Written evidence from Dr. Talita Dias, Junior Research Fellow, Jesus College, University of Oxford, Oxford Programme for International Peace and Security, University of Oxford and Ms. Rhiannon Neilsen, Research Consultant, Oxford Programme for International Peace and Security, University of Oxford (TFP0023)

Foreword

This submission is supported by the Oxford Institute for Ethics, Law and Armed Conflict (ELAC) and its Programme for International Peace and Security (IPS), both housed by the Blavatnik School of Government, University of Oxford. IPS provides a space for research on the critical challenges facing the law, norms, and institutions affecting the maintenance and enforcement of international peace and global security.

Executive Summary

- Control over information and communications technologies (ICTs), as well as online and offline artificial intelligence (AI) systems, are currently shifting the international power landscape. Key global threats occurring in this context include ransomware, information technology (IT) supply chain attacks, cyber influence or information operations, and electronic surveillance.
- To promote responsible business practices online, the FCDO should support corporate compliance with international human rights law, independent verification, standardisation auditing and testing of company IT products, and sector-specific business responsibility awards.
- To encourage internationally accepted norms for the use of social media whilst reaping its benefits for diplomacy, the FCDO should assess the advantages of decentralised approaches to social media, strategize substantive and procedural changes with dominant platforms, promote educational campaigns about human rights-compliant platform standards, and support social media companies facing unlawful government demands to remove, limit or publish content.
- To shape the development of, and promote compliance with, international law applicable to ICTs and artificial intelligence, including by taking advantage of the UK's G7 Presidency, the FCDO should work with relevant government bodies to: i) update and revise the UK's national views on international law in the cyber context, particularly as it pertains to sovereignty and due diligence, ii) propose concrete implementation measures, iii) cooperate with *both* like-minded and non-like-minded governments to seek common ground on how international law governs discrete cyber issues, and iv) develop the UK's national views on the application of international law to emerging technologies, including artificial intelligence.

1. What technologies are shifting power? What is the FCDO's understanding of new technologies and their effect on the UK's influence?

Given their ubiquity, pervasiveness, and dual-use nature, ICTs – also known as 'cyber' technologies – currently dominate the quest for political and economic power in the international system. These technologies comprise the Internet and its various physical, data and logical components (e.g. cables, satellites, online applications, protocols, and Big Data),[1]

[1] Clare Sullivan, 'The 2014 Sony Hack and the Role of International Law', 8 *Journal of National Security Law and Policy* (2015) 437, at 454, fn 88. See also Nicholas Tsagourias, 'The Legal Status of Cyberspace', in

as well as Internet of Things (IoT) devices, such as smartphones, smart watches, sensors, control valves, virtual assistant technologies, and self-driving cars.[2] These technologies have led to immense social, economic, and cultural progress around the world. But their vulnerabilities, which have expanded with our growing dependence on ICTs, have been exploited for a number of malicious ends and caused significant harm to individuals, private entities and States worldwide.[3] Some of the most pressing cyber threats include:

a) **Ransomware**, i.e., malicious software (malware) used to block the availability of data or systems subject to a ransom payment,[4] is often listed as 'the number one cyber threat', given its frequency, pervasiveness, and impact.[5] It is particularly concerning when directed at critical information infrastructure, such as hospitals, oil pipelines, and medical facilities.[6]

b) **IT supply chain attacks**, that is, cyber operations exploiting vulnerabilities in widely used software or hardware. These carry a great risk of systematic harm to individuals, corporations and States,[7] especially when targeted at network management software[8]

Nicholas Tsagourias and Russel Buchan (eds.), *Research Handbook on International Law and Cyberspace* (2015) 13. See also David Johnson and David Post, 'Law and Borders: The Rise of Law in Cyberspace', 48 *Stanford Law Review* (1996) 1367.

[2] Arm, 'What are IoT devices', available at https://www.arm.com/glossary/iot-devices; https://en.wikipedia.org/wiki/Internet_of_things; Wikipedia contributors, 'Internet of Things', available at https://en.wikipedia.org/w/index.php?title=Internet_of_things.

[3] See OEWG, Final Substantive Report, UN Doc A/AC.290/2021/CRP.2, 10 March 2021 ('OEWG Final Substantive Report'), paras 4, 15, 20-21; OEWG, Chair's Summary, UN Doc. A/AC.290/2021/CRP.3*, 10 March 2021, paras 7-8, 25. See also OEWG, 'Draft Substantive Report [Zero Draft], A/AC.290/[DATE], 19 January 2021, available at https://front.un-arm.org/wp-content/uploads/2021/01/OEWG-Zero-Draft-19-01-2021.pdf ('OEWG Zero Draft'), paras 4 and 17.

[4] Josh Fruhlinger, 'Ransomware explained: How it works and how to remove it', *CSO*, 19 June 200, available at https://www.csoonline.com/article/3236183/what-is-ransomware-how-it-works-and-how-to-remove-it.html.

[5] Jason Firch, '10 Cyber Security Trends You Can't Ignore In 2021', *Purplesec*, 31 December 2020, available at https://purplesec.us/cyber-security-trends-2021/#Ransomware.

[6] A recent example of a malicious cyber operation compromising critical infrastructure is the cyber attack (attributed to hacker group *DarkSide*) that shut down Colonial Pipeline, one of the largest oil suppliers in the US. The ransomware attack resulted in 45 per cent of the east coast's fuel supplies being blocked and the US government invoking emergency powers. Please see Erum Salam, 'Cyber-attack forces shutdown of one of the US's largest pipelines', *The Guardian,* 9 May 2021, available at https://www.theguardian.com/technology/2021/may/08/colonial-pipeline-cyber-attack-shutdown. See also Jan Lemintzer, 'Ransomware gangs are running riot – paying them off doesn't help', *The Conversation*, 17 February 2021, available at https://theconversation.com/ransomware-gangs-are-running-riot-paying-them-off-doesnt-help-155254?utm_source=twitter&utm_medium=bylinetwitterbutton; BDO, 'BDO's Fall 2019 Cyber Threat Report: Focus On Healthcare', October 2019, available at https://www.bdo.com/insights/business-financial-advisory/cybersecurity/bdos-fall-2019-cyber-threat-report-focus-on-health. On the impact of the WannaCry ransomware on the NHS in 2017, see Rory Cellan-Jones, 'Ransomware and the NHS -– the inquest begins', *BBC News*, 14 May 2017, available at https://www.bbc.co.uk/news/technology-39917278; Acronis, 'The NHS cyber attack', available at https://www.acronis.com/en-gb/articles/nhs-cyber-attack/..

[7] See Brad Smith, 'A moment of reckoning: the need for a strong and global cybersecurity response', *Microsoft On the Issues*, 17 December 2020, available at https://blogs.microsoft.com/on-the-issues/2020/12/17/cyberattacks-cybersecurity-solarwinds-fireeye/; Kari Paul, 'SolarWinds: company at the core of the Orion hack falls under scrutiny', *The Guardian*, 16 December 2020, available at https://www.theguardian.com/technology/2020/dec/16/solarwinds-orion-hack-scrutiny-technology.

[8] Martin Courtney, 'Digital Doomsday', *Engineering & Technology Magazine*, 8 October 2019, available at https://eandt.theiet.org/content/articles/2019/10/digital-doomsday; Nigel Lawrence and Patrick Traynor, 'Under New Management: Practical Attacks on SNMPv3', *WOOT'12: Proceedings of the 6th USENIX conference on Offensive Technologies*, August 2012, available at https://dl.acm.org/doi/10.5555/2372399.2372416, at 2; Joe Weiss and Bob Hunter, 'The SolarWinds Hack Can Directly Affect Control Systems', *Lawfare*, 22 January 2021, available at https://www.lawfareblog.com/solarwinds-hack-can-directly-affect-control-systems; Sue

used to monitor physical IoT devices, such as sensors used for critical infrastructure, including water distribution, power supply, and nuclear plants.[9]

c) **Cyber influence or information operations**, broadly defined as any coordinated or individual deployment of digital resources for unlawful cognitive purposes, such as disinformation, misinformation, and online hate speech, particularly when these take place during elections, violence and armed conflict.[10] 'Deepfakes' may also present a threat to national security and democracy by eroding trust in institutions, misinforming political decisions, or inciting reactions based on the fabricated depictions of UK leaders or that of its allies.[11]

d) **Electronic surveillance**, comprising the mass or targeted use of software or hardware to intercept private communications, especially through spyware software.[12] The increased use of certain surveillance practices are at risk of breaching data protection, privacy, and equality laws, especially in terms of the employment of facial recognition practices.[13]

A growing number of software applications used for such harmful cyber operations, including malware, are powered by AI technologies. AI is defined as the use of computers, including virtual programmes and robots, to mimic certain human skills, such as perception,

---

Helpern, 'After the SolarWinds Hack, We Have No Idea What Cyber Dangers We Face, *The New Yorker*, 25 Jan 2021, available at https://www.newyorker.com/news/daily-comment/after-the-solarwinds-hack-we-have-no-idea-what-cyber-dangers-we-face; Software Engineering Institute, CERT Coordination Center, 'SolarWinds Orion API authentication bypass allows remote command execution: Vulnerability Note VU#843464', *Carnegie Mellon University*, 26 December 2020, available at https://kb.cert.org/vuls/id/843464.

[9] Cameron Abbott, 'Interlopers in Things? IOT Devices May be used as Backdoors to your Network', *The National Law Review*, 27 August 2019, available at https://www.natlawreview.com/article/interlopers-things-iot-devices-may-be-used-backdoors-to-your-network.

[10] Accenture, '2019 Cyber Threat Landscape Report', *supra* note, at 13, 18-19; Gary Brown, 'Addressing Cyber-Enabled Information Operations', *RUSI*, 1 May 2020, available at https://rusi.org/sites/default/files/20200501_brown_web.pdf; Claire Wardle and Hossein Derakshan, 'Information Disorder: Toward an interdisciplinary framework for research and policymaking', Council of Europe Report, DGI(2007)09, 27 September 2017, available at https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c, at 5, 10-13, 20-2.

[11] Alexa Koenig, '"Half the Truth is Often a Great Lie": Deep Fakes, Open Source Information, and International Criminal Law', *AJIL Unbound* (2019), 113: 250-255; Kelley M. Sayler and Laurie A. Harris, 'Deep Fakes and National Security', *Congressional Research Service,* October 2019; UK House of Commons: Digital, Culture, Media and Sport Committee, "Disinformation and 'fake news': Final Report", House of Commons, 4 February 2019, available at https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf, at 11; Hannah Smith and Katherine Mansted, 'Weaponised Deep Fakes: National Security and Democracy', *Australian Strategic Policy Institute,* April 2020, at 11.

[12] John P. Mello Jr., 'What is spyware? How it works and how to prevent it', *CSO*, 28 March 2019, available at https://www.csoonline.com/article/3384100/what-is-spyware-how-it-works-and-how-to-prevent-it.html; *Software Lab*, 'What is Spyware? Top 5 Types & Examples', available at https://softwarelab.org/what-is-spyware/; Alexander S. Gillis, 'Spyware', *TechTarget*, November 2019, available at https://searchsecurity.techtarget.com/definition/spyware.

[13] For instance, the UK Court of Appeal found the British police force's use of facial recognition was 'unlawful' in August 2020. See Davey Winder, 'Police Facial Recognition Use Unlawful – UK Court of Appeal Makes Landmark Ruling,' *Forbes,* 12 August 2020, available at https://www.forbes.com/sites/daveywinder/2020/08/12/police-facial-recognition-use-unlawful-uk-court-of-appeal-makes-landmark-ruling/?sh=485679a575e0; Jenny Rees, 'Facial recognition use by South Wales Police ruled unlawful', *BBC News,* 11August 2020, available at: https://www.bbc.com/news/uk-wales-53734716.

association, prediction, planning and motor control.[14] The technology uses symbolic, i.e. human-readable, or subsymbolic, also known as machine learning, algorithms, i.e. complex statistical-probabilistic equations, which today dominates the field.[15] In essence, AI algorithms use statistics and probability to make predictions about a variety of subjects, such as the likelihood that an image, word or text belongs to a certain category or will reappear. Its applications range from numerous types of image and speech recognition programmes, such as medical image diagnosis and computer vision, to self-driving vehicles, drones and autonomous weapons systems.[16] A vast amount of AI applications are used online to power search engines, social media feeds and recommendation engines.[17] AI algorithms have also been increasingly used to filter through job applications, identify citizens, predict crime and recidivism, estimate student and teacher performance, calculate credit score and offer numerous social befits, from medical treatment to childcare.[18]

However, all AI algorithms, from good old-fashioned symbolic algorithms to machine and deep learning, are essentially *quantitative*: they make predictions based on the incidence of certain features in the data with which they are trained. Put differently, they learn to make statistical, often non-causal or irrelevant, associations between numerically identified features across their data pool, such as the pixels in an image.[19] As such, they are incapable of making basic *qualitative* judgments that are essential to humans in day-to-day activities, and thus the algorithms do not *innately* account for contextual, abstract, and common-sense knowledge.[20] In the case of machine learning algorithms, numerical weights or parameters are pre-programmed to automatically change based on the vast quantities of data – Big Data – they process over time, which means that their decision-making processes are largely incomprehensible to humans.[21]

This reliance on historical data and quantitative associations have inevitably led to crass errors and the amplification of societal biases,[22] such as image recognition systems that problematically mislabel and misidentify the faces of non-white individuals and women.[23]

---

[14] Margaret A. Boden, *AI: Its Nature and Future* (Oxford University Press, 2016), at 1-3; Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (Pelican Books, 2019), at 7; Access Now, 'Human Rights in the Age of Artificial Intelligence', 8 November 2018, available at https://www.accessnow.org/human-rights-in-the-age-of-AI, at 8.

[15] See Mitchell, ibid, at 9-12, 18-22, 33-36; Access Now, *supra* note 14, at 8-9, 13; Wikipedia contributors, 'Symbolic artificial intelligence', available at https://en.wikipedia.org/wiki/Symbolic_artificial_intelligence; Rhett D'Souza, 'Symbolic AI v/s Non-Symbolic AI, and everything in between?', *Data Driven Investor*, 19 October 2018, available at https://medium.datadriveninvestor.com/symbolic-ai-v-s-non-symbolic-ai-and-everything-in-between-ffcc2b03bc2e.

[16] Mitchell, *supra* note 14, at 12, 90-91, 114; Access Now, *supra* note 14, at 14.

[17] Roger Chua, 'A simple way to explain the Recommendation Engine in AI', *Medium*, 26 June 2017, available at https://medium.com/voice-tech-podcast/a-simple-way-to-explain-the-recommendation-engine-in-ai-d1a609f59d97; Google Cloud, 'Recommendations AI', available at https://cloud.google.com/recommendations.

[18] See Cathy O'Neil, *Weapons of Math Destruction* (Penguin Books, 2016); Access Now, *supra* note 14, at 14-16.

[19] Mitchell, *supra* note 14, at 70-72, 122.

[20] Mitchell, *supra* note 14, at 33-35, 69-70, 108, 136; Boden, *supra* note 14, at 40-44, 56.

[21] Mitchell, *supra* note 14, at 12-22, 27-33,72-88, 109-114; Access Now, *supra* note 14, at 13.

[22] Mitchell, *supra* note 14, at 123-124; Access Now, *supra* note 14, at 11-12, 24.

[23] For instance, black individuals have been mislabelled as 'gorillas' by automated algorithms, and 28 of the 535 members of the US congress were misidentified as 'criminals' by Amazon's Rekognition facial recognition system. See O'Neil, *supra* note 18, at 154-155; Alex Hern, 'Google's solution to accidental algorithmic racism: ban gorillas', The Guardian, 12 January 2018, available at https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people; Russell Brandom, "Amazon's facial recognition matched 28 members of Congress to criminal mugshots," The Verge, July 26, 2018, https://www.theverge.com/2018/7/26/17615634/amazon-rekognition-aclu-mug-shot-congress-

Similarly, AI recommendation algorithms on social media prioritise recurrent content, which results in the amplification of disinformation, division and hatred.[24] At times, such discrimination is a product of intentional, human-controlled exclusion of certain groups (women, minorities, or religion) in programming the AI's targeting algorithms, resulting 'dark ads'.[25] All discrimination on this basis should be condemned. Yet often these issues arise because the AI algorithmic decisions are based on data that is *itself* imbued with systemic or subconscious discriminatory assumptions.[26] For instance, in the case of recidivism in the US, the algorithmic program COMPAS was found to consistently – and wrongly – flag black defendants at a higher risk of re-offending compared to their white counterparts, and were thus denied parole.[27] If the biases embedded within "training data" (the dataset from which algorithms and models learn) are not mitigated, or if the training data is not designed with adequate human input and oversight, automated decision-making algorithms will continue to perpetuate these results, thereby exacerbating systemic bias.[28] The effects of bias in automated decision-making stem from what Joy Buolamwini terms 'the coded gaze' – "reflection of the priorities, the preferences, and also sometimes the prejudices of those who have the power to shape technology".[29]

The FCDO is well aware of this threat landscape and is taking important steps to address it.[30] In particular, the recent Ministerial Declaration ensuing from the latest G7 Digital and

---

facial-recognition.

[24] O'Neil, *supra* note 18, at 180-185; Access Now, *supra* note 14, at 16; Yaël Eisenstat, 'Dear Facebook, this is how you're breaking democracy', *TED*, August 2020, available at https://www.ted.com/talks/yael_eisenstat_dear_facebook_this_is_how_you_re_breaking_democracy; Carole Cadwalladr, 'If you're not terrified about Facebook, you haven't been paying attention', The Guardian, 26 July 2020, available at https://www.theguardian.com/commentisfree/2020/jul/26/with-facebook-we-are-already-through-the-looking-glass; Cathy O'Neil, 'TikTok's Algorithm Can't Be Trusted', Bloomberg, 21 September 2020, available at https://www.bloomberg.com/opinion/articles/2020-09-21/tiktok-s-algorithm-can-t-be-trusted.

[25] Hal Conick, 'The Ethics of Targeting Minorities with Dark Ads', *American Marketing Association,* 21 March 21 2019, available at https://www.ama.org/marketing-news/the-ethics-of-targeting-minorities-with-dark-ads; U.S. Department of Housing and Urban Development, 'HUD Files Housing Discrimination Complaint Against Facebook: Secretary-initiated complaint alleges platform allows advertisers to discriminate', *House and Urban Development*, 17 August 2018, available at https://archives.hud.gov/news/2018/pr18-085.cfm.

[26] Selena Silva and Martin Kenney, 'Algorithms, Platforms, and Ethnic Bias', *Viewpoint* 62, no. 11 (2019), at 37; Heidi Ledford, 'Millions of Black People Affected by Racial Bias in Health-Care Algorithms', *Nature,* 24 October 2019, available at https://www.nature.com/articles/d41586-019-03228-6/; Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (London: Penguin Books, 2017); Nicol Turner Lee, 'Detecting Racial Bias in Algorithms and Machine Learning', *Journal of Information, Communication and Ethics in Society* 16, no. 3 (2018), at 253.

[27] Larson et al., 'How We Analyzed the COMPAS Recidivism Algorithm' *ProPublica,* 23 May 2016, available at https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm; Julia Angwin et al., 'Machine Bias," *ProPublica*, 23 May 2016, available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; Turner Lee, 'Detecting Racial Bias in Algorithms and Machine Learning', at 254; Ed Yong, 'A Popular Algorithm Is No Better at Predicting Crimes Than Random People', *The Atlantic*, 17 January 2018; Julia Dressel and Hany Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism', *Science Advances* 4, no. 1 (2018), at 2.

[28] Solon Barocas and Andrew D. Selbst, 'Big Data's Disparate Impact', *SSRN Electronic Journal* (2016), at 674–75; Silva and Kenney, 'Algorithms, Platforms, and Ethnic Bias', at 37–38; Will Knight, 'AI is Biased. Here's How Scientists are Trying to Fix it', *WIRED,* 19 December 2019, available at https://www.wired.com/story/ai-biased-how-scientists-trying-fix/; David Jacobus Dalenberg, 'Preventing Discrimination in the Automated Targeting of Job Advertisements', *Computer Law & Security Review* 34, no. 3 (2018), at 615–27; John Podesta et al., 'Big Data: Seizing Opportunities, Preserving Values', Executive Office of the President (Washington DC: White House, 2014), at 51.

[29] Joy Buolamwini, 'Compassion through Computation: Fighting the Algorithmic Bias', World Economic Forum Annual Meeting, Davos-Klosters, Switzerland (2019), available at https://www.weforum.org/events/world-economic-forum-annual-meeting-2019/sessions/compassion-through-computation-fighting-algorithmic-bias.

Technology Ministers' meeting under the UK's presidency rightly notes the importance of a) security, resilience and diversity in information technology supply chains; b) industry-led technology standards for the Internet and digital technologies; and c) the safety of Internet users, particularly the most vulnerable ones.[31] Yet more can be done to raise awareness of and address the challenges outlined above together with States, international organisations, technology companies and civil society organisations around the world. To mitigate these risks, the FCDO can require technology corporations to undertake a thorough, external and independent examination of 'training data' before it is used for AI decision-making processes. This is especially for important in the development of AI algorithms that have an impact on individuals' livelihoods – often without their awareness (such as employment, parole, and credit scores).

2. How can the FCDO engage with private technology companies to influence and promote the responsible development and use of data and new technologies?

The UK has already set in motion national strategies to address the cyber threat landscape. Most prominent among these is the recent Government response to the Online Harms White Paper, laying out a proposed legal duty of care on online companies and its ensuing responsibilities.[32] Nevertheless, given the interconnectedness and transboundary nature of ICTs, and the fact that most technology companies are based overseas, a domestic corporate liability legal framework is insufficient on its own to address the root cause of the problem. International(ised) strategies to promote responsible business practices online are thus essential, and may include:

a) Consulting with domestic and foreign technology companies and foreign governments with a view to adopting uniform standards for data protection, content moderation and algorithmic transparency, which are in line with international human rights law instruments, such as International Covenant on Civil and Political Rights[33] and the International Covenant on Economic Social and Cultural Rights,[34] and the United Nations (UN) Guiding Principles on Business and Human Rights.[35] In this regard, the European Union (EU)'s Code of Conduct on Countering Illegal Hate Speech[36]

---

[30] See, e.g., FCDO, 'Foreign Secretary boosts BBC funding to fight fake news', 1 May 2020, available at https://www.gov.uk/government/news/foreign-secretary-boosts-bbc-funding-to-fight-fake-news; FCDO, 'Russia: UK exposes Russian involvement in SolarWinds cyber compromise', 15 April 2021, available at https://www.gov.uk/government/news/russia-uk-exposes-russian-involvement-in-solarwinds-cyber-compromise; FCDO, 'Russia: UK and US expose global campaign of malign activity by Russian intelligence services', 15 April 2021, available at https://www.gov.uk/government/news/russia-uk-and-us-expose-global-campaigns-of-malign-activity-by-russian-intelligence-services.

[31] G7, 'Ministerial Declaration: G7 Digital and Technology Ministers' meeting', 28 April 2021, available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/981567/G7_Digital_and_Technology_Ministerial_Declaration.pdf.

[32] Secretary of State for Digital, Culture, Media and Sport and by the Secretary of State for the Home Department, 'Consultation outcome - Online Harms White Paper: Full government response to the consultation', 15 December 2020, available at http://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response.

[33] 999 UNTS 171 (ICCPR).

[34] 993 UNTS 3 (ICESCR).

[35] UN, 'Guiding principles on business and human rights: implementing the United Nations "Protect, Respect and Remedy" framework', 2011.

[36] EU, 'Code of conduct on countering illegal hate speech online', 30 June 2016, available at https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en.

provides a successful model for public-private partnerships which could be brokered in the context of the other cyber threats described earlier, such as ransomware, disinformation and digital supply chain attacks.

b) Partnering with such corporations to establish independent, international mechanisms for verification, auditing, standardisation, and certification of software and hardware products before their sale and/or use is authorised domestically, bearing in mind the need to protect proprietary rights and trade secrets.[37] As the G7 Digital and Technology Ministers recently recognised,[38] technical standards have the potential to fill regulatory gaps, inform users and promote compliant businesses. Although ISO's expert-driven, internationally agreed standards already apply to several ICT-related areas, such as information security, cybersecurity and privacy protection,[39] similar standards are lacking on IT supply chain integrity and algorithmic transparency. The UK's G7 presidency is an opportunity to push for the adoption of such standards at a global level.

c) Establishing, together with other UK and foreign government bodies, social responsibility awards specific to domestic and foreign technology companies that have a consistent record of compliance with international and domestic rules or standards.[40] While naming and shaming non-compliant behaviour may be an effective deterrent in some instances, corporations also need an incentive to behave responsibly.[41] Like standards, corporate social responsibility awards can enhance user trust in responsible companies and boost their business, thereby setting in motion a cycle of compliance.

3. How can the FCDO engage with private companies to encourage internationally accepted norms for the use of social media as well as to maximise the benefits for diplomacy presented by social media?

The unprecedented and highly concentrated power wielded by social media companies requires targeted, sector-specific measures, in addition to the engagement strategies outlined above. To encourage the adoption of and compliance with internationally accepted norms for the use of social media and leverage its opportunities for diplomacy, the FCDO should:

a) Work together with smaller, non-profit, open-source and decentralised social media platforms, such as Diaspora, Minds and Mastodon,[42] to understand the extent to which

---

[37] Human Rights Council, 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression', 6 April 2018, A/HRC/38/35, para 56; Access Now, *supra* note 14, at 33-36.

[38] G7, 'G7 Digital and Technology Track - Annex 1- FRAMEWORK FOR G7 COLLABORATION ON DIGITAL TECHNICAL STANDARDS, available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/981571/Annex_1__Framework_for_G7_collaboration_on_Digital_Technical_Standards.pdf.

[39] See Barnaby Lewis, 'Safe, secure and private, whatever your business', *ISO News*, 4 May 2020, available at https://www.iso.org/news/ref2495.html.

[40] For a list of current UK Corporate Social Responsibility Awards, please see https://awards-list.co.uk/uk-awards/corporate-social-responsibility-csr-awards/.

[41] For more on naming and shaming in cyberspace and the possible alternative of 'accusation', please see Martha Finnemore and Duncan B Hollis, 'Beyond Naming and Shaming: Accusations and International Law in Cybersecurity', *European Journal of International Law* 31, no. 3 (2020), at 973-74.

[42] Tom Meritt, Top 5 decentralized social networks', 21 March 2019, *TechRepublic*, available at https://www.nytimes.com/2018/03/28/technology/social-media-privacy.html; Kevin Roose, 'Can Social Media be Saved?', The New York Times, 28 March 2018, available at

alternative platform models affording greater user control are more conducive to upholding international human rights law, including the corporate responsibilities laid down in the UN Guiding Principles on Business and Human Rights.[43]

b) Bring together dominant social media companies, such as Facebook, Twitter and TikTok, to strategise changes needed to improve compliance with international human rights law,[44] considering the specific guidance provided by the Special Rapporteur on Freedom of Expression.[45] Such changes likely include:

i. Substantive reforms to platform community standards, which ought to be harmonised and made consistent with international human rights law.[46] In particular, community standards should ban prohibited speech, such as propaganda for war and advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence, but these must be tightly defined, bearing in mind context and intersections with freedom of speech.[47] In the event that initial detection and takedown was automated, prohibited speech that meets a certain threshold of gravity should be subject to meaningful review by human content moderators.[48] Other types of content, which might infringe on the rights and reputations of others, or affect national security interests, public order, health or morals, should be laid down in a clear and accessible manner, preferably with concrete examples.[49] Removal of these types of content should be a measure of last resort once other, less restrictive measures, such as tagging and de-prioritisation, have proven to be ineffective, and subject to meaningful human review.[50] It is also worth noting that, for all types of online hate speech – prohibited or not –, States must ensure access to justice and an effective remedy to affected individuals, whether speakers, targeted individuals, or members of the audience. In the first instance, the FCDO should support social media corporations in flagging online posts containing incitement to violence, or dangerously misleading information.[51]

---

https://www.nytimes.com/2018/03/28/technology/social-media-privacy.html; Margaret Rhodes, 'Like Twitter But Hate the Trolls? Try Mastodon', *Wired*, 4 March 2017, available at https://www.wired.com/2017/04/like-twitter-hate-trolls-try-mastodon/.

[43] See Principles 11-24.

[44] See, e.g. Nathaniel Popper, 'Twitter and Facebook Want to Shift Power to Users. Or Do They?', *The New York Times*, 18 December 2019, available at https://www.nytimes.com/2019/12/18/technology/facebook-twitter-bitcoin-blockchain.html; Mike Masnick, 'Protocols, Not Platforms: A Technological Approach to Free Speech', *Knight First Amendment Institute at Columbia University*, 21 August 2019, available at https://knightcolumbia.org/content/protocols-not-platforms-a-technological-approach-to-free-speech; Adi Robertson, 'Twitter is funding research into a decentralized version of its platform', *The Verge*, 11 December 2019, available at https://www.theverge.com/2019/12/11/21010856/twitter-jack-dorsey-bluesky-decentralized-social-network-research-moderation.

[45] See, e.g., Human Rights Council (HRC), 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression', 6 April 2018, A/HRC/38/35; HRC, 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression', 9 October 2019, A/74/486.s

[46] HRC, A/HRC/38/35, *supra* note 45, paras 10-11 ; HRC, A/74/486.s, *supra* note 45, para 42.

[47] See Article 20, ICCPR; HRC, A/HRC/38/35, *supra* note 45, paras 8, 13, 20, 27-29, 33; HRC, A/74/486.s, *supra* note 45, paras 8-18.

[48] HRC, A/HRC/38/35, *supra* note 45, paras 32-35, 46; HRC, A/74/486.s, *supra* note 45, paras 44-45.

[49] HRC, A/HRC/38/35, *supra* note 45, paras 26, 39-40, 46-47, 52, 63; HRC, A/74/486.s, *supra* note 45, para 47.

[50] HRC, A/HRC/38/35, *supra* note 45, paras 20 and 44; HRC, A/74/486.s, *supra* note 45, paras 19-23.

Inclusive in this effort should be 'redirect method', which prompts users to access accurate and verifiable information provided trusted sources over that of fake news (as evidenced by YouTube, Twitter, and Facebook, during the COVID-19 crisis).[52] Upon the removal of prohibited speech, especially incitement to violence and evidence of war crimes, the FCDO should ensure that such content is stored in encrypted, digital 'evidence lockers'.[53] The FCDO should dedicate resources toward identifying which organisation should be responsible for the storage and maintenance of such lockers.

ii. Procedural reforms to content moderation decision-making and complaint processes. Given the inherent limits of AI content-moderation technologies described above, such as image and text recognition, content removal decisions should always be preceded or promptly confirmed by trained moderators with country-specific contextual knowledge.[54] Affected users, including content authors, addressees and flaggers, should be immediately notified upon the adoption of any limiting measure.[55] Internal complaint or appeal mechanisms against content moderation decisions should be transparent and easily accessible to authors or affected users.[56] Complaints should be decided by independent organs in a transparent manner, with sufficient reasons provided.[57] Recognising the challenges of establishing such a complaint mechanism in a scalable manner, some have proposed the creation of company-specific or industry-wide ombudspersons or social media council.[58] However, to ensure greater representation and impartiality of such decision-making bodies, we propose leveraging existing technologies to decentralise the process and prioritise transparency. Inspired by the jury system, individuals could be randomly selected to sit on country-specific appellate 'juries' and vote on the merits of the decision to maintain, remove or

---

[51] BBC News, Coronavirus: Twitter will label COVID-19 fake news, BBC, 12 May 2020, available at https://www.bbc.com/news/technology-52632909.

[52] Joan Donovan, 'Here's how social media can combat the coronavirus "infodemic"', *MIT Technology Review*, 17 March 2020, available at https://www.technologyreview.com/2020/03/17/905279/facebook-twitter-social-media-infodemic-misinformation/; Carmen Ferri, 'Social media's response to COVID-19 misinformation', *Association for Progressive Communications*, 21 April 2020, available at https://www.apc.org/en/news/social-medias-response-covid-19-misinformation.

[53] For more on social media corporations deleting evidence of atrocity crimes, and the need to develop an independent archival system for the preservation of digital evidence, please see Alexa Koenig, 'Big Tech Can Help Bring War Criminals to Justice', *Foreign Affairs,* 11 November 2020, available at: https://www.foreignaffairs.com/articles/united-states/2020-11-11/big-tech-can-help-bring-war-criminals-justice; Anna Veronica Banchik, 'Disappearing acts: Content moderation and emergent practices to preserve at-risk human rights-related content', *New Media and Society,* (2020), 1-18; Human Rights Watch, '"Video Unavailable": Social Media Platforms Remove Evidence of War Crimes', *Human Rights Watch,* 10 September 2020, available at: https://www.hrw.org/report/2020/09/10/video-unavailable/social-media-platforms-remove-evidence-war-crimes; *Human Rights Center at the University of California, Berkeley, School of Law,* 'Digital, Lockers: Options for Archiving Social Media Evidence of Atrocity Crimes', forthcoming. For more on the collection, analysis, and preservation of digital open source materials for international justice and accountability, please see 'Berkeley Protocol on Digital Open Source Investigations', *Human Rights Center at the University of California, Berkeley, School of Law and Office of the United Nations High Commissioner for Human Rights (OHCHR),* 2020, available at https://www.ohchr.org/Documents/Publications/OHCHR_BerkeleyProtocol.pdf.

[54] HRC, A/HRC/38/35, *supra* note 45, paras 32-35, 56-57; HRC, A/74/486.s, *supra* note 45, paras 34 and 50.

[55] HRC, A/HRC/38/35, *supra* note 45, para 37; HRC, A/74/486.s, *supra* note 45, para 31.

[56] HRC, A/HRC/38/35, *supra* note 45, para 38; HRC, A/74/486.s, *supra* note 45, para 53.

[57] HRC, A/HRC/38/35, *supra* note 45, para 58; HRC, A/74/486.s, *supra* note 45, para 44.

[58] HRC, A/HRC/38/35, *supra* note 45, para 58.

otherwise limit the relevant content.[59] Yet even such popular decisions should always be subject to judicial review, which requires close cooperation between social media platforms and domestic courts.[60] This may operate similarly to the independent Oversight Board, consisting of forty international experts, which became operational in 2020 to assess Facebook's decisions across both Facebook and Instagram.[61] The Board has the binding "authority to decide whether Facebook and Instagram should allow or remove content", including the recent judgement to uphold Donald Trump's accounts suspension from Facebook.[62]

iii. Change in algorithmic design to give users greater choice about the types of content they want to see in their feed.[63] For instance, users could have the right to opt out from platform-curated feeds and set their own, personalised curation standards, such as by sorting content in chronological order or selecting specific interests.[64] To avoid the phenomenon of echo-chambers, whereby users are exposed almost exclusively to like-minded and viral content that captures their attention,[65] platforms should promote respectful dialogues and engagement,[66] periodically prioritise non-like-minded content and counter-narratives, as well as de-amplify prohibited and blatantly false content.[67] For example, since 2018 Facebook has engaged independent fact-checkers to 'rate' the content accuracy of posts, and if fact-checkers "find falsities" they are required to de-prioritise the post.[68] In effect, "a person posting misinformation might find their content cast farther down the News Feed".[69] So long as the misinformation does not incite imminent violence or violate hate speech regulations, the post will remain online, albeit not as

---

[59] The benefit of a country-specific appellate jury is that it allows for decisions to be made by in-country jurors, who will likely be more aware of, and sensitive to, the particular social context and dynamics in which the harmful online content appears.

[60] HRC, A/74/486.s, *supra* note 45, para 57(e).

[61] Oversight Board, 'Ensuring respect for free expression, through independent judgment', 2020, available at https://oversightboard.com.

[62] Ibid and Nick Clegg, 'Oversight Board upholds Facebook's decision to suspend Donald Trump's accounts', *Facebook Newsroom,* 5 May 2021, available at https://about.fb.com/news/2021/05/facebook-oversight-board-decision-trump/; BBC News, 'Facebook's Trump ban upheld by Oversight Board for now', *BBC,* 6 May 2021, available at https://www.bbc.com/news/technology-56985583.

[63] *Supra* note 44.

[64] HRC, A/HRC/38/35, *supra* note 45, paras 60-61.

[65] Tim Wu, *The Attention Merchants: The Epic Scramble to Get Inside Our Heads* (New York, Vintage Books, 2016); Matteo Cinelli and others, 'The echo chamber effect on social media', 118 (9) PNS, 2 March 2021, available at https://www.pnas.org/content/118/9/e2023301118; Zeynep Tufekci, 'We're building a dystopia just to make people click on ads', *TED*, 17 November 2017, available at https://www.youtube.com/watch?v=iFTWM7HV2UI.

[66] Matthew Shaer, 'What emotion goes viral the fastest?', *Smithsonian Magazine*, April 2014, available at https://www.smithsonianmag.com/science-nature/what-emotion-goes-viral-fastest-180950182/.

[67] HRC, A/74/486.s, *supra* note 45, para 54.

[68] Eric Killelea, 'Is Facebook even equipped to regulate hate speech and fake news?' *RollingStone*, 10 August 2018, available at https://www.rollingstone.com/culture/culture-features/facebook-zuckerberg-regulate-hate-speech-fake-news-710009/; Facebook, 'How is Facebook addressing false information through independent fact-checkers?', *Facebook Help Centre,* 2021, available at https://www.facebook.com/help/1952307158131536; Facebook, 'Fact-checking on Facebook', *Facebook Business Centre,* 2021, available at https://www.facebook.com/business/help/2593586717571940.

[69] Killelea, 'Is Facebook even equipped to regulate hate speech and fake news?'.

visible.[70] According to Facebook, this measure "significantly reduces the number of people who see [the false stories]."[71]

iv. The introduction of pilot paid, ad-free version of their platforms, where users have even greater control over their feeds.[72] Alternatively, rather than only having the option of entirely disabling personalised ads (as is the case on Google, for instance), the FCDO should work with social media corporations to allow users to have greater control over the extent of their personalisation.[73] Users may then select which sources of data (location, websites visited, purchases, age, gender, and so on) social media corporations can utilise for advertising purposes.

v. Incorporating greater transparency for platform users regarding targeted advertisements that appear on their feeds.[74] Enabling a function of 'Why am I seeing this?' for each specific advertisement would help foster a deeper understanding of how users' data is being used for advertisement purposes. Google presently has a function which allows users – if the user clicks on the link – to determine 'Why this ad?'; however, this fails in its specificity, as it only reveals broad information such as 'your age group', 'your gender', and 'websites that you've visited'.[75] The FCDO should encourage social media platforms to reveal to its users *precisely* why that user is being targeted with a certain advertisement.

vi. Insisting, as per the recommendation made by the 2019 UK House of Commons Digital, Culture, Media and Sport Committee's final report on "Disinformation and 'fake news'" that social media companies are not 'platforms', thereby bypassing responsibility for content on their sites.[76]

vii. Encouraging other social media corporations to follow Twitter's initiative regarding hate speech online: launched in May 2021, Twitter now prompts users to reconsider their Tweet if it is found to include 'offensive', 'insulting',

---

[70] Ibid. Also in 2018, Facebook claimed that their algorithms now prioritise content posted by friends and family rather than by media outlets and businesses in an effort to curb exposure to viral misinformation. See Julia Carrie Wong, 'Facebook overhauls News Feed in favour of 'meaningful social interactions'', *The Guardian,* 12 January 2018, available at https://www.theguardian.com/technology/2018/jan/11/facebook-news-feed-algorithm-overhaul-mark-zuckerberg.

[71] Facebook, 'How is Facebook addressing false information through independent fact-checkers?', 2021.

[72] Omar Zahran, 'The case for an ad-free social media subscription', *Medium*, 9 December 2020, available at https://omarzahran.medium.com/the-case-for-an-ad-free-social-media-subscription-c921eeeaaf7a.

[73] Google, "Control the ads you see," *Google Account Help,* available at https://support.google.com/accounts/answer/2662856.

[74] UK House of Commons: Digital, Culture, Media and Sport Committee, "Disinformation and 'fake news': Final Report", House of Commons, 4 February 2019, available at https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf, at 5.

[75] Google, "Why you're seeing an ad," *Google Ads Help,* 2021, available at https://support.google.com/ads/answer/1634057?hl=en.

[76] UK House of Commons: Digital, Culture, Media and Sport Committee, "Disinformation and 'fake news': Final Report", House of Commons, 4 February 2019, available at https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf, at 10.

and 'hateful remarks'. The feature is designed to detect such strong language, and prior to posting, Twitter prompts the user: "Want to review this before Tweeting"?[77] According to Twitter, 34% of users either refrained from posting the initial tweet or revised its content.[78] Other social media corporations could adopt a similar approach in order to encourage users from not posting hate speech online.

viii. Whenever feasible, limiting the number of platform users per IP address, whilst ensuring user pseudonymity,[79] to curb harassment and prevent the spread of violent content and disinformation through chatbots and botnets.[80]

c) Collaborate with small and big social media companies to conceptualise and disseminate user awareness-raising or digital literacy campaigns[81] in the UK and abroad. These should seek to educate users about human rights-compliant community standards and promote responsible user behaviour on platforms. Platform membership could be conditional upon attendance of short online courses and questionnaires on basic community standards. Whilst users are prompted to 'agree to' terms and conditions pertaining to community standards, many of terms of service of social media companies are so lengthy that they are indigestible and thus un-read by most users.[82] If a member has consistent evidence of breaching such community standards, that member may be required to revise the online short course before accessing account once again, and be placed on a 'probationary' period.[83]

d) Develop more direct forms of outreach for dissemination of online education regarding respect for international human rights law. This may also assist in 'inoculating' users against hate speech by raising awareness of such risks and thus

---

[77] Anita Butler and Alberto Parrella, 'Tweeting with Consideration', *Twitter,* 5 May 2021, available at https://blog.twitter.com/en_us/topics/product/2021/tweeting-with-consideration.html; Brian Niemietz, 'Twitter prompt warns users to rethink tweeting "insults" and "hateful remarks", *New York Daily News,* 5 May 2021, available at https://www.nydailynews.com/news/national/ny-twitter-insult-tweets-hate-speech-20210505-juup6qkghfbwjfzcd23ag2pfre-story.html.

[78] Butler and Parrella, 'Tweeting with Consideration', 2021.

[79] Sandra Endrez, 'Identity, Pseudonymity, and Social Media Networks', 2018, available at http://networkconference.netstudies.org/2018Bentley/2018/05/04/identity-pseudonymity-and-social-media-networks/

[80] Access Now, *supra* note 14, at 16.

[81] On the importance of education in addressing the root causes of hate speech, see HRC, A/74/486.s, *supra* note 45, para 55; United Nations Secretary-General (UNSG), 'United Nations Strategy and Plan of Action on Hate Speech', May 2019, available at https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf, at 4.

[82] Uri Benoliel and Shmuel I. Becher, 'The Duty to Read the Unreadable', *Boston College Law Review* 60 (2019), at 2255; David Berreby, 'Click to agree with what? No one reads terms of service, studies confirm, *The Guardian*, 4 March 2017, available at https://www.theguardian.com/technology/2017/mar/03/terms-of-service-online-contracts-fine-print; Dustin Patar, 'Most Online "Terms of Service" Are Incomprehensible to Adults, Study Finds', *Vice,* 13 February 2019, available at https://www.vice.com/en/article/xwbg7j/online-contract-terms-of-service-are-incomprehensible-to-adults-study-finds.

[83] This is similar to the suspension of Donald Trump's Facebook accounts, as he has not been entirely banned from the website – although the Oversight Board has recently required Facebook to conclude whether he should be permanently banned from the website). See Elizabeth Culliford, 'Facebook has six months to determine if Trump returns', *Reuters,* 6 May 2021, available at https://www.reuters.com/world/us/facebook-oversight-board-rule-trumps-return-facebook-2021-05-05/.

rendering hate speech less persuasive.[84] In order to ensure such initiatives resonate with target audiences, these educational campaigns ought to be designed in collaboration with local communities, civil rights groups, as well as the small and big social media corporations that would be required to allow their platforms to be used for this purpose.

e) Support social media companies in challenging government demands to remove, limit or publish content that is inconsistent with international human rights law.[85] This could be done through public messages of support as well as diplomatic engagement with the challenged governments.

4. How can the FCDO use its alliances to shape the development of, and promote compliance with, international rules and regulations relating to new and emerging technologies? Is the UK taking sufficient advantage of the G7 Presidency to achieve this?

Direct engagement with companies can be an effective way to promote responsible behaviour in the ICT environment, especially considering that the core of the Internet and most ICT infrastructures are owned or controlled by private entities. However, it is ultimately States that make, interpret and apply international and domestic law. Whilst only international law provides a truly global legal framework applicable to cyber threats worldwide, domestic law, adjudication and enforcement remain essential to give it teeth. Importantly, the ICT environment does not follow territorial boundaries, which means that cyber vulnerabilities in one country can quickly become global threats. Countering those threats in line with the rule of international law requires *all* States to cooperate in the clarification, dissemination and enforcement of international rules applicable to ICTs. The FCDO could play a leading role in this regard by:

a) Cooperating not only with traditional, like-minded allies, such as the European Union, the United States, Canada, New Zealand and Australia but also with developing countries and long-time cyber competitors, including China and Russia. In particular, China is home to some of the biggest tech companies, such as Tencent, Baidu and Huawei, and its manufacturing power is a key component of most IT supply chains. Thus, all efforts to counter global cyber threats will remain ineffective until agreement on key international rules and enforcement arrangements is reached with those States. To find common legal ground, cooperation should start with low-hanging fruits, i.e. discrete issues on which international agreement could be more easily reached, such as cyber operations targeting critical infrastructure, such as the healthcare sector, voting systems, energy, water and food distribution systems, and the core of the Internet. [86]

---

[84] Susan Benesch, 'Countering Dangerous Speech: New Ideas for Genocide Prevention', *United States Holocaust Memorial Museum,* February 2014, available at https://www.ushmm.org/m/pdfs/20140212-benesch-countering-dangerous-speech.pdf, at 13-14; Jonathan Leader Maynard and Susan Benesch, 'Dangerous Speech and Dangerous Ideology: An Integrated Model for Monitoring and Preventing', *Genocide Studies and Prevention,* 9(3): 87; Dara Barlin, 'Can technology prevent genocide? A case for virtual fear-inoculation', *World Policy,* 5 April 2013, available at https://worldpolicy.org/2013/04/05/can-technology-prevent-genocide-a-case-for-virtual-fear-inoculation/.

[85] On State obligations in this regards, see HRC, A/HRC/38/35, *supra* note 45, paras 6-8, 13-21.

[86] See, e.g., On this, see Oxford Institute for Ethics Law and Armed Conflict (ELAC), 'The Oxford Statement on the International Law Protections Against Cyber Operations Targeting the Health Care Sector', 21 May 2020 available at https://elac.web.ox.ac.uk/the-oxford-statement-on-the-international-law-protections-against-cyber-

b) Working with the Government Communications Headquarters (GCHQ) and its National Cyber Security Centre to develop the UK's national position on the application of international law to ICTs. At present, the UK's views on how international law applies to ICTs are found in a 2018 speech by former Attorney General Jeremy Wright QC.[87] A new official, consolidated document laying out the UK's national position on the topic would be an opportunity to review, clarify and further elaborate on how existing international legal obligations apply to ICTs. In particular, the UK should consider revising its position on how the sovereignty applies to ICTs, as well as clearly articulating how existing duties to prevent and redress harm, also known as 'due diligence' obligations, apply to the cyber context. The UK is an outlier when it comes to sovereignty,[88] given its reluctance to acknowledge that this well-established State right can be breached by cyber operations that cause physical or functional effects on a State's territory or which undermine its inherently governmental functions.[89] Likewise, it will lag behind a growing number of States, such as Germany, France,[90] the Netherlands,[91] Finland,[92] Estonia,[93] the Czech Republic,[94] Chile, Ecuador, Guatemala, Guyana and Peru[95] – to name just a few –, until it explicitly agrees that well-recognised duties of due diligence apply to States' use of ICTs. Most prominently among these duties is the

---

operations-targeting-the-hea; ELAC, 'The Second Oxford Statement on International Law Protections of the Healthcare Sector During Covid-19: Safeguarding Vaccine Research', 7 August 2020, available at https://elac.web.ox.ac.uk/article/the-second-oxford-statement#/; ELAC, 'The Oxford Statement on International Law Protections Against Foreign Electoral Interference Through Digital Means', 27 October 2020, available at https://elac.web.ox.ac.uk/the-oxford-statement-on-international-law-protections-against-foreign-electoral-interference-through#/. See also ELAC, 'The Oxford Process on International Law Protections in Cyberspace, 2021, available at https://elac.web.ox.ac.uk/the-oxford-process-on-international-law-protections-in-cyberspace#/.

[87] Cyber and International Law in the 21st Century, Speech by United Kingdom Attorney General Jeremy Wright QC MP', 23 May 2018, available at https://www.gov.uk/government/speeches/cyber-and-international-law-in-the-21st-century.

[88] Ibid, at 5.

[89] Michael Schmitt (ed.), *Tallinn Manual 2.0* on the International Law Applicable to Cyber Operations (Cambridge University Press, 2017), at 11-26, Rules 1-4 and commentary.

[90] France, Ministry of Defence, '*Droit International Appliqué Aux Opérations Dans Le Cyberespace'*, 2019, available at https://www.defense.gouv.fr/content/download/565895/9750877/file/Droit+internat+appliqu%C3%A9+aux+op%C3%A9rations+Cyberespace.pdf, at 10.

[91] The Netherlands, 'Letter of 5 July 2019 from the Minister of Foreign Affairs to the President of the House of Representatives on the international legal order in cyberspace — Appendix: International law in cyberspace', 5 July 2019, available at https://www.government.nl/documents/parliamentary-documents/2019/09/26/letter-to-the-parliament-on-the-international-legal-order-in-cyberspace, at 4-5.

[92] 'International law and cyberspace: Finland's national positions', 15 October 2020 ('Finland's Position'), available at https://um.fi/documents/35732/0/Cyber+and+international+law%3B+Finland%27s+views.pdf/41404cbb-d300-a3b9-92e4-a7d675d5d585?t=1602758856859, at 4.

[93] Estonia, 'President of the Republic at the opening of CyCon 2019', 29 May 2019, available at https://www.president.ee/en/official-duties/speeches/15241-president-of-the-republic-at-the-opening-of-cycon-2019/index.html.

[94] 'Comments submitted by the Czech Republic in reaction to the initial "pre-draft" report of the Open-Ended Working Group on developments in the field of information and telecommunications in the context of international security', 2020, at 2-3.

[95] Organization of American States (OAS), 'Improving Transparency — International Law and State Cyber Operations: Fourth Report (Presented by Prof. Duncan B. Hollis)', OEA/Ser.Q, CJI/doc. 603/20 rev.1, 5 March 2020, para. 58. See also paras 56ff.

rule that the UK itself successfully relied on in its case against Albania before the International Court of Justice, i.e. each State's obligation not to knowingly allow their territory to be used for acts contrary to the rights of other States. [96]

c) Complementing the UK's revised position on how international law applies to ICTs with a detailed roadmap of concrete measures for their implementation, including legal, technical, institutional, capacity-building and cooperative measures. Following the example of Australia, this roadmap could benefit from the input of submissions from academia, civil society and the industry.[97] The FCDO could also partner with other States and leading international institutions doing research on the topic, such as the United Nations Institute for Disarmament Research (UNDIR) and the International Telecommunications Union (ITU), to further investigate which measures are appropriate and necessary to give effect to its international obligations in the ICT environment.

d) Leveraging the UK's G7 presidency to engage with other groups of States, such as G20, clarify how exactly international law applies to current global cyber threats and agree on the necessary implementation measures. As mentioned earlier, the recent G7 Ministerial Declaration on, *inter alia*, Internet safety principles, trust in data free flows and digital technical standards is a positive step in this regard. However, more could be done to link these directly with the existing international legal framework and current efforts seeking to clarify it, such as the invaluable work of the UN Open-ended working group on developments in the field of information and telecommunications in the context of international security[98] and the UN Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security. [99]

e) Studying and formulating an official UK position on how international law applies to AI technologies, focussing, in particular, on the impact of its various applications on internationally recognised human rights.[100]

---

[96] *Corfu Channel Case* (*United Kingdom v Albania*), Judgment, 9 April 1949, ICJ Reports (1949) 4, at 22.

[97] Australian Government, 'Public Consultation: responsible state behaviour in cyberspace in the context of international security - Summary of public submissions on developing best practice guidance on the implementation of the 11 norms of responsible state behaviour in cyberspace articulated in the 2015 GGE Report (A/70/174), as endorsed by the UN General Assembly (A/RES/70/237)', June 2020, available at https://www.dfat.gov.au/sites/default/files/compilation-norm-implementation-guidance.pdf.

[98] OEWG, Final Substantive Report, UN Doc A/AC.290/2021/CRP.2, 10 March 2021.

[99] Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security (GGE), UN Doc. A/70/174, 22 July 2015 ('UN GGE Report 2015'); Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, UN Doc A/68/98, 24 June 2013 ('UN GGE Report 2013'.

[100] See Access Now, *supra* note 14.

May 2021