women and girls can access, or are served up, globally. The content moderation policies of many of the largest social media platforms used in the Global North are often decided within the United States, and then applied to the rest of the world. For example, semi-nudity and sexually

<https://www.ascl.org.uk/ASCL/media/ASCL/Help%20and%20advice/Inclusion/Safer-scrolling.pdf> [accessed 12 November 2025].

²³ Kostantinos Papadamou, et. al., "'How Over is it?'" Understanding the Incel Community on YouTube', *Proceedings of the ACM on Human-Computer Interaction*, 5.2 (2021) 412-25.

<https://doi.org/10.1145/3479556>.

²⁴ Phelia Weiss, Kevin Koban, and Jörg Matthes, 'A Narrow Gateway from Misogyny to the Far Right: Empirical Evidence for Social Media Exposure Effects', *Information, Communication & Society*, 1.19 (December 2024). doi.org/10.1080/1369118X.2024.2445637.

²⁵ 'Written evidence submitted by the Home Office (SDR0015)', *House of Commons, Science, Innovation and Technology Committee*,

<https://committees.parliament.uk/writtenevidence/133688/html> [accessed 11 November 2025].

²⁶ We developed this argument in our written evidence to the House of Commons Science, Innovation and Technology Select Committee's inquiry on social media, misinformation, and harmful algorithms: Hugo Leal, Stefanie Felsberger, and Gian Neff, 'Harmful by design: current approaches to misinformation and how to improve harm mitigation' (December 2024), <https://www.mctd.ac.uk/wp-content/uploads/2025/01/Written-Evidence-Harmful-by-Design-current-approaches-to-misinformation-and-how-to-improve-harm-mitigation.pdf>.

suggestive content may be allowed, while health content such as reproductive or menstrual information is often removed.²⁷ Platforms often prioritise content moderation in English over other languages, with other languages often lacking equivalent content moderation.²⁸

Transparency and Access to Data

Reliable and verifiable research into how AI and algorithms affect gender equality is practically impossible because there is no meaningful transparency in these systems. Researchers have severely limited access to the data that AI systems are using, on how they operate, and how they impact the rights of women and girls.

Social media platforms once provided data access programmes and/or APIs for researchers, but these avenues have been systematically foreclosed.²⁹ We argued that current data access frameworks allow social media companies to be the sole gatekeepers of information, creating information asymmetries 'in which private actors hold more information about the public than public interest institutions'.³⁰ The same arguments can be applied to other AI and digital technology contexts, for instance companies developing Generative AI technologies.

In the UK, the Data (Use and Access) Act 2025 gave the Secretary of State powers to make regulation to require providers of regulated social media platforms, search services, and AI chat bots to 'provide information for purposes related to the carrying out of independent research into online safety, but these regulations have yet to be

²⁷ 'The Digital Gag: The suppression of sexual and reproductive health on meta, tiktok, amazon, and google', *Centre for Intimacy Justice* (2025). <https://www.intimacyjustice.org/report2025> [accessed 23 May 2025]; Kathrin Folkendt, 'Six Women's Health Startups File EU Complaint Against Tech Giants Over Content Censorship' *Feminist Insider*, (11 March 2025). <https://femtechinsider.com/womens-health-startups-file-eu-complaints-against-tech-giants-over-content-censorship/> [accessed 23 May 2025].

²⁸ 'Facebook admits it was used to "incite offline violence" in Myanmar', *BBC*, (6 November 2018). <https://www.bbc.co.uk/news/world-asia-46105934> [accessed 12 November 2025]; Amnesty International, 'Myanmar: Facebook's Systems Promoted Violence Against Rohingya; Meta Owes Reparations', (29 September 2022). <https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/> [accessed 11 November 2025].

²⁹ Hugo Leal, Stefanie Felsberger, Timothy Charlton, and Gina Neff, 'Enabling data-driven research for evidence-based interventions', Written evidence submitted to Ofcom (January 2025). <https://www.mctd.ac.uk/written-evidence-ofcom-data-driven-research-evidence-based-interventions/>.

³⁰ *Ibid.*

published.³¹ We argue that such legal provisions, as well as diligent enforcement of researcher access provisions, are crucial for assessing the scale and nature of algorithmic harms for the rights of women and girls.³²

Roles for Local Jurisdictions: The UK Context

We would like to see increased focus from national governments on ensuring that AI systems do not perpetuate harmful stereotypes and discrimination, that they are trained on responsible data, and that the owners and implementers of these systems are continually assessing impacts on society, altering and improving the design of systems as and when needed.

Under the UK's 2023 Online Safety Act, the UK's regulator for communications services, Ofcom, was required to produce guidance on how online service providers can take action against harmful content and activity that disproportionately affects women and girls. See our evidence in which we welcomed this guidance, but warned that to be more effective it 'needs to rest on sufficiently robust legal codes of practice and risk assessment duties, with adequate enforcement'.³³ Many civil society organisations in the UK have warned that current frameworks do not yet go far enough.³⁴ For example, online dynamics nonetheless make online misogyny more visible is now part of the social media ecosystem.

We argue for better transparency from platforms, including a robust researcher access to data framework,, stringent enforcement and codes of practice for online safety, and content moderation strategies focused on networks of harm, not simply

³¹ 'Researchers access to information from regulated services' *Ofcom* (July 2025).

<https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/call-for-evidence-researchers-access-to-information-from-regulated-online-services> [accessed 11 November 2025].

³² On our arguments for a researcher access to data framework, see: 'Enabling data-driven research for evidence-based interventions' (*Supra*, note 28), on repercussions for keeping women and girls safe online, see: Gina Neff, Hugo Leal, Stefanie Ullmann, Stefanie Felsberger, and Thomas Lacy 'Priorities for keeping women and girls safe online' (May 2025), <https://doi.org/10.17863/CAM.118673>

³³ 'Priorities for keeping women and girls safe online' (*Supra*, note 32).

³⁴ See for example: 'Statement on Ofcom's Illegal Harms Code of Practice', *Online Safety Act Network* (15 January 2025) <https://www.onlinesafetyact.net/analysis/statement-on-ofcom-s-illegal-harms-code-of-practice> [accessed 11 November 2025]; or 'Statement on Ofcom's protection of children codes', *Online Safety Act Network* (06 May 2025) <https://www.onlinesafetyact.net/analysis/statement-on-the-ofcoms-protection-of-children-codes> [accessed 11 November 2025].

on individual instances/posts. These would all be actionable policy solutions for understanding and tackling misogyny online.³⁵

Conclusion

The gendered harms enabled and amplified by AI and digital technologies are neither inevitable nor acceptable. Many of these harms are the direct result of governance failures, misaligned business incentives, and design choices that prioritise profit and engagement over the safety, dignity and rights of women and girls.

Women and girls face multifaceted threats across the digital technology ecosystem: from direct abuse through deepfakes and cyber-stalking, to systemic discrimination embedded in algorithmic systems, to the amplification of misogynistic content that shapes social attitudes and can incite real-world violence. These harms are compounded by the material costs of AI infrastructure development, which disproportionately burden women in global supply chains and communities affected by resource extraction.

The opacity of AI systems and the information asymmetries between technology companies and the public severely limit our collective ability to understand, measure, and address these harms. Without meaningful transparency and better access to data, evidence-based policymaking remains severely constrained.

We call on the UN Working Group to recognise that addressing AI-driven gendered harms requires a coordinated, transnational, rights-based approach. These could include:

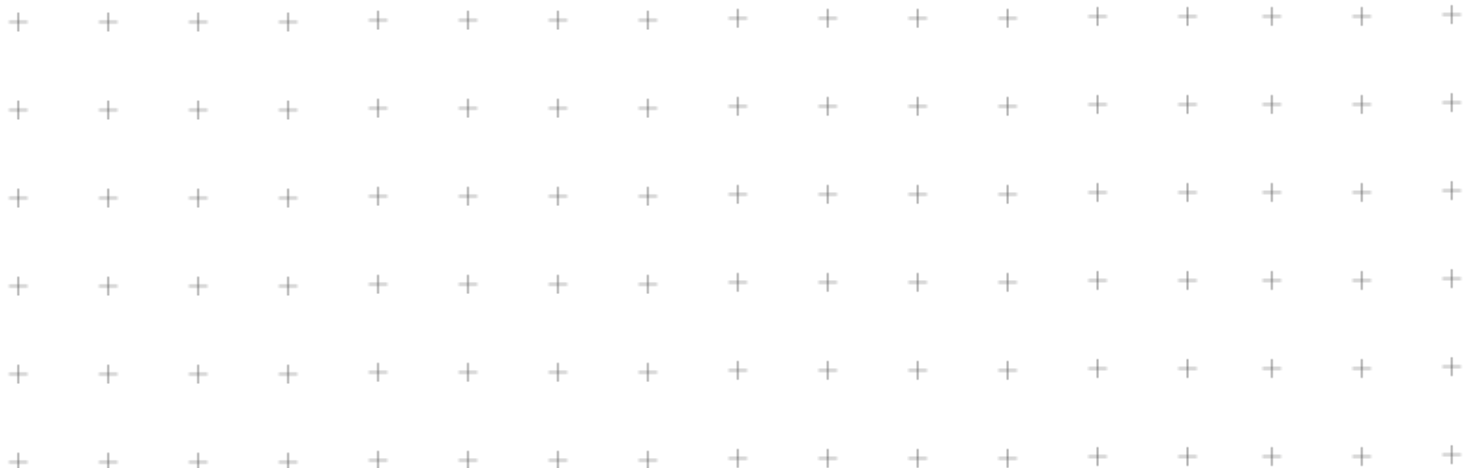
- Mandatory transparency requirements for AI systems and social media platforms, with enforceable researcher access provisions to enable independent scrutiny of algorithmic impacts;
- Standards to prevent and mitigate algorithmic bias and discrimination against women and girls in AI system design, development, and deployment;
- Network-based content moderation strategies that address the structural drivers of misogynistic content amplification, rather than individual posts in isolation;
- Global coordination on technology governance that recognises the transnational nature of digital harms and prevents regulatory arbitrage;

³⁵ House of Commons Women and Equalities Select Committee (2025), *Written evidence from Minderoo Centre for Technology and Democracy, University of Cambridge [MIS0042]*. <https://committees.parliament.uk/writtenevidence/142618/pdf/> [accessed 31 October 2025].

- Participatory design processes that centre the voices and experiences of diverse women and girls in technology development;
- Identification of best practices for business models and design choices that enable fuller participation of women and girls in AI and online systems or mitigate gendered harms.

Women and girls deserve digital environments that support their full participation in society, rather than systems that perpetuate discrimination, amplify abuse, and constrain their rights. Achieving this future requires urgent, coordinated action from governments, international bodies, technology companies, and civil society working together to ensure that digital technologies serve the interests of all people, not just those most powerful in their design and deployment.

[End].



Front Image: Steve Johnson via Unsplash

About the Minderoo Centre for Technology and Democracy

The Minderoo Centre for Technology and Democracy (MCTD) is an independent team of academic researchers at the University of Cambridge who are radically rethinking the power relationships between digital technologies, society, and our planet.

Our research focuses on three key strands:

WORK – Proposing solutions for the harmful impacts of digital technologies on workers' rights.

TRUST – Building informed trust in digital technology and asserting the primacy of democratic values over corporate interests.

ENVIRONMENT – Exposing the global environmental consequences of digital technology.

More information: www.mctd.ac.uk

Get in touch: minderoo@crash.cam.ac.uk

