GP - ORF Series

Digital Debates

CyFy Journal 2022

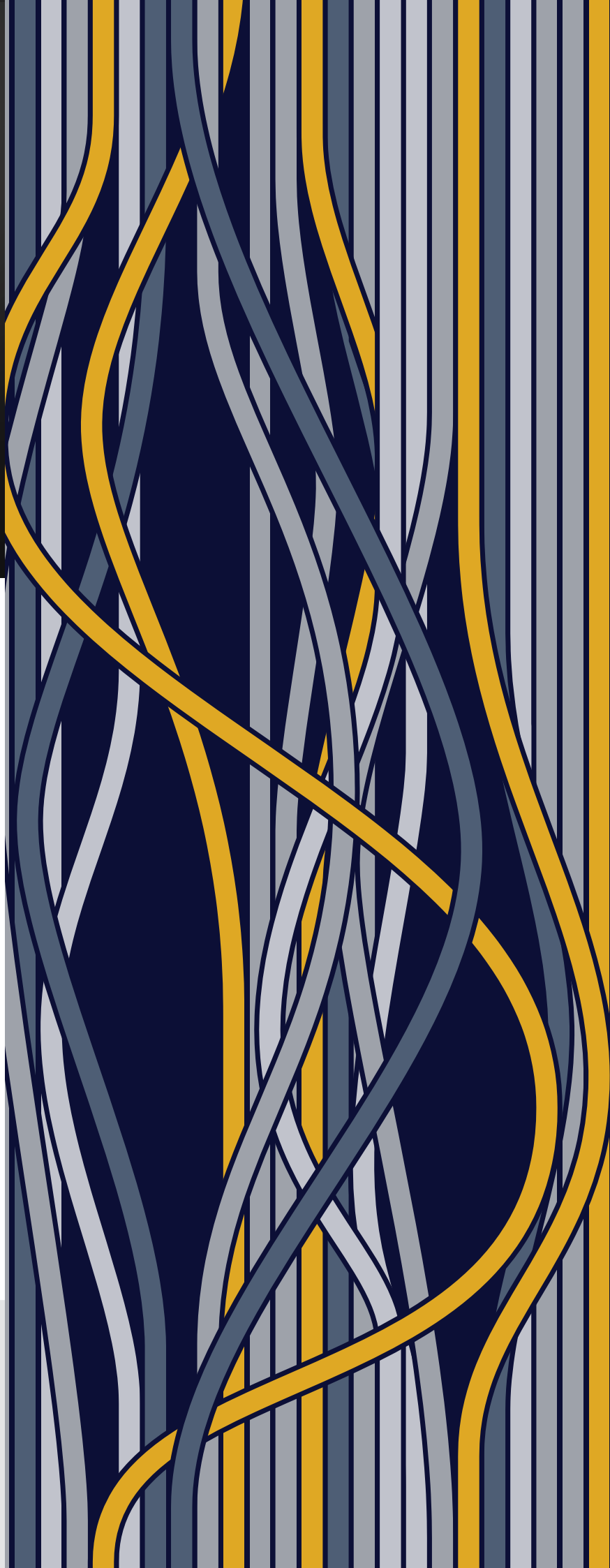
Edited by

Trisha Ray
Rajeswari Pillai Rajagopalan
Pulkit Mohan

CyFy 2022
technology • security • society

Durham University

WILEY-BLACKWELL



Digital Debates

CyFy Journal Volume 09 (2022)

Observer Research Foundation

20 Rouse Avenue, Institutional Area
New Delhi, India 110002
contactus@orfonline.org
www.orfonline.org

ORF provides non-partisan, independent analyses on matters of security, strategy, economy, development, energy and global governance to diverse decision-makers including governments, business communities, academia and civil society. ORF's mandate is to conduct in-depth research, provide inclusive platforms, and invest in tomorrow's thought leaders today.

Design and Layout: Artlab, Chennai
Cover Image Source: J614/Getty/Royalty-free
ISBN: 978-93-90494-72-9
Citation: Trisha Ray, et al., Eds., *Digital Debates: CyFy Journal 2022* (New Delhi: ORF and Global Policy Journal, 2022).

Contents

Editors' Note

The word of the year for 2021 was "vaccine", and it was hardly surprising. The development of the COVID-19 vaccine happened in record time, demonstrating sheer human will and ingenuity. At the same time, as the wealthier nations hoarded supplies and poorer countries were left adrift, we were reminded of the importance of self-reliance, even as the world has become more connected than ever in history.

Technology competition and rivalry is at the heart of this flux, whether in the Indo-Pacific, the Atlantic, or beyond. There are a number of factors contributing to it, and growing risks and challenges from emerging and critical technologies are overlaid upon the inherent complexities present in these regions. Vikram Sharma's essay for this edition of *Digital Debates* highlights one such paradigm-breaking development, that of quantum technologies, and its implications for cybersecurity given the advances in scale, speed, and processing power. Nicolo Andreula and Stefania Petruzzelli, in their piece, cite the many reasons we should be worried about the metaverse, where power can transcend the limits of space and perception. Michael Depp's essay, for its part, interrogates the potential of AI to disrupt the international balance of power.

The US-China tech rivalry is another important factor contributing to the uncertainty. China's efforts to overtake the US to become the global technology leader has prompted successive US administrations to enact measures to ensure their country's dominance in these domains. However, China will not be able to gain proficiency in many of the critical and emerging technologies without sourcing components from the US and its allies. As the US National Strategy for Critical and Emerging Technologies (2020) states, it is Beijing's "targeting sources of United States and allied strength by employing means that include stealing technology, coercing companies to disclose intellectual property, undercutting free and fair markets, failing to provide reciprocal access in research and development (R&D) projects, and promoting authoritarian practices that run counter to democratic values" that have made the US come up with more stringent measures to counter China and Russia.[1]

As the world is broken into camps, countries are turning to likeminded partnerships. Tobias Scholz writes about the resurgent Quad—the "democratic diamond" comprising Australia, India, Japan, and the United States—unified by their view of China as a geopolitical, if not existential, threat to their own states. This grouping, perhaps unlikely given their vastly different political systems, histories, and peoples, has a real opportunity to "shape norms, standards, and institutional mechanisms as well as strategic imperatives for the digital Indo-Pacific in the decades to come."

Also under the scanner are China's efforts at deploying these new and critical technologies for military use in order to make the People's Liberation Army (PLA) a more effective fighting force, as well as Russia's use of legitimate and illegitimate ways to target the US technologies. The more conventional security arenas such as nuclear and space security are also witnessing fierce competition and rivalry, the most evident of which is the new momentum to develop a range of counterspace capabilities. As space security dynamics intensifies and outer space becomes an extension of terrestrial geopolitics, outer space politics is becoming more complicated in various ways.

For one, several countries other than the US, China and Russia are developing counterspace capabilities, with many more establishing dedicated military space institutions, albeit some merely for coordination

purposes. China's PLA Strategic Support Force, for example—which controls space, cyber and electronic warfare—is a much more potent force than India's Defence Space Agency. Unless we are able to earnestly kickstart multilateral discussions, gradually leading to binding global instruments, it is unlikely that we will have uninterrupted, safe and secure access to space. One effort underway is the UN-mandated Open-Ended Working Group (OEWG) tasked to develop norms of behaviour that will address current and emerging space security threats. In her essay, Laetitia Zarkan gives an account of the two sessions of the OEWG held so far, giving the reader a sense of how the discussions are progressing and the likelihood of a consensus at the end of the group's four sessions. For his part, Daniel Porras, in his chapter addresses one counterspace capability that is inherently destabilising—i.e. ASAT weapons. He makes a case for halting the current trend regarding these weapons. Given that the usable orbits in space are limited in nature and ASATs produce large amounts of debris, this is an important issue on which a consensus is overdue. This is of course a subject of the ongoing OEWG and there could be some temporary steps to limit these tests, before more binding measures are formalised.

The salience of nuclear weapons in national security strategies is also growing, which involves modernisation of weapons in both quantitative and qualitative terms. Ruhee Neog in her essay looks at the nuclear dynamics in South Asia, where both politics and technology are important in possibly altering the calculations of India, Pakistan, and China. While nuclear weapons are seen as political tools by these three countries, the relevant query is how they interface with emerging and critical technologies to produce more risks—and this is a focus of Ruhee's essay. Tanvi Kulkarni then takes a close look at the disruptive aspects of emerging and critical technologies and why they tend to threaten strategic stability. Given the destabilising and risk-inducing nature of these technologies, this essay makes a case for strengthened dialogue processes and risk reduction measures as important steps moving forward.

It is evident, however, that the relevance and impact of emerging and critical technologies cannot be seen from a security prism alone. The challenge is in terms of finding the right balance between creativity and intellectual property on one side, and a global system that provides for open and mutual exchange of ideas, capital, and services. This tension between the traditional, open, globalised system and increased protective measures will continue until a certain degree of balance in power dynamics is achieved.

What is even more problematic is that the efforts at developing global rules-of-the-road have fallen victim to great-power contestations, resulting in deadlocks around global governance. Furthermore, with high-level norms being interpreted in practice in divergent ways, Husan Chahal asks, for instance, in relation to AI: "If the ultimate goal is the ethical development and deployment of AI, are efforts towards codifying and devising high-level ethical AI principles even a fruitful exercise?" Additionally, as the contributors to this volume find, some of the older critical technologies such as nuclear and outer space have been governed by rules that were framed in the 1960s and 1970s, and they are clearly showing signs of age. Recent efforts to frame new rules, and political and legal instruments, have not borne results.

Technology has brought in enormous benefits and transformed human lives—from efficient transport systems to gaining greater and safer access to medical services, education, and food as well as digitalisation technologies that have accelerated productivity and competence in many sectors. Sachin Malhan, Smita Gupta and Saurabh Karn, in their piece, see promise in the use of AI to improve access to justice, strengthening the capacity of courts and enabling the more efficient resolution of cases.

With data-intensive technology touching every aspect of our lives, Basu Chandola's essay highlights India's experience developing a data governance framework, citing the country's attempt to strike a balance between data protection—on the grounds of economic growth—on one hand, and data sharing and empowerment, on the other. Viktoria Nagy's chapter brings out another point of tension. She examines technology-based interventions in humanitarian assistance, and how the desire for "efficiency" has contributed to the wholesale exclusion and surveillance of refugees and migrants.

And with this, *Digital Debates* comes full circle to power. The way technologies are used is an expression of power, as Viktoria Nagy's critique so clearly demonstrates. The ability to set the terms by which nations are able to engage with rules, and the choices they make about partners, are a form of power. Indeed, the very act of developing technologies is a manifestation of power as well. Gabriella Skoff and Stuart Rollo's philosophical reflection on quantum includes this stark statement: "It is a society which forms great concentrations of material and technical wealth and concentrates political power in such a way as to direct the use of this wealth towards speculative technological ends. It is clear that under current conditions the power of quantum computing will be far more centralised and controlled than is the computing of today."

As it is with CyFy 2022, the conference we host parallel to the release of this edition of *Digital Debates,* these 13 essays capture the milieu of anxieties, hopes, and questions about our relationship with technology—as individuals, nations, and a global community.

- Trisha Ray, Rajeswari Pillai Rajagopalan, Pulkit Mohan

[1] The White House, National Strategy for Critical and Emerging Technologies, October 2020, https://trumpwhitehouse.archives.gov/wp-content/uploads/2020/10/National-Strategy-for-CET.pdf

Futures

Politics in the Metaverse vs. Politics of the Metaverse

Nicolo Andreula and Stefania Petruzzelli

POLITICS IS DEFINED as the science and art of governing which has as its raison d'être not only the constitution and organisation of a state, but also the administration of public interactions amongst people and their environment. In essence, a coordination of reality—according to direct and indirect practices, legitimate or otherwise—which always presupposes the management of human behaviour and societal interactions.[1]

The physical world has been the only place so far to exercise power, exert political competition, and spread ideologies for centuries. In the past decade, however, we have lived through a significant reorientation of where the drivers of our governance and actions derive from and most often take place in the digital world.

In this context, social media have overhauled earlier norms concerning the dissemination of ideas, the reception of information, and the perception of communities and political affiliations. This phenomenon has produced undoubted advantages on the democratisation of thought, but has also had serious consequences on the tendency towards polarisation, extremism, and the propagation of conspiracy theories.

It is therefore legitimate to ask questions about the fate of politics in the realm of another emerging paradigm: the Metaverse. What will happen in the 'meta-universe' where anyone and anything—even power—can transcend the limits of space, perception, freedom, and human and social barriers? Where our movements can be monitored not only based on our purchasing choices and interactions but, even more importantly, calibrated on corporality? Where 'surveillance capitalism'[2]—already a pervasive political and commercial practice—could amplify, further reducing our free will?

The scenarios, hypotheses, and possible solutions we will illustrate in this reflection may start from varied questions but they are common throughout history. Human

behaviour since time immemorial has taught us that the ability to manage and use new technologies is one of the determining factors of any society. Trying, therefore, to understand the implications, in this field, of the current transformation underway—a revolution that is not only digital but existential and epistemological—is an urgent imperative.

Down with the Influencers, Long Live the Influencers!

ANYONE CAN SEE the enormous impact of influencer marketing in the advertising industry. Unsurprisingly, the stay-at-home era of the COVID-19 pandemic further empowered virtual marketers—those birthed entirely in the digital world—for the second year straight, according to HypeAuditor. These marketers registered decisively higher engagement and influencing rates than their offline competitors.[3]

The reasons behind the digital advertisers' overlap with the physical are easily discernible. The marketer avatars are the end result of accurate market surveys, inevitably intercepting consumer needs more effectively, simulating—using a virtual form of 'empathy'—a real entity perceived by the user as having no computer interface. By programming the humanoid assets *a priori,* but modifying and monitoring them continuously, the creators—mostly anonymous—exercise direct power over this new and peculiar communication channel, free of all the problems and contradictions typical of the human.

The advantage of using virtual influencers grows even further when one wants to advertise not just material products but ideas. Trivially, because of the lower exposure to the risk of any fickleness, discomfort, or fear, anonymity creates an endless possibility of expression, often beyond the limits of freedom of expression, which a physical influencer is hardly willing to exhibit. One just has to think about the removed limitations on propriety and politeness. As avatars, through their anonymity, trendsetters can unleash visceral marketing, which does have its appeal to many.

What can stop them? Legal constraints are absent, for the time being. Avatars are not obligated to abide by any code of ethics that, in several countries, require the declaration of remuneration on the dissemination of content by marketers. In the future, a simple approach by policymakers could be to extend current regimes, in force for the marketing industry, to the digital space, as we have seen already happening in the United States (US).[4] Apart from the obvious advantages of saving money, the omission of remuneration will likely create a greater emotional bond between the consumer and the avatar and, by extension, with the product it advertises.

One can immediately understand that due to the high possibility of profiling untraceable behavioural data on the web, creators can build avatars capable of influencing and determining consumer action with precise techniques of cognitive manipulation. Here we are not talking about science fiction, but neuroscience, what in the industry is called 'character merchandising'.[5] It is more than plausible that political parties might soon decide to personify their brand with a virtual influencer, just like fashion houses.

The use of this kind of influencer marketing in politics could raise concerns, especially considering that these para-social relationships based on anonymity but displayed in a hyper-realistic form, can induce a higher release of oxytocin—our brains' chemical key to trust. This in turn can strengthen propagandist messaging and mystify extremism, with the outcome of spreading distrust in democratic societies.

Crypto-Contributions Unlocked

CRYPTOCURRENCIES HAVE BEEN adopted in national politics, with major US parties implementing the technology, especially in support of their candidates. A few examples include the following:

- On 25 February 2022, a bill was introduced in the state of Louisiana allowing parties to receive donations in cryptocurrencies in support of their election campaigns, provided they are converted and declared in fiat.[6]

- For the 2022 mid-term elections in the US, a new fundraising platform exclusively in cryptocurrencies, Engage Raise, has been established, which will allow political campaigns to manage donations.

Both developments could greatly transform election governance, but only if supported by precise strategies and policies, some of which are already in the pipeline. Several countries have begun systematically regulating digital assets and tackling the regulatory void caused by anonymity. For example, in the US, although digital currencies allow anonymous transactions, platforms such as Engage Raise bind donors to declare themselves to the Federal Election Commission.[7]

Yet, not all countries view such technical measures as sufficient to stem the dangers of corruption and cyber vulnerabilities. Ireland, for example, understandably concerned over the war between Russia and Ukraine, has prevented political parties from receiving donations in cryptocurrencies to curtail foreign entities' reach and influence in the country's democratic processes.[8]

Another way of crypto-financing political parties directly linked to the Metaverse could be with NFTs—non-fungible tokens, or cryptographic assets representing real-world items—that offer the buyer not only the opportunity to contribute to the donation, but also to obtain a series of benefits which consolidate a sense of belonging to the cause they are devoting to.

A pioneer in this regard was the Democratic Party of South Korea (DPK). In January 2022, on the occasion of the presidential elections, DPK announced the issuance of its own token to raise funds to support the election campaign and, at the same time, attract supporters by distributing extra content—mainly pictures and videos related to the political programme—linked to NFTs.[9] By going down this route, parties could associate NFTs with other functional created content and consolidate communities around their values. For example, through VR (virtual reality) technology, voters could immerse themselves in simulations of political debates or election rallies.

It is not difficult to imagine the political evolution of this scenario: i.e., the creation of new parties represented by a token that reflects the values of a community stimulating its growth. Consequently, also foreseeable is a Decentralised Autonomous Organisation (DAO)— essentially blockchain-fuelled organisations [10] – based on tokens, and thus reflecting the nature of activists.

In the best of worlds, politics could use these tools to engage citizens and enhance the resonance of its best ideals. Perhaps this can be achieved by regulating the DAOs, or by fostering greater transparency in terms of funding and the dissemination of ideas, with the establishment of digital citizenship, which would undoubtedly convey wider traceability. However, in the worst of worlds, such a transformation could cause a democratic degeneration due to a decentralisation which makes accountability impossible and coding errors all too frequent, due to an objective digital divide which remains an urgent need to be addressed.

With Great Power Comes Great Responsibility

THE INSTITUTION MOST at risk with the Metaverse is democracy. We know that the main aim of all kinds of marketers is to conquer new spaces to reach a wider audience and position their message wherever the recipients are. There is no doubt, therefore, that the place to be in the coming decades, where to advertise goods and opinions, is and will likely be the Metaverse.

In Italy, amidst the hype of their 2022 political election cycle, the Metaword[11] group developed a digital service that provides virtual environments in the Metaverse for the election campaign. Up to this point, nothing is worrisome; in fact, much of it is praiseworthy.

Yet, any tool can be weaponised.

A series of experiments by a group of researchers at Stanford University[12] found that by modifying the characteristics of an unknown individual, a politician, for example, to make him look like a voter—in a process akin to deepfaking, whereby viewers' emotions are triggered—the audience were led to rate the subject more favourably. Consequently, the assumption that political parties in the Metaverse could make use of a manipulation which has no parallel in reality raises fears and unforeseen complexities.

A Democratic Universe

SOCIAL MEDIA SHOWS high potential in promoting emotionally intense content to engage a specific target audience, personalising a message, and deploying appeal mechanisms based on social identity and categorisation. The challenge to democracy is now well established if one considers, that based on recent precedents, fringe ideas (including extreme political views) have benefited precisely from these digital amplifiers. On the one hand, individuals generally tend to not follow the same social norms in the virtual dimension as they do in the physical, developing more aggressively oriented attitudes online.[13] On the other hand, digital behemoths such as Facebook and Twitter have made it clear that it is almost impossible in the digital realm to stem the spread of hatred, despite being aware of it.[14]

It is therefore legitimate to be alarmed about the entry of politics into the Metaverse, especially for the following reasons:

First, the possibility of influencing and altering user behaviour in a far more pervasive way. This is achieved by tracking body movements and detecting reactions to more stimuli than the web. For instance, some headsets already used to access augmented realities are equipped with sensors that monitor brain activity. VR does the same with eye movements.[15] By using far more accurate and sensitive data, propaganda campaigns—even those by government—could not only become more powerful and potentially more dangerous, but also create a stronger bias.

Second, the risk of polarisation is also based on social and economic status. The gradual evolution of the Metaverse will be matched by the need for users to purchase more technologies and be competent in using them. This is a barrier to entry for anyone without sufficient financial means and digital literacy.[16] If politics were to expand into this new universe, electing it as the main communication space, it could certainly attract new segments of the population (think Generation Z/Alpha), but it could more or less tactically exclude others.

Third, the security dangers are already evident. Think of the recorded radicalisation movements, such as the building of proto-fascist states; or the risk of terrorism, which is so real that it has been reported by the European Union Counter-Terrorism Agency and the National Counterterrorism Innovation, Technology and Education Centre (NCITE).[18]

Some avant-garde groups are using social media and cryptocurrencies to raise funds and taint the hearts and minds of their targeted audiences. So far, authorities have largely stayed abreast.[19] Through the Metaverse, terrorist cells could spread disinformation, incite hatred, and heighten anti-democratic attitudes.[20] For example, newly radicalised proselytes would be able to simulate attacks through the virtual reproduction of existing urban spaces via augmented reality and avatars.

These trends will only accelerate in the Metaverse—in a potentially deadly cat-and-mouse game where the aggregation of technologies is the plat du jour, adding ever more importance to safeguarding cybersecurity. With this in mind, a more significant distribution of cyber policing would certainly favour more efficient supervision, as well as pivoting on new techs—such as AI—to analyse behavioural patterns across networks. [21]

Conclusion: Counteracting a Heavy Immateriality

EVOLUTION SHOULD NOT be held back. Certainly, however, it must be examined, monitored, and regulated.

It is only through the analysis of the problems that solutions can be outlined and proposed. On them, scaffolding can be built to erect a strong edifice so that it does not collapse under the excessive weight of its immateriality. To this end, a few notable trends are already taking place: the emergence and consolidation of stronger data protection policy frameworks;[22] learning from others' mistakes can often provide easy stepping stones;[23] and finally, the increased need for identification and authentication by social media will likely be mirrored in further developments, such as with Meta's Metaverse.[24]

The Metaverse is still an organism in the making, and potential risks must be considered to prevent them, ensuring that its evolution does not take place to the detriment of people. This issue is urgent for politics and leaders because it concerns the exercise and freedom of thought.

Endnotes

1. Adrian Leftwich, "What is Politics?: The Activity and its Study," *Wiley*, September 2004, https://www.wiley.com/en-br/What+is+Politics%3F:+The+Activity+and+its+Study-p-9780745630557#content-section

2. Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (London: Public Affairs, 2019)

3. HypeAuditor, *State of Influencer Marketing 2021*, https://hypeauditor.com/s/resources/US_State_of_IM_2021.pdf

4. Shawn Maria Estrada et al, "What Fintech and Digital Marketing Companies Need to Know Now About the CFPB's Expanding Jurisdiction," *JDSupra*. August 29, 2022.

5. World Intellectual Property Organization, *Character Merchandising,* December 1994, https://www.wipo.int/export/sites/www/copyright/en/docs/wo_inf_108.pdf

6. Heather Morton, "Cryptocurrency 2022 Legislation," *NCSL,* 2022, https://www.ncsl.org/research/financial-services-and-commerce/cryptocurrency-2022-legislation.aspx

7. Bria Schwartz, "New Crypto Fundraising Start-up Will Take Political Donations in Digital Currencies as 2022 Midterms Heat Up," *CNBC,* June 16, 2022, https://www.cnbc.com/2022/06/16/new-crypto-fundraising-start-up-will-take-political-donations-in-digital-currencies-as-2022-midterms-heat-up.html

8. Morwenna Coniam, "Crypto Donations Face Ban in Ireland to Avert Russian Meddling," *Bloomberg,* April 19, 2022, https://www.bloomberg.com/news/articles/2022-04-19/crypto-donations-face-ban-in-ireland-to-avert-russian-meddling

9. Paek Jae-hyuk, "Ruling Party to issue NFTs for Fundraising in Presidential Election," *The Korea Times,* February 1, 2022, https://www.koreatimes.co.kr/www/biz/2022/01/488_321539.html

10. "Decentralized Autonomous Organizations: Beyond the Hype," World Economic Forum, June 2022, https://www3.weforum.org/docs/WEF_Decentralized_Autonomous_Organizations_Beyond_the_Hype_2022.pdf

11. Metaword, "Home," https://www.metaword.cloud

12. Jeremy N. Bailenson et al., "Facial Similarity between Voters and Candidates Causes Social Influence, Department of Communication," *Public Opinion Quarterly vol. 72*, pp 953-961 (2008), https://stanfordvr.com/mm/2008/bailenson-facial-similarity-.pdf

13. Tanya Basu, "The Metaverse Has a Groping Problem Already," *MIT Technology Review,* December 15, 2021, https://www.technologyreview.com/2021/12/16/1042516/the-metaverse-has-a-groping-problem

14. Emmanuel Akinwotu, "Facebook's Role in Myanmar and Ethiopia Under New Scrutiny," *The Guardian,* October 7, 2021, https://www.theguardian.com/technology/2021/oct/07/facebooks-role-in-myanmar-and-ethiopia-under-new-scrutiny

15. Adi Robertson, "MindMaze's Hand-tracking, Mind-reading Virtual Reality Headset Is Just as Complicated as It Sounds," *The Verge,* March 4, 2019, https://www.theverge.com/2015/3/3/8136405/mind-maze-mind-leap-thought-reading-virtual-reality-headset

16. "The Metaverse Needs to Keep an Eye on Privacy to Avoid Meta's Mistakes," *Cointelegraph.* April 23, 2022, https://cointelegraph.com/news/the-metaverse-needs-to-keep-an-eye-on-privacy-to-avoid-meta-s-mistakes

17. Cecila D'Anastasio, "How Roblox Became a Playground for Virtual Fascists," *Wired,* June 10, 2021, https://www.wired.com/story/roblox-online-games-irl-fascism-roman-empire/

18. Joel S. Elson et al., "The Metaverse Offers Much Potential for Terrorists and Extremists," *Defense One,* January 10, 2022, https://www.defenseone.com/ideas/2022/01/metaverse-offers-much-potential-terrorists-and-extremists/360503/

19. US Department of Justice, *Global Disruption of Three Terror Finance Cyber-Enabled Campaigns,* August 13, 2020, https://www.justice.gov/opa/pr/global-disruption-three-terror-finance-cyber-enabled-campaigns

20. Richard Lawler, "Second Life Joins the Metaverse Discussion with the Return of Its Founder – and Some Key Patents," *The Verge,* January 13, 2022, https://www.theverge.com/2022/1/13/22881864/metaverse-second-life-decentralized-moderation-patent-virtual-reality

21. Abeed Rashid, "Cybersecurity and the Metaverse: Patrolling the New Digital World," *IBM,* August 4, 2022, https://securityintelligence.com/posts/metaverse-cybersecurity-concerns/

22. European Commission, *Data Protection in the EU,* https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

23. Daniele Marinelli, "The Metaverse Needs to Keep an Eye on Privacy to Avoid Meta's Mistakes," *Cointelegraph.* April 23, 2022, https://cointelegraph.com/news/the-metaverse-needs-to-keep-an-eye-on-privacy-to-avoid-meta-s-mistakes

24. Andrea Vittorio, "Metaverse Technology Opens Up a Wider World of Privacy Concerns," *Bloomberg Law,* https://news.bloomberglaw.com/privacy-and-data-security/metaverse-technology-opens-up-a-wider-world-of-privacy-concerns

Quantum and the Cybersecurity Imperative

Vikram Sharma

THE DAWN OF the Fourth Industrial Revolution (4IR) is ushering in a dramatic convergence of technologies and synthesising the physical, digital, and biological worlds.[1] The 4IR is rapidly and fundamentally transforming our everyday experience, from enhancing how we connect and interact with each other to operating enterprises in more sustainable ways. It may even remedy the environmental damage of previous revolutions and allow us to rethink how we organise our societies.

In contrast to the first three industrial revolutions, each of which was enabled by single technologies—steam, electricity, and digital—the fourth is powered by a synthesis of significant scientific and technological innovations. It leverages developments in an array of technology domains, including artificial intelligence, 5G communications, the Internet of Things, and quantum, and promises step-change technological advances through 2050 and beyond. With these innovations comes a whole new wave of risks and vulnerabilities that in themselves offer advanced opportunities to innovate, whether for good actors or adversaries alike.

This article explores the important role of quantum as a defining technology of the 4IR and the consequent cybersecurity imperative given the advances in scale, speed, and sheer processing power that quantum computing will make possible. Establishing timely actions to safeguard, while we innovate, are critical to our global economy and to assure societal benefits of the second wave of quantum technology are captured to the fullest. Delays will have costly consequences and setbacks that we have not seen to date.

Emergence of Quantum Technology

THE FIRST WAVE of quantum technology, heralded by the invention of the transistor by Walter Brattain, John Bardeen, and William Shockley in 1947 (for which they would be awarded the Nobel Prize in Physics in 1956), has transformed our lives and societies over the past 75 years. The first wave of quantum technology harnessed

effects innate at the most fundamental scales of nature to deliver step-change technical capability advances. It drew on the foundations of quantum theory developed in the early 20th century by Niels Bohr, Max Planck, and Albert Einstein—all of whom were also awarded Nobel Prizes for Physics, in different years, for their seminal contributions to the field.

From lasers to GPS to MRI to cell phones, technologies of the first quantum wave are firmly entrenched in our everyday lives. Each day we almost unknowingly carry on our persons billions, if not trillions, of transistors courtesy of mobile devices, smartwatches, and other digital wearables. These technological advances have intrinsically reshaped our interpersonal relationships, how we conduct business, and indeed, geopolitics and military strategy.

The promise of the second phase of advancements in quantum is upon us, of technological advances and innovations as significant as those in the first wave. In contrast to the first, where we passively leveraged quantum effects that routinely exist in nature, in the second phase, we are actively engineering quantum states that do not naturally occur. The development of superalloys and materials offers an apt analogy—creating these alloys and materials, which do not exist in nature, has enabled us to build skyscrapers, advanced jet engines, and sophisticated pollution control equipment. In like manner, the ability to engineer novel quantum states opens exciting possibilities for extraordinary advances in computational capability, imaging, sensing, and secure communications.

Exponential Power of Quantum Computers

IT IS IMPORTANT to briefly explore the transformative impact of quantum on computing so as to better value what we are putting at risk if we do not safeguard sensitive information as we innovate. Traditional computers encode information in a sequence of zeros and ones, or 'bits'—each bit storing either a zero or a one at any point during a calculation. Quantum computers, in contrast, use 'qubits'. Each qubit can concurrently store a zero, a one, or anything in between, courtesy of a quantum property known as superposition. This property enables a sequence of qubits to encode much more information than an equivalent sequence of conventional bits, thereby allowing quantum computers to process certain classes of problems exponentially faster than traditional computers.[2]

A range of industries is braced for revolution, as previously intractable problems become solvable when tackled by quantum computers and algorithms. Use cases will include drug discovery, financial optimisation, artificial intelligence, and cryptanalysis.

Researchers in Canada published in July 2021 a helpful heuristic to identify and classify problems at which quantum computers would excel. They suggest that problems which rely on combinatorics would benefit dramatically from quantum computational capabilities.[3] Such problems typically require finding an optimal solution from a myriad possible permutations. The computational complexity to find a solution often grows exponentially for each new permutation value added.

These problems are frequently encountered in chemical and biological engineering, cryptanalysis, artificial intelligence, financial services, and complex manufacturing. Advanced companies operating in these sectors are already trialing early quantum algorithms on emergent quantum computers.[4]

Reflecting the tremendous potential and impact of this rapidly developing technology, Honeywell, for example, is bullish and expects the quantum computing industry to top US$1 trillion by 2050.[5] Meanwhile,

IBM and Google, amongst others, have publicly released roadmaps offering vignettes of their expected progress in quantum computing efforts.[6,7]

However, for all the remarkable benefits quantum computing will afford, it will imperil the very foundations of the cybersecurity protecting our digital ecosystem. As our lives are increasingly integrated with and lived in the digital realm, data has become our most precious resource and asset. Safeguarding valuable information has never been more critical.

From Heightened Cyber Risks to a Quantum-enabled Adversary

THE TIMING OF this pending transformation with quantum looms as our global digital economies face an exponentially-escalating cyber threat environment composed of capable and well-funded cybercriminals and malicious state and non-state actors with an aim to hack systems and steal valuable data at scale. The current asymmetry of costs in launching cyberattacks means large potential economic gains can be made following a successful breach.

As a result, the frequency and sophistication of hacking attempts have escalated dramatically in recent years, with some 4,145 data breaches reported globally in 2021, exposing some 22 billion records.[8] Unsurprisingly, cyber-attacks are among the critical long-term risks to global economic stability and social cohesion identified by the World Economic Forum in its *2022 Global Risks Report*.[9] The race to quantum technology for such adversaries can mean a strategic advantage and highly lucrative pay days for years to come.

Quantum Computing Threat to Current Cryptography

FOR ALL THE transformational benefits offered by quantum computers, they will also threaten the foundational tenets of trust in our digital platforms—challenging the verification of the authenticity of a virtual person or organisation, and the ability to exchange information with them securely.

Today's secure communications technologies, such as data encryption and identity authentication, rely on mathematical complexity to protect data exchanges. Frequently, they employ mathematical techniques that are easy to perform in one direction but hard to compute in the reverse. For example, if we wish to multiply two large prime numbers—say 165,181 and 417,953—we can compute the product in a couple of seconds on a basic calculator. However, performing the reverse operation, where you are given the product (referred to as a *semi-prime*)—69,037,894,493 in our example—and wish to find the numbers that were multiplied together to produce that result, is a difficult task.[10] Most contemporary secure communication protocols rely on this technique or similar mathematical constructs for their security.

As foreshadowed, and somewhat counterintuitively, quantum computers will compromise much of today's cryptography. As a case in point, Peter Shor in 1994 at Bell Labs developed an algorithm[11] that can leverage a quantum computer to rapidly find the factors of a large number, such as a semi-prime in the above example. This would break all encryption methods reliant on this mathematical technique for their security. Additional quantum algorithms, such as those developed by Lov Grover[12] and Daniel Simon,[13] will similarly impact other cryptographic technologies in widespread use today.

While quantum computers at the scale required to factorise the large semi-prime numbers used to secure today's e-commerce and sensitive communications are still some years away, rapid advances in science and engineering are driving a growing cyber risk. One of the technologies undergirding current global e-commerce systems, RSA Key Exchange, will be significantly impacted and likely broken.[14] How we conduct our online lives and share sensitive information would be disrupted, putting the viability of key industries and security of critical infrastructure at risk.

Q-Day and Harvest Now Decrypt Later Risk

THE TERM Q-DAY (also tagged as Y2Q by the Cloud Security Alliance[15]) has been coined to mark the date that quantum computers will compromise current cryptographic technologies. As significant advances are needed to deliver quantum computers at a cryptographically-relevant scale, we have an estimated five to 10 years to implement alternatives.

An annual poll of leading academics and industry experts conducted by EvolutionQ offers an informative heatmap of the likelihood of quantum cyber risk classified by five-year intervals.[16] It is instructive to consider the quantum cyber threat through the lens of risk to determine the timeframe within which cyber risk breaches (organisation-specific) risk tolerance thresholds. This offers essential guidance on the time available to prepare for a quantum-resilient cybersecurity posture—to understand the nature and breadth of quantum cyber risk, evaluate and pilot potential mitigation technologies, develop deployment roadmaps, implement solutions, and finally, integrate quantum-resilient technologies into applications, platforms, and shared infrastructure.

A further, perhaps more pressing, cyber risk looms. State actors, and potentially criminal adversaries, are actively eavesdropping on sensitive information being communicated today and storing them. The motivation is the expectation that a cryptographically-relevant quantum computer will become available within the utility period of the data collected. This would expose the data in unencrypted form, allowing it to be leveraged for malicious financial or strategic gain. Data with sensitivity periods beyond a few years are particularly at risk.

This type of *Harvest Now Decrypt Later* attack (or HNDL, pronounced 'Handle') has become a clear and present risk. Technology consulting firm Booz Allen suggests that state actors will have collected extensive datasets of economically valuable intellectual property, in encrypted form, during the course of this decade.17 Protagonists anticipate that quantum computing capability at the scale required to break the encryption will be realised within the useful lifetime of the data. Such a capability would be exploited to decrypt the data and monetise the previously harvested intellectual property—potentially to massive financial gain. This risk applies equally to sovereign secrets, certain financial data, and sensitive personal information.

The Cybersecurity Imperative

GIVEN THE CRITICALITY of the foregoing, we must develop timely and coherent risk management strategies to ensure trust is sustained in our ever-more digital lives. To secure sensitive data, we must establish roadmaps to transition to new quantum-resilient technology before Q-Day arrives. If we match the timelines on these roadmaps with our risk-weighted assessment of time available for the transition, we can effectively mitigate the threat of a quantum-enabled adversary.

Considerable progress has already been made in provisioning quantum-resilient measures. Key pillars of this defensive arsenal are post-quantum cryptographic algorithms, cryptographic agility, and quantum random number generation and key distribution.

The US Department of Commerce's National Institute of Standards and Technology (NIST) in 2015 launched a competition to find post-quantum cryptographic algorithms that are expected to resist both conventional and quantum attacks. In July 2022 they announced the first group of quantum-resistant algorithms that will become part of its Post-Quantum Cryptography (PQC) standards. The four recommended algorithms for encryption and digital signatures are CRYSTALS-Kyber, CRYSTALS-Dilithium, FALCON, and SHPINCS+. [18] The two CRYSTALS algorithms are expected to be widely deployed to replace encryption and digital signatures commonly used today, and can be implemented on current computers and cybersecurity infrastructure.

As novel encryption algorithms are released, it is vital that our digital platforms and IT systems can rapidly integrate them. Agility in the management of cryptographic credentials will allow this to occur as seamlessly as possible. Many contemporary systems have specific encryption methods hardwired—these will need to be upgraded to cryptographically agile solutions.[19] Further, it is unlikely that this upgrade will be a one-time exercise. Cryptographic agility will be an important enabler for an expected rolling series of updates to cryptographic algorithms over the foreseeable future.

While quantum computing presents a risk to current encryption methods, quantum technologies, in equal measure, offer solutions. These solutions come in the form of quantum random number generation (QRNG) and quantum key distribution (QKD). Random numbers are the foundational building blocks of cryptographic credentials. Today we mostly use software-generated pseudorandom numbers; QRNGs deliver true random numbers, thereby enabling high-quality encryption.[20] Similarly, the secure sharing of encryption keys between two parties wishing to securely exchange information is dependent, even with PQC, on complex mathematics coded into software. QKD turns this paradigm on its head by leveraging the laws of quantum physics, not mathematics, to assure secure key exchange.[21] Many see QKD as an important technology securing high-value communications links—complementing more broadly deployed PQC.[22]

A Quantum-Resilient Organisation Rises

RECOGNISING THE QUANTUM cybersecurity threat, policymakers, technology consulting firms, and research organisations are developing guidance and regulations to support government agencies and industries transitioning to a quantum-resilient cybersecurity posture. Earlier this year, the US president signed a National Security Memorandum setting out the administration's plan to mitigate the cybersecurity risk from quantum computers.[23]

The World Economic Forum, in its report *Transitioning to a Quantum-Secure Economy,*[24] and the US Quantum Economic Development Consortium,[25] alongside other respected organisations have published guidelines to help organisations understand and transition to quantum-resilient cybersecurity. Technology consulting and cybersecurity firms have also been actively publishing white papers on the topic.[26]

Many of the recommended strategies for achieving a quantum-resilient cybersecurity posture share the following elements:

- Map and classify information held by the organisation.
- Identify systems and platforms exposed to quantum risk.
- Conduct threat modelling to understand the consequences of compromise for each data category.
- Evaluate quantum-resilient solutions—PQC, crypto-agility, and QRNG/QKD.
- Develop a roadmap to transition to a quantum-resilient cybersecurity posture.
- Implement the roadmap and integrate new systems with existing infrastructure.

Deployments of new systems and processes must gain community trust as they are rolled out. Trust is crucial for maximising the adoption of new technologies and for society to reap the benefits. Achieving and sustaining such trust requires assurance that these new systems are secure enough to protect sensitive information, such as personal details, financial transactions, intellectual property, and national security secrets.[27] Well planned, timely action will ensure this is the case.

Building Trust in Our Digitalised Future

BUILDING ON DECADES of scientific breakthroughs in research laboratories worldwide, we are seeing emerging technologies of the second quantum wave find adoption at forward-looking organisations. Deployments, and pilot projects, are delivering step-change advances in capabilities in computation, sensing, imaging, and cybersecurity, to name a few of the early domains to benefit. Over the coming decades, these nascent technologies will shape our lives and geopolitics.

For all the dramatic innovations these technologies will offer, they present clear risks to the security of our digital ecosystem and, consequently, our trust in them. As set out in this essay, quantum computers at a cryptographically-relevant scale will break the algorithms that assure us of the security of our e-commerce platforms and the exchange of sensitive information. The good news is that we already possess measures to meet the challenge effectively.

The quantum cyber threat is likely to materialise within the lifecycles of many IT/OT systems being deployed today; the cyber risk, however, particularly for long-lived data, is clear and present. Further, the timeframe required to transition the current cybersecurity infrastructure to a quantum-resilient posture is likely to be material.

We must prioritise implementing appropriate policy and regulatory settings, gaining active support from enterprise leadership and cyber decision-makers, and developing quantum cyber readiness metrics. Collectively, these will drive timely action on safeguards to ensure that societal benefits of the second wave of quantum technology are captured to the fullest.

Endnotes

1. Klaus Schwab, "The Fourth Industrial Revolution: What It Means and How to Respond," *World Economic Forum,* January 14, 2016, https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/

2. IBM, "What Is Quantum Computing?," 2020, https://www.ibm.com/in-en/topics/quantum-computing

3. Francesco Bova, Avi Goldfarb, and Roger Melko, "Quantum Computing Is Coming. What Can It Do?," *Harvard Business Review,* July 16, 2021, https://hbr.org/2021/07/quantum-computing-is-coming-what-can-it-do

4. Francesco Bova, Avi Goldfarb, and Roger G. Melko, "Commercial Applications of Quantum Computing," *EPJ Quantum Technology volume 8,* 2 (2021), https://epjquantumtechnology.springeropen.com/articles/10.1140/epjqt/s40507-021-00091-1

5. Honeywell, "Honeywell Quantum Solutions and Cambridge Quantum Computing Will Combine to Form World's Largest, Most Advanced Quantum Business," 2021, https://www.honeywell.com/us/en/press/2021/06/honeywell-quantum-solutions-and-cambridge-quantum-computing-will-combine-to-form-worlds-largest-most-advanced-quantum-business

6. IBM, "Our New 2022 Development Roadmap," https://www.ibm.com/quantum/roadmap

7. Google, "Our Quantum Computing Journey," https://quantumai.google/learn/map

8. Inga Goddijn and Ashley Allocca, *Data Breach QuickView Report,* Flashpoint, 2022, https://go.flashpoint-intel.com/docs/2021-Year-End-Report-data-breach-quickview

9. World Economic Forum, *Global Risks Report 2022,* 2022, https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2022.pdf

10. Vikram Sharma and Hans Bachor, "Rethinking Cybersecurity for a Quantum World," *Australian Academy of Science,* February 22, 2021, https://www.science.org.au/curious/policy-features/rethinking-cybersecurity-quantum-world

11. Peter W. Shor, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer," *SIAM Journal on Computing 26,* no. 5 (1997)

12. IBM Quantum Composer, "Grover's Algorithm," https://quantum-computing.ibm.com/composer/docs/iqx/guide/grovers-algorithm

13. Thomas Santoli and Christian Schaffner, "Using Simon's Algorithm to Attack Symmetric-Key Cryptographic Primitives," *Quantum Information and Computation 17,* no. 1–2 (2017)

14. John Loeffler, "How Peter Shor's Algorithm Dooms RSA Encryption to Failure," *Interesting Engineering,* May 2, 2019, https://interestingengineering.com/innovation/how-peter-shors-algorithm-dooms-rsa-encryption-to-failure

15. "Cloud Security Alliance Sets Countdown Clock to Quantum," *Cloud Security Alliance,* March 9, 2022, https://cloudsecurityalliance.org/press-releases/2022/03/09/cloud-security-alliance-sets-countdown-clock-to-quantum/

16. Michele Mosca and Marco Piani, *2021 Quantum Threat Timeline Report,* Global Risk Institute, 2022, https://globalriskinstitute.org/publications/2021-quantum-threat-timeline-report/

17. Booz Allen Hamilton, *Chinese Threats in the Quantum Era,* 2021, https://www.boozallen.com/expertise/analytics/quantum-computing/chinese-cyber-threats-in-the-quantum-era.html

18. Chad Boutin, "NIST Announces First Four Quantum-Resistant Cryptographic Algorithms," *National Institute of Standards and Technology,* July 5, 2022, https://www.nist.gov/news-events/news/2022/07/nist-announces-first-four-quantum-resistant-cryptographic-algorithms

19. Encryption Consulting, "What Is Cryptographic Agility? How Do You Get Crypto-Agility?," https://www.encryptionconsulting.com/education-center/what-is-crypto-agility/

20. Xiongfeng Ma, Xiao Yuan, Zhu Cao, Bing Qi, and Zhen Zhang, "Quantum Random Number Generation," *Quantum Information 2,* 16021 (2016), https://arxiv.org/abs/1510.08957

21. Quantum Flagship, "Quantum Key Distribution (QKD)," https://qt.eu/discover-quantum/underlying-principles/quantum-key-distribution-qkd/

22. Victor Lovic, "Quantum Key Distribution: Advantages, Challenges and Policy," *Cambridge Journal of Science & Policy 1,* no. 2 (2020), https://www.repository.cam.ac.uk/handle/1810/311529

23. The White House, *National Security Memorandum on Promoting United States Leadership in Quantum Computing While Mitigating Risks to Vulnerable Cryptographic Systems,* 2022, https://www.whitehouse.gov/briefing-room/statements-releases/2022/05/04/national-security-memorandum-on-promoting-united-states-leadership-in-quantum-computing-while-mitigating-risks-to-vulnerable-cryptographic-systems/

24. Jeremy Jergens, Isaac Kohn, and Colin Soutar, "Transitioning to a Quantum-Secure Economy," *World Economic Forum,* September 13, 2022, https://www.weforum.org/whitepapers/transitioning-to-a-quantum-secure-economy/

25. "A Guide to a Quantum-Safe Organization," Quantum Economic Development Consortium, December 6, 2021, https://quantumconsortium.org/guide-to-a-quantum-safe-organization/

26. QuintessenceLabs, "Living in a Quantum-Safe World," 2020, https://info.quintessencelabs.com/quantum-safe-world-whitepaper; Alireza Shabani and Ramana Kompella, "Towards a Unified Internet for Classical and Quantum Communication," *Cisco Tech Blog,* June 15, 2022, https://techblog.cisco.com/blog/making-a-quantum-ready-internet; Niko Mohr, Mateusz Masiowski, Matija Zesko, and Henning Soller, "Quantum Technology Monitor," *McKinsey & Company,* 2022, https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/quantum%20computing%20funding%20remains%20strong%20but%20talent%20gap%20raises%20concern/quantum-technology-monitor.pdf; Lucian Comandar, Jean-François Bobier, Michael Coden, and Stefan Deutscher, "Ensuring Online Security in a Quantum Future," *Boston Consulting Group,* March 30, 2021, https://www.bcg.com/publications/2021/quantum-computing-encryption-security; Adriano Poloni, Mathangi Sandilya, Jim Ricotta, Will Finigan, and Prineha Narang, "Untangling the Future of Quantum Communications," *Accenture,* 2021, https://www.accenture.com/_acnmedia/PDF-167/Accenture-Future-Quantum-Communications.pdf

27. Vikram Sharma and William Dixon, "We Need to Build a Quantum Security Coalition. Here's Why," *World Economic Forum,* August 11, 2020, https://www.weforum.org/agenda/2020/08/we-need-to-build-a-quantum-security-coalition/

Nuclear Dynamics in Southern Asia: Politics and Tech as Mediators

Ruhee Neog

SOUTHERN ASIA INCLUDES three nuclear-armed neighbours—China, India, and Pakistan. The three states are modernising and expanding their nuclear enterprises,[1] even as the military applications of dual-use technologies[a] are evolving.[2] In what ways do these two developments intersect to shape strategic dynamics between China, India, and Pakistan, especially when all three ultimately view nuclear weapons as political tools whose greatest utility rests in their nonuse? How could emerging technologies mediate nuclear balances—or imbalances—in such a complex regional security environment?

As of January 2022, based on open-source data, China's nuclear inventory is estimated at 350, India's at 160, and Pakistan's at 165.[3] Developments in their respective nuclear enterprises are aimed at increasing their deterrent value so no adversary considers the use, or threat of use, of nuclear weapons against them. They have each successfully achieved this aim to varying degrees, with broad agreement on the existence of a certain "maturity"[4] in the India-Pakistan and China-India dyadic nuclear equations, albeit for different reasons. Maturity,however, does not mean an absence of risk.

Strategic competition in Southern Asia, as in the rest of the world, is being rewritten by developments in the technological domain.[5] Emerging technologies—or technologies whose battlefield uses are still evolving—could disrupt the maturity and level-headed approach to nonuse displayed so far. Their incorporation can lower the threshold for nuclear use, create new pathways for crisis escalation, and be intentionally or inadvertently used to manipulate nuclear vulnerabilities.

---

[a] Technologies that have both civilian and military application potential.

Further, in Southern Asia, nuclear equations and emerging technologies intersect in a complex security environment. Any analysis of the politics of nuclear weapons must account for Chinese, Indian, and Pakistani perceptions of their broader threat landscape because these interpretations will inevitably colour nuclear decision-making. Islamabad's 'full spectrum deterrence'[6] is poised against New Delhi; New Delhi's nuclear calculations are primarily driven by the Chinese force posture;[7] and Beijing's nuclear purpose is extra-regional, deriving from US nuclear developments.[8]

Nuclear weapons anchor the India-Pakistan security discourse. Beijing cites New Delhi and Islamabad's exclusion from the Nuclear Non-Proliferation Treaty as the reason for the absence of official Chinese nuclear dialogue with the two countries. Nuclear weapons feature nominally in the China-India security equation, although China's nuclear assistance to Pakistan is central to international perceptions of their bilateral relationship. The converse—that nuclear weapons are in a way one focal point of India's relationship with China because of their importance to the China-Pakistan equation—is, therefore, also true.

Additionally, the three countries' foreign policy towards others, if viewed as threatening to any of their respective core national interests, will impact their nuclear decision-making in direct and indirect ways. For instance, China building more missile silos,[9] ostensibly to make its deterrent more credible to the US,[10] could shape security decision-making in India.[11] Any outcome of these decisions could have consequences for the Pakistani security mindset.[12] Beijing may not consider New Delhi a nuclear peer,[13] but new analytical lenses could emerge from the growing proximity between the US and India, their joint emphasis on the Indo-Pacific, and announcements such as the Australia-UK-US (AUKUS) trilateral security pact. These complex—sometimes convergent and often divergent—capabilities, intentions, and perceptions have contributed to characterisations of nuclear equations in Southern Asia as "dyads", a "trilemma",[14] or a "strategic chain".[15]

Emerging Tech as Mediator

EMERGING TECHNOLOGIES CAN weaken deterrence by enmeshing nuclear and non-nuclear components. They can reduce the threshold for nuclear use through unanticipated and, therefore, unplanned-for escalation spirals, termed by some scholars as "wormholes"[16] or "entanglements".[17] Using two examples from the Southern Asian context—nuclear submarines and cyber risk, and anti-satellite (ASAT) capability framed within the nuclear discourse—this section discusses how advancements in nuclear enterprise could be put at risk by technological developments, contributing to instability.

- ### Nuclear submarines and cyber risk

The operationalisation of sea-based deterrents are aimed at enhancing survivability and assured retaliation through further dispersal and submersion under water, making these nuclear weapons harder to find. It also, however, takes nuclear dynamics into the maritime domain. This can create command-and-control issues, and complicate China and India's no-first-use policies.[18] A malicious state or non-state actor's cyber ability—and the decision to wield it—to hijack, disable, or confuse nuclear command, control, and communications systems will have serious consequences for deterrence breakdown. An incident at sea characterised by miscommunication, misinformation, and a lack of clarity on who or what perpetrated the incident could create or compound inter-state hostilities. It can also foment crises.

The regional nuclear maritime environment makes such a scenario plausible. On the nuclear submarine front, China is in a league of its own, with Jin-class submarines armed with JL-2 submarine-launched

ballistic missiles that have a 7,400-km range.[19] India has one operational ballistic missile nuclear submarine, the INS Arihant, which carries the nuclear-capable K-15 Sagarika missile with a range of 750 km.[20] Pakistan does not have a nuclear-powered, nuclear-armed submarine yet, but may seek such capability[21] as a natural progression of its plans to mount the Babur-3 cruise missile, with a range of 450 km, on conventional submarines.[22]

- ### ASAT in the Nuclear Discourse

China conducted its ASAT test in 2007.[23] In 2019, India announced[24] the successful completion of Operation Shakti, its own ASAT test, in which an indigenously-developed ballistic missile interceptor was launched against an Indian satellite in lower Earth orbit.[25] According to India's foreign ministry, the test "provides credible deterrence against threats to our growing space-based assets from long range missiles, and proliferation in the types and numbers of missiles".[26] Observers agree that the test was primarily intended as a technology demonstration to Beijing, which retains conventional and nuclear superiority in the region.[27]

What are the potential regional ramifications of this development? Pakistan has historically used India's ballistic missile defence (BMD) programme as one justification for its 'full spectrum deterrence'.[28] India's ASAT capability could incentivise Pakistan into seeking such capability of its own, for which help could be forthcoming from China,[29] as with the country's nuclear programme. Pakistan could rationalise further acquisitions to its nuclear portfolio by arguing that a full operationalisation of Indian BMD "could embolden India's decision makers in a future conflict with Pakistan, to contemplate a pre-emptive counterforce strike".[30] This hypothetical scenario would involve Indian targeting of Pakistani command-and-control assets, which "could add pressure on the defender to consider the early use of its nuclear weapons before these could be neutralized on the ground".[31]

Social media is a crisis wildcard that is not often acknowledged as an emerging technology, but it can play a significant role in both cases explored above because of its still-developing role in communication. Real-time disinformation, misinformation, and political signalling via social media can accelerate and/or compound crises. Social media's reach, actor diversification beyond the government and traditional news media platforms, and extraordinarily swift and mostly unchecked diffusion of content gives it immense power to constrain and enable crisis decision-making.[32] It is increasingly becoming the conduit through which people and governments communicate. Its influence can limit or encourage crisis escalation, contributing to a puncturing of firewalls between the sub-conventional, conventional, and nuclear realms.

Nuclear Weapons and Regional Security Complexities

A STATE'S NUCLEAR choices are made within its national security and foreign policy calculus. It stands to reason that the geopolitical environment that states occupy will contribute to their nuclear decision-making. China's efforts to credibly deter a nuclear-armed US may not be aimed at India, but will still have repercussions for India's threat perception. As a domino effect, any developments in India's nuclear weapons programme, even if in response to changes in the Chinese nuclear posture, will impact—or be used to rationalise—Pakistani nuclear calculations. These developments take place in a wider environment of regional inter-state tensions and evolving regional and global geopolitics. This section gives an overview of how these factors interact with nuclear thinking in Beijing, New Delhi, and Islamabad and Rawalpindi.

- ## Inter-state tensions in Southern Asia

China-India and India-Pakistan are divided over long-standing territorial disputes. The Chinese People's Liberation Army and the Indian Army have faced off along the Line of Actual Control since June 2020. China incurred into and occupied Indian territory in Ladakh, amid the raging COVID-19 pandemic.[33] Clashes in Galwan Valley, involving combat with fists, stones, and poles, contributed to 20 casualties on the Indian side.[34]

Nuclear weapons are not—and not expected to be—relevant to the India-China dispute. While there were fears of the Galwan stand-off escalating into war, most observers agree that it is unlikely to spiral into a full-blown conventional conflict, and certainly not to the nuclear level.[35] The reason for this is simple: while India's nuclear deterrent is driven by China, the latter's is to deter the US. India's nuclear weapons are peripheral in Chinese security thinking, except in the context of the New Delhi-Islamabad bilateral, or South Asia.[36] Southern Asia includes China, thus denoting it as a regional (and not extra-regional) power at the same table as its South Asian neighbours. India and Pakistan, for their own reasons, find the Southern Asia framing to be an accurate reflection of regional trilateral dynamics, particularly in the nuclear domain.

India and Pakistan have maintained somewhat consistent levels of low-intensity conflict, with fluctuating annual ceasefire violations across the Line of Control.[37] Pakistan's 'full spectrum nuclear deterrence' is aimed at India, which it perceives as an existential threat. New Delhi's main security concern vis-à-vis Islamabad is the latter's support for sub-conventional conflict against India in the nuclear shadow. It involves state-sponsored terrorism against India, while threatening the use of nuclear weapons should New Delhi decide to respond militarily or violate any of Pakistan's very broadly-defined redlines.[38] This effort by a conventionally inferior Pakistan to seek strategic parity with India has long been successful in holding New Delhi to ransom.

More recently, the Indian political leadership has shown a greater appetite for risk-taking by exploring options for limited military responses against terror attacks by groups in Pakistan. In 2019, for example, the Jaish-e-Mohammad (JeM) claimed an attack on an Indian paramilitary convoy near Pulwama in Jammu and Kashmir.[39] The incident caused at least 49 casualties.[40] India responded with military force; for the first time since the 1971 Indo-Pak War, India breached Pakistani territory, with the Indian Air Force reportedly striking a JeM camp in Balakot in the Khyber Pakhtunkhwa region.[41]

Tellingly, careful official statements and caveats clarifying that their objectives had been met—lest hostilities spiral to the nuclear level—have typically been issued in the immediate aftermath of such potentially destabilising incidents or instances of political grandstanding. For instance, this was seen after India's Balakot strike, and by Pakistan after India's accidental misfiring of the BrahMos cruise missile in 2021.[42]

- ## Regional and Global Political Competition

China and India are locked in a regional political face-off for primacy among the other South Asian states. China's rising influence across the world, chiefly through strategic gains made via the purportedly collaborative and development-oriented Belt and Road Initiative (BRI),[43] its military forays into the Indian Ocean,[44] and attempts to rewrite international institutions[45] form part of India's security calculations. Further, the China-Pakistan relationship—with Pakistan's nuclear development enabled by Chinese assistance, among other factors—contributes to New Delhi's consideration of a two-front war.[46] At the

same time, India's growing proximity to the US, and what this entails in terms of political and material engagement, is a cause for concern for both China and Pakistan.[47]

The language and vocabulary of the Indo-Pacific, which India champions, distinguishes itself from the characteristics associated with China's global ascendance, often encapsulated in the BRI. Some of the terminology that the US[48] and India[49] use frequently in official communications related to the Indo-Pacific include 'rules-based international order', 'fair', 'open', and 'balanced', traits they believe the Chinese system does not embody.[50] One of its manifestations is the Quadrilateral Security Dialogue (Quad), which includes Australia, Japan, India, and the US. While it is framed as a cooperative framework for non-military collaboration, it is widely understood to be motivated by a collective interest in checking China's burgeoning influence.[51]

Of the four countries, New Delhi's emphasis on the Quad's non-traditional security aims is the most emphatic—a testament to its efforts to balance its partnership with Washington while minimising Beijing's antagonisms. India's balancing act acknowledges that global and extra-regional alignments could negatively impact its regional equations. It takes place against the geopolitical realities of initiatives such as AUKUS, under which the UK and the US will assist Australia in acquiring nuclear-powered submarines. Significantly, AUKUS also includes collaboration on emerging tech such as artificial intelligence, cyber, quantum technologies, and hypersonic capabilities. Other recent churns that will figure in the Chinese security narrative, and have knockdown effects for India and Pakistan, are the US's political positioning on Taiwan and Russia's invasion of Ukraine.

Conclusion

CHINA, INDIA, AND Pakistan understand that nuclear weapons are most effective when not used on the battlefield. Regional nuclear equations are thus stable to the extent that Beijing, New Delhi, and Islamabad view the utility of nuclear weapons through a political lens. They are each evolving their nuclear deterrents to make them more credible to different adversaries. But nuclear deterrence is not perfect. Its evolution can open it up to new vulnerabilities even as others are foreclosed. These very advances in nuclear enterprise could intersect dangerously with emerging technologies that are developing rapidly and simultaneously. Any fallout will have far-reaching implications for regional and global security.

Endnotes

1. Dinakar Peri, "China, India, Pakistan expanding nuclear arsenal, says Swedish think tank," *The Hindu,* June 14, 2021, https://www.thehindu.com/news/national/china-india-pakistan-expanding-nuclear-arsenal-says-swedish-think-tank/article34814508.ece.

2. Kelley M. Sayler, "Emerging Military Technologies: Background and Issues for Congress," *Congressional Research Service,* April 6, 2022, https://sgp.fas.org/crs/natsec/R46458.pdf.

3. SIPRI, *Global nuclear arsenals are expected to grow as states continue to modernize–New SIPRI Yearbook out now,* June 13, 2022, https://www.sipri.org/media/press-release/2022/global-nuclear-arsenals-are-expected-grow-states-continue-modernize-new-sipri-yearbook-out-now.

4. Manpreet Sethi, "Contemporary Nuclear Dynamics in Southern Asia: Many Challenges, Few Possibilities," *Asia-Pacific Leadership Network Policy* Brief No. 82, August 26, 2022, https://www.apln.network/analysis/policy-briefs/contemporary-nuclear-dynamics-in-southern-asia-many-challenges-few-possibilities.

5. Tuneer Mukherjee, "Sino-Indian Maritime Competition: Shadow Fighting in The Indian Ocean," *Stimson Center,* June 19, 2020, https://www.stimson.org/2020/sino-indian-maritime-competition-shadow-fighting-in-the-indian-ocean/.

6. Sannia Abdulla, "Pakistan's Full-Spectrum Deterrence: Trends and Trajectories," *South Asian Voices,* December 13, 2018, https://southasianvoices.org/pakistan-full-spectrum-deterrence-trends-trajectories/#:~:text=The%20full%2Dspectrum%20deterrence%20doctrine,against%20India's%20massive%20retaliation%20doctrine.

7. Vijay Gokhale, *The Long Game: How the Chinese Negotiate with India* (Gurugram: Penguin Random House, 2021); Sharad Joshi, "Nuclear Proliferation and South Asia: Recent Trends," *Nuclear Threat Initiative,* July 31, 2007, https://www.nti.org/analysis/articles/nuclear-proliferation-south-asia/.

8. Eric Heginbotham, Michael S. Chase, Jacob L. Heim, Bonny Lin, Mark Cozad, Lyle J. Morris, Christopher P. Twomey, Forrest E. Morgan, Michael Nixon, Cristina L. Garafola, and Samuel K. Berkowitz, "*China's Evolving Nuclear Deterrent: Major Drivers and Issues for the United States," RAND* Corporation, 2017, https://www.rand.org/pubs/research_reports/RR1628.html.

9. Suyash Desai, "An Expert Explains: Why China seems to be building three missile silos," *Indian Express,* September 2, 2021, https://indianexpress.com/article/explained/china-missile-silo-test-nuclear-weapons-7471044/.

10. SD Pradhan, "China constructs new 119 nuclear missile silos along with plans to double its nuclear arsenal: Objectives and implications," *Times of India,* July 14, 2021, https://timesofindia.indiatimes.com/blogs/ChanakyaCode/china-constructs-new-119-nuclear-missile-silos-along-with-plans-to-double-its-nuclear-arsenal-objectives-and-implications/.

11. Harsh Pant and Kartik Bommakanti, "Keeping an eye on China's expanding nuclear stack," *The Hindu,* August 19, 2021, https://www.thehindu.com/opinion/op-ed/keeping-an-eye-on-chinas-expanding-nuclear-stack/article35987126.ece.

12. Lora Saalman and Petr Topychkanov, "South Asia's Nuclear Challenges: Interlocking Views from India, Pakistan, China, Russia and the United States," *SIPRI,* 2021, https://www.sipri.org/sites/default/files/2021-03/2104_south_asias_nuclear_challenges_0.pdf.

13. Ruhee Neog, "Nuclear Suppliers Group: Why India will be Kept Out," *South Asian Voices,* June 1, 2016, https://southasianvoices.org/nuclear-suppliers-group-why-india-will-be-kept-out/.

14. Tanvi Kulkarni, "Managing the China, India, and Pakistan Nuclear Trilemma," *Asia-Pacific Leadership Network,* July 2022, https://www.apln.network/projects/china-india-pakistan-nuclear-trilemma/special-report-managing-the-china-india-and-pakistan-nuclear-trilemma.

15. Robert Einhorn and Waheguru Pal Singh Sidhu, *The strategic chain: Linking Pakistan, India, China, and the United States," Brookings,* 2021, https://www.brookings.edu/research/the-strategic-chain-linking-pakistan-india-china-and-the-united-states/.

16. Rebecca Hersman, "Wormhole Escalation in the New Nuclear Age," *Texas National Security Review* 3, no. 3 (Autumn 2020): 90-109, http://dx.doi.org/10.26153/tsw/10220.

17. James M. Acton, "Why Is Nuclear Entanglement So Dangerous?," C*arnegie Endowment for International Peace, J*anuary 23, 2019, https://carnegieendowment.org/2019/01/23/why-is-nuclear-entanglement-so-dangerous-pub-78136 ; PR Chari, "India, Pakistan and the Nuclear Race: The Strategic Entanglement," *Institute of Peace and Conflict Studies,* April 11, 2013, http://www.ipcs.org/comm_select.php?articleNo=3879.

18. Tanvi Kulkarni, "Managing the China, India, and Pakistan Nuclear Trilemma," *Toda Peace Institute,* July 2022, https://toda.org/assets/files/resources/policy-briefs/134.managing-the-cip-nuclear-trilemma.pdf.

19. Matthew P. Funaiole, Joseph S. Bermudez Jr., and Brian Hart, "A Glimpse of Chinese Ballistic Missile Submarines," *Center for Strategic and International Studies,* August 4, 2021, https://www.csis.org/analysis/glimpse-chinese-ballistic-missile-submarines.

20. Dinakar Peri, "India successfully test-fires 3500-km range submarine-launched ballistic missile K-4," *The Hindu,* January 19, 2020, https://www.thehindu.com/news/national/india-successfully-test-fires-3500-km-k-4-slbm/article30601739.ece.

21. Zia Mian, M.V. Ramana, et al., "Nuclear Submarines in South Asia: New Risks and Dangers," *Journal for Peace and Nuclear Disarmament* 2, no. 1 (2019): 184-202, https://doi.org/10.1080/25751654.2019.1621425.

22. Rajat Pandit, "India behind China and Pakistan in nuclear-warheads but not worried," *Times of India,* June 15, 2021, https://timesofindia.indiatimes.com/india/india-behind-china-pakistan-in-nuclear-warheads-but-not-worried/articleshow/83524404.cms.

23. Carin Zissis, "China's Anti-Satellite Test," *Council on Foreign Relations,* February 22, 2007, https://www.cfr.org/backgrounder/chinas-anti-satellite-test.

24. Narendra Modi, "The successful test of the Anti-Satellite (ASAT) Missile," YouTube video, March 27, 2019, https://www.youtube.com/watch?v=0v6eRj7HI0s.

25. Ashlyn Still, Júlia Ledur and Ally J. Levine, "India shoots down own satellite," *Reuters,* March 27, 2019, https://graphics.reuters.com/INDIA-SATELLITE-WEAPON/0100918Q1RV/index.html.

26. Media Center, "Frequently Asked Questions on Mission Shakti, India's Anti-Satellite Missile test conducted on 27 March, 2019," Ministry of External Affairs, Government of India, March 27, 2019, https://www.mea.gov.in/press-releases.htm?dtl/31179/Frequently_Asked_Questions_on_Mission_Shakti_Indias_AntiSatellite_Missile_test_conducted_on_27_March_2019.

27. Pranab Dhal Samanta, "ET Analysis: India's Space Shakti sends a Signal to Chinese Satellites," *The Economic Times,* March 28, 2019, https://economictimes.indiatimes.com/news/politics-and-nation/et-analysis-indias-space-shakti-sends-a-signal-to-chinese-satellites/articleshow/68607275.cms?from=mdr.

28. Baqir S. Syed, "Pakistan to retain full spectrum deterrence policy," *Dawn,* December 22, 2017, https://www.dawn.com/news/1378106; Zafar Nawaz Jaspal, "Full Spectrum Deterrence: Capability and Credibility," June 7, 2018, https://pakistanpolitico.com/full-spectrum-deterrence-capability-and-credibility/.

29. MK. Narayanan, "Down to earth on the ASAT test," *The Hindu,* April 23, 2019, https://www.thehindu.com/opinion/lead/down-to-earth-on-the-asat-test/article62109913.ece.

30. Adil Sultan, "India's ASAT and Nuclear Entanglement in South Asia," *STRAFASIA*, March 29, 2019, https://strafasia.com/south-asia-and-the-nuclear-entanglement/.

31. Adil Sultan, "India's ASAT and Nuclear Entanglement in South Asia," *Strategic Foresight For Asia,* March 29, 2019, https://strafasia.com/south-asia-and-the-nuclear-entanglement/.

32. Ruhee Neog, "Self-Referencing the News: Media, Policymaking, and Public Opinion in India-Pakistan Crises," *Investigating Crises: South Asia's Lessons, Evolving Dynamics, and Trajectories* (Sameer Lalwani and Hannah Haegeland Ed., Washington, D.C.: Stimson Center, 2018), https://www.stimson.org/wp-content/files/file-attachments/InvestigatingCrises.pdf.

33. Rajeswari Pillai Rajagopalan, "China's PLA upgrades its forces along the disputed border with India," *ORF,* May 29, 2021, https://www.orfonline.org/research/chinas-pla-upgrades-its-forces-along-the-disputed-border-with-india/.

34. Suhasini Haidar, Ananth Krishnan, and Dinakar Peri, "Indian Army says 20 soldiers killed in clash with Chinese troops in the Galwan area," *The Hindu,* June 16, 2020, https://www.thehindu.com/news/national/indian-army-says-20-soldiers-killed-in-clash-with-chinese-troops-in-the-galwan-area/article61668218.ece.

35. Saheli Roy Choudhury, "Wider armed conflict between India and China unlikely after 'violent' border clash," *CNBC,* June 16, 2020, https://www.cnbc.com/2020/06/17/india-china-border-standoff-analysts-say-war-is-unlikely.html.

36. Toby Dalton and Tong Zhao, "At a Crossroads? China-India Nuclear Relations After the Border Clash," *Carnegie Endowment for International Peace,* August 19, 2020, https://carnegieendowment.org/2020/08/19/at-crossroads-china-india-nuclear-relations-after-border-clash-pub-82489.

37. Surya Valliappan Krishna, "Bordering on Peace: Evaluating the Impact of the India-Pakistan Ceasefire," *Carnegie Endowment for International Peace,* February 24, 2022, https://carnegieindia.org/2022/02/24/bordering-on-peace-evaluating-impact-of-india-pakistan-ceasefire-pub-86513.

38. Shalini Chawla, "Drumming up N-hysteria," *The Tribune,* February 26, 2020, https://www.tribuneindia.com/news/comment/drumming-up-n-hysteria-47256.

39. Hakeem Irfan Rashid, "Jaish-e-Mohammed claims responsibility for Pulwama attacks," *The Economic Times,* February 15, 2019, https://economictimes.indiatimes.com/news/politics-and-nation/jaish-e-mohammed-claims-responsibility-for-pulwama-attacks/articleshow/68003116.cms?from=mdr.

40. Mudasir Ahmad, "J&K: At Least 49 CRPF Jawans Killed in Deadliest Militant Strike on Security Forces," *The Wire,* February 15, 2019, https://thewire.in/security/kashmir-pulwama-crpf-attack-jem.

41. "Pushing boundaries: on Balakot air strikes," *The Hindu,* February 27, 2019, https://www.thehindu.com/opinion/editorial/pushing-boundaries-on-balakot-air-strikes/article26379272.ece?homepage=true/.

42. Ruhee Neog, "An Accidental Missile Launch and a Lesson for Indian Communications," *Asia-Pacific Leadership Network,* April 29, 2022, https://www.apln.network/analysis/commentaries/an-accidental-missile-launch-and-a-lesson-for-indian-communications.

43. Peter Cai, "Understanding China's Belt and Road Initiative," *The Lowy Institute,* March 22, 2017, https://www.lowyinstitute.org/publications/understanding-china-s-belt-road-initiative.

44. Zack Cooper, "Security Implications of China's Military Presence in the Indian Ocean," *Center for Strategic and International Studies,* April 2, 2018, https://www.csis.org/analysis/security-implications-chinas-military-presence-indian-ocean.

45. "'Creeping capture': How China is trying to 'control' global bodies like UN, WHO," *Times of India,* July 2, 2021, https://timesofindia.indiatimes.com/world/china/creeping-capture-how-china-is-trying-to-control-global-bodies-like-un-who/articleshow/84062355.cms.

46. Sushant Singh, "The Challenge of a Two-Front War: India's China-Pakistan Dilemma," *Stimson,* April 19, 2021, https://www.stimson.org/2021/the-challenge-of-a-two-front-war-indias-china-pakistan-dilemma/.

47. Anway Iqbal, "US-India defence pact to impact Pakistan, China," *The Dawn,* August 30, 2016, https://www.dawn.com/news/1280873/us-india-defence-pact-to-impact-pakistan-china; Ananth Krishnan, "China's Foreign Minister Wang Yi says U.S. wants 'Indo-Pacific NATO,'" *The Hindu,* March 7, 2022, https://www.thehindu.com/news/international/chinas-foreign-minister-wang-yi-says-us-wants-indo-pacific-nato/article65201008.ece.

48. The White House, *Indo-Pacific Strategy of the United States,* February 2022, https://www.whitehouse.gov/wp-content/uploads/2022/02/U.S.-Indo-Pacific-Strategy.pdf.

49. Media Center, "Prime Minister's Keynote Address at Shangri La Dialogue (June 01, 2018)," Ministry of External Affairs, Government of India, June 1, 2018, https://mea.gov.in/Speeches-Statements.htm?dtl/29943/Prime_Ministers_Keynote_Address_at_Shangri_La_Dialogue_June_01_2018.

50. Media Center, "Official Spokesperson's response to a query on participation of India in OBOR/BRI Forum," Ministry of External Affairs, Government of India, May 13, 2017, https://mea.gov.in/mediabriefings.htm?dtl/28463/Official+Spokespersons+response+to+a+query+on+participation+of+India+in+OBORBRI+Forum; Quint Forgey and Phelim Kine, "Blinken calls China 'most serious long-term' threat to world order." *Politico*, May 26, 2022, https://www.politico.com/news/2022/05/26/blinken-biden-china-policy-speech-00035385.

51. Zaheena Rasheed, "What is the Quad and can it counter China's rise?," *Al Jazeera,* November 25, 2020, https://www.aljazeera.com/news/2020/11/25/what-is-the-quad-can-us-india-japan-and-australia-deter-china.

Emerging 'Disruptive' Technologies and the Threat to Strategic Stability

Tanvi Kulkarni

SCIENCE, TECHNOLOGY, AND human intellect often meet at the intersection of the impacts on human society. Scientific discoveries and technological innovations affect the sociocultural, economic, and political life of human societies, and, more fundamentally, their perceptions and experiences of security. Technology[a,1] has been a driver of economic growth, social change, and national power, including military power. In the digital age, the understanding of national security and power is shaped, more than ever, by scientific advancements and access to key technologies. In turn, that understanding shapes the technological priorities for the present and future.

Technology is an important element of modern human societies' state-making and security-making endeavours. The term 'emerging technologies' is frequently used in the context of states' military capabilities and weapon systems. In that context, emerging technologies are those that influence military competition between states, and impact national and international peace and security. The assumption that the role of the 'state' is pervasive—directly or indirectly—in the emergence of new technologies makes emerging technologies a topic of great interest for policymakers worldwide. It is not unusual for global leaders and government officials to express concerns about existing and future technologies while assessing threats, challenges, and opportunities to their state's national interests.[2]

With that premise, this essay examines why emerging technologies are generally seen as 'disruptive' in the context of nuclear weapons and strategic stability, and how this

[a] Science and technology are not synonymous. Science consists of understanding and explaining (through laws and concepts) the complexities of the universe, whereas technology is the practical manifestation of the ability to change the material existence of human conditions; technology gives science its 'social character'. If science is curiosity, technology is strategy. The primary difference between science and technology relates to their scale, nature of output, impact on society, and accuracy to predict outcomes. Relative to science, technology has a lower accuracy of prediction of outcomes.

understanding shapes nuclear risk-reduction measures. It highlights that serious and urgent dialogues between states on the disruptive and destabilising effects of emerging technologies and strategic stability are crucial steps toward strategic and nuclear risk reduction.

Why are Emerging Technologies Seen as 'Disruptive'?

EMERGING TECHNOLOGIES ARE generally considered to be those innovations whose potential effects and impacts on defence and security are not entirely clear,[3] or remain undemonstrated and recessed.[4] The most serious concerns about emerging technologies are commonly associated with their potential disruptive effects. This is a possible reason why the phrases 'emerging technologies'[b,5] and 'disruptive technologies' are sometimes used interchangeably. Among those who employ these phrases, however, there is no consensus on what the terms 'emerging' and 'disruptive' specifically constitute. Some experts prefer to use the catch-all phrase 'emerging and disruptive technologies' to denote "weapons, support systems and subsystems that have been significantly improved or recently deployed—or could be rapidly developed in the near future".[6] Certain technologies are already in the process of being prominent and can be called as 'emergent technologies'. For instance, artificial intelligence (AI), robotics, nanotechnology, blockchains, quantum computing, 3D printing, and hypersonic systems are all emergent technologies. Their military applications in the nuclear sphere, however, are still being explored.

In the realm of nuclear weapons, the uncertainty and ambiguity about the nature and scale of effects of emerging technologies beget the questions: Will emerging technologies cause nuclear war?[7] Do they erode the state's ability to launch a pre-emptive and retaliatory nuclear strike precisely? Do emerging technologies or their impacts warrant changes to states' nuclear doctrines and postures? Emerging technologies, in different ways and to varying extents, can affect how states conduct their nuclear operations, how their command-and-control systems function, how they manage deterrence stability, crisis communication and escalation control, and the prospects for arms control and disarmament.[8] A technology, its application, or their combination, is considered consequential, particularly if it affects the three strands of strategic stability—deterrence stability, arms race stability, and crisis stability.

Of the range of emerging technologies that now operate along with conventional and nuclear weapons systems, the military applications of hypersonic systems, anti-satellite weapons, directed energy weapons (DEWs), offensive cyber capabilities, AI applications for information warfare, lethal autonomous weapon systems, and dual-use weapons systems are considered to be most destabilising to strategic stability.[9] Hypersonic weapon systems and DEWs are especially worrying because they significantly shrink decision-making time during crises and incentivise escalation. Their 'use-it-or-lose-it' character incentivises first-mover advantage, making them very high-impact weapons. Although not entirely new to nuclear operations, the applications of cyber and AI for offensive uses have grown exponentially. These applications can interfere with nuclear weapons systems to constrain, confuse, or malfunction them, increasing the vulnerability of nuclear forces to preemptive, inadvertent, and accidental launches. Whereas AI-related threats are especially difficult to predict, the effects of cyber operations are exceedingly difficult to control.[10]

---

[b] NATO defines a disruptive technology as, "Those technologies or scientific discoveries that are expected to have a major, or perhaps revolutionary, effect on NATO defence, security or enterprise functions in the period 2020-2040".

[c] Directed Energy Weapons can absorb a nuclear attack and degrade an adversary's command-and-control systems. They are, therefore, seen to be destabilising because they challenge the adversary's second-strike capability.

Emerging technologies are also prone to be considered disruptive by states because the pace of technological innovations far outdoes that of policymaking.[11] Governments are simply unable to keep abreast with the rate of innovation taking place outside their direct jurisdiction, especially in the private and commercial sectors.[12] This is the case, for instance, with AI, biotechnology, and cyber activities. Managing the societal, political, and normative effects of these technologies is a tremendous challenge for states and international institutions. Unless governments can regulate access and exercise some control over their potential effects, the emergent and emerging technologies are deemed to be 'disruptions of legal and regulatory orders'.[13] During relative peacetime, government interest in the military applications of these technologies can be significantly high to gain comparative defence and security advantages on the battlefield during crises.

Rapid advancements in technologies not only impact national security, but they also render existing instruments and institutions of global order anachronistic. The Nuclear Non-Proliferation Treaty (NPT), for instance, which has been at the cornerstone of the global nuclear order for over 70 years, has struggled in recent decades to keep up with the speed and spread of technological innovations. The basic skills and the know-how required to make nuclear weapons are also more widely accessible now than in the 1960s and 1970s when the NPT was designed. Several non-nuclear weapons states within the NPT have made advances in the technical know-how and technological capabilities that have implications for nuclear proliferation.[d] These changes have made the treaty's definition of 'nuclear capabilities' redundant. Another component of the global nuclear order that has been blunted by emerging technologies over the last several decades is traditional arms controls. Even in the 1970s and 1980s, traditional or structural arms controls, which focus on the quantitative aspects of military forces (through restrictions on the numbers of forces) were unable to keep pace with the ambiguous nature and rapid changes in conventional and nuclear military technology.[14] Today, with highly decentralised and dual-use technologies like AI and omni-use capabilities [e,15] like cyber systems, traditional arms controls are at the risk of becoming obsolete.

Finally, popular narratives also accentuate the disruptive quality of emerging technologies, particularly in the context of nuclear weapons. In June 1983, then US President Ronald Reagan panicked[16] when he watched *WarGames,*[17] a science-fiction Hollywood movie. The film showed an American whiz-kid accidentally hack into the North American Aerospace Defence Command's computing programme. The programme mistakes this as an imminent Soviet nuclear attack and prepares to launch a preventive US nuclear first attack, triggering an inadvertent nuclear war. Quite disturbed by what he saw, Reagan authorised a series of White House meetings and studies, which led to congressional hearings and the first US national policy on cyber threats.[18] The episode brought to light serious concerns and troubled conversations among US defence bureaucrats with regard to threats like hacking, cyberattacks and cyberwars[19] from then-emerging automation and computing technologies. American movies may have inspired Reagan[f,20] but even today, science-fiction thinking is a widely used approach for imagining the implications of emerging technologies and the future of war and conflict.[21] The science-fiction imaginations of nuclear crises are generally and reasonably that of dystopia and oblivion.[22]

---

[d] For instance, many countries have acquired relevant dual-use technologies or conventional weapons systems powered by fissile material, like the nuclear-propelled submarines.

[e] Omni-use capabilities are those which "could be used for a range of purposes simultaneously, from improvements in healthcare and infrastructure to exceptionally efficient surveillance and military operations".

[f] American science-fiction movies had a significant influence over Ronald Reagan's thinking and his administration's policies. More famously, Reagan's Strategic Defense Initiative was inspired from the notional orbital laser missile shield from the Star Wars movies.

Scholarly studies have noted, however, that emerging technologies do not produce a monolithic effect when combined with nuclear strategy. These technologies have varying, sometimes even contradictory, effects on strategic stability and nuclear deterrence.[23] Different applications of the same technology may also produce varying and contradictory effects.[g,24] Moreover, these effects may differ during peacetime, crisis, and wartime. While some technologies undermine crisis stability, others undermine arms race stability; some can produce escalatory pressures and raise risks of inadvertence and accidents; and some others create "dual-use security dilemmas",[h,25] blurring the boundaries between conventional and nuclear war. Some nuclear researchers and technology experts have argued that the same technologies that are perceived as disruptive during crises can have a stabilising effect on nuclear deterrence during relative peacetime, but such potential of technologies is overlooked to emphasise the negative impacts.[26] For instance, AI applications related to intelligence, surveillance, and reconnaissance, situational awareness, early warning, command and control, human-machine coordination, and network empowerment can help to increase survivability of nuclear forces, collect accurate information, effectively process battlefield conditions, and enhance transparency, all of which have a relatively positive impact on strategic stability.[27]

Whether the impacts of these technologies on strategic stability will be mostly negative or positive also depends on many allied factors, such as the social and institutional contexts in which these technologies are embedded,[28] the levels of technological readiness, increased geopolitical competition,[29] attitudes of leaders,[30] and the social, cultural, economic, and geopolitical contexts in which norms, policies and regulations around applications of technologies are designed.[31] The tendency to excessively focus on technological drivers of instability and escalation overlooks the critical role of political and strategic choices of governments in shaping the impact of technology.[32]

Need for Dialogue and Risk Reduction

AS MORE AND more technologies creep into the domain of military competition, there is growing concern among scientists, experts, and policymakers alike about the implications of the military-technology nexus on strategic stability.[33] There has been a sharp rise in the literature on advancing the conversations about which technologies impact nuclear stability and how.[34] But despite the growing scholarship on the strategic consequences of emerging technologies and serious concerns vis-à-vis the potential challenges and opportunities of these technologies, governments have been slow to discuss the implications of emerging technologies on strategic stability and nuclear escalation through systematic multilateral channels of dialogue.[j,35] The United Nations' Under-Secretary-General and High Representative for Disarmament Affairs has noted that while "developments in a variety of technologies are diminishing predictability, shared understandings and trust, while raising the risks of misperception, arms races, and

[g] This is especially the case with AI applications.

[h] This concept was introduced by researcher Amir Lupovic: "A dual-use security dilemma occurs when actors face an opponent that wields technologies with both civil and military/harmful applications".

[i] Technology readiness level is the method for determining the maturity of a technology during the production, acquisition, deployment, and employments phases. Maturity levels are affected by technical, legal, regulatory, ethical, normative, and state-specific barriers to technology development.

[j] In August 2022, at a Tenth Non-Proliferation of Nuclear Weapons review conference side event, the US National Nuclear Security Administration noted, "The emergence of new technologies that significantly lower the barrier to proliferation, an explosion of open-source research, and the proliferation of new types of warheads and delivery vehicles all require reciprocal advancements in verification capabilities" and called for "the international community to work together to address this daunting challenge". This was followed by a brief mention in the Tenth Non-Proliferation of Nuclear Weapons review conference's draft final document (in para 37a on nuclear weapons states' commitment): "to intensify regular dialogue among and between the nuclear-weapon States, and with the non-nuclear weapon States, on nuclear doctrines and arsenals, … , as well as on the potential implications of emerging technologies."

potential escalation through miscalculation,…none of the nuclear weapons-related forums are discussing the intersection between technology and nuclear risk, adding to decreasing transparency and a climate of misperception".[36] The Under-Secretary General also pointed that this dearth of dialogue was especially problematic because "some nuclear doctrines now carry the possibility of a nuclear response to any attack on critical infrastructure, potentially including a cyberattack, and where space-based assets are considered critical infrastructure".[37] The ominous warning is especially apposite against the backdrop of the Russian invasion[38] and nuclear signalling in Ukraine.[39]

There is a lot yet to be understood about the stabilising and destabilising effects on emerging technologies in the nuclear realm. It is, therefore, imperative to first determine the implications of these technologies on different components of strategic stability and then to explore new governance structures, tools, approaches, and processes to regulate or constrain those impacts. For this, governments will have to narrow, if not remove the organisational silos in which government officials currently look at emerging technologies and nuclear weapons. The major military and nuclear powers must be at the forefront of multilateral dialogues and conversations on the destabilising effects of emerging technologies because their potential to acquire, access, produce, and deploy highly complicated emerging technologies far exceeds those of other powers. This task, however, should not be left to their governments alone, since most forward-looking technological expertise usually lies outside. Therefore, government engagement with the private and commercial sectors is critical to determine and address the destabilising and disruptive implications of emerging technologies.[40]

The answers to technological implications on strategic stability are embedded in the overarching questions of power and international conflict. As geopolitical tensions rise, a serious and systematic international dialogue on military and particularly dual-use applications of emerging technologies should be a key part of the path of risk reduction. In an environment of deep mistrust and frequent crises, however, reorienting conversations towards employing emerging technologies for risk reduction is as challenging as it is urgent. Public engagement toward policies governing and regulating military applications of emerging technologies is crucial is this regard.

Finally, conversations and dialogues on emerging technologies and risk reduction are futile without strategic stability dialogues. Parity and mutual vulnerability can no longer be determined simply by the quantity and quality of nuclear arsenals. The strategic potentials of emerging technologies, including the 'entanglement' between commercial, conventional military, and nuclear weapon systems, and their ability to incentivise escalation, are forcing a redefinition of the core tenets of strategic stability in the contemporary global order. Strategic stability dialogues are, therefore, crucial for determining the policy responses to the disruptive effects of emerging technologies.

Endnotes

1. Sumit Bhaduri, "Science, Society and Technology: Three Cultures and Multiple Visions," *Journal of Science Education and Technology* 12, no. 3, (September 2003): 303-308, https://link.springer.com/article/10.1023/A:1025037108006.

2. Prabhjote Gill, "India's foreign minister explains why the MEA needs a separate division to monitor emerging technology," *Business Insider,* January 16, 2020, https://www.businessinsider.in/tech/news/indias-foreign-minister-explains-why-the-mea-needs-a-separate-division-to-monitor-emerging-technology/articleshow/73288153.cms.

3. NATO Science & Technology Organization, *Science & Technology Trends 2020–2040: Exploring the S&T Edge,* 2020, https://www.nato.int/nato_static_fl2014/assets/pdf/2020/4/pdf/190422-ST_Tech_Trends_Report_2020-2040.pdf.

4. Christopher F. Chyba, "New Technologies & Strategic Stability," *Daedalus* 149, no. 2 (April 2020): 150–170, https://doi.org/10.1162/daed_a_01795.

5. NATO Science & Technology Organization, *Science & Technology Trends 2020–2040: Exploring the S&T Edge,* 2020, https://www.nato.int/nato_static_fl2014/assets/pdf/2020/4/pdf/190422-ST_Tech_Trends_Report_2020-2040.pdf.

6. Andrew Futter, "Explaining the Nuclear Challenges Posed by Emerging and Disruptive Technology: A Primer for European Policymakers And Professionals," *EUNPDC Non-Proliferation and Disarmament Paper 73* (2021): 2, https://www.sipri.org/sites/default/files/2021-03/eunpdc_no_73_0.pdf

7. Matthew Kroenig, "Will Emerging Technology Cause Nuclear War?: Bringing Geopolitics Back In," *Strategic Studies Quarterly* 15, no.4 (Winter 2021): 59-73, https://www.airuniversity.af.edu/Portals/10/SSQ/documents/Volume-15_Issue-4/D-Kroenig.pdf.

8. Andrew Futter, "Explaining the Nuclear Challenges Posed by Emerging and Disruptive Technology: A Primer for European Policymakers And Professionals," *EUNPDC Non-Proliferation and Disarmament Paper 73* (2021): 2, https://www.sipri.org/sites/default/files/2021-03/eunpdc_no_73_0.pdf.

9. Michael Onderco and Madeline Zutt, "Emerging technology and nuclear security: What does the wisdom of the crowd tell us?," *Contemporary Security Policy* 42, no.3(2022): 288, https://www.tandfonline.com/doi/pdf/10.1080/13523260.2021.1928963.

10. Michael Onderco and Madeline Zutt, "Emerging technology and nuclear security: What does the wisdom of the crowd tell us?" *Contemporary Security Policy* 42, no.3(2022): 288, https://www.tandfonline.com/doi/pdf/10.1080/13523260.2021.1928963.

11. Marina Favaro, "Emerging Technologies and Nuclear Stability," *Asia Pacific Leadership Network*, July 19, 2021, https://www.apln.network/analysis/commentaries/emerging-technologies-and-nuclear-stability.

12. Camino Kavanagh, "New Tech, New Threats, and New Governance Challenges: An Opportunity to Craft Smarter Responses?," *Carnegie Endowment for International Peace*, August 28, 2019, https://carnegieendowment.org/2019/08/28/new-tech-new-threats-and-new-governance-challenges-opportunity-to-craft-smarter-responses-pub-79736.

13. Roger Brownsword, Eloise Scotford, and Karen Yeung, eds., *Oxford Handbook of Law, Regulation and Technology,* (Oxford: Oxford Handbooks Online, 2017).

14. Hans Gunther Brauch, "Confidence Building Measures and Disarmament Strategy," *Current Research on Peace and Violence* 2, no.3/4 (1979): 117.

15. Brigitte Dekker and Maaike Okano-Heijmans, "The US–China Trade–Tech Stand-Off and the Need for EU action on Export Control," Clingendael Report, August 2019, *https://www.clingendael.org/sites/default/files/2019-08/Report_US-China_stand-off.pdf.*

16. James M. Lindsay and Margaret Gach, "Five Movies Worth Watching About the Threat of Nuclear War," *Council on Foreign Relations,* August 7, 2020, https://www.cfr.org/blog/five-movies-worth-watching-about-threat-nuclear-war.

17. Wikipedia contributors, "WarGames," *Wikipedia, The Free Encyclopedia,* https://en.wikipedia.org/w/index.php?title=WarGames&oldid=1113175292

18. The White House, *National Policy on Telecommunications and Automated Information Systems Security (U),* National Security Decision Directive Number 145, Washington D.C, September 17, 1984, https://irp.fas.org/offdocs/nsdd145.htm.

19. Fred Kaplan, *Dark Territory: The Secret History of Cyber War* (New York: Simon & Schuster, 2016), 7, https://books.google.co.in/books?id=tZxvDgAAQBAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=snippet&q=much%20worse&f=false.

20. Kevin Bankston, "How Sci-Fi Like 'WarGames' Led to Real Policy During the Reagan Administration," *New America Weekly,* October 11, 2018, https://www.newamerica.org/weekly/how-sci-fi-wargames-led-real-policy-during-reagan-administration/.

21. Kevin Bankston, "How Sci-Fi Like 'WarGames' Led to Real Policy During the Reagan Administration."

22. Tricia Mawire, "Oblivion & 9 Other Best Dystopian Films About Nuclear War," *Screen Rant,* August 31, 2021, https://screenrant.com/oblivion-best-dystopian-films-about-nuclear-war/.

23. Marina Favaro, "Emerging Technologies and Nuclear Stability," *Asia Pacific Leadership Network,* July 19, 2021, https://www.apln.network/analysis/commentaries/emerging-technologies-and-nuclear-stability.

24. Cai Cuihong, "The shaping of strategic stability by artificial intelligence," in *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk: Volume II East Asian Perspectives,* ed. Lora Saalman (Stockholm International Peace Research Institute, 2019), 54-77, http://www.jstor.com/stable/resrep24532.16.

25. Amir Lupovici, "The dual-use security dilemma and the social construction of insecurity," *Contemporary Security Policy* 42, no.3 (2021): 257-285, https://www.tandfonline.com/doi/abs/10.1080/13523260.2020.1866845?journalCode=fcsp20#:~:text=A%20dual%2Duse%20security%20dilemma,and%20constructivist%20approaches%20to%20security.

26. Jessica Cox and Heather Williams, "The Unavoidable Technology: How Artificial Intelligence Can Strengthen Nuclear Stability," *The Washington Quarterly* 44, no. 1(2021): 69–85, https://www.tandfonline.com/doi/abs/10.1080/0163660X.2021.1893019.

27. Cai Cuihong, "The shaping of strategic stability by artificial intelligence," in *The impact of Artificial Intelligence on strategic stability and nuclear risk: Volume II East Asian Perspectives,* ed. Lora Saalman (SIPRI, 2019), 54-77, https://www.jstor.org/stable/resrep24532.16.

28. Melvin Kranzberg, "Technology and History: "Kranzberg's Laws"," *Technology and Culture 27*, no.3 (1986): 544–560, cited in Marina Favaro, Neil Renic, Ulrich Kuhn, "Negative Multiplicity: Forecasting the Future Impact of Emerging Technologies on International Stability and Human Security," *Institute For Peace Research and Security Policy,* September 2022, 11, https://ifsh.de/file/publication/Research_Report/010/Research_Report_010.pdf.

29. Paul van Hooft, Lotje Boswinkel and Tim Sweijs, "Shifting sands of strategic stability Towards a new arms control agenda," *The Hague Centre for Strategic Studies,* February 2022, https://hcss.nl/wp-content/uploads/2022/02/Arms-Control-Shifting-sands-of-strategic-stability-2022-HCSS.pdf.

30. Michael Onderco and Madeline Zutt, "Emerging technology and nuclear security: What does the wisdom of the crowd tell us?," *Contemporary Security Policy* 42, no.3 (2021): 286-311, https://doi.org/10.1080/13523260.2021.1928963.

31. Marina Favaro, Neil Renic, Ulrich Kuhn, "Negative Multiplicity: Forecasting the Future Impact of Emerging Technologies on International Stability and Human Security," *Institute For Peace Research and Security Policy,* September 9, 2022, https://ifsh.de/file/publication/Research_Report/010/Research_Report_010.pdf.

32. Caitlin Talmadge, "Emerging technology and intra-war escalation risks: Evidence from the Cold War, implications for today," *Journal of Strategic Studies* 42, no. 6 (2019): 864-887, https://doi.org/10.1080/01402390.2019.1631811.

33. Brad Roberts, "Multi-Domain Complexity and Strategic Stability in Peacetime, Crisis, and War – Annotated Bibliography," *European Leadership Network Report,* March 2, 2021, https://www.europeanleadershipnetwork.org/report/multi-domain-complexity-and-strategic-stability-in-peacetime-crisis-and-war-annotated-bibliography/.

34. Marina Favaro, "Emerging Technologies and Nuclear Stability," *Asia Pacific Leadership Network,* July 19, 2021, https://www.apln.network/analysis/commentaries/emerging-technologies-and-nuclear-stability.

35. National Nuclear Security Administration, "NNSA Administrator Hruby's remarks at the High-Level Briefing on United States Nuclear Policy Side Event of the Tenth NPT Review Conference," August 5, 2022, https://www.energy.gov/nnsa/articles/nnsa-administrator-hrubys-remarks-high-level-briefing-united-states-nuclear-policy and 2020 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons, *Draft Final Document,* NPT/CONF.2020/CRP.1/Rev.2, August 25, 2022, https://reachingcriticalwill.org/images/documents/Disarmament-fora/npt/revcon2022/documents/CRP1_Rev2.pdf.

36. Izumi Nakamitsu (Keynote Speech, Virtual UK Project on Nuclear Issues, June 10, 2020), 2020, Annual Conference Royal United Services Institute for Defence and Security Studies, https://front.un-arm.org/wp-content/uploads/2020/06/10-June-High-Representative-Keynote-at-RUSI-UK-PONI-Annual-Conference-2020.pdf.

37. Izumi Nakamitsu (Keynote Speech, Virtual UK Project on Nuclear Issues, June 10, 2020), 2020, Annual Conference Royal United Services Institute for Defence and Security Studies, https://front.un-arm.org/wp-content/uploads/2020/06/10-June-High-Representative-Keynote-at-RUSI-UK-PONI-Annual-Conference-2020.pdf.

38. International Atomic Energy Agency, "Update 103 – IAEA Director General Statement on Situation in Ukraine," September 17, 2022, https://www.iaea.org/newscenter/pressreleases/update-103-iaea-director-general-statement-on-situation-in-ukraine.

39. Associated Press, "Putin's path: From pledges of stability to nuclear threats," *The Indian Express,* October 7, 2022, https://indianexpress.com/article/world/putins-path-pledges-stability-nuclear-threats-ukraine-russia-8195207/.

40. Camino Kavanagh, "New Tech, New Threats, and New Governance Challenges: An Opportunity to Craft Smarter Responses?" *Carnegie Endowment for International Peace,* August 28, 2019, https://carnegieendowment.org/2019/08/28/new-tech-new-threats-and-new-governance-challenges-opportunity-to-craft-smarter-responses-pub-79736.

Ethics

Humanitarian surveillance of connected migrants

Veronika Nagy

Introduction

MIGRANTS AND REFUGEES—who must rely on humanitarian aid for their survival—are considered among the most vulnerable populations in the world. In recent years, service strategies that use mobile technologies have been developed to provide fast, transparent, and efficient services to displaced populations.[1] After the migration crisis in 2015, with the influx of newcomers in the Schengen region, various governance initiatives have been undertaken to support civil society in their tasks to provide digitised infrastructure that will enable low-access services not only in the destination countries, but also outside their geographical area.[2]

A number of international NGOs like UNHCR, Red Cross, and Doctors Without Borders receive funding from the UN and other bodies such as the European Union (EU) for the provision of immediate support for migrants and their integration in their new community, and to bridge all levels of government (national, regional, local) for the sake of "effective migration polices".[3] New tools are being provided to collect, store, share and analyse personal information of migrants and their relatives, which could then be shared with local or national authorities. Subsequently, authorities became dependent on the data collection of NGOs and associations that were instrumentalised and activated to create an integrated information system.[4]

Though basic registration was always part of the tasks of humanitarian organisations, their presence and first-hand contact with refugees or asylum seekers has also contributed to their complex role in digitised administrative decision-making.[5] Neoliberal governments, including the EU, create pressures on NGOs to increasingly

reinforce the need of efficient and sustainable service provisions by rapidly increasing data processing and identification practices, including biometric authentication systems.[a,6]

Engagement with automatised decision-making has become a necessity for humanitarian organisations that are dependent on structural funds. In some cases, collaboration with data sharing has been normalised in their practices, such as the service provision of UNHCR and Red Cross.[7] Preserving the anonymity of clients, political independence, and providing aid without conditionalities, are the essential principles of civil society. However, data collection, storage, and exchange and in some cases even digital tracking of clients become part of everyday practices of humanitarian network.[8] Social, financial, and emotional support become provisional based on personal data sharing and collaboration with local, national, or even with third-party migration agencies. Humanitarian service providers, due to their position in the field, are increasingly being included into networks of digital border control—this carries the risk of exposing refugees and migrants to administrative and penal repercussions.[9]

Indeed, there are growing concerns over data sharing and privacy issues of migrants as humanitarian aid has not been safeguarded from the pressures of surveillance culture. As most humanitarian organisations are generally funded by supranational governments, even "governmental and humanitarian biometric data collection from refugees and migrants has been normalised with the additional risk of putting already-marginalised groups in greater danger of having their personal data used against them."[10] While digitised data systems are growing and AI technologies slowly take over humanitarian services, increasing populations of migrants become unintended victims of unethical data sharing practices.

This article highlights three core elements that should be challenged in considering the technology-based interventions applied in humanitarian assistance. It focuses on digitised software and data systems and excludes the subjects of hardware technologies like drones.[11] It highlights issues of informed consent; data collection such as biometric data use; and digital litter.[b] In order to prevent future harms of humanitarian interventions, there is a need for a better understanding of how NGOs operate with the online data of their clients and how they inform them about the potential consequences of the abovementioned issues.

Is there 'informed consent' in humanitarian services?

DATA COLLECTION ON vulnerable groups whose lives depend on immediate humanitarian aid triggers several questions regarding principles of free choice in information sharing and the role of transparency in the process and administration of informed consent.[c] Can we expect that people in emergency situations

---

[a] "At a basic level, the state's retreat from welfare creates the conditions for NGOs to serve as providers of services that people desperately need. This in itself has come to justify the ubiquitous presence of NGOs, particularly in the face of humanitarian crisis (Edmonds, 2012; Klein, 2008; Krause, 2014; Schuller, 2013, 2017). In times of economic precarity, NGOs appear as relatively stable sources of employment and NGO experts represent the 'intellectual leadership' of an instrumental and technocratic variety (Moore and Moyo, this issue). The corporatised, professionalised and specialised NGO reframes movements and struggles to fit within an apolitical 'global policy language' (Mannan, 2015)"

See: Ismail, F., & Kamat, S. (2018). NGOs, Social Movements and the Neoliberal State: Incorporation, Reinvention, Critique. Critical Sociology, 44(4–5), 569–577. https://doi.org/10.1177/0896920517749804

[b] "Digital litter" refers to poor-quality information or outdated data, links,spread online or through digital tools, apps, and social media that can undermining refugee and migrant decision-making , increasing their risks of social and legal harms

[c] "Since 2018 the UN High Commissioner for Refugees (UNHCR) has registered hundreds of thousands of Rohingya refugees in Bangladeshi camps and the Bangladesh government has issued them identity cards, which are needed for essential aid and services. Bangladesh then used the information, including analog photographs, thumbprint images, and other biographic data to submit refugee details to the Myanmar government for possible repatriation." See: https://www.hrw.org/sites/default/files/media_2021/06/UNHCRLetter_Bangladesh.pdf

are able to objectively consider and possess the sense to reject or withdraw their consent for data collection? And in terms of data sharing, do they fear repercussions, such as exclusions from services; to become the subject of suspicion, or even to lose their right to asylum elsewhere if they do not give full consent? Unfortunately, latest empirical studies show that these are realistic expectations. As the cases of UNHCR have illustrated in the recent years, potential beneficiaries of humanitarian aid are excluded if they do not cooperate with intrusive requests for their data.[12] The datafication of NGO clients in the field of migration support has well-known risks regarding their asylum applications in host countries, risks of identification by politically hostile authorities, and also being tracked by their social networks. As Theodora Gazi (2020) has noted:

> **Although the need of the NGO to record humanitarian needs and displacement patterns is evident, there is not a "one size fits all" legal basis for data collection. In reality, determining the legal grounds for processing information is not straightforward. First, while consent is considered one of the most common bases for processing, in many cases, it could be invalid during humanitarian interventions. Consent must be freely given, specific, informed and unambiguous (article 7 and recital 32 of the GDPR).**[13]

Communication with applicants has been the core argument for clear agreements on consent, but the issue is far more complex than providing information and ensuring that one can withdraw their consent even after availing humanitarian services. Data collection of NGOs has become more complex in the last decade as governments are demanding access to data collected by humanitarian organisations, justified by security needs such as to counter a terrorist threat or espionage. This puts NGOs under pressure with new accountability policies. With these developments, humanitarian organisations have had to invest in digitised administrative systems of third parties, often big tech companies, to keep their activities sustainable and to remain transparent for legal monitoring. Such securitisation processes tend to reduce the discretionary power of NGOs and increase their control over personal information. Consequently, unconditional support has become almost impossible for unidentified beneficiaries in the field of migrant humanitarian aid. Consent on data sharing seemed to have turned into a precondition for saving lives.

Although it should be a matter of choice, when providing vital assistance to people in need, the importance of informed consent is neglected. NGOs must refrain from using this legal basis when collecting data from vulnerable people to provide assistance. In 2020, Gazi wrote:

> **Moreover, when data is collected based on consent, assessing a person's vulnerability involves understanding the social and cultural norms, so that the choice is not made for the population. In any case, it is imperative that goods and services are not withheld, even if individuals are not willing to consent to the processing of their information.**[14]

Organisations should rely on vital interest such as distribution of food, water, and medical assistance as legal basis, when data processing is necessary to protect someone's life, health or security.[15] Simply put, legalities, structural pressures, and the cultural context should not shape the decisions around informed consent of refugees as a requirement for basic provisions. Besides the challenges of informed consent, there is an additional role for NGOs in the surveillance mechanisms of migration control. Providers of humanitarian aid are not only mandated to collect and share the data of their beneficiaries but are also forced to check and screen them. While they need to monitor their identities, they must also assure their funders about the legitimacy of their service use, regularly monitored by trustees and external authorities.

The neoliberal funding structure and security narratives on terrorist threats have made migrant-rights NGOs slowly turn into the extensions of border policing agencies, exposing their beneficiaries to third-party interests. A 2021 study by the International Society Centre noted:

> **NGO data is sometimes also shared with donors to obtain funding or as part of specific data-sharing agreements. Each contractual arrangement with commercial service providers, be it outsourced IT services, such as the storage and data analysis, or simple bank and mobile pay transfers, disclose identity and information about aid recipients. Those who manage the personal data, the senior NGOs managers, and data protection officers, are aware of the issues at stake.**[16]

NGOs testify as to how they must adhere to migration polices due their own vulnerabilities and the urge to remain sustainable. Service providers are experiencing reduced autonomy amidst increasing regulations, fines, and monitoring. As revealed in surveillance studies, INGOs even tend to implement extra control measures on personal data collection of their beneficiaries to ensure their reliability and to protect their position in the field of NGO services.[d] In other words, consent for data sharing becomes merely tokenistic in the process of providing humanitarian aid to refugees and migrants. One intrusive development in this process-which illustrates the complexity of state, corporate, and NGO relationships—is the increasing use of biometric authentication systems.[17]

The Biometric price of humanitarian aid provision

THE FIRST NOTABLE case of humanitarian agencies applying biometric data collection was that of UNHCR and WFP in Jordan, in 2013. IrisGuard, as a contracted data management partner, participated in experiments with iris scans to govern the registered refugees and their financial transactions in refugee camps in Jordan. While refugees seem to have had no choice but to enrol into the system as a precondition for payments, there were no real considerations to refugee consent in terms of data sharing, and the registration of biometric data of refugees became a subject of corporate interest. Beneficiaries of UNHCR could access support only if they unconditionally cooperate with their registration practices and hand over their personal biometric data in exchange for basic provisions such as food. Meanwhile, this digitised control technology is increasingly being implemented by the World Food Programme (WFP), which does not have a privacy policy included in the contract with IrisGuard on data storage.[18] Such private companies collaborating, supporting or funding humanitarian organisations often remain out of sight, while their lobby activities and market interests deeply impact these administrative practices of NGOs.[19]

---

[d] See also discussion on the limitation of national legislations by Karl Steinacker and Katja Lindskov Jacobsen, 2022:
"It can be said that NGOs fall under the jurisdiction of the country where they operate. That might reassure the public in the countries of the North where data protection laws have been passed and are being enforced. However, on a global scale, as more and more governments enact data protection laws, we are in the paradoxical situation that NGOs often face new obligations to disclose personal data in support of national sovereignty and security issues. Countries like Turkey and Rwanda have modelled their data protection laws following the example of the European GDPR and demand that personal data of their citizens and residents is kept exclusively on servers on their territory. But NGO data is sometimes also shared with donors to obtain funding or as part of specific data-sharing agreements. Each contractual arrangement with commercial service providers, be it outsourced IT services, such as the storage and data analysis, or simple bank and mobile pay transfers, disclose identity and information about aid recipients."
Data Protection and Biometrics: Scanned by International Aid Organisations – The International Civil Society Centre (icscentre.org)

Yet, corporations like Palantir,[20] Amazon, or IrisGuard carry a shared responsibility to respect and promote users' rights, and collection of data, independent of local legislation. Their work must be in line with the principles of necessity and proportionality. Therefore, the need for transparency on NGO and state-corporate collaborations is essential to ensure accountability when serving those forcibly displaced, like refugees struggling to exercise their fundamental rights.[21]

Other cases in Yemen and Bangladesh have also raised awareness on the complexities of biometric data use by WFP and local governments.[22] Though the UN has warned about the increasing shift of goals in their food assistance practices, with the growing dependency on corporate funds, INGOs experience pressure of companies using their facilities as a human laboratory. According to research by Reuters, "WFP had been unable to implement agreements with the Houthis on registration of people in need, including biometric system, using iris scanning, fingerprints or facial recognition, to support aid delivery."[23] With 10 million people per month dependent on WFP, its officials could not give an assurance that all the assistance goes to those who need them most, because WFP is not allowed to operate independently and aid is being diverted for profit and tech development purposes.[24]

With these dependencies, humanitarian organisations not only lose their autonomy, but also the trust of their partners in the NGO networks.[25] Beneficiaries raise awareness about the unintended effects of these practices via social media and rapidly learn that no information is provided about the risks of written consent for biometric data collection. Many grassroots organisations warn INGOs about the limitations of transparency, that beneficiaries have often limited or no access to their own data or to query that data, and there is no recourse open to them.[26] Often the deletion of personal data is not even possible and refugees as pure data subjects are excluded from information about data breaches and data sharing arrangements of their aid providers, especially in cases when they collaborate with nation states.[27] These concerns widen the fissures between different humanitarian organisations and increase the barrier against cooperation between NGO networks.

To be sure, contracted NGOs make many attempts to justify their practices and show their reliability through financial audits. However, data audits of NGOs remain an exception. Operations, security processes and strategies for protecting critical and sensitive data are rare within organisations and absent at an inter-organisational level. Self-regulation is the norm, rather than independent oversight. While many NGOs are concerned about their role in data collection for corporate and governmental purposes, they are also part of the securitisation processes of migration control measures. They also agree to disclose detailed information on their tasks and arrangements, including their digital waste.

Data waste or wasted lives

ONLINE APPLICATIONS, INFORMATION platforms, and internal data systems become the heart of law enforcement practices in migration management. With the digital solutionism, many humanitarian service providers collect and sometimes leave their data unprotected and expose refugees to unintended risks. For instance, in cases of emergency in conflict countries, or due to unexpected environmental or financial matters, they may leave the administration behind. For instance, in the latest case of Afghanistan, as Steinacker and Jacobsen were quoted:

> **"When western military and civil organisations evacuated their personnel from Afghanistan, large amounts of sensitive personal data, including biometric data, was left behind. Only time will tell whether that data has been adequately protected and**

> cannot be abused. Closer at home, the German Red Cross received the 2018 edition of the Big Brother Award from a civil society organisation for its digital system of asylum shelter management. The Red Cross software instituted humanitarian surveillance and total control of the asylum seekers and refugees by movement tracking to and within the shelter, detailed recording of medical checks, food consumption, relationships, religious and ethnic affiliations and much more."[28]

As has been pointed out since what is called the "migration crisis" of 2015, there were many similar or even overlapping systems developed to support refugees in different settings and geographies, but it remains unclear how many of them still exist or are up to date. The impetus to offer better organised, authoritative information for refugees represents, somewhat ironically, one factor behind the spread of digital litter. Much of the enthusiasm for tech solutions in the wake of the European crisis quickly dissipated: some tech startups tried to reinvent the wheel or designed apps without enough attention to privacy and security issues. Others lost funding and folded. As Meghan Benton wrote in 2019 in an article on 'digital litter':

> One particular problem was a proliferation of apps whose whole raison d'être was to be the "default" one-stop shop for refugees, for instancing by consolidating information about local services or job opportunities.…most of the 169 civic tech projects for refugees launched in 2015 and 2016 had become inactive as of July 2018….But many of these offline deaths did not receive the online burial they needed.[29]

One of the biggest platforms designed to help migrants and refugees navigate complex admissions systems, MiGreat, folded in 2016 because of funding issues. Its website—full of old but undated information with no message that it was no longer being updated—lingered for years. Most of the problems associated with the digital litter has been associated with the consequences of the neoliberal funding culture. Benton argued:

> "Part of the blame lies at the hands of policymakers, NGOs, and civil-society organisations that seized on the tech hype to create maps or digital platforms aimed at migrants and would-be asylum seekers. Default platforms to consolidate information were sometimes based on time-limited funding, leaving no one to perform updates after the project ended. In addition, in academia research reports often reference old initiatives or create an insular effect where they all refer to one another, regardless of their reliability."[30]

In sum, not all information should be shared, digitised and distributed for the sake of migration assistance.

Conclusion

THE FUNDAMENTAL PRINCIPLES underlying data collection and informed consent need to change. First, beneficiaries of humanitarian aid organisations should be included in the decision-making processes and in the development of protocols.

To be sure, there have been many new policies introduced for the protection of migrants in terms of data protection and privacy, including the privacy notices of Oxfam,[31] policies of bigger international organisations such as Save the Children,[32] Médecins Sans Frontières,[33] or the Swedish Organization for Individual Relief.[34] In practice, however, NGOs appear to be less than keen to comply with these rules, or to provide informed consent.

Empirical studies show[35] that promoting digitalisation in countries where humanitarian aid is needed seems to be more important than data protection. This is often explained by financial pressures and the conflict of interests with the EU and local authorities. This is not to say that the problem is unknown and that there is no constructive engagement of NGOs to protect digitised private data. Many civil society actors recognise the emergence of unforeseen risks related to the use of personal data that have been collected in many different contexts. They now advocate for intensified discussions of approaches to responsible uses of personal, in particular biometric data and for safe storage of information provided by their beneficiaries.[36]

Meanwhile, the trend towards surveillance and biometric overkill continues, including even the collection of DNA data. Data subjects should be protected from such protocols and should receive proper information in all stages of their administrative process. It is essential that humanitarian aid workers are not forced to get clients registered and that international NGOs, like UN agencies, cannot escape public scrutiny because they are immune from challenge before a national court. All data collectors in civil society should institute appropriate oversight bodies and provide recourse procedures for their data subjects. Auditing and monitoring should be subject to new protocols including the right to be forgotten and the right to reject data sharing or data storage for all beneficiaries.

If countries, civil society, and education institutions are serious about expanding protection to those in greatest need rather than those who already have considerable resources or pre-existing social networks in destination countries, they should consider ways by which to better tailor information to people who may have limited institutional knowledge and understanding of how to navigate unfamiliar systems. At the same time, governments, social enterprises, foundations, and others focused on the provision of information or services to refugees and migrants would do well to clean up their digital litter and develop new strategies for sustainability. A few rules-of-the-road can be identified to prevent the problem from escalating during the next crisis.

Endnotes

1. Tom Scott-Smith, "Humanitarian Dilemmas in a Mobile World." *Refugee Survey Quarterly* 35, no. 2 (2016): 1–21, https://www.jstor.org/stable/48503278.

2. Koen Leurs, "Migration Infrastructures" in *The SAGE Handbook of Media and Migration* (2019), 91-102.

3. Aaron Martin et al, "Digitisation and sovereignty in humanitarian space: Technologies, territories and tensions," *Geopolitics* (2022): 1-36, https://doi.org/10.1080/14650045.2022.2047468.

4. S.Vannini, R.Gomez, and B.C. Newell, "Mind the five: Guidelines for data privacy and security in humanitarian work with undocumented migrants and other vulnerable populations", *Journal of the Association for Information Science and Technology* 71, no.8 (2020): 927-938.

5. Mirca Madianou, "Technocolonialism: Digital innovation and data practices in the humanitarian response to refugee crises", *Social Media+ society* 5, no. 3 (2019),1-13, https://doi.org/10.1177/2056305119863146.

6. Mirca Madianou, "The biometric assemblage: Surveillance, experimentation, profit, and the measuring of refugee bodies," *Television & New Media* 20, no.6 (2019): 581-599, https://doi.org/10.1177/1527476419857682.

7. Mirca Madianou, "Nonhuman humanitarianism: when 'AI for good' can be harmful", *Information, Communication & Society* 24, no.6 (2021): 850-868, https://doi.org/10.1080/1369118X.2021.1909100.

8. Veronica Nagy, *Crime prevention, migration control and surveillance practices: Welfare bureaucracy as mobility deterrent* (Routledge, 2018).

9. Koen Leurs, *Migration infrastructures* (The SAGE handbook of media and migration, 2019), 91-102.

10. The Ongoing Digitisation of Europe's Borders – Digital Freedom Fund

11. Maria Gillespie, Soud Osseiran, and Margie Cheesman, "Syrian refugees and the digital passage to Europe: Smartphone infrastructures and affordances", *Social media+ society* 4, no.1(2018), https://doi.org/10.1177/2056305118764440.

12. M.H. Abdirahman, "Identity management systems at UNHCR: from paper registration to biometric data management" (Doctoral dissertation). Marie Smith, "Between control and care: UNHCR and the use of biometrics" *(Thesis).*

13. Theodora Gazi, "Data to the rescue: how humanitarian aid NGOs should collect information based on the GDPR", *Int J Humanitarian Action* 5, 9 (2020). https://doi.org/10.1186/s41018-020-00078-0.

14. Theodora Gazi, "Data to the rescue: how humanitarian aid NGOs should collect information based on the GDPR".

15. Theodora Gazi, "Data to the rescue: how humanitarian aid NGOs should collect information based on the GDPR".

16. Karl Steinacker and Katja Lindskov Jacobsen, "Data Protection and Biometriccs:Scanned by International Aid Organisations", The International Society Centre, 2021, https://icscentre.org/2021/10/20/data-protection-and-biometrics-scanned-by-international-aid-organisations/.

17. Catarina Kinnvall and Jennifer Mitzen, "Anxiety, fear, and ontological security in world politics: thinking with and beyond Giddens", *International Theory* 12, no.2 (2020): 240-256,

18. https://www.unhcr.org/blogs/wp-content/uploads/sites/48/2018/01/article_1.pdf

19. Special mentioning is necessary of the large, specialised UN agencies, such as IOM, WFP and UNHCR. These organisations have, like no other non-governmental bureaucracy, amassed personal data files of tens of millions of people around the globe. Their data subjects for example, surrender their biometric imprints (commonly a fingerprint or an iris scan) for a bar of soap, a sack of rice or a cash transfer, but also for a residence permit, or the opportunity to be resettled in another country.

20. "Palantir Technologies Contracts Raise Human Rights Concerns before NYSE Direct Listing", Amnesty International, https://www.amnestyusa.org/press-releases/palantirs-contracts-with-ice-raise-human-rights-concerns-around-direct-listing/

21. "Iris Scanning of refugees is disproportionate and dangerous-What is happening behind Iris Guard's closed doors," Access Now, https://www.accessnow.org/irisguard-refugees-jordan/.

22. "Palantir Technologies Contracts Raise Human Rights Concerns before NYSE Direct Listing", Amnesty International.

23. "U.N. food chief warns aid suspension in Yemen likely to start this week", Reuters, https://www.reuters.com/article/us-yemen-security-un/u-n-food-chief-warns-aid-suspension-in-yemen-likely-to-start-this-week-idUSKCN1TI1X7.

24. U.N. food chief warns aid suspension in Yemen likely this week (trust.org)

25. Karl Steinacker and Katja Lindskov Jacobsen, "Data Protection and Biometriccs: Scanned by International Aid Organisations", The International Society Centre, 2021, https://icscentre.org/2021/10/20/data-protection-and-biometrics-scanned-by-international-aid-organisations/.

26. Meghan Benton and Alex Glennie, "Digital Humanitarianism: How Tech Entrepreneurs Are Supporting Refugee Inclusion," *Migration Policy Institute* (2016), https://www.migrationpolicy.org/research/digital-humanitarianism-how-tech-entrepreneurs-are-supporting-refugee-integration. Betterplace Lab, "Digital Refugee Projects", https://docs.google.com/spreadsheets/d/1t82LzxBH5GL2HOnEySZE6irLXLAI6rogJ-r8Cf573yo/edit#gid=1052587333. European Resettlement Network, "Supporting Refugees to Access Higher Education", http://www.resettlement.eu/page/supporting-refugees-access-higher-education. European University Association, "Refugees Welcome Map", http://refugeeswelcomemap.eua.be/Editor/Visualizer/Index/48.

27. "UN shared Rohingya data without informed consent," Human Rights Watch, https://www.hrw.org/news/2021/06/15/un-shared-rohingya-data-without-informed-consent.

28. Karl Steinacker and Katja Lindskov Jacobsen,"Biometrics data and the Taliban: What are the risks?", Interview by Irwin Loy, *The New Humanitarian,* September 2, 2021, https://www.thenewhumanitarian.org/interview/2021/2/9/the-risks-of-biometric-data-and-the-taliban

29. Meghan Benton, "Digital Litter: The Downside of Using Technology to Help Refugees", Migration Policy Institute, June 20, 2019, https://www.migrationpolicy.org/article/digital-litter-downside-using-technology-help-refugees

30. Meghan Benton, "Digital Litter: The Downside of Using Technology to Help Refugees".

31. "Privacy Policy", Oxfam, https://www.oxfam.org.uk/privacy-and-security/full-privacy-policy

32. "Privacy and Cookie Policy," Save the Children, https://www.savethechildren.org.uk/misc/privacy-cookie-policy

33. "Privacy Policy", MSF, https://www.msf.org/privacy-policy

34. https://www.imsweden.org/om-im/integritetspolicy

35. Paragi Beata, "Digital4development? European data protection in the Global South," *Third World Quarterl*y 42, no.2(2021): 254-273, DOI: 10.1080/01436597.2020.1811961

36. Ben Hayes and Massimo Marelli, "Facilitating innovation, ensuring protection: the ICRC Biometrics Policy", *ICRC Blog,* October 18, 2019, https://blogs.icrc.org/law-and-policy/2019/10/18/innovation-protection-icrc-biometrics-policy/.

Open AI for Justice

Sachin Malhan, Smita Gupta and Saurabh Karn

IN THE LATE 1800s, large cities around the world depended on thousands of horses to transport people and goods. This massive number of horses generated between 15 to 35 pounds of manure per day, a problem captured in the news in 1894 when *The Times of London* predicted that "in 50 years every street in London will be buried under nine feet of manure".[1] The 'great horse manure crisis of 1894',[2] as the situation eventually came to be known, was debated again in 1898 at the world's first International Urban Planning Conference in New York, but no solution was found. It seemed urban civilisation was doomed to drown in horse faeces. But even as planners struggled to solve the crisis, automobile technology had matured to the point where it was affordable at scale. In the early 1900s, it became cheaper to own a motor vehicle than a horse-drawn carriage, and by 1917 horse-drawn vehicles vanished almost entirely. What was once thought to be an insurmountable threat to humanity's existence disappeared in little over a decade, and the entire incident is now a barely remembered footnote in human history.

India now has its own version of the great horse manure crisis—the great case pendency crisis. Pendency in Indian courts is at an alarmingly high level, with ~71,000 cases in the Supreme Court (as of 22 August 2022), ~59 lakh cases in 25 high courts, and 4.13 crore cases in subordinate courts (as of 29 July 2022 for both).[3] Sixty-six percent of all civil disputes in India are related to property/land issues;[4] approximately 20 million people (or less than 2 percent of citizens) have availed of legal aid in the last 25 years, even though 80 percent of the population is entitled to it;[5] and the overwhelming majority of Indian citizens are not aware of their basic rights and entitlements. Notably, pendency has become worse, not better, in the last five years.

Quite like the urban planners in the late nineteenth century, we have looked at this problem with the same mindsets that created it, asking largely misguided questions—How can we resolve cases more efficiently *(using largely the same processes)*? How can we prioritise more important cases *(within the same broken process)*? How can we strengthen the capacity of the courts *(assuming it's a capacity problem)*?

Instead, like the disruptive automobile innovators of the early twentieth century, we must ask ourselves the more fundamental questions related to the needs of citizens, businesses, and society that are not contained within the answers we have traditionally relied upon. Questions such as: How can a citizen receive an empowering answer to their legal query instantly and in their own language? How can certain types of disputes be mitigated early and, if that is not possible, resolved amicably and quickly? For disputes best resolved in courts, how can priority matters that concern lives and livelihoods be quickly identified? How can we quickly prevent fraud and exploitation through more secure systems of property and worker rights?

Legal empowerment in the 2020s cannot merely be restricted to providing legal literacy but must also involve a reimagining of our public and private justice systems to empower citizens to improve their lives by resolving conflicts, securing their assets, accessing their rights and entitlements, and building trusted relationships. Artificial intelligence (AI) can help make this little hop towards genuine legal empowerment a leap.

AI for Legal Empowerment

A.I. IS LARGELY made possible because of the underlying meanings, patterns, and relationships in information. The pillars of law and justice—the laws, judgements, documents, and titles—are often forms of code, be they laws as lofty as the Constitution of India or as pedantic as the Standards of Weights and Measures Act 1976, or contracts as simple as a non-disclosure agreement or as complex as shareholder agreements. Indeed, developing AI models that can understand legal language and information is not just possible but highly probable, and it will be able to do truly remarkable things. These include enabling the translation and transliteration of speech into text and vice versa to ensure greater accessibility of legal solutions in the vernacular languages;[a,6] helping create smart summaries of laws and judgements[7] to support both lawyers and citizens; understanding legal concepts within laws and judgements to create legal aid chatbots; and helping give a rough prediction of the duration and outcome of a case. These are some capabilities of AI that already exist and are being used to ensure greater access to justice.

The leap lies in going beyond solving these known issues and making new things possible. To imagine these in one shot would defeat the purpose, but one can envision a future where citizens are guided in their commercial and property dealings on the most secure ways to proceed, where it is possible to prevent unintended outcomes of bad laws by testing law in a social context, where disputes are mitigated by intelligent systems that guide transacting parties towards conflict avoidance, and where judicial officers are ably assisted through intelligent inputs pertaining to the laws and the impact of their decisions.

Making this happen in an empowering, inclusive, transparent, and collaborative way is our generation's task. Even in the harshest of conditions, innovators and entrepreneurs find a way. India also has a generation of entrepreneurs building AI for law and justice applications. We have already seen innovative solutions such as AI-led contract drafting and review tools, smart prison and legal Enterprise Resource Planning systems, comprehensive legal search across multiple sources, document automation software-as-a-service

[a] The Supreme Court has developed a dedicated open-source judicial domain language translation tool called SUVAS (Supreme Court Vidhik Anuvaad Software) to translate judicial documents from English to nine vernacular languages (Marathi, Hindi, Kannada, Tamil, Telugu, Punjabi, Gujarati, Malayalam, and Bengali)

platforms, lawyer-client match-up services, automated stamping registrations and compliances, and AI-led digital property due diligence. We have also seen Indian courts experimenting with AI technologies to provide translations to increase access to the judgements. Their work is precious and necessary.

Open AI: Digital Public Goods

THERE ARE TWO problems with the current approach to AI-based solutions—the tools being developed are usually proprietary, and creators rarely collaborate. Consequently, AI tools require significant private investment, and, as has happened in India, the best tools are built behind closed doors with no pathway to sharing. Tools for public purposes, including judicial system analytics, research, and citizen services, are non-existent. This holds back the creation of new and improved solutions with modest investments, most seriously impacting the creation of tools for non-commercial purposes. Rather than rapidly building on each other's work, each innovator begins from scratch by training models for similar functionalities.[b,8]

The answer is the creation of essential AI public goods, such as baseline AI models, datasets, benchmarks, and reference solutions. These are expensive and time-consuming artefacts but have a high repeat value, as the same AI model can be one of the critical components of different products. For instance, an AI model that helps identify the different components of a judgement and structure it accordingly (facts, arguments, analysis, precedent, decision, ruling of previous courts, and so on) can help operationalise solutions like automated bail recommendations, case lifecycle prediction, automatic charge identification and punishment calculation, precedent search and the like. This is not to say that these solutions will not require many more foundational technologies, apart from judgement structuring, to become operational.

Without these public goods, the cost of innovation and experimentation increases drastically, creating an environment where highly funded and resourced individuals and institutions are the only ones that can harness the value of AI. Conversely, with these public goods, we could see tremendous innovation for different actors, including citizens, lawyers, judges, and governments. We would also see a faster experimentation cycle of different products for various problems, and a proportion of these experiments turning into successful and critical solutions.

A burst of innovation has always followed the emergence of such AI public goods. In 2008, American computer scientist Fei-Fei Li and researchers at the University of Illinois-Urbana Champaign and Princeton University began work on ImageNet to categorise millions of images that could improve image and character recognition by machines.[9] By April 2010, there were more than 11 million images in 15,000+ synsets categorised through crowdsourcing on Amazon's Mechanical Turk platform. The ImageNet database now contains more than 14 million annotated images. Since being launched, ImageNet has given researchers a common set of images to benchmark their models and algorithms. In turn, this has driven research in machine learning and deep neural networks, making it easier to classify images and complete other tasks associated with computer vision. In just seven years, the accuracy of classification models rose from 71.8 percent to 97.3 percent.[10] These models, a large part of which is open-sourced, now form the basis of the Facebook system that tags your photos, navigation systems of self-driving cars, and even

[b] To find more ways to spur co-creation in a largely proprietary-models-led ecosystem, we need to come up with different incentive structures. One such solution has been proposed by attorney Rahul Matthan when he says, "We could consider contributory royalties that will allow those on whose intellectual property a new innovation was based to receive a proportion of the revenue that appropriately recognises their contribution but at the same time provides enough of a monetary incentive for new inventors to invest the time and effort required to innovate."

AI diagnostic tools to find anomalies in X-rays, Basically, any AI computer vision application that has to do with images and videos uses this model.

Digital public goods, such as open datasets, open-source software, and open AI, are also having a huge impact in India. In August 2022, the United Payment Interface (UPI)—the foundational technology being used by several digital wallets for real-time financial transactions—processed over six billion transactions worth over INR 10 trillion.[11] UPI, a public good, has been able to democratise access to digital payments. It has revolutionised payments for the retail sector and even attempted to bridge the digital divide to some extent.

Another instance is Bhashini, an initiative of India's Ministry of Electronics and Information Technology (MeitY) and private partners like data sciences firm ThoughtWorks, to develop an open repository of Indian language models and datasets (universal language contribution application programming interface). They have solved for text-to-text and speech-to-text translation, automatic speech recognition, and even real-time speech-to-speech machine translation of different Indian languages. It is truly championing an open ecosystem of co-creation by inviting everyone to contribute to the 'Bhasha Daan', a crowdsourced initiative that allows everyone to contribute to the dataset of text and speech in various Indian languages.

Similarly, digital public goods in law and justice can greatly impact democratising access to justice.

Collaborative and Open Legal AI Models: Digital Public Goods

VARIOUS KINDS OF AI goods can enable multiple downstream applications in AI for justice. Many legal tasks underpinning downstream applications depend on certain legal document types—contracts, laws, property titles, and court judgements. The language contained in these documents is unique—coding for legal concepts, legal system culture, and a wide variety of interlinked data. Court judgements, for instance, contain an interlinked amalgam of system data (court, city, type of case), legal concepts (a vast universe from constitutional to commercial and beyond), facts, and relationships. Understanding this 'language' is no trivial effort. But even partial understanding, leading to some fundamental tasks, can unlock various possibilities downstream. For instance, an AI model that identifies the distinct parts of a court judgement—such as facts, relevant law, interpretations, or decision (also called 'rhetorical roles')—can significantly improve document search and legal research, and also enable new use cases such as judgement summarisation and document validation. Another example of a fundamental AI model can identify the important 'entities' in any judgement, such as the court, judge, petitioner, place, addresses, statutes, and so on. Such a 'named entity recognition' could, like rhetorical roles, support multiple tasks, including legal search, court administration, and case classification.

Importantly, critical digital public goods do not spring up in isolation; they result from the presence of vital public-minded actors such as governments, universities, large foundations, and arms of companies. More recently, public-minded technologists and domain experts (such as iSpirit, Beckn, Bhasini, OpenNyAI) have come together to create quasi-institutions that have led to the emergence of these public goods. Creating these public goods often requires large teams, digital infrastructure, and financial resources. For example, building datasets for justice applications requires gathering legal experts and building datasets on which AI models can be trained. The AI models are also very compute-hungry, especially when dealing with text data, which increases the cost of compute infrastructure. Lastly, given the multidisciplinary nature of the field, collaboration between experts across technology, academia, and industry is needed,

requiring a large team to synthesise and execute these projects. For instance, AI4Bharat, Indian Institute of Technology (IIT) Madras, IIT Bombay, IIIT Hyderabad, and MeitY's Centre for Development of Advanced Computing came together as the dataset and model contributors for the Bhashini initiative. Additionally, Tarento Technologies, ThoughWorks, and AI4Bharat were the code developers, initial funding came from the EkStep Foundation, and Microsoft provided the initial platform credits.[12]

An example of this in the justice space is the OpenNyAI initiative,[13] an open and collaborative mission started in 2021 to develop essential AI public goods for justice (models and datasets) to bridge the gap between lawyers and technologists working in the sector, and develop a long-lasting community to shepherd the field in the years to come. The OpenNyAI mission—founded by education platform EkStep,[14] justice sector catalyst Agami,[15] National Law School of India University, and ThoughtWorks, and supported by an expert community that includes professors from leading universities such as IIT Kanpur and IIT Kharagpur, and entrepreneurs in the legal tech space—has already published a judgement rhetorical roles model and a named entity recognition model, and released reference solutions to trigger the uptake of the models.

In AI for justice, as in the models and datasets mentioned above, the process should be as open as possible to ensure auditability and transparency of AI creation. OpenNyAI relies on wide community consultation to generate the schemas for dataset creation. This is then turned into a massive open online course opportunity where university law students get to learn and contribute to an open-source AI for justice project. Various quality checks are done before the model is trained. One of the key challenges in the process is that despite being very technical- and rules-driven fields, experts and other actors often do not have a shared vocabulary or an understanding of the problem. Law students go through a steep learning curve while learning how to work on Label Studio (an open-source data labelling tool), understand the basics of annotation, and follow a process of learning, calibration, annotation, and adjudication. New processes and systems must also be established from the ground up for a crowdsourced annotated data activity for a legal AI model tailored for Indian legal text.[c]

Our AI Future

IN *THE INNOVATORS,* a book on the people behind the birth of computers, Walter Issacson writes: "But the main lesson to draw from the birth of computers is that innovation is usually a group effort, involving collaboration between visionaries and engineers, and that creativity comes from drawing on many sources. Only in storybooks do inventions come like a thunderbolt, or a lightbulb popping out of the head of a lone individual in a basement or garret or garage."[16] Initiatives such as ImageNet, Bhashini, and OpenNyAI are focussed on doing a thousand little things to support such group efforts.

If India can develop a robust open ecosystem for AI in law and justice, it will also end up securing itself against many of the challenges associated with AI in sensitive areas such as law enforcement, decision-making, and entitlements. This is one of the reasons that an open approach is a must have and not a just a good-to-have for AI in law and justice. The open approach allows many actors to participate who might not be otherwise represented, especially the youth. It also enables full data and model auditability, ensuring that institutions employing these systems can ask questions about how accurate it is, what dataset it has been trained on, and so on.

---

[c] Data scientists from within the OpenNyAI community are participating at SemEval 2023, the international workshop on semantic evaluation where their fundamental model of judgement structuring, entity recognition, and judgement prediction will be opened up to data scientists, researchers, and linguists from around the world who will participate to build better and more accurate models on the reference datasets provided. This task given to researchers is called LegalEval.

The US experiment with their case management and decision support tool, Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), highlights the need for open and collaboratively developed algorithms of models operating in the law and justice ecosystem. In 2016, an appeal was filed in the Wisconsin Supreme Court questioning the six-year sentence awarded to a convict based on an assessment by COMPAS that the convict was at high risk of becoming a reoffender.[17] The grounds of the appeal were that in considering the outcome of an algorithm whose inner workings were secretive and could not be examined, the judge violated due process.[18] While the appeal was dismissed on the grounds that the sentencing was appropriate regardless of the COMPAS assessment, the ensuing public debate[19] went on to show the importance of open and transparent designs of critical AI tools, and the challenges of over-dependence on a single AI tool in key public areas. The presence of multiple open AI models, collaboratively developed and evolved by a diverse community, will go a long way in mitigating the concerns around the application of AI in sensitive areas such as law and justice.

Justice has long been thought of as something that an institution like government, courts, or lawyers 'give' individuals. The future tools built on AI for law and justice technologies will empower every actor to make justice happen. By giving agency to citizens, particularly young citizens who feel the most alienated by institutions, AI can become a creative force, seen less as one magical solution, and more as a platform for endless creativity and problem-solving to close the justice gap once and for all. For this to happen, our AI future must be open, inclusive, and collaborative.

Endnotes

1. Utrecht University, "The Great Manure Crisis", Urban Futures Studio, https://www.uu.nl/en/research/urban-futures-studio/initiatives/mixed-classroom-techniques-of-futuring/mobility-museum-2050/the-great-manure-crisis

2. Stephen Davies, "The Great Horse-Manure Crisis of 1894", Foundation for Economic Education, September 1, 2004, https://fee.org/articles/the-great-horse-manure-crisis-of-1894/

3. Department of Justice, Ministry of Law and Justice, Government of India, *Pending Cases of Civil and Criminal Nature in Various Courts,* 2022, https://pqars.nic.in/annex/257/AU2180.pdf

4. Namita Wahi, "Understanding Land Conflict in India and Suggestions for Reform", Centre for Policy Research, June 26, 2019, https://cprindia.org/understanding-land-conflict-in-india-and-suggestions/

5. Nupur, Radhika Jha, Devika Prasad, Devyani Srivastava, Madhurima Dhanuka, Sugandha Mathur, Shruthi Naik, Leah Verghese, Prof. Vijay Raghavan, Ameen Jauhar, Chitrakshi Jain, Lakhwinder Kaur, Niyati Singh, *India Justice Report 2020: Ranking States on Police, Judiciary, Prisons and Legal Aid,* New Delhi, Tata Trusts, 2021, https://www.tatatrusts.org/Upload/pdf/ijr-2020-overall-report-january-26.pdf

6. "Software developed to translate SC judgments in 9 vernacular languages: Law Minister RS Prasad", *Business Standard,* December 12, 2019, https://www.business-standard.com/article/pti-stories/software-developed-to-translate-sc-judgments-in-9-vernacular-languages-law-minister-rs-prasad-119121200851_1.html.

7. Such as the summarisation provided by the OpenNyAI AI Models. See: "Judgment Summarise", OpenNyAI, https://summarizer-fer6v2lowq-uc.a.run.app/

8. Rahul Matthan, "Open Data Sharing is the Best Way to Step Up Innovation", *Live Mint,* September 13, 2022, https://www.livemint.com/opinion/columns/open-data-sharing-is-the-best-way-to-step-up-innovation-11663085808314.html

9. Dave Gershgorn, "The data that transformed AI research - and possibly the world", Quartz, July 26, 2017, https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/

10. Dave Gershgorn, "Data that Transformed AI Research".

11. National Payments Corporation of India, "UPI Product Statistics", NPCI, https://www.npci.org.in/what-we-do/upi/product-statistics

12. Pritam Bordoloi, "India's Project Bhashini: Breaking the language barrier with AI", *Analytics India Magazine,* July 14, 2022, https://analyticsindiamag.com/indias-project-bhashini-breaking-the-language-barrier-with-ai/

13. OpenNyAI, https://opennyai.org/

14. EkStep Foundation, www.ekstep.org

15. Agami, www.agami.in

16. Walter Isaacson, *Innovators: How a Group of Hackers, Geniuses, and Geeks Created the Digital Revolution,* (New York: Simon & Schuster, 2014).

17. State of Wisconsin v Eric Loomis 2016 WI 68, 371 Wis. 2d 235, 881 N.W., https://www.scotusblog.com/wp-content/uploads/2017/02/16-6387-op-bel-wis.pdf

18. Ed Yong, "A Popular Algorithm is No Better at Predicting Crimes than Random People", *The Atlantic,* January 18, 2018, https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/

19. Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias", *ProPublica,* https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; and Cynthia Rudin et al., "The Age of Secrecy and Unfairness in Recidivism Prediction", *Harvard Data Science Review,* 2, no. 1 (2020), https://doi.org/10.1162/99608f92.6ed64b30; Eugene Jackson and Christina Mendoza, "Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not", *Harvard Data Science Review,* 2, no. 1 (2020), https://doi.org/10.1162/99608f92.1b3dadaa

Ethics of AI: Principles, Rules and the Way Forward

Husanjot Chahal

THE ONGOING GROWTH in artificial intelligence (AI) research and development is prompting constant calls for addressing the accompanying ethical challenges. Indeed, over the past few years, various research institutions, government bodies, and private entities in different countries have issued principles and guidelines for the ethical use of AI. There remains little consensus, however, over universal ethical principles and how to implement them. The imperative therefore is to understand the similarities and differences in AI ethics discussions across different geographies, and explore the existing gaps in these debates. Crucially, if the ultimate goal is the ethical development and deployment of AI, are efforts towards codifying and devising high-level ethical AI principles even a fruitful exercise?

This article examines current efforts towards developing ethical principles related to the development and use of AI. It describes the landscape of discussions on ethical AI principles and the different stakeholders in this space. It also examines the points of convergence and divergence between the various AI ethical guidelines, as well as their gaps. The article closes with specific recommendations for the way forward.

Discussions on AI Ethics: Current Landscape

A.I. IS BEING deployed in ways that touch people's lives, whether in healthcare, financial transactions, or the criminal justice system. Advances in AI can have profound impacts across varied societal domains, and in recent years, this realisation has sparked ample debate about the values that should guide its development and use.

States and international organisations have reacted to these societal concerns in various ways. Some have formed ad-hoc committees tasked to deliberate and provide recommendations on the subject. Examples include the United States National

Artificial Intelligence Advisory Committee (NAIAC) that dispenses advice to the president and various federal officials; the expert group on AI at the Organisation for Economic Co-operation and Development (OECD); the High-Level Expert Group on AI formed by the European Commission; and the Select Committee on AI appointed by the UK Parliament's House of Lords.[1] These bodies have either drafted or are currently drafting policy documents on the ethical, economic, and social implications of advances in AI.

Similar efforts are underway in the private sector. Companies that are at the forefront of AI development like Google, IBM, Intel, Microsoft, and Sony, have released guidelines for developing ethical AI.[2] Some analysts have propounded that these private entities desire to shape the AI ethics domain in ways that either eschew regulation or else meet their own business priorities.[3] That said, non-profit organisations and professional associations, such as the Institute of Electrical and Electronics Engineers (IEEE), Internet Society, OpenAI, and the World Economic Forum have also issued declarations and recommendations on AI principles and policies. The multitude of efforts across such diverse stakeholders reflects the need for guidance in AI development. Apart from the types of organisations that have produced ethical guidelines on AI, the content of such documents is equally wide ranging. Several empirical studies of AI ethical principles have attempted to examine the various topics under discussion across sectors and countries, and to propose how such principles can be implemented in practice.[4] A review of the findings across these studies can offer insights into the scope and potential for a global agreement on the subject of AI ethics, as well as the disagreements therein.

Points of Convergence

RESEARCH SHOWS THAT most of the available ethical guidelines adopted by nations, international organisations, and companies include a discussion of the following five ethical principles: transparency, justice and fairness, responsibility and accountability, privacy, and non-maleficence.[5] These themes were referenced in at least half of the documents analysed across different studies and could indicate some convergence in global thinking on ethical AI.

- *Transparency.* The principle of transparency, or the need to have transparent processes in the development and design of AI algorithms, reflects a commitment to increase interpretability, explainability, or other acts of disclosure. It is one of the most prevalent principles in literature.[6]

- *Justice and fairness.* This principle is expressed mainly in terms of fairness and mitigation of unwanted bias, as a caution to the global community that AI may increase inequality and reinforce societal biases if they are not addressed adequately.[7]

- *Responsibility and accountability.* Despite the widespread references to "responsible AI," responsibility is rarely defined. Recommendations centred on responsibility include clarifying legal liability, focusing on underlying processes that may cause potential harm, or whistleblowing in case of potential harm.[8] Responsibility seems to be intertwined with the principles of transparency and justice such that promoting both these themes can increase responsibility and accountability by AI developers and deployers.

- *Privacy.* While often undefined, privacy is viewed both as a value to uphold and as a right to be protected in ethical AI, and gets presented commonly in relation to data protection and data security.[9]

- *Non-maleficence.* The mention of non-maleficence (encompassing calls for safety and security) exceeded that of beneficence, indicating the precedence of moral obligation to preventing harm over

the promotion of good.[10] This could be due to a negativity bias in characterisation of ethical values concentrating more on negative issues and events rather than positive ones.[11] For instance, existing guidelines do not generally discuss how ethical principles could be promoted through responsible innovation in AI.

Points of Divergence

THERE ARE SUBSTANTIVE divergences across various ethical AI guidelines as analysed by scholars. Most of them relate to following three main factors:

- Interpretation

There are significant differences in how the same principles are interpreted across various guideline documents and the requirements considered important for their realisation. For instance, the need for more datasets to "unbias" AI—to ensure that AI models are trained on representative data in order to avoid flawed or biased conclusions and recommendations—appears to be in conflict with the need to give individuals greater control over their data and ensure privacy. Some guidelines emphasise the need to balance risks and benefits in AI development while others talk of avoiding harm at all costs.[12]

- Attribution

There are also divergences in attribution—interpreting which domain, actor, or issue these ethical principles pertain to. For instance, does the European guideline on privacy (encompassing protection of individual's data from both state and commercial entities) also apply to China where privacy guidelines only target private companies and citizens are accustomed to living in a protected society with high trust in their government?[13] Different perspectives, interpretations, and priorities in ethical AI are of course to be expected given that these documents are developed by a broad range of countries, international organisations, and companies. That said, such divergences could undermine attempts to develop a global ethical AI agenda because varied perspectives, for example risk-benefit evaluations, will lead to different results based on whose well-being they are developed for or the actors involved in developing them.[14]

- Implementation

Finally, there are differing opinions on how ethical AI principles should be implemented—through government organisations, inter-governmental organisations, industry leaders, individual users or developers, or by harmonising AI agendas across the board. If harmonisation is a goal, then how does one account for moral pluralism and cultural diversity across countries, considering that AI is a general-purpose technology operating in varied contexts and cultures?

Persistent Gaps

DISCUSSIONS ON THE ethical development and use of AI are ongoing, and as such, there are gaps that remain unaddressed. For one, themes such as sustainability and solidarity are sparsely referenced across documents.[15] Sustainability appears more commonly in public sector documents versus private sector or non-governmental organisations (NGOs).[16] AI deployment today requires massive computational resources, and hence high energy consumption, and this need will only expand with time. This makes the broader underrepresentation of sustainability-related principles particularly concerning, and calls into question the possibility of harnessing the benefits of AI for the entire biosphere.[17] Solidarity—a concept

mostly referenced in relation to the consequences of AI for the labour market—is also absent in most discussions. There are very few guidelines that pay attention to promoting solidarity by exploring the use of AI expertise for redistributing the augmentation of prosperity for all, and solving socio-economic challenges such as job losses, inequality, and unfair sharing of burdens. Sharing prosperity could mean, for example, compensating humans whose actions provide data for training AI models.[18]

Integrity—meaning being explicit about best practices and disclosure of errors—is another theme that is missing across guideline documents.[19] Current documents place crucial focus on propagating the values of accountability and responsibility, but hardly any emphasise the duty of all stakeholders to develop and deploy AI with integrity. Similarly, the discussion of lack of diversity within the AI community is mostly absent, which is problematic because such dearth of diverse thought could result in flawed AI systems that perpetuate gender and racial biases.[20]

Several initiatives, particularly those offered by industry, are generally criticised as mere virtue signalling designed to debate on abstract problems and delay regulation.[21] In relation to this, it has been observed that many guidelines, especially those produced by the private sector, indicate that technical solutions exist for several of the identified issues, such as privacy and non-maleficence. However, very few guidelines have offered, or at least acknowledged, technical explanations at all; and when they do, they are sparse.[22] While one cannot expect guidelines to be exhaustive about all problems AI could cause, issues pertaining to political abuse of AI systems—generating election fraud, fake news, and propaganda, which are widely acknowledged as critical problems of today—are also an oversight.

Furthermore, shifting the focus from principle-development to implementation is an important next step. However, existing discussions lack clarity on which ethical principles should be emphasised and how existing conflicts in interpretation can be resolved. Moreover, there is a need to determine how conflicts between ethical principles can be resolved and who should enforce oversight and ensure researchers and institutions comply with ensuing guidelines.

Factors for the Convergence, Divergence, and Gaps

THE FIELD OF A.I. ETHICS is expanding. Convergences across the five ethical principles is understandable as it could be a testimony to the significance of those principles; divergences likely reflect the diversity in viewpoints, and gaps could result because most of the work in this domain is still in progress. Having said that, it is crucial to consider other factors possibly influencing these results.

A significant question pertains to equality of participation in the ongoing global discussion on AI ethics. Some scholars have indicated that the current AI ethics discourse is mostly dominated by countries in the Global North.[23] Of the 506 AI-related documents listed in Council of Europe's data visualisation of AI initiatives (as of October 2022), only 10 percent come from countries outside Europe and North America.[24]. Additionally, research indicates that there is a dearth of reference to key terms associated with gender within AI ethics documents and the ratio of female-to-male authors across these documents is about 31 percent.[25] Therefore, like other parts of AI research, the discourse on AI ethics is also primarily shaped by men. The absence of an inclusive AI ethics landscape means that mainstream discussions are reinforcing certain viewpoints while possibly neglecting other risks and ethical considerations of importance to women and countries beyond Europe and North America.

Consensus or dissensus among AI ethics documents could also result due to the provenance of literature. Different types of organisations—public, private, and NGOs—have differing priorities, audiences,

motivations, and scope of responsibility. The public sector is known to emphasise questions related to unemployment and economic growth, while the private sector focuses more on ethical issues with technical fixes (such as transparency and algorithmic bias); for their part, NGOs address a broader range of topics such as accountability and misinformation.[26] In comparison to the private sector, NGOs and public sector entities are reportedly more similar to each other in their approach to AI ethics—they have more participatory processes in creation of guidelines, greater engagement with issues of regulation and law, and more depth and ethical breadth.[27] Consequently, depending on the corpus of documents and types of organisations at hand, an assessment of AI ethics could indicate meaningful variations or similarities in the choice of topics.

The Way Forward

IN A.I. ETHICS, what forms "AI for good" is under negotiation through dialogues among people or organisations impacted by AI development and other intergovernmental initiatives. If calls for more technology access and multi-stakeholder participation are followed, the field is likely to become even more diverse. Narrower versions of the existing themes are likely to emerge with respect to particular geographies and stakeholder groups.[28] This strengthens the case for putting more effort into clarifying the variations that exist within themes and also undertaking measures to resolve differences in interpretation or attribution where possible. If the goal is to have a better articulated ethical AI landscape, the current discourse should be enriched through evaluation of critical but underrepresented principles, such as sustainability and solidarity underlining social and ecological costs of AI.

Beyond a principled approach to AI ethics

While 'principlism' has been the underlying framework to influence the development of safe and beneficial AI, many have questioned its effectiveness. Some critics have pointed out that the field of AI ethics has produced largely vague and high-level principles and value statements. A 2018 study by McNamara et al. reviewed the idea that ethical guidelines serve as a basis for ethical decisions made by developers.[29] The study found that the effectiveness of guidelines is almost negligible since it does not change the behaviour of students or technology professionals.

Relatedly, scholars have indicated that there are other reasons to be concerned about the future impact of AI ethical guidelines.[30] Certain characteristics of AI development indicate that any principled efforts at ethics might not have significant impact on AI's governance and design.

First, the fundamental aims of AI developers, users, and affected parties do not align, and a unified regulatory framework does not exist yet in the field that establishes clear fiduciary duties towards data subjects and users. This means that users cannot trust that developers will act in their best interests when implementing ethical principles in practice. Reputational risks may compel companies, and personal moral conviction may press AI developers towards good behaviour. However, any righteous actions that place public interests before the company and that do not align with company incentive structures are unlikely.[31]

Second, the situation gets further complicated given that AI development lacks a homogenous professional culture, history, moral obligations, and professional standards of what it means to be a "good" AI developer. AI ethics initiatives try to address this gap by offering broadly acceptable guidelines for AI development across radically different contexts of use.[32] But this results in principles or values that

are abstract and based on vague concepts that are not specific enough to guide action and are left to developers to interpret as they see fit.

Third, outside of academic contexts, any principled approach to AI ethics does not have proven methods to transform principles into practice. For instance, the field of medicine has numerous professional societies, accreditation and licensing boards, ethics review bodies, codes of conduct, peer self-governance, and other mechanisms reinforced by strong institutions that ensure ethical conduct on a daily basis.[33] AI development lacks comparable structures to translate guidelines into practice to ensure that this technology, developed behind closed doors, is value-conscious.

Finally, a key weakness for AI is the relative lack of professional and legal accountability mechanisms to redress misbehaviour and ensure that standards are upheld. Research indicates that the existence of mere codes of ethics is not sufficient, and they are often viewed as "checklists" that get pursued in letter rather than spirit.[34] Broader guidelines and self-regulatory efforts alone cannot prevent AI development from failures or misuse, and existing norms and requirements will not be able to set matters right. What makes matters more complicated is that setting up strong accountability mechanisms in AI appears unlikely in the future given that AI is not a unified profession operating in a single sector with a long history of harmonised aims. All of this questions the need for high-level principles as a tool to effect change.

Conclusion

A PLETHORA OF national, international, and commercial AI guidelines in recent years have paved the way for some progress on the development of a principles-led approach to AI. However, one should not celebrate limited consensus on high-level guidelines that conceal deep normative and political disagreements.[35] Instead, it is time to move forward in defining clear long-term pathways, setting explicit professional standards tailored towards specific applications, and building accountability structures that are not only country-specific but also sector- and organisation-specific. Mechanisms should also be set up to license developers of applications with elevated risks, such as facial recognition tools or other systems trained on biometric data.

It will also be interesting to see any future AI principles-based discussions geared toward particular applications of AI, like autonomous vehicles, credit scoring services, recruitment procedure software, or other high-risk AI. There have been instances where ethically motivated efforts have been undertaken to improve AI systems, and most of them have been in specific fields where technical fixes exist for particular problems. For example, many privacy-preserving techniques, like homomorphic encryption or federated learning, or other methods using differential or stochastic privacy, have been developed for the use of data and learning algorithms.[36] A deeper assessment of these context-specific cases to underline guidelines for AI principles could be a way forward.

Admittedly, principles are difficult to translate into practice. However, they still play a crucial role in building awareness and acting as catalysts for building beneficence and a culture of responsibility among AI developers. Internalised norms and values have a role in influencing extrinsic measures, and how individual developers conceptualise, communicate, and enforce extrinsic measures will be crucial in facilitating their implementation. Principles alone cannot govern AI, but nor can rules and requirements.[37] An effective AI governance strategy will require both—principles encouraging cultural change in the AI community, and explicit rules and regulations buttressing them.

Endnotes

1. U.S. National Artificial Intelligence Initiative, *The National AI Advisory Committee (NAIAC),* (Washington, D.C: 2022), https://www.ai.gov/naiac/; Organisation for Economic Co-operation and Development, *OECD creates expert group to foster trust in artificial intelligence,* (2018), https://www.oecd.org/innovation/oecd-creates-expert-group-to-foster-trust-in-artificial-intelligence.htm; European Commission, *High-level expert group on artificial intelligence,* https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai; United Kingdom Parliament Select Committee on Artificial Intelligence, (London: 2017), https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10001.htm.

2. Google AI, "Artificial Intelligence at Google: Our Principles," Google, https://ai.google/principles/; IBM Think Blog, "Transparency and Trust in the Cognitive Era," 2017, https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/; Intel, "Artificial Intelligence: The Public Policy Opportunity," 2017, https://community.intel.com/legacyfs/online/files/Intel-Artificial-Intelligence-Public-Policy-White-Paper-2017.pdf; Microsoft, "Responsible AI," https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6; Sony, "Sony Group's Initiatives for Responsible AI," https://www.sony.com/en/SonyInfo/sony_ai/responsible_ai.html.

3. Daniel Greene et al., "Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning" (paper presented at the Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019), https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/849782a6-06bf-4ce8-9144-a93de4455d1c/content.

4. Fjeld et al. compared 36 documents side by side to identify trends that suggest the earliest emergence of sectoral norms. Zeng et al. collected 27 proposals of AI principles and introduced Linking Artificial Intelligence Principles (LAIP), a platform to link and analyze them. Jobin et al. conducted a scoping review of the existing corpus and analyzed 84 documents of AI ethical guidelines in their paper; Jessica Fjeld et al., "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI," Berkman Klein Center for Internet & Society, 2020, https://dash.harvard.edu/bitstream/handle/1/42160420/HLS%20White%20Paper%20Final_v3.pdf?sequence=1&isAllowed=y; Yi Zeng et al., "Linking Artificial Intelligence Principles" (paper presented in the Proceedings of the AAAI Workshop on Artificial Intelligence Safety, AAAI-Safe AI, 2019), https://arxiv.org/pdf/1812.04814.pdf; Anna Jobin et al., "The global landscape of AI ethics guidelines," *Nature Machine Intelligence,* 389-399 (2019), https://www.nature.com/articles/s42256-019-0088-2.

5. Jobin et al., "The global landscape of AI ethics guidelines"

6. It featured in 73 out of 84 sources analyzed by Jobin et al. and 94 percent of the documents in Fjeld et al.'s dataset; Fjeld et al., "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI"

7. Jobin et al., "The global landscape of AI ethics guidelines"

8. Official Microsoft Blog, "Responsible bots: 10 guidelines for developers of conversational AI," Microsoft Corporation, 2018, https://www.microsoft.com/en-us/research/publication/responsible-bots/; Christina Demetriades and Tom McLaughlan, "Responsible AI and Robotics: An ethical framework", Accenture, https://www.accenture.com/gb-en/company-responsible-ai-robotics.

9. Australian Government, Commonwealth Scientific and Industrial Research Organisation, *Artificial In-*

telligence: Australia's Ethics Framework CSIRO Data61 report: Artificial Intelligence: Australia's Ethics Framework, (Canberra, Australia: 2019), https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjV4Yziiqn6AhWQM1kFHS3RC2kQFnoECAkQAQ&url=https%3A%2F%2Fwww.industry.gov.au%2Fpublications%2Faustralias-artificial-intelligence-ethics-framework&usg=AOvVaw3wz-VNTEhbMSjjq6uW7Acm; Sony, *Sony Group's Initiatives for Responsible AI; Intel, Artificial Intelligence: The Public Policy Opportunity.*

10. Harm is generally interpreted across documents as discrimination, violation of privacy, or physical harm.

11. Thilo Hagendorff, "AI ethics and its pitfalls: not living up to its own standards?" *AI and Ethics,* (2022), https://link.springer.com/article/10.1007/s43681-022-00173-5

12. Jobin et al., "The global landscape of AI ethics guidelines"

13. Pascale Fung and Hubert Etienne, "Can China and Europe find common ground on AI ethics?," *World Economic Forum,* 2021, https://www.weforum.org/agenda/2021/11/can-china-and-europe-find-common-ground-on-ai-ethics/.

14. Jobin et al., "The global landscape of AI ethics guidelines"

15. Some of the documents that mention sustainability (protecting the environment, improving biodiversity, minimizing ecological footprint, creating fairer and equal societies, etc.) are by the Future of Life Institute, Green Digital Working Group, and the French Parliamentary mission. Solidarity is referenced in varied contexts (implications of AI for labor market, calls for a strong safety net, etc.) by the Norwegian Data Protection Authority, U.S. National Science and Technology Council, and by academic researchers at various universities.

16. As per Cathy Roche et al. (2021), the term "sustainable" has been referenced 18/31 times in documents by the public sector, 8/35 times by NGOs, and 2/18 times in private sector documents; Cathy Roche, Dave Lewis and P. J. Wall, "Artificial Intelligence Ethics: An inclusive global discourse?" *arXiv* (2021), https://arxiv.org/pdf/2108.09959.pdf.

17. Sergio Genovesi and Julia Maria Mönig, "Acknowledging Sustainability in the Framework of Ethical Certification for AI," *Sustainability,* 14(7), 4157, (2022), https://doi.org/10.3390/su14074157.

18. Miguel Luengo-Oroz, "Solidarity should be a core ethical principle of AI," *Nature Machine Intelligence* 494, 2019, https://www.nature.com/articles/s42256-019-0115-3.

19. Valerie Carey, "AI Integrity: Leadership Lessons from Other Industries," *Towards Data Science,* February 4, 2022, https://towardsdatascience.com/ai-integrity-leadership-lessons-from-other-industries-82e3d6af2e95.

20. Kelsey Snell, "Lack of diversity in AI development causes serious real-life harm for people of color," NPR, February 13, 2022, https://www.npr.org/2022/02/13/1080464162/lack-of-diversity-in-ai-development-causes-serious-real-life-harm-for-people-of-.; Maria Klawe, "Why Diversity in AI Is So Important," *Forbes,* July 16, 2020, https://www.forbes.com/sites/mariaklawe/2020/07/16/why-diversity-in-ai-is-so-important/?sh=1c64456c7f2b.

21. Brent Mittelstadt, "Principles alone cannot guarantee ethical AI," *Nature Machine Intelligence,* Volume 1, 501-507 (2019), https://www.nature.com/articles/s42256-019-0114-4.

22. Thilo Hagendorff, "The Ethics of AI Ethics: An Evaluation of Guidelines," *Minds and Machines,* 30, 99-120 (2020), https://link.springer.com/article/10.1007/s11023-020-09517-8.

23. The United States and the United Kingdom cumulatively contributed to 40 percent of all the 84 documents analysed by Jobin et al. study. Cathy Roche et al., "Artificial Intelligence Ethics: An inclusive global discourse?"

24. Council of Europe, *AI initiatives,* https://www.coe.int/en/web/artificial-intelligence/national-initiatives

25. Roche et al., "Artificial Intelligence Ethics: An inclusive global discourse?"; Thilo Hagendorff, "The Ethics of AI Ethics: An Evaluation of Guidelines."

26. Daniel Schiff et al., "AI Ethics in the Public, Private, and NGO Sectors: A Review of a Global Document Collection," *TechRxiv* (2021), https://www.techrxiv.org/articles/preprint/AI_Ethics_in_the_Public_Private_and_NGO_Sectors_A_Review_of_a_Global_Document_Collection/14109482/1.

27. Daniel Schiff et al., "AI Ethics in the Public, Private, and NGO Sectors: A Review of a Global Document Collection."

28. Jessica Fjeld et al., "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI"

29. Andrew McNamara et al., "Does ACM's code of ethics change ethical decision making in software development?" (paper published in Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, 2018), https://dl.acm.org/doi/10.1145/3236024.3264833.

30. Brent Mittelstadt, "Principles alone cannot guarantee ethical AI"

31. Jennifer J. Kish-Gephart et al., "Bad apples, bad cases, and bad barrels: meta-analytic evidence about sources of unethical decisions at work," *Journal of Applied Psychology,* 95, 1–31 (2010), https://www.researchgate.net/publication/41087509_Bad_Apples_Bad_Cases_and_Bad_Barrels_Meta-Analytic_Evidence_About_Sources_of_Unethical_Decisions_at_Work; Daisuke Wakabayashi and Scott Shane, "Google will not renew Pentagon contract that upset employees," *The New York Times,* June 1, 2018, https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html.

32. Brent Daniel Mittelstadt, "The ethics of algorithms: Mapping the debate," *Big Data & Society* (2016), https://journals.sagepub.com/doi/10.1177/2053951716679679.

33. Stephen Toulmin, "How medicine saved the life of ethics," *Perspectives in Biology and Medicine,* 25, 736–750 (1982), https://muse.jhu.edu/article/404227.

34. Livia Iacovino, "Ethical principles and information professionals: theory, practice and education," *Australian Academic & Research Libraries,* 33, 57–74 (2002), https://www.tandfonline.com/doi/abs/10.1080/00048623.2002.10755183; Rosamond Rhodes, "Good and not so good medical ethics," *Journal of Medical Ethics,* 41, 1 (2015), https://jme.bmj.com/content/41/1/71.info.

35. Brent Mittelstadt, "Principles alone cannot guarantee ethical AI"

36. John C, Duchi et al., "Privacy Aware Learning," *arXiv* (2013), https://arxiv.org/pdf/1210.2085.pdf; Benjamin Baron and Mirco Musolesi, "Interpretable Machine Learning for Privacy-Preserving Pervasive Systems," arXiv (2020), https://arxiv.org/pdf/1710.08464.pdf.

37. Elizabeth Seger, "In Defence of Principlism in AI Ethics and Governance," *Philosophy & Technology*, 45 (2022), https://link.springer.com/article/10.1007/s13347-022-00538-y

Techno-social Futures: Trapped or Transformative

Gabriella Skoff and Stuart Rollo

THE IMPLICIT ASSUMPTION in technological progress is that each innovation improves lives. Yet, technologies are imbued with politics, sometimes encoding the intentions of their creators. At times too, as Langdon Winner argued in his seminal 1980 paper, the technology is political and favours certain kinds of regimes.[1] What values are implicitly embedded in our current conceptions of emerging tech and tech-driven progress, and do they simply entrench existing political/social structures and problems thereof?

To begin with, developments in science and technology are intimately linked with society and politics. The pathway of techno-progress—often associated with enlightenment values, freedom of enquiry, liberation from drudgery, human interconnectedness, and the provision of plenty—must also be appraised for its darker effects, the displacement of organic human-centred social ends with technical, organisational, and productive efficiency, and a demonstrable record of replicating and reinforcing systems of subjugation, control, and inequity—favouring the few over the many.

Philosopher and psychologist Shoshanna Zuboff's critique of surveillance capitalism highlights some of the ways in which the fourth industrial revolution has birthed technologies that reinforce asymmetries of power.[2] While the intentions of the engineers and programmers themselves may not be as explicit as Jeremy Bentham's notorious panopticon concept for enforcing social control through schematic design, certain technological innovations in the digital space today fall into similar patterns. The observation and interrogation of these patterns of technological progress and their social effects inevitably gives rise to questions of agency: of whether the design process can be shaped so as to overcome the deleterious impacts of technology on society, or whether it is a matter of the very spirit of a technological society that determines these trajectories, regardless of the discrete interventions or characteristics of particular technological projects.

Critiques of Technological Progress

ENQUIRY INTO THE role of technological revolution on social change abounds across different traditions of political and social theory. Marxists have emphasised the role of technological change in altering the dominant modes of economic production as the motive force of historical materialism. Liberals have traced the way that technology fosters and amplifies the spread and interchange of ideas, values, and goods that underpins international cooperation. Realists too, have reconciled their focus on the immutable characteristics of human nature with the role of technology, and particularly military technology, in shaping human society and international competition. Even Sigmund Freud's psychoanalytic theory of civilisation and its discontents highlighted the role of technology in creating many of the problems it purported to solve.[3]

Discussions on the effects of today's new and transformative technologies on human social organisation bring to the fore some of the core debates of the history and philosophy of science. Many prominent 20th-century theorists of technology have pointed to its power to create its own logic of progress, which subsumed human agency in social organisation. Marshal McLuhan, for instance, believed that technology would ultimately determine the shape of human society, not just in character but through the pace and rhythms of life.[4] For his part, Harold Innis observed the centrality of the dominant means of communication to the civilisation that utilised them, with a particular emphasis on their bias towards either 'time', being heavy, durable, and lasting, or 'space', being light, portable, and easily transmittable, determining the structure and characteristics of the social organisation.[5]

Neil Postman, in the earliest years of the digital revolution, foresaw the transition of the West, led by the United States: from a technocracy—within which the traditions of the premodern social and cultural worlds were subordinated but not rendered entirely ineffectual, into a 'technopoly'—which eliminates all alternatives to the instrumentally rational and technological approach to the understanding of humanity and its place in the world.[6] Postman believed that under conditions of technopoly, decisions bearing on the future of humanity would be made along lines that furthered the spread of technology and humanity's reliance on it, rather than on the basis of their own self-formed interests and aspirations, as a result of reliance on technology itself to solve human problems.[7] All three emphasised the centrality of information technology in social organisation, and to varying degrees world order itself, and all appreciated how the temporal characteristics of technology—how it accelerated the transfer of information, the movements of things, what could be accomplished in a period of time—determined its socio-political effects.[8]

Postman's work built on many of the ideas of Lewis Mumford, who had posited in his magisterial work, *Technics and Civilisation,* that technology had, by the mid-20th century, played such an enormous role in shaping the social, political, and cultural practices in the West, that it had come to take the form of a 'technological society'.[9] Other civilisations and cultures had reached high degrees of technical proficiency, but it was only in the modern West that technology had come to dominate all else in shaping the trajectory of human social organisation.[10] Mumford believed that technology could ultimately be tamed and harnessed by a clear-sighted and effective social organisation, under the pro-social principles of what he labelled 'organic ideology' which would be directed towards humanistic rather than instrumental-technological goals.[11]

The surrender of culture and social organisation to technology rendered humanity machine-like, regimented around techno-industrial production and consumption. Martin Heidegger saw the roots of the enslavement of humans to technology in our collective focus on its instrumental rather than essential

characteristics, and our inability to clearly observe that technology was not itself value-neutral.[12] Despite the extreme danger that Heidegger foresaw technology posed to humanity, he glimpsed the potential for a 'saving power', reached by way of understanding the true essence of technology. This, he believed, could free humanity of the binary compulsions of either pushing on blindly with an all-consuming technological 'progress', or rebelling against all technology as the work of the devil.[13]

Challenging the Norms of Techno-progress

SINCE THE 1990s, addressing concerns for accountability of novel technologies has become an increasingly institutionalised process. These concerns can be mapped onto deeply entrenched relations between knowledge and power that have consumed philosophers and social scientists for generations. The governance of novel technologies occurs at many different levels, involves a variety of stakeholders, and encompasses a diversity of approaches. Technology and risk assessments, standardisation, cost-benefit analysis, climate modelling, and the development of ethical technology accords, frameworks, and guidelines are all useful pieces of a complex puzzle. Yet, these retrospective modes of technology governance do not question the notion of techno-progress, nor do they challenge the structure of systems through which novel technologies are developed.

As Prof. Sheila Jasanoff, a prominent scholar of Science and Technology Studies, writes: "The analytic ingenuity of modern states has been directed toward refining what we may call the 'technologies of hubris'. To reassure the public, and to keep the wheels of science and industry turning, governments have developed a series of predictive methods… that are designed, on the whole, to facilitate management and control, even in areas of high uncertainty".[14] As a salve to the 'technologies of hubris' resulting from current systems of innovation, Jasanoff prescribes the concept of 'technologies of humility'. These technologies are created by methods and modes that can serve "to make apparent the possibility of unforeseen consequences; to make explicit the normative that lurks within the technical; and to acknowledge from the start the need for plural viewpoints and collective learning."[15]

While Jasanoff's conceptualisation may appear optimistic, recent years have witnessed a shift in how a broader segment of society understands and responds to the potential and real harms that technologies pose. This offers a glimmer of hope for Heidegger's vision of how society can engage with the concept of technological progress in a more nuanced way and even direct its trajectory toward beneficent ends. Many such cases related to Big Tech have received sustained public attention. From the Cambridge Analytica scandal that made use of Facebook's platform to obtain user's personal data for political profiling that played a role in both Brexit and the election of Donald Trump, to the harms wrought on Amazon's employees through the use of their AI systems.[16] These cases have led not only to legal action on the part of Facebook and a unionising movement for Amazon workers but also to a thriving public conversation around data privacy and the integration of AI into workplace systems.

In 2018, a number of Google employees resigned in relation to the company's contract with the Pentagon to develop AI technology for a defence project known as Project Maven. Google later withdrew from the contract as public pressure mounted.[17] More recently, in 2020, Timnit Gebru, a computer scientist and AI ethics researcher who was co-leader of Google's ethical AI team, was fired for attempting to publish a paper that highlighted the potential harms of the large language models Google was developing. Gebru's public announcement of the situation led many to question the approach of 'self-regulation' that Big Tech has taken. It also led Gebru, along with other prominent and diverse professionals in the AI field, to start a foundation for AI research that aims to understand and uncover the harms that AI can pose to marginalised communities.[18]

These examples, a few of many, illustrate a new focus on the relationship between the researchers and designers of new technologies, the companies they work for, and accountability. They demonstrate how the political power of techno-progress is being challenged by individuals and groups with an understanding of the system from the inside. It also demonstrates the value of public engagement, as the inner workings of advanced technological systems that touch the lives of the vast majority of the world in a multitude of quotidian ways are shrouded in opacity, require detailed technological knowledge to penetrate. The work done by these individuals illuminates the structural biases and systemic inequities that impede the development of fairer and more egalitarian technologies within a technopoly that has been accelerated blindly in recent years by the 'move fast and break things' culture of Big Tech.

Winner observes that "in the processes by which structuring decisions are made, different people are differently situated and possess unequal degrees of power as well as unequal levels of awareness."[19] Peeling back the narrative of de-risked techno-progress reveals the underlying architecture of power and disempowerment that lies beneath systems of the commercial innovation landscape.

Public education and participation, structural incorporation of a greater diversity of perspectives, experiences, and expertise, reflexive learning and action, formalised lines of communication between each of these nodes, and decommodification—all incorporated by design from the very initiation of the research and development process of novel technologies—may create the conditions for technologies that are embedded in and shaped by human social values. However, these transformations must occur at the ground-level of technological development, not as a virtue-signalling band-aid but through unravelling the threads of the complex tapestry of technological interventions into human social, political, and economic life that lie at the heart of techno-progress.

Quantum: The Next Technological Frontier

IN HIS SEMINAL essay on the subject, 'Do Artifacts Have Politics?', Langdon Winner engages with the tradition of taking the determinative characteristics of technics seriously, arguing that, consciously or not, societies choose structures for technologies that influence almost every aspect of our lives.[20] The moment with the greatest potential for exercising human and social agency in making these choices is, according to Winner, at the very beginning of the introduction of a new technology, "because choices tend to become strongly fixed in material equipment, economic investment, and social habit, the original flexibility vanishes for all practical purposes once the initial commitments are made."[21] Thus certain types of technologies can be shaped and directed from the outset of their invention and incorporation into human activity to serve democratic or humanistic, rather than simply instrumental-technocratic, purposes.

Quantum computing, communications, and sensing technologies offer a relevant case study for the interrogation of this position. Still in the early stages of research and development, the many ways that the third quantum revolution is likely to reshape military and economic affairs is becoming clearer as the technology develops. A more complex, and equally important, task lies in enquiring into the essential manner by which this revolution is likely to effect human social organisation. From the understanding here formed, we can then gauge its impact on human security, and human flourishing. Many areas promise to be influenced by this new technology: social inequality, civil rights and privacy, civic discourse and democracy, and the patterns of economic production and the distribution of wealth, both within and between states. If the quantum era is to avoid being defined by its negative social and geopolitical consequences, and demonstrate that there is indeed a 'saving power' through which a technological society can exert pro-social and humanistic influence over potent new technologies, serious effort must be made in its

nascent stages to form normative practices and make material interventions into the design of quantum technology's applications and availability that promote its peaceful, equitable, and ethical use.

Perhaps the first key task in directing a different course for quantum lies in parsing the effects of the technology itself, from pre-existing political and social problems, which may be exacerbated by quantum, but which are not necessarily caused by it. The military applications of quantum technology are clearly technical, and it is conceivable that, much as has occurred with arms control agreements like the Intermediate-Range Nuclear Forces treaty, the Chemical Weapons Convention, and the Comprehensive Nuclear-Test-Ban treaty, accords could be reached which limit the immediate destructive potential of these technologies. However, many of the higher-order strategic implications of quantum, ones concerning the configuration and power distribution of world order, human security, inequality, and issues of state power, are not technical problems themselves generated by quantum. Rather they are political and social problems that underpin current systems of technological innovation. Should these foundations remain unchallenged, they are likely to exacerbate existing inequities in both expected and unexpected ways as quantum begins to shape our world.

Is it Too Late for Quantum?

THE MUCH ANTICIPATED next generation of quantum technologies remain, for the most part, in the experimental phases of development. Quantum's vast implications for technological advantage, in combination with an international climate of escalating geopolitical tension, does not bode well for its collaborative, egalitarian, and peace-oriented development.

The challenges thrown up by contemporary issues of international security and world order are exacerbated by the material underpinnings of quantum technology itself. In 'Do artifacts have politics?', Winner expands upon a particularly troubling claim, raised by the thinkers mentioned above and many others—it is that some technologies are by their nature political, and that their adoption will unavoidably lead to the orientation of society towards certain social and political structures—authoritarian or democratic, egalitarian or inegalitarian, repressive or liberatory.[22] The determining power of such technologies can be found in both their capacity to disrupt established social and economic order (such as the steam engine, the cotton gin, or the telegraph of the industrial revolutions), as well as in the complexity and resource intensiveness of their own invention and operation.

It is only a highly organised and technologically advanced society that is capable of producing, for example, a quantum computer—which requires large amounts of capital investment, supercooling facilities shielded from even minuscule levels of ambient environmental interference, the supply of critical raw materials, and intellectual capital in the form of mastery of quantum theory, scientific, and engineering expertise. It is a society which forms great concentrations of material and technical wealth and concentrates political power in such a way as to direct the use of this wealth towards speculative technological ends. While there is some potential in distributing the computational power of quantum computing systems through cloud-based networks, it is clear that under current conditions the power of quantum computing will be far more centralised and controlled than is the computing of today, or of the earlier era of internet and personal computing pioneers.

Overcoming these structural and material characteristics of quantum technology to assert democratic, egalitarian, and liberatory political effects is a tall task. If it can be done at all, it must, as Winner observes, be prioritised immediately, while the technology is still being developed. The attempt to assert a meaningful quantum ethics reflects the qualified optimism of Mumford, and even allows for incorporating

the emphasis on art and the artist as interlocutor into the nature of technology and its interaction with humanity posited by Heidegger.

Producing quantum technologies responsive to the demands of a more egalitarian society necessitates the far deeper understanding of the social impact of new technologies exemplified by the work of many scholars over the past century or so. This work will require both those inside and outside of the modes of technological production to "imagine new institutions, processes, and methods for restoring to the playing field of governance some of the normative questions that were sidelined in celebrating the benefits of technological progress."[23] It also requires an accounting of how these quantum effects will differ globally, and what this will mean for the formulation of agreements and accords that will seek to limit or mediate the technology in ways that benefit humanity.

Endnotes

1. Langdon Winner, "Do Artifacts Have Politics?," *Daedalus 109,* no. 1 (1980).
2. Shoshana Zuboff, *The Age of Surveillance Capitalism,* (London, England: Profile Books, 2019)
3. Sigmund Freud, *Civilization and Its Discontents,* (Broadview Press, 2015), pp. 26
4. Marshall McLuhan, "The Medium Is the Message," in *Understanding Media: The Extensions of Man* (London & New York: Signet, 1964), pp. 7
5. Harold Adams Innis, *Empire and Communications* (Rowman & Littlefield, 2007), pp. 20
6. Neil Postman, *Technopoly: The Surrender of Culture to Technology* (New York: Vintage Books, 1992), pp. 56-58
7. Postman, *Technopoly: The Surrender of Culture to Technology*
8. McLuhan, "The Medium Is the Message", pp. 7; Innis, *Empire and Communications,* pp. 16; Postman, *Technopoly: The Surrender of Culture to Technology,* pp. 45
9. Lewis Mumford, *Technics and Civilization* (University of Chicago Press, 2010)
10. Mumford, *Technics and Civilization,* pp. 4
11. Mumford, *Technics and Civilization,* pp. 368
12. Martin Heidegger, "The Question Concerning Technology," in *The Question Concerning Technology and Other Essays* (New York & London: Garland Publishing, 1977), pp. 4
13. Heidegger, "The Question Concerning Technology," pp. 25-26
14. Sheila Jasanoff, "Technologies of Humility: Citizen Participation in Governing Science," *Minerva 41,* 223–244 (2003), pp. 240
15. Jasanoff, "Technologies of Humility: Citizen Participation in Governing Science," pp. 240

16. Rosalie Chan, "The Cambridge Analytica Whistleblower Explains How The Firm Used Facebook Data To Sway Elections," *Business Insider,* October 6, 2019, https://www.businessinsider.in/tech/news/the-cambridge-analytica-whistleblower-explains-how-the-firm-used-facebook-data-to-sway-elections/articleshow/71461113.cms; Jay Greene, "Amazon's employee surveillance fuels unionization efforts: 'It's not prison, it's work'," *The Washington Post,* December 2, 2021, https://www.washingtonpost.com/technology/2021/12/02/amazon-workplace-monitoring-unions/

17. Daisuke Wakabayashi and Scott Shane, "Google Will Not Renew Pentagon Contract That Upset Employees," *The New York Times,* June 1, 2018, https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html

18. Nitasha Tiku, "Google fired its star AI researcher one year ago. Now she's launching her own institute," The *Washington Post,* December 2, 2021, https://www.washingtonpost.com/technology/2021/12/02/timnit-gebru-dair/

19. Winner, "Do Artifacts Have Politics?," pp. 127

20. Winner, "Do Artifacts Have Politics?," pp. 127

21. Winner, "Do Artifacts Have Politics?," pp. 127

22. Winner, "Do Artifacts Have Politics?," pp. 128

23. Jasanoff, "Technologies of Humility: Citizen Participation in Governing Science," pp. 226

Rules

Assessing India's Position on Data Protection

Basu Chandola

THE IMPORTANCE OF data and data processing cannot be overstated. Data has become an extremely valuable commodity, and data processing is now used in almost all sectors of the economy, from business to governance. But the true potential of data is not yet fully realised. To ensure that this capacity of data is harnessed, it is essential to have the appropriate regulatory infrastructure.

India has made several attempts to regulate personal and non-personal data in recent years. However, the country still does not have a comprehensive legal framework on data. This paper assesses all data-related laws, policies, and studies undertaken by the Indian government to better understand India's position on data. It only looks at general data policies and does not cover specialised data (such as financial, corporate and compliance, or healthcare data).

India's Data-Related Legal Landscape

THE INFORMATION TECHNOLOGY Act, 2000 (IT Act) and the supplementary legislations issued thereunder form the data protection framework in India. In its original form, the IT Act did not have any provisions for data protection. However, an Expert Committee was set up in January 2005 to review the provisions of the Act,[1] and recommend suitable legislation for data protection under it. Based on the committee's recommendations, the government introduced the Information Technology (Amendment) Bill, 2006.[2] The Bill was referred to the Standing Committee[a] on IT, which submitted its report in September 2007,[3] and its recommendations became the basis of the Information Technology (Amendment) Act, 2008 (IT Amendment Act).[4]

The IT Amendment Act incorporated Sections 43A and 72A, which now form the basis for data protection in the country—Section 43A provides compensation for the failure of corporates to protect sensitive personal data, while Section 72A provides

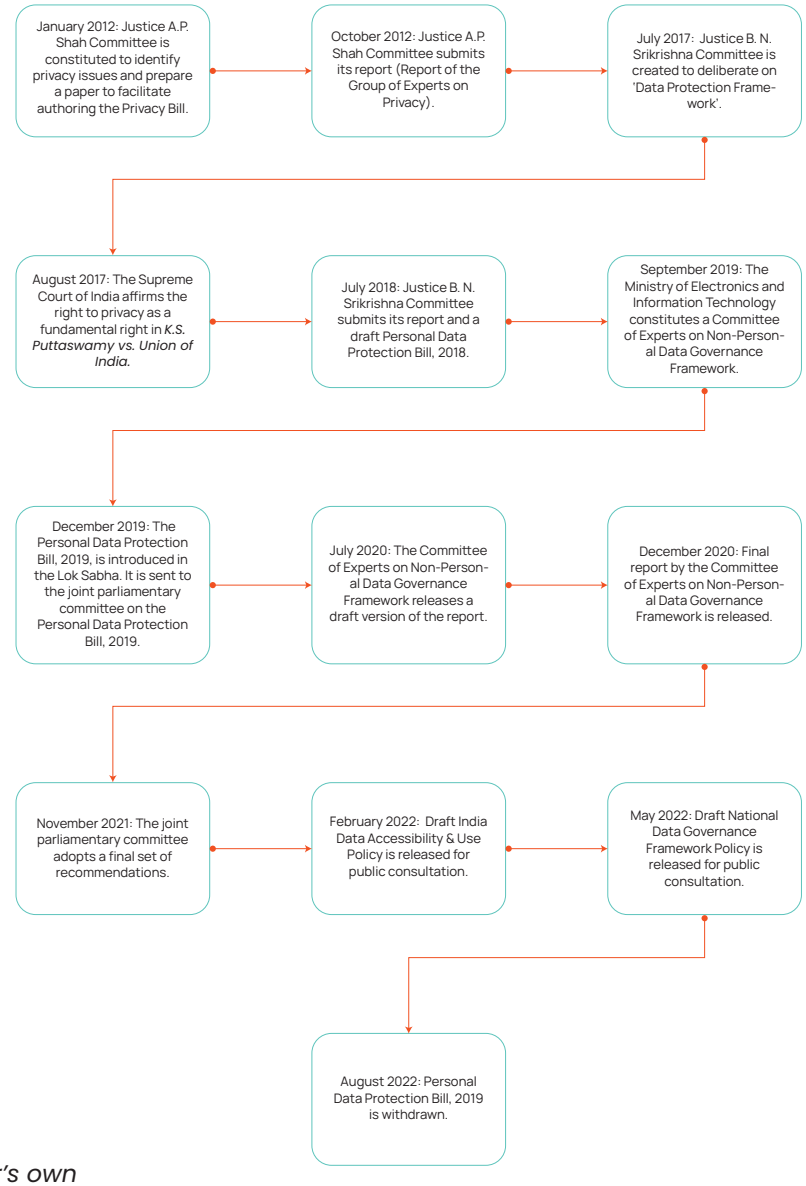
[a] Standing Committees are permanent and regular committees constituted in pursuance of the provisions of an Act of parliament, rules of procedure, and conduct of business in the Lok Sabha.

for punishment for the disclosure of information in breach of lawful contract. In addition, using the rule-making power under Section 87(2)(ob) read with Section 43A of the IT Act, the government enacted the Information Technology (Reasonable security practices and procedures and sensitive personal data or information) Rules, 2011 (IT Rules)[5] that impose additional responsibilities on businesses regarding the collection and disclosure of sensitive personal data and information. The IT Rules explain the scope of 'sensitive personal data' and impose obligations on corporates to provide privacy policies on their website. They also expand on the procedures for collecting, disclosing, and transferring data, and specify reasonable security practices that a corporate must comply with while collecting data.

Recent Efforts on Data Protection

INDIA IS IN the midst of a massive digital transformation, with citizens becoming increasingly dependent on digital services and creating a significant quantum of data. It is one of the fastest-growing data-generating countries, producing approximately 150 exabytes of data annually.[6] Despite this massive transition, the country lacks a uniform and comprehensive data protection law. While there have been multiple attempts in recent years to establish comprehensive legislation, these efforts have not been very fruitful (see Figure 1).

Figure 1: Timeline of Data-Related Developments in India



Source: Author's own

- ## Justice A.P. Shah Committee

The Justice A.P. Shah Committee was established in 2012 with three objectives: to study privacy laws in other countries, assess the impact of the Indian government's programmes on privacy, and make recommendations on the draft privacy law.[7] In its final report,[8] the committee recommended a detailed framework that could act as the conceptual foundation for India's privacy law. It acknowledged that privacy is multidimensional, and that any framework on the right to privacy should include privacy-related concerns around data protection on the internet. The committee also acknowledged the economic importance of data and the risks to privacy in transborder data flows.

- ## K.S. Puttaswamy vs. Union of India

In *K.S. Puttaswamy v Union of India,*[9] a nine-judge constitutional bench of the Supreme Court had the opportunity to determine if the Indian Constitution encompasses a right to privacy, and whether privacy is a constitutionally protected value. Through six different opinions, the bench held that the "right to privacy is protected as an intrinsic part of the right to life and personal liberty under Article 21 and as a part of the freedoms guaranteed by Part III of the Constitution".[10] Thus, the right to privacy has been declared a fundamental right and can be traced to Articles 14, 19, and 21[b] of the Indian Constitution. While debating the issue, the Supreme Court also acknowledged and considered the needs, opportunities, and dangers posed to liberty in a global information-based society.

- ## Justice B. N. Srikrishna Committee

In July 2017, the Ministry of Electronics and Information Technology (MeitY) constituted an expert committee chaired by Justice B.N. Srikrishna to study and identify key data protection issues.[11] One of the major motivations for establishing the committee was that data protection could provide a big boost to India's digital economy. The government recognised the importance of data protection, and the need to balance the growth of the digital economy with protecting citizens' personal data.[12]

The committee suggested that a strong data protection framework is necessary for India to "harness the benefits of the digital economy and mitigate the harms consequent to it".[13] It opined that data protection is closely related to informational privacy, and that privacy must be seen as a legal right over personal information.[14] In its final report, the committee noted that the "protection of personal data holds the key to empowerment, progress, and innovation".[15] It added that data has the potential to benefit citizens immensely and that the digital economy has massive transformative potential to improve lives. But it also added a caveat that data could harm citizens as it could be used for discrimination and exclusion. The committee noted that there is a need for a "free and fair digital economy that empowers the citizen" based on individual autonomy and to maximise the common good.

- ## Committee of Experts on Non-Personal Data Governance Framework

In September 2019, MeitY constituted a Committee of Experts on Non-Personal Data Governance Framework to study the various issues pertaining to non-personal data and make recommendations to the government on regulating such data.[16] The committee released a draft report in July 2020[17]and its final report in December 2020.[18]

---

[b] Articles 14, 19, and 21 provide for the fundamental rights of equality, freedom, and life and liberty, respectively.

The committee noted that data creates economic value and wealth in addition to social and public value, and that data plays a central role in all economic sectors. It observed that data offers intrusive information and has the potential to cause harm to individuals and communities, but data is necessary for social and political interests. It also noted that welfare losses are caused when such data is closed to the public.
The committee suggested creating a framework that "establishes rights of India and its communities over its non-personal data" and allows the realisation of "economic benefit from non-personal data for India and its people". It added that the benefits from processing non-personal data must not be limited to the entities collecting such data but must be accrued to the communities that produce the data. This would allow for the realisation of economic benefits for citizens and communities, and unlock the data's social, public, and economic value. The committee recommended creating a single national-level regulation to control and establish rights over the non-personal data collected and created across the country.

- ## The Personal Data Protection Bill, 2019, and the Joint Parliamentary Committee Report

The Personal Data Protection Bill, 2019 (PDP Bill)[19] was introduced in the Lok Sabha in 2019. The PDP Bill was referred to the joint parliamentary committee (JPC), which released its report[20] in December 2021. The JPC recommended major changes to the PDP Bill, including rechristening the regulation as the Data Protection Bill, 2021, and expanding its purview to regulate non-personal data. The JPC proposed more than 80 drafting changes in the original Bill. Following the JPC report, the government withdrew the PDP Bill in August 2022, stating that it would introduce a revised legislation that provides a comprehensive legal framework on personal data.[21]

Even though the PDP Bill has been withdrawn, an assessment of it and the JPC's report can help better understand India's approach to data. The statement of objectives in the PDP Bill summarises that the protection of personal data is an essential facet of informational privacy; that the use of data has expanded in the digital economy; and that there is a need to create a culture facilitating "free and fair digital economy, respecting the informational privacy of individuals, and ensuring empowerment, progress, and innovation through digital governance and inclusion".

The JPC report considers data to be the fuel of the new economy and an "asset of national importance". It noted that data can be used for global socioeconomic transformation since it provides insights and allows for the individual, group, and global activity prediction. The JPC also noted that the true potential of data is yet to be tapped comprehensively in India and that it is necessary to have the right infrastructure and data governance mechanisms to unleash this. The committee also noted that data is a significant enabler of digital governance.

- ## Draft India Data Accessibility & Use Policy, 2022

The MeitY released the Draft India Data Accessibility & Use Policy, 2022[22] (IDAUP) for public comments in February 2022. The legislation aims to enhance "access, quality, and use of data" to meet the requirements of current and emerging technologies. The policy intends to transform "India's ability to harness public sector data for catalysing large-scale social transformation". Severe criticism emerged amid the public consultations, resulting in several changes to the draft.[23] Despite these changes, the policy drew widespread criticism for its plan to monetise peoples' non-personal data.[24] There have been no further updates about the IDAUP.

- Draft National Data Governance Framework Policy

In May 2022, MeitY released the Draft National Data Governance Framework Policy (NDGFP),[25] considered the successor to the IDAUP. A data governance framework essentially provides the rules, policies, standards, and processes to manage the various aspects of data in an entity's system, including the usability, availability, integrity, and security of data. Although government data is stored and managed haphazardly (with data management protocols varying from department to department), the NDGGP aims to provide a uniform system for data governance across different government entities.

The NDGFP aims to revamp the government's data collection and management processes and provide a data-led governance approach to improve the delivery of services. It also seeks to create a repository of Indian datasets to help with artificial intelligence (AI) and data-led research and bolster the AI and analytics ecosystem. It also seeks to provide rules and standards to ensure data security and informational privacy. The NDGFP will apply to all central government departments and cover all personal and non-personal data with the government. While the policy does not apply to state governments, they are encouraged to adopt it.

India's Position on Data

INDIA'S POLICY POSITION on data can be summarised as:

- India considers data a national asset.
- The true potential of data is yet to be tapped comprehensively in the country, and the government is looking to fully realise the economic and social welfare value of data.
- India considers personal data protection an essential facet of informational privacy.
- India is likely to continue including data localisation requirements in future data regulations.
- India sees data as a catalyst for digital transformation and innovation, and is increasingly looking to adopt data-supported digital governance.
- Indian data regulations aim to balance two main objectives: maximising the benefits of data and the expansion of digital markets; and minimising privacy challenges.

In addition to the previously discussed policies, India has several ancillary policies on data. The 2019 Draft National E-Commerce Policy[26] notes that India and its citizens have a sovereign right over their data, and that the data is "best thought of a collective resource, a national asset, that the government holds in trust, but rights to which can be permitted". Similarly, the 2018-19 Economic Survey noted that data should be "of the people, by the people, for the people",[27] and that the government can create data as a 'public good'. The Economic Survey recommended that data be used for social welfare or monetised, while considering privacy concerns. A similar opinion was echoed in 2020 by former IT Minister Ravi Shankar Prasad who described data as a 'national asset' that needs to be utilised properly to achieve prosperity in various sectors.[28]

Treating data as a 'national asset' can have crucial implications: the regulations can mandate that data generated by Indian citizens must necessarily be stored within the national boundaries, and the country

can reserve the right to use that data.[29] India's efforts towards data localisation further indicate its position on viewing data as a collective resource over which citizens have a sovereign right.[30] India considers data a "foundational raw material for enabling domestic data businesses as well as data-driven governance by state actors".[31] This approach of considering data a 'national asset' is in stark contrast to the rights-based approach of the EU's General Data Protection Regulation (GDPR).[32]

Rajeev Chandrasekhar, India's Minister of State for Electronics and IT, noted that the GDPR is "a little bit more absolutist in terms of how they approach data protection" and that this is not feasible for India.[33] He added that India's requirements differ from that of the EU, and that India must arrive at a solution that addresses the "issues of security and consumers' rights to data protection".[34]

India is also in the midst of "a concerted diplomatic mission to convince global players of the vitality of its "data sovereignty" vision".[35] For instance, India refused to sign the G20's 2019 Osaka Declaration as it conflicted with its policy on data localisation. Indian policy efforts also show "growing assertions of technological self-reliance and sovereignty in data governance".[36]

India's experience developing a data governance framework shows that while data protection is important, data sharing and empowerment are integral to its approach.[37] The country's approach is based on the premise that data is an important tool for economic and social growth and national security; therefore, data sovereignty and localisation will continue to feature in future data governance frameworks.

After withdrawing the PDP Bill, the government has already stated it will introduce a revised and comprehensive data protection law. While there have been many speculations on the upcoming law,[38] what is almost certain is that it will contain some essential elements—such as considering data a national asset, featuring data localisation requirements, and developing policy to maximise the benefits of data while minimising privacy challenges.

Endnotes

1. Ministry of Communications, Government of India, https://pib.gov.in/newsite/erelcontent.aspx?relid=6372
2. The Information Technology (Amendment) Bill, 2006, http://www.cyberlawclinic.org/Amendments.pdf
3. Lok Sabha Secretariat, *Fiftieth Standing Committee Report on Information technology (amendment) Bill, 2006*, New Delhi, 2007, https://prsindia.org/files/bills_acts/bills_parliament/2006/scr1198750551_Information_Technology.pdf
4. The Information Technology (Amendment) Act, 2008, https://eprocure.gov.in/cppp/rulesandprocs/kbadqkdlcswfjdelrquehwuxcfmijmuixngudufgbuubgubfugbububjxcgfvsbdihbgfGhdfgFHytyhRtMT-k4NzY=
5. Ministry of Communications and Information Technology, *The Information Technology (Reasonable security practices and procedures and sensitive personal data or information) Rules, 2011,* https://www.meity.gov.in/writereaddata/files/GSR313E_10511%281%29_0.pdf
6. Joint Committee on the Personal Data Protection Bill, 2019, *Report of the Joint Committee on*

the Personal Data Protection Bill, 2019, Delhi, 2021, http://164.100.47.193/lsscommittee/Joint%20Committee%20on%20the%20Personal%20Data%20Protection%20Bill,%202019/17_Joint_Committee_on_the_Personal_Data_Protection_Bill_2019_1.pdf

7. Planning Commission, Government of India, https://pib.gov.in/newsite/PrintRelease.aspx?relid=88503

8. Planning Commission, *Report of the Group of Experts on Privacy, Delhi,* 2012 https://cis-india.org/internet-governance/blog/report-of-group-of-experts-on-privacy.pdf

9. *K.S. Puttaswamy v Union of India,* (2017) 10 SCC 1

10. *K.S. Puttaswamy v Union of India*

11. Ministry of Electronics & IT, Government of India, https://pib.gov.in/newsite/PrintRelease.aspx?relid=169420

12. Ministry of Electronics & Information Technology, Government of India, https://www.meity.gov.in/writereaddata/files/MeitY_constitution_Expert_Committee_31.07.2017.pdf

13. Ministry of Electronics & Information Technology, *White Paper of the Committee of Experts on a Data Protection Framework for India,* Delhi, 2017, https://www.meity.gov.in/writereaddata/files/white_paper_on_data_protection_in_india_171127_final_v2.pdf

14. Ministry of Electronics & Information Technology, *White Paper of the Committee of Experts on a Data Protection Framework for India*

15. Ministry of Electronics & Information Technology, *A Free and Fair Digital Economy Protecting Privacy, Empowering Indians*, by Justice B.N. Srikrishna Committee, Delhi, 2018, https://www.meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf

16. Ministry of Electronics & Information Technology, Government of India, https://www.meity.gov.in/writereaddata/files/constitution_of_committee_of_experts_to_deliberate_on_data_governance_framework.pdf

17. Ministry of Electronics & Information Technology, *Report by the Committee of Experts on Non-Personal Data Governance Framework,* Delhi, 2020, https://ourgovdotin.files.wordpress.com/2020/07/kris-gopalakrishnan-committee-report-on-non-personal-data-governance-framework.pdf

18. Ministry of Electronics & Information Technology, *Final Report by the Committee of Experts on Non-Personal Data Governance Framework,* Delhi, 2020, https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf

19. The Personal Data Protection Bill, 2019, http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf

20. Joint Committee on the Personal Data Protection Bill, 2019, *Report of the Joint Committee on the Personal Data Protection Bill, 2019*

21. Soumyarendra Barik, "Explained: Why the Govt has withdrawn the Personal Data Protection Bill, and what happens now", *Indian Express*, August 6, 2022, https://indianexpress.com/article/explained/explained-sci-tech/personal-data-protection-bill-withdrawal-reason-impact-explained-8070495/

22. Ministry of Electronics & IT, Government of India, *Draft India Data Accessibility & Use Policy 2022*, https://www.meity.gov.in/writereaddata/files/Draft%20India%20Data%20Accessibility%20and%20Use%20Policy_0.pdf

23. "The Government Wants To Sell Your Data | #SaveOurPrivacy", *Internet Freedom Foundation*, 2022, https://internetfreedom.in/the-government-wants-to-sell-your-data/

24. Apar Gupta, "Why draft data accessibility policy is dangerous", *The Indian Express,* March 05, 2022, https://indianexpress.com/article/opinion/columns/draft-data-accessibility-policy-privacy-suveillance-7801714/

25. Ministry of Electronics & IT, Government of India, *Draft National Data Governance Framework Policy*, https://www.meity.gov.in/writereaddata/files/National%20Data%20Governance%20Framework%20Policy_26%20May%202022.pdf

26. Ministry of Commerce & Industry, Draft National e-Commerce Policy, https://dpiit.gov.in/sites/default/files/*DraftNational_e-commerce_Policy*_23February2019.pdf

27. Ministry of Finance, *Economic Survey 2018-19*, Delhi, 2019, https://www.indiabudget.gov.in/budget2019-20/economicsurvey/doc/echapter.pdf

28. "Data nation's asset; must be secured: Ravi Shankar Prasad", *Economic Times*, July 20, 2020, https://government.economictimes.indiatimes.com/news/digital-india/data-nations-asset-must-be-secured-ravi-shankar-prasad/77060909

29. Vijay Govindarajan, Anup Srivastava, and Luminita Enache, "How India Plans to Protect Consumer Data", *Harvard Business Review*, December 18, 2019, https://hbr.org/2019/12/how-india-plans-to-protect-consumer-data

30. Debashree Mukherjee, Kanwarpreet Singh and Goldie Dhama, "Data localisation norms: A key pillar for privacy protection", *PWC*, 2019, https://www.pwc.in/assets/pdfs/consulting/cyber-security/data-privacy/data-localisation-norms-a-key-pillar-for-privacy-protection.pdf

31. Amba Kak and Samm Sacks, "Shifting Narratives and Emergent Trends in Data-Governance Policy: Developments in China, India, and the EU", *Yale Law School – Paul Tsai China Center*, 2021, https://law.yale.edu/sites/default/files/area/center/china/document/shifting_narratives.pdf

32. Mayuran Palanisamy and Jignesh Oza", "An introductory analysis of the Joint Parliamentary Committee's report on Personal Data Protection Bill, 2019", *KPMG*, March 01, 2022 https://home.kpmg/in/en/blogs/home/posts/2022/03/joint-parliamentary-committee-personal-data-privacy-mantra.html

33. Soumyarendra Barik, "Rajeev Chandrasekhar: EU's GDPR more absolutist, not possible for us … our law will be clear on data misuse", *The Indian Express*, September 26, 2022, https://indianexpress.com/article/business/rajeev-chandrasekhar-eus-gdpr-more-absolutist-not-possible-for-us-our-law-clear-data-misuse-8172858/

34. Soumyarendra Barik, "Rajeev Chandrasekhar: EU's GDPR more absolutist, not possible for us … our law will be clear on data misuse"

35. Amber Sinha and Arindrajit Basu, "The Politics of India's Data Protection Ecosystem", *EPW Engage*, No. 54 (49) 14 (2019) https://www.epw.in/engage/article/politics-indias-data-protection-ecosystem

36. Smriti Parsheera, "What's Shaping India's Policy on Cross-Border Data Flows?" in Data Governance, Asian Alternatives: How India and Korea Are Creating New Models and Policies Evan, ed. A. Feigenbaum and Michael R. Nelson (Carnegie Endowment for International Peace, 2022)

37. Rahul Matthan and Shreya Ramann, "India's Approach to Data Governance" in Data Governance, Asian Alternatives: How India and Korea Are Creating New Models and Policies Evan, ed. A. Feigenbaum and Michael R. Nelson (Carnegie Endowment for International Peace, 2022)

38. Andy Mukherjee, "India's new rules for data privacy may be more like China's than Europe's", *Business Standard*, August 9, 2022, https://www.business-standard.com/article/economy-policy/india-s-new-rules-for-data-privacy-may-be-more-like-china-s-than-europe-s-122080900115_1.html

Destructive ASAT Testing: A Risk to India's Space Interests

Daniel Porras

INDIA MAY OFTEN be called an "emerging space power", but it has proven many times that it is already an established power in this domain. Its feats in the recent years include the "most payloads in a single launch",[1] and the most cost-effective mission to Mars in human history.[2] Like the other formidable space actors—i.e., the United States (US), Russia, and China—India is admired around the world for its place among the stars; and like those three, India is facing myriad challenges related to outer space safety and security.

One challenge comes from the threat to space objects from space debris. It is no secret that low-Earth orbit (LEO, roughly 100-1,500km) is quickly being populated by more satellites than ever before. Operating in LEO could soon become far more difficult due to the accumulation of trash, traveling at high-speed, in that orbit. A single collision with some of these objects can damage or completely destroy a space object. It is critical, therefore, to not only slow down, but reverse the proliferation of debris to ensure the long-term sustainability of outer space activities.

Due to the shared nature of earth's orbits, it is vital that global actors seek international solutions to the challenges in space. Failing to do so could lead to a "tragedy of the commons", as being seen here on Earth. For this reason, United Nations (UN) agencies are examining various options to minimise and mitigate the spread of space debris. One option currently being examined is the voluntary cessation by states of destructive anti-satellite (ASAT) testing. A resolution was meant to be tabled before the UN General Assembly this fall, seeking to commit states to not conduct such tests, but was not eventually presented.

India should support this Resolution to stop the destructive testing of ASAT weapons; after all, such a policy will help India overcome the physical, political, and legal challenges to its own space aspirations. A policy of not conducting destructive

ASAT tests will mitigate threats from physical debris to its space assets, cast India as a leader and responsible actor in space, and avoid unnecessary legal liability under international space law. Moreover, India will still be free to pursue counterspace capabilities and remain one of the world's preeminent space powers. Given the scope of India's current space ambitions, it has more to lose from destructive ASAT testing than it does to gain.

India's Space Capabilities

INDIA'S SPACE POWER comes from its own satellites and launch capabilities. In terms of its fleet, there are roughly 60 satellites registered as having Indian origin, nearly all of which are government-operated.[3] These satellites provide critical services including navigation, communications, and earth imaging and science. Without them, key sectors of public infrastructure including health, sanitation, and finance will be hampered. Indeed, it has been said that the average Indian depends on 11 satellites per day. This makes space-based services as essential to daily life as public transportation or electricity. Moreover, India has one of the strongest launch programmes in the world. Prior to the COVID-19 pandemic, India was keeping pace with the countries of Europe—[4] a considerable feat considering it has less financial resources available for its own national space programmes. This achievement can be attributed to the fact that India is one of the most cost-effective launch-service providers available today, and still has one of the best records to date.[5]

Its space activities provide it with financial, scientific, and security benefits. Yet all of these activities are dependent upon a stable and predictable space environment in which actors are safe to operate. The imperative, therefore, is to address existing challenges, among which is that of space debris.

The Debris Challenge

WHILE THE WORLD is witnessing unprecedented growth in outer space, there are equal challenges to match the progress. The largest of these is that of space debris. Coming from old rocket bodies, broken satellites, or even tools lost by the astronauts on space stations, these pieces of refuse remain in orbit for varying periods of time, depending on their rate of orbital decay. At the time of writing this article, there were more than 34,000 pieces of debris in Earth's orbits larger than 10 cm that are trackable.[6] However, there are probably close to one million pieces of junk smaller than 10 cm that human instruments are unable to track. Any of these objects, traveling at roughly 29,000 km/h, could have a catastrophic collision with a functioning satellite, thereby creating even more debris.

The challenge of dealing with debris belongs to all space actors because debris does not discriminate and are uncontrollable: they can hit civilian or military satellites, and the consequences would be the same. Likewise, debris can strike the satellites of any country, or even living crews of spacecraft such as the International Space Station or the Tiangong Station. Left unchecked, the spread of space debris could make certain orbits completely unusable for years or even generations to come.[7] Debris traveling at an altitude of the ISS (roughly 500 km) takes 25 years to deorbit. Debris traveling at 1,200km—the upper end of LEO—can take up to 2,000 years to deorbit. As such, any significant debris-generating events that take place in some of the most populous orbits will remain there for two millennia unless removed. India has stakes in this situation, as most of its satellites are in this orbit and are susceptible to impacts from debris.

Yet, despite the risk to safety posed by debris, some of the largest debris-generating events in history were intentional. These events involved the destructive testing of anti-satellite weapons.

ASAT Testing: A Brief History

THE SINGLE LARGEST event in space that generated the most debris so far is the 2007 destruction of the Chinese FengYun satellite, which generated more than 3,500 pieces of debris.[8] While destructive ASAT testing began years ago, it was not until recently that the world became aware of the disastrous consequences.

ASAT tests were first developed during the Cold War, when the US and Soviet Union began experimenting with different types of weapons that could cripple their rival's space capabilities, which were still limited at that time.[9] The two countries were following distinct pathways. The Soviet Union tested a type of exploding co-orbital vehicle, which could approach a target satellite and detonate within a close distance.[10] The resulting shrapnel would incapacitate or destroy the other satellite. The Soviets tested this system nine times between 1968 and 1994, creating relatively small amounts of debris (the largest instance creating 253 pieces of debris, some of which remains in orbit to date).[11]

For its part, the US, following a few brief and unsuccessful tests with nuclear detonations in orbit, sought direct-ascent missiles that could hit a target in LEO.[12] These missiles would physically strike a satellite, the same way that a missile interceptor would. The US successfully used this approach only once during the Cold War, in 1985. The test created 285 pieces of debris that remained in orbit for over 18 years.

After the end of the Cold War, testing in this field largely ceased. However, in 2007, China demonstrated its own ASAT capabilities by shooting down the defunct FengYun weather satellite. In addition to creating a considerable amount of debris, it reignited and propagated global efforts to develop and possess ASAT capabilities in a number of forms. This included jamming, hacking and, more sinister destructive methods as well.

Following China's demonstration, in 2008, the US used a destructive ASAT during Operation Burnt Frost, destroying a failing NASA satellite at just over 200 km in altitude. This test resulted in less than 200 pieces of debris that remained in orbit for roughly 18 months. In 2019, India carried out Mission Shakti and destroyed a satellite target using a direct-ascent missile. The test was carried out at roughly 300 km in altitude and generated a relatively small amount of debris (128 pieces), though these pieces did spread out over a considerably wider range than was anticipated.[13]

The most recent ASAT test was carried out by Russia in November 2021 and could arguably be considered the most disruptive in history. The test, executed at roughly 500 km, generated nearly 1,400 pieces of debris. The cloud of debris generated by the test endangered the International Space Station as well as the astronauts and cosmonauts on-board.[14] Moreover, Russia conducted this test in one of the most populous orbits, putting a slew of commercial satellites at great risk.[15] In one instance, debris from the test created a debris "squall", involving more than 6,000 close-approach incidents (within 10 km) with nearly 850 SpaceX spacecraft in just a single day. It is not yet clear how long the debris will remain in orbit, but the debris from Operation Solwind (conducted by the US only 50 km higher in altitude) remained in space for over 18 years.

Current Physical, Political, and Legal Risks

Fifteen years of renewed destructive ASAT testing have resulted in historic amounts of debris orbiting the Earth, a notable percentage of them created by a few space actors deliberately. This presents certain

physical, political, and legal risks for India and any other country aiming to become a major space player. The physical risks are the most obvious. Many of India's satellites are located in LEO, providing critical services both to civilians and the military. Further widespread testing of destructive ASATs will create more clouds of debris, circling the orbit and forcing functioning objects to avoid collision. Depending on the altitude, these clouds could render entire orbital planes hazardous because they lie in the direct path of the cloud's trajectory. Such a new hazard is not only problematic for existing space objects but also future ones. Projects and operations that might come into contact with such a debris cloud will have to account for this hazard throughout its entire life-cycle. This creates greater uncertainty not only for engineers and physicists, but also for investors and insurers. A more hostile space environment is, therefore, not only worse for the Indian space objects, but for the many clients that depend on Indian space-based services.

The political risks of destructive ASAT tests are much less obvious, but no less problematic. India is presently holding itself out not only as one of the big four space powers but also as a responsible actor dedicated to ensuring stability and security in outer space.[16] Yet the growing international viewpoint of destructive ASAT weapons—among both established spacefaring nations and emerging space actors—is that their use in testing situations is an irresponsible act that threatens the safety of neighbouring operators.[17] If the momentum of such sentiment continues, the majority of nations will look at states that generate debris intentionally as "irresponsible". This could lead to a loss of international influence and prestige, as well as fewer partnership possibilities on space activities.

The political risks of being labelled an "irresponsible" actor might appear to be superficial, but there are likely to be legal consequences for the testing of destructive ASAT weapons in the future as well. To date, few actors have invoked the Convention on International Liability for Damage Caused by Space Objects (Liability Convention), and no court has ever sought to apply the rule of fault-based liability. This will likely change soon: as the number of both space objects and debris continues to grow, so too, does the likelihood of an attributable collision. Under the Liability Convention, the court's deliberation will necessarily require an assessment of "duty of care" from one party to another, whether there was breach of that duty of care, and if the breach led to the damage.[18] In order to determine whether there was a breach of duty, given the lack of precedent in this field, the courts will have to look to secondary sources for guidance.

One such source will be established norms of responsible behaviour. As noted above, if international sentiment towards destructive ASAT testing continues, it will provide a compelling argument that any country conducting such tests is not only irresponsible, but also legally liable for any damage caused by resulting debris. That there are ways to develop ASAT weapons, even destructive ones, without generating debris (eg. use of virtual targets or fly-bys) further strengthens this argument. As such, testing ASATs in populated orbits, particularly LEO, will create legal liabilities for the offending country. The risk of hitting a particularly high-value target, such as the ISS, makes destructive ASAT testing in that orbital plane extremely costly.

Physical, Political, and Legal Solutions (and Opportunities)

DESPITE THE DIPLOMATIC trend towards *not testing destructive* ASAT weapons, there is no doubt that there is a growing desire by modern military forces to possess counterspace capabilities. It is also clear that India, as a preeminent global military power, will necessarily need to navigate the physical, political and legal challenges discussed above in order to establish its own space puissance. Fortunately, there are pathways to achieving India's space priorities—including counterspace capabilities—that do not entail the

physical testing of destructive ASATs.

First, while destructive ASATs do well as a physically visible demonstration, it is not the most effective counterspace capability, nor the most efficient. Non-kinetic or electronic counterspace capabilities are far more effective in disabling or disrupting a satellite's functions, without the messy consequences of creating additional debris. Cyber-attacks can also be more effective than kinetic weapons, particularly where numerous satellites are providing a particular capability. In the case of any of the global space powers, multiple satellites provide nearly all strategically critical services, making a destructive attack extremely difficult to coordinate and carry out in a meaningful way. Nevertheless, if a destructive capability is desirable, testing can still be done using virtual targets or by conducting a fly-by of the target. Either of these options are available to develop a capability, and still comply with global norms of responsibility.

Choosing a policy of non-destructive testing of ASATs also opens up a political avenue to retain standing as a responsible actor and leader in space activities. A country in need of counterspace capabilities could continue to develop and leverage the necessary technology to protect its space-based interests while also ensuring the safety of its own spacecraft, as well as those of other operators. For India, this approach would mean gaining considerable prestige among both established space actors and emerging ones. It also sends a powerful signal about India's overall intentions to balance defence and a desire for peace and security in outer space.

Finally, destructive ASAT testing should not be carried out largely because it can lead to exposure for legal liability. This will become increasingly true as space-tracking capabilities continue to improve, enabling attribution to offending actors. Given the number of high-value objects that are flying in space today—not to mention the incalculable human lives—a collision caused by ASAT-generated debris could be extremely costly and not only in monetary ways.

Towards a Commitment to Not Conduct Destructive ASAT Tests

AS THINGS NOW stand, there is no law or rule to prevent a country from conducting destructive ASAT tests. This is worrying, since a future in which there is widespread testing of destructive ASATs would likely lead to a future space environment where it will be more difficult to operate. Adoption of a commitment not to conduct destructive ASATs by a country such as India could be a watershed that puts sufficient political will behind creating a norm of responsible behaviour. Whilst it is unlikely that other global space powers (namely, Russia and China) may not adopt the ASAT commitment due to historical and geopolitical differences, adoption by a power such as India would send a signal to the rest of the world that the negative view of destructive ASAT testing is truly widespread. Such support could prove critical in bringing much of the world, particularly non-aligned countries, to the table. In this sense, India stands as a critical vote in the UN and at large. The world will be watching to see what it does.

Endnotes

1. Department of Space, ISRO, Government of India, https://www.isro.gov.in/pslv-c37-successfully-launches-104-satellites-single-flight .

2. Jonathan Amos, "Why India's Mars mission is so cheap - and thrilling", *BBC News*, September 24, 2014, https://www.bbc.com/news/science-environment-29341850 .

3. Union of Concerned Scientists Satellite Database, last updated May 1, 2022, https://www.ucsusa.org/resources/satellite-database.

4. Gunter's Space Page, Chronology of Space Launches, last visited September 18, 2022, https://space.skyrocket.de/directories/chronology.htm .

5. Cyrus John, "SpaceX or ISRO, Who's Winning the Race to Space", *The Quint*, April 2, 2018, https://www.thequint.com/tech-and-auto/tech-news/isro-vs-spacex-where-does-indias-premier-space-agency-stand#read-more.

6. "Space Sustainability – An Infographic," Secure World Foundation, January 25, 2022, https://swfound.org/media/207308/secure-world-foundation_space-sustainability_updated-data_v2.pdf.

7. "Falling to Earth takes a long Time", European Space Agency – UN Office for Outer Space Affairs, February 17, 2021, https://www.esa.int/Space_Safety/Space_Debris/ESA_UNOOSA_space_debris_infographics_and_podcast

8. "Counterspace Capabilities: An Open-Source Assessment," Secure World Foundation, April 2022, p. 5-1, https://swfound.org/counterspace/.

9. Daniel Porras, "Towards ASAT test guidelines" *UNIDIR,* February 2018, p. 4, https://unidir.org/publication/towards-asat-test-guidelines.

10. "Counterspace Capabilities: An Open-Source Assessment," Secure World Foundation, April 2022, p. 2-1, https://swfound.org/counterspace/.

11. "Counterspace Capabilities: An Open-Source Assessment," Secure World Foundation, April 2022, p. 5-1, https://swfound.org/counterspace/.

12. "Counterspace Capabilities: An Open-Source Assessment, "Secure World Foundation, April 2022, p. 1-2, https://swfound.org/counterspace/.

13. Manu Pubby, "India tests first anti-satellite missile system, codenamed Mission Shakti," *The Economic Times*, March 28, 2019, https://economictimes.indiatimes.com/news/politics-and-nation/pm-modis-big-announcement-india-successfully-tests-anti-satellite-weapon/articleshow/68592702.cms?from=mdr.

14. Tea Kvetenadze, "Debris From Russian Satellite Forces Astronauts Aboard ISS To Take Shelter ", *Forbes*, November 25, 2021, https://www.forbes.com/sites/teakvetenadze/2021/11/15/debris-from-russian-satellite-forces-astronauts-aboard-iss-to-take-shelter/?sh=3caa4b93e53c.

15. Jeff Foust, "Starlink satellites encounter Russian ASAT debris squalls", *Space News*, August 9, 2022, https://spacenews.com/starlink-satellites-encounter-russian-asat-debris-squalls/.

16. Narendra Modi, "Mission Shakti'', (speech, March 27, 2019), MEA, https://www.mea.gov.in/Speeches-Statements.htm?dtl/31180/Speech+by+Prime.

17. Up to now, five countries have voluntarily committed not to conduct destructive ASAT tests: the US, Canada, New Zealand, Japan and Germany. Others have expressed their support for such a policy - such as France, Brazil and the UK – but have not yet gone so far as enshrining this commitment in a policy. See UN Open Ended Working Group, Statements by France, Brazil and the UK. https://meetings.unoda.org/meeting/oewg-space-2022/ The US has now proposed a UN General Assembly resolution calling on all States to commit not to conduct destructive ASAT tests, though exact wording of the text has yet to be seen. Jeff Foust, "US to introduce UN Resolution on ASAT testing ban", *SpaceNews*, September 10, 2022, https://spacenews.com/u-s-to-introduce-u-n-resolution-on-asat-testing-ban/

18. Joel A Dannerly, "State Liability for Space Object Collisions: The Proper Interpretation of 'Fault' for the Purposes of International Space Law", *The European Journal of International Law* 29, no. 1 (2018), https://doi.org/10.1093/ejil/chy003.

One Step Closer to Space Security: The Role of Multilateral Discussions

Laetitia Cesari Zarkan

CREATING WAYS BY which state and non-state actors can conduct outer-space activities in a stable and responsible manner can mitigate emerging threats. After all, in outer space, the smallest disruption can increase the risk of accidental collisions. These collisions can cause the destruction of space objects used for strategic or military purposes, which could in turn compromise relations between States and give rise to military concerns.[1] Compounding the threat is the uncertainty created by the increasing number of private and public space operators, and thereby, the number of objects being launched into orbit each year. This article describes the various multilateral discussions that took place in 2022 as part of efforts to find solutions to emerging security threats in outer space.

Space safety and space security are essential to the sustainability of space activities. Space security is the domain of the Committee of Peaceful Uses of Outer Space (COPUOS) in Vienna, and space security is under the purview of the Conference on Disarmament (CD) in Geneva.

During discussions that involve many states, country delegations are keen to make a strict distinction between two notions: 'space safety' is understood as a "result of measures precluding inherent malfunction and mitigating the risks of accidental damage that would be caused by or undergone by a space object, including its component parts;"[2] and 'space security' is "the protection of a space object, including its component parts, against the threat of intentional actions undertaken by external or unauthorized actors."[3]

Ever since the 'Prevention of An Arms Race in Outer Space' item has been put on the Conference on Disarmament agenda in 1984,[4] member states have not found a common position on the suitable ways by which to reduce threats to space infrastructure, both Earth-based and in orbit.

Space security concerns are addressed in the Outer Space Treaty, a legally binding instrument negotiated in the 1960s. From its inception, the treaty has been working on the premise that space is a strategic domain and that there is a need to reduce the risk of conflict. At the core of the arms control provisions contained in Article IV of the Outer Space Treaty lie two limitations on space activities. First, Article IV contains a commitment not to place a space object in orbit around the Earth, nor install them on the moon or any other celestial body, or station in outer space, nuclear weapons or any other weapons of mass destruction. It restricts the use of the moon and other celestial bodies exclusively to peaceful purposes and expressly prohibits their use for the establishment of military bases, installations, or fortifications, as well as the testing of weapons of any kind and the performance of military manoeuvres.[5]

Despite the strong principles contained in the Outer Space Treaty, however, some analysts believe that its provisions fail to address certain space security challenges. For example, it does not elaborate on the deployment of other types of weapons in outer space nor does it expressly prohibit the launching of weapons from Earth to targets in outer space or the use of outer space for certain hostile purposes against targets on Earth. With the establishment of the work around the concept of the Prevention of an Arms Race in Outer Space (PAROS), States are attempting to find a common understanding on current and future threats to minimise mutual tensions and reduce threats.

In this context, an Open-Ended Working Group on reducing space threats through norms, rules and principles of responsible behaviours (OEWG on space threats), chaired by Chilean Ambassador Hellmut Lagos, has been established.[6]In December 2021, the United Nations General Assembly resolution 76/231 on "Reducing Space Threats through Norms, Rules and Principles of Responsible Behaviour" mandated the OEWG on space threats to, *inter alia,* "take stock of the existing international legal and other normative frameworks concerning threats arising from State behaviours with respect to outer space; consider current and future threats by States to space systems, and actions, activities and omissions that could be considered irresponsible; make recommendations on possible norms, rules and principles of responsible behaviours relating to threats by States to space systems, including, as appropriate, how they would contribute to the negotiation of legally binding instruments, including on the prevention of an arms race in outer space; submit a report to the General Assembly at its seventy-eighth session."[7]

To this end, the States have agreed on the broad lines of the discussions planned within the framework of the group. Thus, after an organisational session in February 2022, the OEWG has been divided into four sessions: May 2022, September 2022, January 2023, and August 2023.

The OEWG Chair invited experts to inform the discussions on various topics and assist delegations in preparing for the sessions. The experts are drawn from the academic, civil, commercial and scientific domains and are tasked with informing the delegations on specific aspects of the various issues raised in the OEWG. To date, two sessions have been held—in May and September 2022.

In May 2022, the first session focused on international legal and other normative frameworks concerning threats related to States' conduct in outer space, and addressed five topics: (1) existing international law; (2) international law relating to the use of force in international affairs; (3) protection of civilians, civilian objects and the natural environment; (4) applicable elements of the legal regimes governing aviation and the sea; and (5) voluntary mechanisms and regimes applicable to outer space.

During these discussions on the first topic, the experts clarified a number of issues. First, they addressed the regulation of military activities in outer space and how gaps currently lead to uncertainty about the proper use of space. They provided a review of international law applicable to the non-weaponisation of

outer space.[8] The question of dual-use satellites was raised, considering that based on the Law of Armed Conflicts, if military and civilian assets are too closely intertwined, a degree of precision in launching an attack would be impossible.[9] Additionally, there was a discussion on how to apply the principle of due regard to address threats arising from State behaviours with respect to outer space.[10]

Drawing on international law relating to the use of force, an expert presented the prohibition of the use of force as a necessary guarantee for space security and underscored that defining the notion of "armed attack" in space is a sensitive process, especially concerning the threshold that would trigger this qualification.[11] Several points were made on the legal gaps that allow for the placement of weapons in outer space, including the deployment of space-based anti-satellite capabilities. An expert stated that the termination of the Anti-Ballistic Missiles (ABM) Treaty could pave the way for the development of space weapons for missile defence, underscoring that with the collapse of the ABM Treaty in 2002, the commitments not to develop or place space-based ABM systems and their components ceased to exist.[12]

Following these points, experts also addressed the protection of civilians, civilian objects, and the natural environment. After a description of how these International Humanitarian Law and the Law of Armed Conflicts could apply to outer space, including the principles of distinction, proportionality and neutrality,[13] and experts shared the perspective of the International Committee of the Red Cross on the constraints of international law on military operations in or relating to outer space during armed conflict, urging States to consider the potential humanitarian consequences when deciding on military operations in or relating to outer space, whether at the national or multilateral level.[14]

To provide delegations with more clarity regarding the existing legal frameworks, presentations also addressed the applicable elements of the legal regimes governing airspace and the sea, including how they are often compared to outer space due to shared characteristics. Experts set out the main elements from the legal regimes governing air law and the law of the sea that could apply to outer space.[15]

On the last day of the first session, after a clarification on how to apply more efficiently agreed transparency and confidence-building measures (TCBMs) to address threats arising from State behaviours with respect to outer space, an expert highlighted how TCBMs could be precursors to advancements in space security governance and potentially lay the foundation for non-binding measures and legally-binding treaties. Following this statement, an expert summarised how to utilise TCBMs effectively to address space threats.[16] Another expert then highlighted the important role of soft law instruments, referring to the Long-Term Sustainability guidelines issued by the Committee on the Peaceful Uses of Outer Space (COPUOS)[17] and the recommendations in the report of the Group of Governmental Experts on TCBMs[18] and stated the characteristics of good TCBMs.[19]

This first session of the OEWG on space threats provided States with a detailed understanding of the international legal and other normative frameworks relating to threats associated with State behaviour in outer space.[20] During the general exchange of views on the agenda, delegations relied on the expert presentations to discuss what they considered to be the most pressing topics and how to move forward. These discussions would not have had the same impact without the active participation of civil society and the contributions read and submitted in writing through the OEWG secretariat. Thus, not only did civil society demonstrate its expertise in clarifying certain issues related to reducing tensions in space, but it had also conducted extensive research on situations that could trigger instabilities in space.

Delegations debated how to strengthen the existing international framework applicable to outer space to address space security concerns and rapid technological developments. Overall, some delegations agreed that while norms, rules and principles could serve as a basis for a legally binding instrument, non-binding measures could also lead to elaborating a legally binding instrument. Furthermore, delegations considered how any State could act in a responsible or irresponsible manner, whether intentionally or inadvertently. For some delegations, the approach to enhance space security is twofold: agreeing on the parameters of responsible behaviours that foster trust and confidence while identifying irresponsible behaviours that cause mistrust, and misperceptions, subsequently increasing tensions between States. On this basis, delegations seemed to agree on the importance of building a mutual understanding on undesirable acts which fall below the threshold of the use of force. In parallel, delegations exchanged views on how to take some of the most relevant points from the law of the sea and air law and adapt them to current space activities, while at the same time putting in place effective TCBMs and keeping in mind the protection of civilians and civilian objects.

Based on the observation that this type of system has extremely destructive effects and undermines the peaceful use of outer space, the United States worked on a mechanism to ban this type of operation. For this purpose, US Vice President Kamala Harris stated that the United States would commit to not conducting such operations in the future, and called on other States to do the same.[21] Focusing on in-orbit activities and, in particular, the placement of weapons systems in space, the Russian Federation pointed out the potential danger caused by the deployment of such technologies and emphasised the need to commit not to initiate space-to-ground threats or to develop, test or deploy weapons in space. To this end, a Russian-Chinese draft treaty is being designed to prevent the placement of weapons in outer space and of the threat or use of force against outer space objects.[22]

While these two threats, outlined by some delegations, are not to be taken lightly, they are part of a broader picture that also includes cyber, electromagnetic and non-kinetic physical interference, as well as other technologies that fall into other categories of "conventional weapons". To discuss these different threats, delegations met again in Geneva for a second session, with the participation of other experts who were able to share their knowledge on this subject.

In September 2022, delegations met for a second session on the subject of current and future threats to space systems. During the first day, a first set of experts were called upon to brief delegations on these issues, including the nature and uses of the outer space environment and space systems in relation to current and future threats by States to space systems.[23] On this topic, an expert clarified what falls under the ordinary meaning of 'space threat' and, for instance, excluded safety issues based on unintentional risks, such as collision or break-up in orbit, uncontrolled re-entry, and space debris mitigation. In order to strengthen States' understanding of such threats, some experts also described the state of the space industry and the economic stakes that a rise in tensions in space would imply.

Concerning earth-to-space threats, discussions started with a description of how the development of kinetic counterspace and anti-satellite (ASAT) weapons could lead to further escalation and create long-term risks in outer space.[24] On this topic, delegations underlined their concerns about such threats, and some joined—and are still joining—the US-led initiative aiming at committing not to conduct destructive direct-ascent anti-satellite missile tests. Drawing on the available public information, an expert explained how electromagnetic, cyber and non-kinetic physical interference could compromise the confidentiality, integrity and availability of space systems and what it would mean for the future of space activities.[25] Harmful interference could cause the premature cessation of a mission and increase the number of

malfunctioning assets in orbit. In addition to these presentations, members of the group discussed the means available to States with new and expanding space programmes for detecting and reacting to threats by States emanating from the Earth, including as a result of possible collateral harms caused by actions between other States.[26]

Following these points, experts addressed the issue of space-to-space threats, starting by the description of how co-orbital counterspace systems were placed into orbit and later manoeuvred to approach and attack a target satellite through different means, using as examples demonstrations of close approach and rendezvous and proximity technologies.[27] An expert also stressed the importance of distinguishing between dual-use and dual-purpose space systems when determining the nature of the assets.[28] Thereafter, representatives from civil society and commercial actors shared their experience on various topics concerning the steps that commercial actors could take to promote responsible behaviour in connection with rendezvous and proximity operations, including standards for transparency.[29]

Discussions focused on the question of whether lasers designed for non-hostile means, such as communications, could be used to intentionally interfere with, disrupt, damage, or destroy satellites. In this case, such lasers would qualify as dual-purpose assets. On this point, an expert concluded that three characteristics of laser use in space play an important role in the discussions about their responsible use in outer space. First, laser systems can be used for both malicious and peaceful purposes, a capability that increases the risk of disguising the original intent of use between an accident or a malicious act. Second, in many cases the use of lasers can be very hard to detect and thus, to verify. Hence, attributing damages that stem from the use of lasers is not an easy process. Third, while lasers can have either a temporary or permanent effect on a satellites function, the threshold between what can cause temporary and permanent damage is technically hard to establish. As a consequence, due to the lack of norms and common understanding, the intentions to temporarily laser blind a satellite sensor can be considered the same as a permanent laser damage to a satellite, which increases risk of unintentional escalation.[30]

On the topic of space-to-earth threats, experts explained how these concerns are important, as the deployment of space-based strike assets would bring a strategic advantage to space power.[31] However, a presentation was made on the fact that the deployment and maintenance of space-based technologies is expensive and that it would be unsustainable to operate such space-based weapons aiming at targets on the ground.[32] On this point, the Russian Federation explained how important the prevention of placement of weapons in outer space is, stressing the importance of not using space assets as a means of destroying any targets on Earth, in the air, or in outer space. Afterwards, an expert stated that military space systems are not a singular and direct cause of tension, but rather the symptom of terrestrial political tensions and insecurity, affirming that what counts is the point when routine weapons development and deployment conflated reaches the threshold of arms racing.[33]

Lastly, the discussions that took place during the last day of the second session focused on earth-to-earth threats, a representative from a company described how the growing importance of space data and the emergence of large constellations of satellites in low-Earth orbit brings space assets closer to the battlefield.[34] From the International Committee of the Red Cross' perspective, space threats have a significant impact on commercial and industrial actors as well as civilian populations,[35] an affirmation confirmed by an expert who exposed the concerns related to cyber operations against the space infrastructure.[36]

During the week, delegations raised concerns about threats to critical national infrastructure, particularly when it is placed in orbit.[37] Delegations were also able to discuss issues they considered to be priorities,

ranging from the use and testing of direct-ascent anti-satellite weapons to the use of harmful non-kinetic interference, or the use of space-based weapons to strike targets on Earth and the threats posed by uncontrolled launches or re-entry operations.

Overall, these exchanges of views enabled all delegations to gain a global perspective on each other's concerns, which will ultimately facilitate future discussions. Scheduled for the end of January 2023, the third session will serve as a launch pad for further discussions to enable the international community to agree on solutions to ensure space security through diplomacy and the rule of law.

The involvement of the different delegations in the debates shows that the type and number of space systems a State possesses and the nature of its space activities is not what matters most. The significance of these discussions demonstrates that the involvement that each State, private and commercial entity and member of civil society puts in, on its individual level, will have collective positive consequences.

Furthermore, space security issues are no longer the sole purview of States: with the increasing number of private and commercial space entities, the obligation of States to provide continuing supervision of their national space activities is essential.[38] Therefore, the development of minimum protection and safety requirements and risk mitigation and preventive measures at the national level will contribute to reducing risks in space.

Rather than working in silos with the COPUOS on the one side and the Conference on Disarmament on the other, parallel discussions would benefit from exchanging on the most pressing issues and identifying safety and security concerns, especially now that an increasing number of private and commercial actors are operating in outer space.

In this context, international cooperation will have an important role in ensuring that all States have equivalent measures of protection applying to their space assets and by extension, act in a responsible manner when carrying out space activities. This would enhance the sustainability of space activities and, consequently, nurture more peaceful relations between States and greater stability in outer space. In the long run, building bridges between the COPUOS and the Conference on Disarmament while ensuring that their respective mandates are fulfilled would not only allow for efficient work on the use and exploration of outer space but also for a better overall knowledge of the space domain, which will most certainly ensure the long-term sustainability of space activities.

Endnotes

1. Laetitia Cesari Zarkan, *What's in a word? Notions of 'security' and 'safety' in the space context,* UNIDIR (2020), https://unidir.org/commentary/whats-word-notions-security-and-safety-space-context

2. Laetitia Cesari Zarkan, *What's in a word? Notions of 'security' and 'safety' in the space context*.

3. Laetitia Cesari Zarkan, *What's in a word? Notions of 'security' and 'safety' in the space context.*

4. *See* General Assembly Resolution 39/59, 39th Sess., on the Prevention of an Arms Race in Outer Space, (12 December 1984), https://www.unoosa.org/pdf/gares/ARES_39_59E.pdf

5. *See* art. IV of the Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies, 27 January 1967, 18 UST 2410; 610 UNTS 205; 6 ILM 386 [hereinafter "Outer Space Treaty" or "OST"].

6. UN News, "Space talks seek to bring global security into their orbit", Interviews, 11 May 2022, https://news.un.org/en/audio/2022/05/1118022

7. *See* General Assembly Resolution 76/231, 76th Sess. (24 December 2021), https://undocs.org/Home/Mobile?FinalSymbol=A%2FRES%2F76%2F231

8. Mr. Kuan-Wei (David) Chen, Topic 1: Existing international law concerning threats arising from State behaviours with respect to outer space, Agenda item 6(a), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, First session, Geneva, 9 May 2022.

9. Dr. David Koplow, Topic 1: Existing international law concerning threats arising from State behaviours with respect to outer space, Agenda item 6(a), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, First session Geneva, 9 May 2022.

10. Dr. Setsuko Aoki, Topic 1: Existing international law concerning threats arising from State behaviours with respect to outer space, Agenda item 6(a), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, First session Geneva, 9 May 2022.

11. Dr. Guoyu Wang, Topic 2: International law relating to the use of force in international affairs in the context of threats arising from State behaviours with respect to outer space, Agenda item 6(a), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, First session Geneva, 10 May 2022.

12. Dr. Andrey Malov, Topic 2: International law relating to the use of force in international affairs in the context of threats arising from State behaviours with respect to outer space, Agenda item 6(a), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, First session Geneva, 10 May 2022.

13. Dr. Cassandra Steer, Topic 3: Protection of civilians, civilian objects and the natural environment in relation to threats arising from State behaviours with respect to outer space, Agenda item 6(a), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, First session Geneva, 11 May 2022.

14. Dr. Wen Zhou, Topic 3: Protection of civilians, civilian objects and the natural environment in relation to threats arising from State behaviours with respect to outer space, Agenda item 6(a), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, First session Geneva, 11 May 2022.

15. Ms. Almudena Azcarate Ortega, Mr. Charles Stotler, Topic 4: Applicable elements of the legal regimes governing aviation and the sea in the context of threats arising from State behaviours with respect to outer space, Agenda item 6(a), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, First session Geneva, 12 May 2022.

16. Ms. Nivedita Raju, Topic 4: Applicable elements of the legal regimes governing aviation and the sea in the context of threats arising from State behaviours with respect to outer space, Agenda item 6(a), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, First session Geneva, 13 May 2022.

17. Guidelines for the Long-term Sustainability of Outer Space Activities of the Committee on the Peaceful Uses of Outer Space, Report of Committee on the Peaceful Uses of Outer Space, A/74/20, Annex II (20 August 2019), available online at https://www.unoosa.org/oosa/en/oosadoc/ data/documents/2019/a/a7420_0.html

18. Report of the Group of Governmental Experts on Transparency and Confidence-Building Measures in Outer Space Activities, A/68/189 45 (June 2013), available online at https://digitallibrary.un.org/record/755155

19. Dr. Peter Martinez, Topic 4: Applicable elements of the legal regimes governing aviation and the sea in the context of threats arising from State behaviours with respect to outer space, Agenda item 6(a), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, First session Geneva, 13 May 2022.

20. Chair's Summary of discussions under agenda items 5 and 6 (a), A/AC.294/2022/3 (20 May 2022), available online at: https://documents.unoda.org/wp-content/uploads/2022/07/A-AC.294-2022-3-Chairs-summary-Advanced-Unedited-Version.pdf

21. *See* The White House, Remarks by Vice President Harris on the Ongoing Work to Establish Norms in Space (18 April 2022), available online at https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/04/18/remarks-by-vice-president- harris-on-the-ongoing-work-to-establish-norms-in-space/; Almudena Azcárate Ortega, Laetitia Cesari Zarkan, *The road to a moratorium on kinetic ASAT testing is paved with good intentions, but is it feasible?*, Note de la FRS n°22/2022, Fondation pour la Recherche Stratégique, 23 May 2022, available at: https://www.frstrategie.org/en/publications/notes/road-moratorium-kinetic-asat-testing-paved-good-intentions-it-feasible-2022

22. *See* A new edge in global stability: What does space security entail for states?, supra note 4

23. Dr. Kazuto Suzuki, Mr. James Black, Dr. Guoyu Wang and Mr. Clinton Clark, Topic 1: Nature and uses of the outer space environment and space systems in relation to current and future threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 12 September 2022.

24. Ms. Victoria Samson, Topic 2: Current and future earth-to-space threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 13 September 2022.

25. Ms. Laetitia Cesari Zarkan, Topic 2: Current and future earth-to-space threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 13 September 2022.

26. Ms. Victoria Valdivia, Topic 2: Current and future earth-to-space threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 13 September 2022.

27. Dr. Rajeswari (Raji) Pillai Rajagopalan, Topic 3: Current and future space-to-space threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 14 September 2022.

28. Ms. Almudena Azcárate Ortega, Topic 3: Current and future space-to-space threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 14 September 2022.

29. Mr. Petr Boháček and Ms. Aya Iwamoto, Topic 3: Current and future space-to-space threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 14 September 2022.

30. Boháček, *see* note 27.

31. Mr. Dmitry Stefanovich, Topic 4: Current and future space-to-earth threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 15 September 2022.

32. Dr. Laura Grego, Topic 4: Current and future space-to-earth threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 15 September 2022.

33. Dr. Bleddyn Bowen, Topic 4: Current and future space-to-earth threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 15 September 2022.

34. Mr. David Bertolotti, Topic 5: Current and future earth-to-earth threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 16 September 2022.

35. Mr. Mauro Vignati, Topic 5: Current and future earth-to-earth threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 16 September 2022.

36. Ms. Elina Morozova, Topic 5: Current and future earth-to-earth threats by States to space systems, Agenda item 6(b), Open-ended working group on reducing space threats through norms, rules and principles of responsible behaviours, Second session, Geneva, 16 September 2022.

37. Chair's Summary of discussions under agenda item 6 (b), A/AC.294/2022/4 (5 October 2022), available at: https://documents.unoda.org/wp-content/uploads/2022/10/A_AC294_2022_4_Chairs-Summ-2nd-Session-2022-au.pdf

38. See art. VI of the Outer Space Treaty.

Dynamic Stability: How AI will Reinforce, Not Overturn the Balance of Power

Michael Depp

ARTIFICIAL INTELLIGENCE (AI) IS likely to reshape the politics of the world in many ways that are yet unpredictable. To do that, however, AI must move forward from its current state as an "emerging" technology, into one that has in fact emerged. This has proved difficult for AI, which has been in consistent development since the mid-1950s, albeit with far more success since the rise of deep learning in the 21st century.[1] AI development continues to face numerous constraints such as its need for a specialised workforce and the massive costs in time and financing.

It is likely, therefore, that only those states and companies that are already powerful will be able to field the AI systems that can materially affect world politics; this is especially true for military AI systems. In turn, we are unlikely to see a future where AI becomes the fulcrum by which the balance of international power will be radically overturned. AI deployment tends to be uneven and gradual, with small systems coming online one at a time in a way that does not easily upend international politics. This is again particularly pertinent to military systems where rudimentary AI applications have been slowly rolled out over the decades.

With the costs and difficulty of developing AI limiting the number of viable participants in the international AI competition and the specificity of its application preventing individual breakthroughs from having a wider appeal, AI will struggle to uniformly change the world. Even if it is being developed around the world, AI technology will not allow burgeoning tech hubs in Israel, India, Japan, South Korea, and the UK to supplant the United States and China as leading powers, as the latter have too many built-in advantages. However, despite its inability to overturn the balance of international power overnight, AI will still have a significant effect overall by spurring greater competition between the United States and China. Strategic deployments and clever use of AI systems could give a military enough of an edge

to overtake its international competitors, though its value will be most pronounced for those at the upper echelons of the balance of power and in tipping the scales of close competitions where the slightest change could have drastic consequences. Because of its numerous constraints and gradual deployment, AI will have the effect of creating more stability in the international balance of power writ large, but will also introduce great dynamism between individual competitors, creating a dynamic stability.

Barriers to Entry

ONE OF THE LARGEST largest barriers to the emergence and deployment of AI has been cost. The development costs of AI systems can easily reach into the millions, and this is even before they are tested, evaluated, validated, and deployed for actual use.[2] The costs for developing an AI system are two-fold: there is a great deal of computational power required to train the system, and there is also a need for large sets of data to train it on.

The development of modern AI systems is iterative: they learn principally through repeated interaction: a chess program may play thousands or millions of games of chess in order to build the "expertise" necessary to play against humans. Naturally, for the system to carry this process out, it requires a great deal of computational power, a limited and often expensive resource. For example, Google used USD 1.5 million worth of computation cycles to develop its Meena chatbot.[3] This computational power must either be found in the lab doing the AI research (in which case it must be diverted from other work at a cost, such as with Google) or rented from one that has available computing time (which comes with a more direct monetary cost), or often some combination of both which the MIT labs use given that its research needs require five times more computational capacity than they have available.[4] This naturally gives AI laboratories with access to their own computing power a significant edge in the race to produce a viable AI system. It also winnows the number of organisations that can even produce AI systems: there is a limited amount of computational time that can be used or rented and if a firm has no access, it is automatically loses out on the AI race.

At the same time, massive datasets are required to train systems: for instance, in order for a computer to learn how to identify and target an incoming missile, it not only needs practice, but it also needs examples of what a missile is and what it looks like when one is launched. The precise amount of data required varies depending on both the task and the design of the system, but the larger the dataset, the better it is.[5] Like computational cycles, these datasets must either be collected by the researchers, or purchased from those with actual datasets. This acts as yet another barrier to entry for potential AI development.[6]

Another cost, and one that has been continually increasing, is talent. AI research is a complex and esoteric field that requires both specialised training and experience to carry out effectively. Trained and experienced talent is important in the development of modern AI systems, particularly those created through deep learning mechanisms, because the systems themselves are opaque: there is a general lack of codification of AI knowledge which is then difficult to pass on without practical experience. Thus, increasing the talent pipeline is difficult and not something that can be quickly scaled where it does not exist. As a result, existing talent can command a premium in price. It has already been reported that large technology firms are luring AI researchers away from universities in the United States with salaries in the hundreds of thousands of dollars.[7] And even if they are not lured away by these large salaries, they tend to congregate in areas that have the resources to accommodate their research.[8]

All of these effects are cumulative. Access to computational power, large datasets, and human talent are critical for the development of AI systems and all three are in short supply and are costly. This naturally

creates an edge for the labs that have either existing access to them, or the money to purchase that access. This has led to what has been dubbed the "de-democratisation of AI"[9] where a smaller and smaller number of large firms and universities are able to hoard the resources necessary for AI development and thus constrain the ability of others to develop these systems. Not only are these capabilities concentrated in specific institutions, they are also concentrated geographically: a recent analysis published in the Harvard Business Review noted how AI research resources are concentrated in about 50 cities around the world (most of which are in the United States and China).[10] Firms and universities in these cities have a significant advantage when it comes to developing AI technology, and the governments of those countries will be the ones to reap the political benefits of AI.

Finally, even after all of the costs and difficulties of developing and deploying an AI system have been completed, the final product will be a system that is specific to an individual task or goal. In order to tackle a new problem or adapt to an unexpected change, an entirely new AI system must be developed. AI development is not iterative; except for the trained staff, none of the components can necessarily be reused for a future project. These costs will have to be repeated for a significant amount of time in order to create enough systems to have a broad effect on either military or economic standing, further cementing the power of the leading states.

This specificity is barrier to preventing middle-tier global powers from using AI to catapult themselves beyond their rivals. Despite the cost and difficulty of AI development, great advancements have been made, and not only in the most prolific firms and universities. West Asian powers have used it to improve their military capabilities, with Iran improving its drone capabilities through swarming[11] and Israel using AI to make the Palestinian occupation less costly.[12]

Meanwhile, Japan is using AI to improve its healthcare capacities and ability to take care of its aging population[13] and India has used it to alleviate concerns about its labour market through improved automation.[14] Many of these initiatives will likely bear fruit given enough time and effort despite the costs, but they will lack a broader significance because they will not easily translate to other fields: Japan's efforts in healthcare will not spill over into improving its military drone technology and Israel's use of AI for crowd control will not advance its data analytic industry. Upending the balance of international power will require much more than one breakthrough in a technology where every problem requires the reinvention of the wheel.

The Gradual Distribution of the Future

DESPITE THE CHALLENGES facing the development of AI systems and the many pitfalls that may be found on the road to functional development, AI already exists and surrounds us. AI does not just refer to far future androids or fully self-driving cars, but also far more mundane systems and programs that already have significant impact. Numerous AI systems already permeate daily life, such as the algorithms that deliver social media or search content, the computer opponents that populate online video games, and advanced tools in spreadsheets and tax software that make accounting many times easier.

And it is not only daily civilian life that has been altered by the gradual implementation of AI systems, whether rudimentary or what may come in the future; military systems have also been changed significantly by the rollout of AI and automated systems. Aircraft that automatically modulate the actions of pilots, missiles and bombs that guide themselves to targets while accounting for weather conditions, systems that filter out noise to make targets easier to spot in radar or early warning systems, and sonar systems that help differentiate between the noise of a blue whale or a Typhoon-class submarine have

all already altered the way that militaries plan and conduct operations. Despite the myriad systems that have already been used and deployed by militaries, they have yet to upend any balances of power precisely because only the most advanced militaries have been able to develop large amounts of them unaided.

However, there is an even bigger implication for the international balance of power for already deployed AI systems that was touched upon before: the specificity of AI systems. This will necessitate a gradual evolution of AI deployment as more advanced systems slowly come online. Unlike nuclear weapons, jet propulsion, or dreadnoughts which burst onto the international scene suddenly and upended the then-balances of power, AI's graduality will be more akin to improvements to existing weapons systems that are regular and marginal. Even as new breakthroughs are taken advantage of and we move beyond the existing rudimentary AI systems, we are likely to see this same pattern borne out moving forward. It is very hard for a gradual but ever improving process to radically alter the world quickly. There is unlikely to be a tipping point on such a shallow ramp, especially one that is only accessible to a select few.

Dynamic Stability

HOWEVER, THIS IS not to say that AI will have no effect on the balance of international power or international politics generally. If powerful AI will likely be clustered in and controlled by the United States and China, it will further entrench their positions as leading states to an even greater degree, but it can also alter the balance between them. Instead of moving middle powers into the upper echelons of the balance, AI has the potential to create a stronger barrier to those upper echelons but change the nature of competition within them.

This is primarily the fear that animates the world's global powers today. In particular, it can be seen clearly in the United States where the debate over the correct usage and development of AI has reached a fever-pitch: the National Security Commission on Artificial Intelligence notes: "China possesses the might, talent, and ambition to surpass the United States as the world's leader in AI in the next decade."[15] The United States is not concerned that AI will create a new cast of near-peer competitors or reduce American standing in the world; its great fear is that an already formidable adversary will be able to use AI to become an even graver threat. At the same time, China is no passive observer of these trends; it too, recognises the strategic promise of AI systems and the danger of falling behind an American advantage, as their development plan clearly outlines the importance of AI development: "AI is a strategic technology that will lead in the future; the world's major developed countries are taking the development of AI as a major strategy to enhance national competitiveness and protect national security."[16]

These fears are not far-fetched: it is entirely possible that a Chinese breakthrough in AI could threaten the American position in the Western Pacific or an American one could cement its already formidable global strength and counter Chinese military modernisations. These advancements could range from lethal autonomous robots that can greatly improve manpower and reduce risk to human forces,[17] to loitering munitions that are much cheaper to produce and more lethal than existing rocket forces,[18] to systems that can hasten military planning and decision-making to get inside an opponent's OODA loop.[19] However, with that being said, AI's specific effect on the international balance of power will remain difficult to predict: as Michael Horowitz correctly points out, it will not be the systems themselves that change international politics but how they are used and adopted that will.[20]

These predictions about the importance of AI to American and Chinese power in the future are likely to hold a large degree of truth, but they also add another potential source of dynamic in the balance of international power that is even more insidious: the effect that AI deployment itself can have on

perceptions. Since AI systems are so specifically designed, they are unable to deal with unexpected situations which makes them particularly brittle in a military context.[21] This flaw could prove debilitating to their use on the battlefield, especially when they are untested by real world events: it will be difficult to know if a system will work as intended before it is actually used in a real-world environment.

At the same time, prudence should dictate that opponents treat those systems as operational and working even in cases where they may not be, creating a classic security dilemma.[22] This can cause a dangerous situation where the deployment of an AI system provokes a strong reaction from an opponent while creating caution for the side deploying it, creating a crisis while offering no actual benefit. Like all technologies, AI can have a significant effect beyond its actual use, which may create dangerous instability in the international balance of power, but this effect's contagion will still be limited to those who actually have the capacity to deploy these technologies.

Conclusion

TECHNOLOGICAL CHANGE CREATES unpredictability for international relationships. AI has a particularly potent possibility for creating these changes, but specific features of its development will drastically diminish its ability to upend the international balance of power. Its high cost in money and time, and the human talent required to develop it will narrow the pool of potential creators of AI; the de-democratisation of AI development leaves more power in the hands of relatively few states. At the same time, because it is a diffuse technology affecting all aspects of human life, but also a specific one that must be designed for an individual problem, its deployment will be gradual. This will prevent it from any widespread upending of the balance of power: any state liable to be dethroned will have time to marshal its own capacity to meet the new challenge over time, and that capacity is likely to be well suited to do so.

That is not to say that AI will have no importance in the international balance of power. Many of the most capable states will leverage their capacity to create it to improve their own position and are likely to reap large rewards as a result. Overall, because of the high barrier to entry, but with the allure of using it to alter the international balance of power, AI's effect will be to introduce pockets of dynamism into the otherwise far more stable international balance of power that it helps create.

Endnotes

1. "From Not Working to Neural Networking." *The Economist*, June 23, 2016, https://www.economist.com/special-report/2016/06/23/from-not-working-to-neural-networking.

2. Fabian Westerheide, "The Artificial Intelligence Industry and Global Challenges," *Forbes*, November 27, 2019, https://www.forbes.com/sites/cognitiveworld/2019/11/27/the-artificial-intelligence-industry-and-global-challenges/.

3. Bryan Walsh, "Stanford Experts Call for National Resource for AI Research," *Axios*, April 1, 2020, https://www.axios.com/2020/04/01/ai-research-stanford.

4. Kim Martineau, "What a Little More Computing Power Can Do," *MIT News*, September 16, 2019, https://news.mit.edu/2019/what-extra-computing-power-can-do-0916.

5. For example, a commonly used benchmark for image identification for birds is the CUB-200-2011 dataset which contains 11,788 images of about 200 species of birds. Catherine Wah el al., The Caltech-UCSD Birds-200-2011 Dataset. Computation & Neural Systems Technical Report, 2010-001, California Institute of Technology, Pasadena, CA (2011), https://authors.library.caltech.edu/27452/1/CUB_200_2011.pdf.

6. Facebook is a prime example here having its own access to images to training its own identification software, but also in packaging these resources for sale for use by other researchers. Kevin Lee and Xiaodong Wang, "The next Step in Facebook's Ai Hardware Infrastructure," *Engineering at Meta*, April 22, 2021, https://engineering.fb.com/2018/03/20/ml-applications/the-next-step-in-facebook-s-ai-hardware-infrastructure/.

7. Cade Metz, "Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent," *New York Times*, October 22, 2017, https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html.

8. Bhaskar Chakravorti et al., "50 Global Hubs for Top Ai Talent," *Harvard Business Review*, January 21, 2022, https://hbr.org/2021/12/50-global-hubs-for-top-ai-talent.

9. Nur Ahmed and Muntasir Wahed, "The De-democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research," (2020) https://arxiv.org/ftp/arxiv/papers/2010/2010.15581.pdf.

10. Bhaskar Chakravorti, "50 Global Hubs for Top Ai Talent".

11. Evan Omeed Lisman, "Iran's Bet on Autonomous Weapons," *War on the Rocks*, August 30, 2021, https://warontherocks.com/2021/08/irans-bet-on-autonomous-weapons/.

12. Sophia Goodfriend, "How the Occupation Fuels Tel Aviv's Booming AI Sector," *Foreign Policy,* February 21, 2022, https://foreignpolicy.com/2022/02/21/palestine-israel-ai-surveillance-tech-hebron-occupation-privacy/.

13. Euma Ishii et al., "The advent of medical artificial intelligence: lessons from the Japanese approach," *Journal of Intensive Care* 8, (2020), https://jintensivecare.biomedcentral.com/articles/10.1186/s40560-020-00452-5.

14. Aarti Betigeri, "India's AI Conundrum," *The Interpreter,* October 20, 2021, https://www.lowyinstitute.

org/the-interpreter/india-s-ai-conundrum.

15. "Final Report: National Security Commission on Artificial Intelligence," *National Security Commission on Artificial Intelligence,* May 4, 2021, https://www.nscai.gov/2021-final-report/, pg 7.

16. Graham Webster et al. (transl.), "Full Translation: China's 'New Generation Artificial Intelligence Development Plan," *New America*, August 2017.

17. Paul Scharre, "Autonomous Weapons and Operational Risk," *Center for a New American Security,* February 2016. https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS_Autonomous-weapons-operational-risk.pdf?mtime=20160906080515&focal=none.

18. Kelsey Atherton, "Loitering Munitions Preview the Autonomous Future of Warfare," *Brookings Institution*, August 4, 2021, https://www.brookings.edu/techstream/loitering-munitions-preview-the-autonomous-future-of-warfare/.

19. Michael C. Horowitz, et al., "Strategic Competition in an Era of Artificial Intelligence," *Center for a New American Security,* July 2018, https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS-Strategic-Competition-in-an-Era-of-AI-July-2018_v2.pdf?mtime=20180716122000&focal=none. pg 15.

20. Michael C. Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power," *Texas National Security Review* 1, no. 3 (May 2018), https://tnsr.org/2018/05/artificial-intelligence-international-competition-and-the-balance-of-power/.

21. Edgar Jatho and Joshua A. Kroll, "Artificial Intelligence: Too Fragile to Fight?" *Proceedings* 148, no. 2 (February 2022), https://www.usni.org/magazines/proceedings/2022/february/artificial-intelligence-too-fragile-fight.

22. Robert Jervis, "Cooperation Under the Security Dilemma," *World Politics* 30, no. 2 (1978), https://doi.org/10.2307/2009958.

Quad Vadis? A Risk Assessment of the Quad's Emerging Cybersecurity Partnership

Tobias Scholz

VARIOUS COUNTRIES IN the Indo-Pacific region are witnessing an increase in the number and intensity of their cybersecurity threats. Emerging challenges include threats to national security and international supply-chain vulnerabilities as well as the weaponisation of cyberspace by state and non-state actors. Responding to these challenges, the Quadrilateral Security Dialogue (Quad) is emerging as a key leader in shaping security norms and alignments. As one of several new plurilateral formats in the region, the Quad was initiated to meet growing international challenges in the Indo-Pacific, particularly those pertaining to China's rise as a great power.[1]

Choosing plurilateral cooperation as a means to achieve national cybersecurity interests bears as many promises as potential pitfalls. The fundamental reasons why plurilateral platforms can prove effective include their flexibility in shaping norms and standards, a relative simplicity of knowledge exchange, the presence of trust-building opportunities, and their possible deterrence effects. On the downside, however, earlier plurilateral platforms have rarely succeeded in shaping international politics in a significant way. From IBSA[2] to the Shanghai Cooperation Organisation (SCO), plurilateral cybersecurity efforts have had little geopolitical impact, if at all. Such previous experiences raise pertinent questions on the resilience of the Quad cybersecurity cooperation and possible challenges the platform is likely to encounter in the next years.

This article analyses geopolitical and geoeconomic risks in the Quad's cybersecurity cooperation. It begins with an overview of the dynamic cybersecurity threat landscape for Quad partners, followed by a brief history of Quad cybersecurity cooperation

emphasising political decisions and underlying motivations. Building on the experience of other plurilateral platforms, a conceptual section then presents four ideal type risks for plurilateral diplomatic engagement. The article then discusses four challenges to the Quad's future cybersecurity cooperation, and argues that addressing them requires the Quad to utilise risk prevention and risk management capabilities.

The Indo-Pacific: An Ocean of Cyber Challenges

VARYING NORMS AND national interests have resulted in different understandings of what constitutes the Indo-Pacific region. Despite these differences, nations that have subscribed to the concept share an understanding of the Indo-Pacific as a geopolitical and geoeconomic space. Among the first nations to officially recognise the Indo-Pacific concept were the four Quad countries, i.e., Australia, India, Japan, and the United States.[3] While each of them was affected by different cyber challenges during the 2010s, all four identified China's heightened belligerence in cyberspace as a significant cybersecurity threat.

Until 2015, the United States strategy to deal with China's rising cyber capabilities and willingness to use them for coercive purposes was geared towards offering China an active role in a rules-based order. This idea was most significantly incorporated in the US-China Cyber Agreement between Barack Obama and Xi Jinping.[4] Months after the US had revealed that Advanced Persistent Threat Actor 30 (APT 30) had spied on Indian state secrets for 10 years, then President Obama still hoped to contain Chinese ambitions in global espionage.[5]

However, national cybersecurity challenges between China and all four countries mounted in the following years. Chinese cyber espionage and cyber-attacks on key industrial enterprises and government agencies in Japan reinforced the political and economic distrust between the two nations.[6] In 2016, revelations that the Chinese threat actor Bronze Butler had attempted and successfully intruded Japanese companies for at least ten years, added to Japan's threat perception.

Australia's cybersecurity concerns in its 2017 Foreign Policy White Paper reached far beyond issues of industrial and political espionage by calling for protecting critical infrastructures, fighting misinformation, and media manipulation.[7] After the White Paper made a case for building a stronger cyber defence, then Prime Minister Scott Morrison in 2020 warned the Australian people of a possible state-sponsored cyber-attack targeting Australia's national institutions.[8] India's alertness towards Chinese cybersecurity threats increased significantly after the border clashes in the Galwan Valley that began in May 2020. In the months after the initial skirmishes, India started experiencing power cuts. In Mumbai, where the greatest of such attacks took place in October 2020, local authorities identified the power outage as a result of malicious software.[9] Meanwhile, the US, of all Quad countries is the only one with a public attribution system. Since its launch in 2017, the Cybersecurity & Infrastructure Security Agency (CISA) has declared Chinese threat actors as being responsible for various cyber-attacks on US territory, including espionage, attacks at critical infrastructure, and influence operations.[10]

China's rising assertiveness in cyberspace is the one central element that united Quad nations before the group came together for closer cooperation in 2020. The following section introduces the progress that Quad cybersecurity cooperation has made over the past two years.

The Emergence of the Quad as a Cybersecurity Actor

CYBERSECURITY COOPERATION HAS been a built-in feature of the Quad since the first informal gathering of foreign ministers at the sidelines of the United Nations General Assembly in September 2019. Only months after a serious cyber-attack on the Australian Parliament, which was believed to have originated in China,[11] cybersecurity cooperation was jointly seen as a political priority. The evolution of Quad cybersecurity cooperation has since been stimulated by three high-level meetings.[12]

During the first Quad Leaders' summit which took place virtually in March 2021, heads of government agreed on stronger cooperation on new technologies within the emerging strategic principle of a "Free and Open Indo-Pacific".[13] The meeting's key cybersecurity achievement was establishing a Quad Critical and Emerging Technologies Working Group. The group was formed to pave the way for planning a joint Quad approach as it was meant to develop a set of shared norms and standards as well as to identify shared priorities in telecommunications, supply chains, and public-private partnerships.[14]

Six months after the initial summit in March 2021, Quad leaders met again and reviewed the recommendations of the established working group.[15] They also designed the Quad Senior Cyber Group to coordinate cybersecurity matters within the Quad. The group now forms the core of the new Quad Cybersecurity Partnership and has two central functions. First, it provides a space in which it consults with different public and private stakeholders on matters like cyber standards, supply chains, and critical infrastructure resilience. Second, it is meant to utilise knowledge exchange mechanisms to drive the future cybersecurity agenda of the Quad.[16]

While the Quad Senior Cyber Group is coordinated by senior officials within the four governments, the Quad summit further resulted in the launch of several multistakeholder projects to foster cooperation on specific technologies. Most significantly, leaders agreed on a Track 1.5 dialogue on 5G standards and Open RAN technology to enhance security, interoperability, and openness in the telecommunications sector. The Quad nations also created a Semiconductor Supply Chain Initiative and two Technical Standards Contact Groups on Advanced Communications and Artificial Intelligence (AI), respectively. Finally, Quad leaders identified policy coordination in multilateral organisations such as the International Telecommunications Union (ITU) as important areas. Including the ITU in their joint statement stands out, as this organisation was dominated by China in previous years and has recently played only a marginal role in US foreign policy.

The Quad Senior Cyber Group met for the first time in March 2022 and was led by member states' senior officials coordinating national cyber security. The US statement following the meeting reflected a focus on protecting critical infrastructures and indicated an interest in extending cooperation in the Indo-Pacific region.[17] Subsequently, the recommendations were provided as input to the most recent Quad Leaders' Meeting in May 2022.

In the meeting, Quad leaders announced their intent on "improving the defense of our nations' critical infrastructure by sharing threat information, identifying and evaluating potential risks in supply chains for digitally enabled products and services, and aligning baseline software security standards for government procurement, leveraging our collective purchasing power to improve the broader software development ecosystem so that all users can benefit."[18] The declaration marks the final step of the Quad to establish direct lines of communication and cooperation between the relevant national nodal agencies dealing with cybersecurity.

Cooperating directly through CERTs and ministries instead of setting up joint facilities or secretariats illustrates the Quad's continuing self-perception as a platform. The Quad nations addressed the issue of increased areas of cooperation without introducing institutional mechanisms by defining leadership roles for each country. The meeting assigned Australia the responsibility for critical-infrastructure protection, India is tasked to coordinate supply-chain resilience and security, Japan's key focus is on workforce development, and the US will lead efforts on software security standards.[19]

The Quad leaders not only increased horizontal cooperation, but also introduced vertical partnerships. The meeting confirmed that the Quad will extend people-to-people connectivity through a Quad Cybersecurity Day and capacity building programs which shall be open to partners in the Indo-Pacific region.

Finally, the Quad increasingly shows a willingness to deepen its cooperation towards specific cybersecurity challenges. At the sidelines of UNGA 2022, Quad foreign ministers announced that the platform will now more closely coordinate its efforts in fighting ransomware.[20] While concrete measures remained missing in the statement, senior officials of Quad members now have a strong mandate that could potentially evolve into institutional mechanisms.

Risks of Plurilateral Diplomacy

PLURILATERAL PLATFORMS SUCH as the Quad consist of "at least three but a limited number of sovereign nation states that jointly coordinate political or economic demands toward the international system or parts of it while maintaining a minimum degree of institutionalisation and delegation."[21] This section discusses past challenges faced by other plurilateral platforms to identify potential categories of political risk that they are susceptible to.

A central problem for plurilateral platforms is internal credibility. The SCO offers a suitable example for this instance. Led by China and Russia, the platform made quick advances on cybersecurity cooperation in the early 2010s. Russia proactively pushed for mechanisms and norms of international information security which elevated the SCO to a serious institution in this field.[22] However, with India's and Pakistan's accession to the SCO, the organisation was not able to maintain this momentum. Instead of further shaping the international cybersecurity debate, the SCO's international security aspirations proved to be without credibility for an environment in which key geopolitical competitors China, India, and Pakistan were involved in the decision-making process. Distrust and hedging tactics among member states have locked in the SCO as a credible actor for cybersecurity cooperation.

Plurilateral platforms also fail to reach their goals when they lack external credibility. Dissatisfied with a US-dominated Internet, the IBSA nations India, Brazil, and South Africa pitched an alternative. The platform was initially well positioned to balance the US through South-South cooperation as it was formed by three key rising powers in Africa, Asia, and South America. IBSA realised that a Committee for Internet-Related Policy (CIRP) could position the platform as a leader of the Global South by calling for "improving the quality of peoples' lives everywhere."[23] However, IBSA nations underestimated the effort needed to convince other countries of CIRP. After CIRP failed to find the necessary support in the UN, the proposal had to be withdrawn. Excluding other nations from developing an alternative model to Internet governance and supposing international support of the Global South without including many of its nations had a negative effect on IBSA's credibility.

BRICS, for its part, provides evidence of how internal conflict among members in one policy area can limit the overall productivity of the platform. The group that includes Brazil, Russia, India, China, and later South Africa came together mostly to challenge international financial and trade institutions. In its first years, the group had a modicum of success even though some of its members had fundamental bilateral issues with another. After the Sino-Indian border crisis at Doklam in 2017, however, BRICS countries found it increasingly difficult to agree on shared points of view. Animosities had outgrown the will to seriously cooperate on global challenges and significantly impaired the ability of BRICS to formulate policy solutions in global governance.

Finally, plurilaterals can lose or change their function as a result of external shocks. One example that stands out is the Group of Eight (G8), a platform that had been designed by Western powers to find solutions to global challenges. In the early 2000s, the G8 remained a space for informal exchange on international issues such as trade, energy, and terrorism. Yet, when the global financial crisis erupted in 2007, neither the G8 nor other countries believed that the platform would be suitable to address the situation. A more inclusive G20 platform was formed and the G8 lost its status as a platform for international financial and trade issues.

The Quad's Four Biggest Cyber Risks

INITIATING THE QUAD gave member countries a sense of optimism about maximising their national cybersecurity through cooperation. However, there are good reasons for Quad countries to proceed with caution and foresight. Applying the previously introduced ideal types of risks for plurilateral platforms, Quad countries must be aware of the following potential risk scenarios.

Risk 1: Internal credibility. Multi-alignment strategies of Quad countries may decrease credibility.

The United States is by far the most powerful cybersecurity actor among the Quad countries. Besides its membership in the Quad, the United States is also part of AUKUS[24] and the Five Eyes intelligence alliance, while maintaining strong bilateral security partnerships with countries like Australia and the Philippines. As the United States is opting for a multi-alignment strategy, India might also deprioritise further Quad integration and instead seek to diversify its regional cybersecurity partnerships outside of Quad. If regional cybersecurity cooperation continues to fragment, the Quad can become susceptible to losing its status as a security coalition in the Indo-Pacific.

As the Quad is the only security platform that has the potential to be a regional leader in setting cybersecurity norms and standards, a solution could be to transform it into a cybersecurity alliance. While a security alliance in conventional domains appears unrealistic for various reasons, the promise of mutual assistance in defending against international cyber-attacks can significantly increase internal and external credibility. In the beginning, this can include a rapid response mechanism through which Quad nations immediately support an attacked member through satellite Internet access, when Internet availability is affected through the attack.

Risk 2: External credibility. Regional distrust of Quad's geopolitical intentions can lead to a legitimacy deficit.

Many countries in the region regard the Indo-Pacific concept as a diplomatic instrument designed to isolate China. For some, such a call for division might even stand in paradigmatic opposition to the powerful narrative of the "Asian century".[25] If Quad countries are aiming to succeed in convincing regional players of their version of a free, open, and secure Indo-Pacific, they must make an offer to countries in the region. In other words, Quad countries have to prove to their regional partners that they are not securitising the Indo-Pacific for their own good, but instead offer cybersecurity solutions for all of them.

The Quad can increase its regional recognition by establishing Track 1.5 and Track 2 networks with other regional forums, most significantly ASEAN.[26] There is a growing recognition within ASEAN, that cybersecurity capacity building is a quality that must be encouraged beyond national borders. An ASEAN-Quad cybersecurity capacity building programme could be a starting point for regional outreach. Establishing a cybersecurity warning mechanism among ASEAN and Quad CERTs could be central for trust and confidence building.

Risk 3: Internal conflict. Disagreements over the role of Big Tech may affect trust.

All four Quad countries are currently dissatisfied with how its partners wish Big Tech companies to be regulated. India has in recent years shown great aversion to US-based social media companies' norms on online freedom of speech and even more significantly on the access of local data and standards for international routing.[27] In one incident, the Twitter country office in New Delhi was even raided as part of a police investigation. In another incident in 2020, Australia clashed so massively with Facebook, that the company threatened to pull out of Australia. Two observers of the 2021 Quad summit pointedly summarised, "While the unity on display at the Quad summit was an impressive show of strength, it was also an incomplete picture that masked growing friction among members on several elemental technology issues such as cross-border data flows, data privacy, payments, digital taxation, competition, e-commerce, and law enforcement."[28] If such disagreements remain unsolved, they may cause hardly reconcilable trust deficits among the member states.

Diverging opinions on the role of US-based technology companies in other Quad markets are essentially a matter of trade as well as technology norms and practices. To deal with such issues, the European Union (EU) has already set up a bilateral Trade and Technology Council with the United States and India, respectively. The Track 2 Quad Tech Network offers an existing space that can be elevated to a more formal forum on technology and trade issues.[29]

Risk 4: External shock. A more assertive China could probe the effectiveness and belie expectations.

As the institutional mechanisms of the Quad are still evolving and have yet to be tested, a major cyber-attack originating in China could catch the partners by surprise and evoke internal tensions. Such a scenario is, e.g., imaginable when an attacked Quad country wants other Quad nations to attribute a cyber-attack, but in response its partners prefer to first gain additional cyber-forensic confirmation. If a delayed attribution confers a heavy security, economic, or status cost on the impacted country, this Quad member might lose trust in its partners and their platform.

In their further integration, Quad members must remain cautious and serious over the aspirations that their decisions set free. For example, a joint attribution system of cyber-attacks could work properly only if the rules of attribution are clear. Instead of relying too heavily on cyber defence cooperation, the Quad could first turn to capacity-building solutions that increase the resilience of its members.

Outlook

ADDRESSING EACH OF these risks requires political will, timing, and a shared awareness of how to deal with challenges and potential setbacks. The Quad must act decisively and self-reflectively if it is to maximise the benefits of a unified and sustainable security partnership.

What distinguishes the Quad from other plurilateral platforms such as BRICS and IBSA is its unified belief in China as a geopolitical threat. As this shared belief is not likely to completely disappear within the next 50 years, the Quad countries have a realistic opportunity to shape many of the norms, standards, and institutional mechanisms as well as strategic imperatives for the digital Indo-Pacific in the decades to come. To succeed, the Quad must remain aware of its internal diversity and external pressures. If it wants to be a regional leader on cybersecurity efforts, the Quad Cybersecurity Partnership must make a credible offer for peace, prosperity, and stability in the entire Indo-Pacific region.

(Author's note: I would like to thank Sameer Patil and Sarah Bressan for their most helpful and constructive comments on earlier versions of this essay. All shortcomings are my own.)

Endnotes

1. A plurilateral platform is a "political instrument by at least three but a limited number of sovereign nation states that jointly coordinate political or economic demands toward the international system or parts of it while maintaining a minimum degree of institutionalization and delegation"; Harsh V. Pant and Tobias Scholz, "BRICS: Expiring Political Relevance and Inspiring New Coalitions," in *Handbook on Global Governance and Regionalism,* ed. Jürgen Rüland and Astrid Carrapatoso (Cheltenham, UK & Northhampton, MA, USA: Edward Elgar Publishing, 2022), 149–60, forthcoming.
2. IBSA is a plurilateral platform consisting of India, Brazil, and South Africa.
3. Australian Government, 2017 *Foreign Policy White Paper: Opportunity, Security, Strength* (Canberra: Department of Foreign Affairs and Trade, 2017), https://www.dfat.gov.au/sites/default/files/2017-foreign-policy-white-paper.pdf
4. Mark Bryan F. Manantan, "Advancing Cyber Diplomacy in the Asia Pacific: Japan and Australia," *Australian Journal of International Affairs* 75, no. 4 (2021): 432–59, https://www.tandfonline.com/doi/full/10.1080/10357718.2021.1926423
5. *Council on Foreign Relations*, "APT 30," https://www.cfr.org/cyber-operations/apt-30

6. Stefan Soesanto, "A One-Sided Affair: Japan and the People's Republic of China in Cyberspace (Hotspot Analysis)," ETH Zürich, 2020, https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/389371/Cyber-Reports-2020-01-A-one-sided-Affair.pdf

7. Australian Government, 2017 *Foreign Policy White Paper: Opportunity, Security, Strength.*

8. "Australia Cyber Attacks: PM Morrison Warns of 'sophisticated' State Hack", *BBC News*, June 19, 2020, https://www.bbc.com/news/world-australia-46096768

9. Ritvick AB, "Maha Min Speaks in Assembly, Says Mumbai Blackout Was Cyber Attack," *TheQuint*, March 4, 2021, https://www.thequint.com/news/india/maharashtra-minister-nitin-raut-speaks-in-assembly-says-mumbai-blackout-was-cyber-attack

10. CISA, *China Cyber Threat Overview and Advisories* (Washington, D.C.: CISA, 2022), https://www.cisa.gov/uscert/china

11. Peter Hartcher, "Farewell Tech Utopia: How Governments Are Readying the Web for War", *The Sydney Morning Herald*, February 18, 2019, https://www.smh.com.au/national/farewell-tech-utopia-how-governments-are-readying-the-web-for-war-20190218-p50yhh.html

12. For the purpose of this article, the following paragraphs do only point to the Quad cybersecurity partnership and do not include cooperation in other areas relating to emerging technologies.

13. The White House, *Quad Leaders' Joint Statement: 'The Spirit of the Quad'* (Washington, D.C.: The White House, 12 March 2021), https://www.whitehouse.gov/briefing-room/statements-releases/2021/03/12/quad-leaders-joint-statement-the-spirit-of-the-quad/

14. The White House, *Fact Sheet: Quad Summit* (Washington, D.C.: The White House, 12 March 2021), https://www.whitehouse.gov/briefing-room/statements-releases/2021/03/12/fact-sheet-quad-summit/

15. Rajeswari Rajagopalan, "The Growing Tech Focus of the Quad", *The Diplomat*, July 9, 2022, https://thediplomat.com/2022/07/the-growing-tech-focus-of-the-quad/

16. The White House, *Fact Sheet: The Quad Summit.*

17. The White House, *Statement by National Security Council Spokesperson Emily Horne on Quad Senior Cyber Group Meeting* (Washington, D.C.: The White House, 25 March 2022), https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/25/statement-by-national-security-council-spokesperson-emily-horne-on-quad-senior-cyber-group-meeting/

18. The White House, *Quad Joint Leaders' Statement* (Washington, D.C.: The White House, 24 May 2022), https://www.whitehouse.gov/briefing-room/statements-releases/2022/05/24/quad-joint-leaders-statement/

19. The White House, *FACT SHEET: Quad Leaders' Tokyo Summit 2022* (Washington, D.C.: The White House, 24 May 2022), https://www.whitehouse.gov/briefing-room/statements-releases/2022/05/23/fact-sheet-quad-leaders-tokyo-summit-2022/

20. Ministry of External Affairs, Government of India, *Quad Foreign Ministers' Statement on Ransomware* (New Delhi: Government of India, 23 September 2022), https://www.mea.gov.in/bilateral-documents.htm?dtl/35749/Quad+Foreign+Ministers+Statement+on+Ransomware

21. Pant and Scholz, "BRICS: Expiring Political Relevance and Inspiring New Coalitions".

22. Eneken Tikk and Mika Kerttunen, "Parabasis: Cyber-Diplomacy in Stalemate," *Norsk Utenrikspolitisk Institutt,* 2018, https://www.nupi.no/en/publications/cristin-pub/parabasis-cyber-diplomacy-in-stalemate

23. Dushyant Singh, "India's Proposal for a United Nations Committee for Internet-Related Policies (CIRP)" (speech, New York City, 2011), IT for Change, https://itforchange.net/indias-proposal-for-a-united-nations-committee-for-internet-related-policies-cirp

24. A security platform between Australia, the United Kingdom, and Australia.

25. Among the authors that have contributed in popularising such narratives are Parag Khanna's *The Future Is Asian* (New York: Simon and Schuster, 2019) and Kishore Mahbubani's *Has China Won?: The Chinese Challenge to American Primacy* (United Kingdom: Hachette, 2020). Their approaches vocalise an aspirational future in Asia without geopolitical conflict. While Mahbubani builds his argument on a predicted Western decline and the need for South-South solidarity to reduce global inequalities, Khanna's work exclusively focuses on the economic growth trajectories in a connected and economically liberal Asia.

26. Aakriti Bachhawat, Danielle Cave, Jocelinn Kang, Rajeswari Pillai Rajagopalan and Trisha Ray, *Critical Technologies and the Indo-Pacific : A New India-Australia Partnership*, ASPI & ORF (Policy Brief Report No. 39/2020), 2020, https://www.orfonline.org/research/critical-technologies-and-the-indo-pacific-policy/ have already suggested Quad Plus formats in the field of technology cooperation, an idea complementary to the ideas above.

27. Sameer Patil, "Flexing Quad Strength on Fintech and Cybersecurity", *Financial Express*, December 10, 2021, https://www.financialexpress.com/opinion/flexing-quad-strength-on-fintech-and-cybersecurity/2348178/

28. Mark Linscott and Anand Raghuraman, "How to Leverage the Quad to Counter China's Digital Sinosphere", *Atlantic Council*, May 17, 2021, https://www.atlanticcouncil.org/blogs/new-atlanticist/how-to-leverage-the-quad-to-counter-chinas-digital-sinosphere/

29. See e.g., Trisha Ray et al., *The Digital Indo-Pacific: Regional Connectivity and Resilience* (QTN Series) (Acton: National Security College at the Australian National University, 2021): 39.

About the Editors & Authors

Nicolo Andreula

Nicolò Andreula is Founder and Managing Director, Disal Consulting, and visiting lecturer at the Chinese University of Hong Kong, Nanyang Business School in Singapore.

Husanjot Chahal

Husanjot Chahal is a Research Analyst at Georgetown University's Center for Security and Emerging Technology (CSET).

Basu Chandola

Basu Chandola is an Associate Fellow at ORF, New Delhi.

Michael Depp

Michael Depp is Junior Fellow and Program Coordinator at the Cyberspace Cooperation Initiative, ORF America.

Smita Gupta

Smita Gupta is Weaver at Agami, which aims to radically increase innovation and changemaking in systems of law and justice in India.

Saurabh Karn

Saurabh Karn is Curator at Agami and Co-Founder of Sampatti Card.

Tanvi N Kulkarni

Tanvi Kulkarni is a Policy Fellow at APLN. She is also a Visiting Fellow at the Institute of Peace and Conflict Studies, New Delhi and visiting lecturer at the Department of Defence and Strategic Studies at the Savitribai Phule Pune University.

Sachin Malhan

Sachin Malhan is the co-founder of Agami.

Veronika Nagy

Dr. Veronika Nagy is an Assistant Professor in Criminology at the Law Department of Utrecht University.

Ruhee Neog

Ruhee Neog is the Director of the Institute of Peace and Conflict Studies (IPCS) in New Delhi, India and the coordinator of its Nuclear Security Program.

Daniel Porras

Daniel Porras is the Director of Strategic Partnerships and Communications at the Secure World Foundation.

Stefania Petruzzelli

Stefania Petruzzelli, PhD. is an Italian expert in Modern and Contemporary Literature and in the way human beings approach other worlds, such as the Metaverse and departures.

Stuart Rollo

Stuart Rollo is a Postdoctoral Research Fellow at University of Sydney.

Tobias N Scholz

Tobias Scholz is a PhD Candidate at King's College London and National University of Singapore.

Vikram Sharma

Vikram Sharma founded and leads QuintessenceLabs, based in Australia, which is at the forefront of the global quantum cybersecurity industry.

Gabriella Skoff

Gabriella Skoff is a Researcher on Quantum Security and Social Impact at Project Q Sydney.

Laetitia Cesari Zarkan

Laetitia Cesari Zarkan is a Consultant at UNIDIR. She is also a doctoral researcher at Luxembourg University.

Editors

Pulkit Mohan

Pulkit Mohan is an Associate Fellow with the Centre for Security, Strategy and Technology (CSST) at ORF, New Delhi

Rajeswari Pillai Rajagopalan

Dr Rajeswari Pillai Rajagopalan is the Director of CSST at ORF, New Delhi.

Trisha Ray

Trisha Ray is the Deputy Director of CSST at ORF New Delhi.