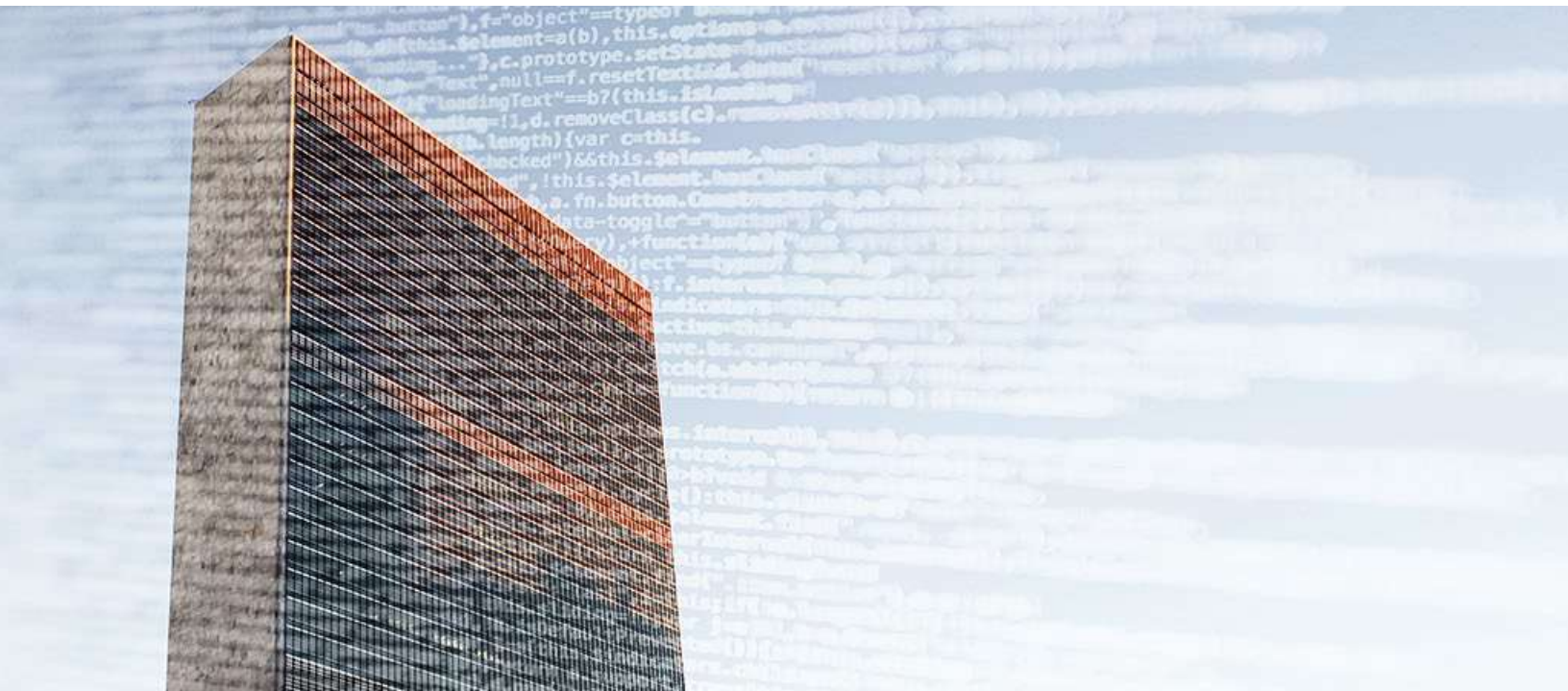


THE NEW GEOPOLITICS OF CONVERGING RISKS

THE UN AND PREVENTION IN THE ERA OF AI

Eleonore Pauwels

Research Fellow on Emerging Cybertechnologies
United Nations University Centre for Policy Research



**UNITED NATIONS
UNIVERSITY**

Centre for Policy Research

THE AUTHOR



Eleonore Pauwels is Research Fellow on Emerging Cybertechnologies at United Nations University Centre for Policy Research, focusing on Artificial Intelligence. Prior to joining the Centre, Ms Pauwels was Director of the Anticipatory Intelligence Lab with the Science and Technology Innovation Program at the Woodrow Wilson International Center for Scholars.

A former official of the European Commission's Directorate on Science, Economy and Society, Ms Pauwels is an international science policy expert, specializing in the governance and democratization of converging technologies.

Her research analyzes and compares how emerging technologies, such as AI, genome-editing, cyber and biotechnologies, raise new opportunities and challenges for health, security, economics and governance in different geo-political contexts. She examines the promises and perils that will likely arise with the development of AI civil and military technologies, the Internet of Bodies and Genomes, and the convergence of cyber and bio-security.

Acknowledgements

The author wishes to express her appreciation and gratitude to mentors and colleagues at UNU for their intelligent insights and sharp-witted, enlightened guidance, notably Dr James Cockayne, Adam Day, Anthony Dursi, Cale Salih, Alexandra Cerquone, Jessica Caus and Dr Kelly Gleason. The author also owes a debt of gratitude to UNU's Rector, Dr David Malone, for his trust and encouraging support. Warm thanks go to experts in converging technologies, dear Sarah Denton, Dave Rejeski and Garrett Dunlap. Finally, conversations with bright visionary minds at EOSG, DPPA, UN Global Pulse, UNODA, UNIDIR and UNCTED have made this journey even more captivating.

ISBN 978-92-808-6504-2

©United Nations University, 2019. All content (text, visualizations, graphics), except where otherwise specified or attributed, is published under a Creative Commons Attribution-Noncommercial-Share Alike IGO license (CC BY-NC-SA 3.0). Using, re-posting and citing this content is allowed without prior permission.

Citation: Eleonore Pauwels, "The New Geopolitics of Converging Risks: The UN and Prevention in the Era of AI", United Nations University Centre for Policy Research, 29 April 2019.

CONTENTS

Executive Summary	i
Section I: Introduction	1
Section II: The Paradigm Of AI Convergence	5
Section III: Scenarios Of AI Convergence	13
<i>The Deception Machine: Degradation Of Truth And Trust</i>	13
<i>The Internet Of Bodies, Genomes, And Minds</i>	17
<i>Inside Smart But Vulnerable Cities, Factories, And Laboratories</i>	24
<i>AI For Prevention – A Positive Scenario On AI Convergence</i>	27
Section IV: Key Scenario Takeaways For Global Security	33
Section V: Responsible Governance In The Era Of AI Convergence	39
Section VI: Prevention In The Era Of AI Convergence	47
Conclusion	57
Annex 1: Mechanisms Within The Global Foresight Observatory	59
Annex 2: Methodology For Scenarios Of AI Convergence	61
References	63

EXECUTIVE SUMMARY

We are entering an era of hybrid opportunities and threats generated by the combination of artificial intelligence (AI) and other powerful dual-use technologies, with implications for nearly every aspect of daily lives. The convergence of AI and affective computing, cyber and biotechnologies, robotics and additive manufacturing raises complex global implications that are poorly understood, leaving the multilateral system with limited tools to anticipate and prevent emerging risks. At the same time, the spread of AI convergence across a wide range of States, non-State and transnational actors and entities means that the challenges of tomorrow must be addressed collectively and innovatively.

How can the multilateral system better understand and anticipate risks as AI convergence with dual-use technologies intrudes increasingly into the political, social, economic, and security spheres, creating new potential for systemic vulnerabilities and distributive inequalities? How can actors within the multilateral system build better anticipation and prevention capacities in the face of these risks?

This report is the first step in developing a common understanding of the emerging impacts of AI convergence on the United Nations' (UN's) prevention agenda. It provides: (i) an analysis of current trends in AI convergence; (ii) scenarios that examine emerging opportunities and risks; (iii) principles to guide how innovation should be deployed responsibly by actors in the multilateral system; and (iv) a recommendation for a foresight capacity housed within the UN and shared across key communities.

Section I introduces the main themes of the report, providing an overview of AI convergence and its implications for the UN's prevention agenda. It traces the major trends in AI convergence, flagging the huge potential for human wellbeing alongside the existential risks of large-scale crises that are already emerging. The resulting loss of trust within and among States, and between public and private sectors, creates a risk that the multilateral system's response may be driven by competition and isolation rather than cooperation. Here, the UN has a potential role in providing a bridge between stakeholders and a common vision for preventing major crises, but continues to be limited by its inflexible, State-centric ways of working.

Section II offers a paradigm of AI convergence and explains why this paradigm represents a time of technological rupture with implications for large-scale crisis prevention. Within this, it traces six fundamental shifts: (i) the impact of the combinatorial nature of technologies today; (ii) technological decentralization and democratization; (iii) civil-military fusion, particularly

around weaponized technologies; (iv) the erosion of proportionality in the context of geopolitical conflicts; (v) the rise of new forms of major power competition that are accelerating inequalities; and (vi) a shift away from traditional, military forms of conflict, towards the use of bio-power and attacks on social cohesion. Taken together these shifts present a complex, often chaotic, landscape for the UN's prevention agenda and one where rigorously applied tools of foresight can offer meaningful strategic insights and potential ways forward.

Section III builds on a set of in-depth scenarios involving AI convergence and complex socio-technical systems. These scenarios do not predict the future but rather offer an evidence-based framework for understanding the ways in which technologies can converge, building on the trends identified in previous sections.

- The first scenario – **The Deception Machine** – describes how AI and converging technologies could manipulate and simulate information, amplifying hybrid cyberthreats, political and social engineering, to the point of turning manifestations of distrust into a global epidemic. Here, the lines between reality and deception become blurred, and the potential for large-scale mobilization of people, resources and weapons around false narratives creates significant global risks.
- The second scenario – **The Internet of Bodies, Genomes and Minds** – explores how bio-power is drastically magnified under AI convergence, driving far more intrusive and coercive forms of surveillance and use of personal data, and the eventual “cyber colonization” of States in the Global South. Far beyond today's privacy concerns, AI convergence will soon allow for control and manipulation of bio-data that could open the door for weaponization and substantial risks to large populations.
- The third scenario – **Smart, Vulnerable Cities** – illustrates how new forms of virtual, physical and urban attacks, crime and violence could drastically increase human and civilian insecurity. At the same time, urbanization and AI convergence create drastic shifts in the future of work, where urbanized populations face new and rapidly changing socio-economic risks.
- The final scenario – **AI for Prevention** – focuses on the empowering potential of AI convergence to address many of the crisis risks in the prior scenarios, offering a vision of innovative, large-scale prevention and the opportunity for mass human flourishing. Here, technological

optimization allows for progress on the UN Sustainable Development Goals (SDGs) related to health, famine and justice (among others), creating a positive feedback loop that rapidly diminishes some of the most acute crisis risks worldwide.

Drawing from the scenarios, **Section IV** provides a **geopolitical analysis** of the implications of AI convergence for human security and potential large-scale crises. Across all of the scenarios, common themes emerge, including the disempowerment and marginalization of the massive communities across the globe unable to access the benefits of AI convergence; the emergence of “vulnerable links” in the global security order; and the resulting rise of a small number of AI superpowers that may leave all others behind. By mapping AI convergence globally, this section demonstrates how scenario-based foresight could lead to a multilateral response.

Section V argues that the concept of “**responsible innovation**” is crucial to prevent the kinds of large-scale risks present across the major scenarios above. To this end, it offers a “**Matrix of Emerging Threats**,” classified by separate security domains, and cross-referenced by the different technologies that converge with AI. This matrix builds a bridge between the scenarios and the

responses, identifying sectors, actors and entities that would need to be involved in responsible innovation. Three tensions emerge from this analysis: “transparency versus secrecy,” “predictive accountability versus liability,” and “competition versus shared prosperity.” Any effective prevention strategy will need to involve the key actors identified, and address the inherent tensions associated with their possible role.

Section VI offers a recommendation to implement the UN’s prevention agenda in the era of AI convergence. It proposes a path for the UN to build, guide and lead a **Global Foresight Observatory for AI Convergence**, drawing from the methodology and evidence supplied in this report. The Observatory would be a constellation of key public and private sector stakeholders convened by a strategic foresight team within the UN to implement a shared foresight methodology. The Observatory would equip the UN to (i) articulate tailored and robust scenarios from which innovative strategies can emerge; (ii) map and involve key stakeholders that reflect the unique ways in which technologies are converging; and (iii) develop coherent and responsible approaches to leverage innovation and technology for prevention. To illustrate this process, Section VI offers two use-cases that could become models for the UN’s leadership to support multi-stakeholder-driven processes.

KEY STRATEGIC MESSAGES

This report offers a path towards a radically improved foresight capacity for the UN and its partners with respect to AI convergence, raising the need for multilateral innovation in the face of unprecedented opportunities and risks across the world. The report is laid out as a series of related arguments/messages about AI convergence, as follows:

- At a time of technological rupture, the risks of global insecurity are heightened by trends of isolationism and lack of collective responsibility.
- A new geopolitics of inequality, vulnerability and potential human suffering is emerging.
- To meet these challenges, a common understanding of opportunities and risks across the international community is needed, driven by responsible innovation and incentives for a shared approach to prevention.
- An inclusive foresight tool, housed at the UN and shared across key stakeholders, can be the locus for “preventive innovation,” offering new opportunities for stability and security worldwide.



INTRODUCTION – A NEW ERA FOR CYBER-AI PREVENTION

Across the world, at any given moment, there are pervasive cognitive-emotional conflicts being waged for the control of populations' thoughts, emotions and attitudes. These battles of influence do not tend to occur in wartime, but rather in peacetime, infiltrating homes and smart cities. They sow disinformation, affective manipulation and forgeries as new means of undermining social cohesion and trust. They exacerbate societal tensions and amplify public polarization. They increasingly condition and limit notions of self-determination, and could continue to do so with the future generations to come.

The rise of cognitive-emotional conflicts and the subsequent "trust-deficit disorder" they unleash is born out of the entanglement of technology, data, and geopolitics. The convergence of AI with other emerging technologies creates the potential for deception and subversive attacks that manipulate populations' perceptions. In such "AI convergence," AI optimizes data, processes and techniques that are part of physical, digital and biological lives.

AI is combining with an extraordinary array of other technologies, from cyber and biotechnologies, affective computing and neurotechnologies, to robotics and additive manufacturing. Computer scientists are developing deep learning algorithms that can recognize patterns within massive amounts of data with superhuman efficiency and, increasingly, without supervision. At the same time, geneticists and neuroscientists are deciphering data related to genomes and brain functioning, learning about human health, well-being and cognition.

The result? Functional capabilities for averting crises that were previously unimaginable are now real, and they are upgrading efforts from precision medicine and food security to conflict prevention.

For example, deep learning algorithms are diagnosing retinopathy in patients living in rural India where there is a shortage of ophthalmologists. The same algorithms can identify malign biomarkers among large swaths of genomics data from human populations to design blood-tests for various cancers. Portable genomics sequencers bring the lab to the jungle, allowing for the diagnosis of Ebola viruses in "hot zones." In a community bio-lab in the Bay Area, a fifteen-year-old is using additive bio-manufacturing to print a molecular patch that could one day help treat thousands of people who, like her brother, suffer from a rare lung disease. In Shenzhen's Open Innovation Lab, young inventors have designed wearable devices that rely on image recognition to help farmers detect diseases on crops. Companies like

Zipline are using AI technology in autonomous drones to deliver critical medical supplies, such as vaccines, to rural hospitals in Africa.

AI could also become a powerful tool for the international development efforts by multilateral organizations. The World Bank, in collaboration with other global partners including the UN, is building a Famine Action Mechanism, which relies on deep learning systems, developed by Microsoft, Google and Amazon, to detect when food crises will turn into famines. The same tool allows agile financing to be connected directly to areas of food insecurity. UN Global Pulse is developing algorithms to characterize, quantify and automatically detect elements of hate speech, from heterogeneous data sources. UNICEF is collaborating with MIT on deep learning expertise to simulate images of major global cities "in ruin," to help empathize and connect with the suffering of those who have experienced bombing, loss and war.

The combined optimization of biometrics, genomics, behavioural, and physical systems' data is giving rise to "affective computing" – algorithms that can successfully analyse, nudge, and communicate with us. This form of emotional analysis will improve human-machine interactions in applications that could empower underserved populations, from precision medicine to targeted education. For instance, Affectiva, one of the leading affective computing companies, is interested in quantifying emotions to "get a complete understanding of individuals' overall wellbeing" as a suicide prevention tool.¹ Affectiva has also developed Peppy Pals,² a series of education apps that teach children about social and emotional intelligence by learning from situations in an online world.

As was emphasized in a report from Finland's Parliamentary Committee for the Future, societies are facing a plethora of other changes in addition to technological change: "Parallel to technological change, there is change in societal and social structures, business models, knowledge, skills and professions, cultures and the ways of life."³

While these trends may unlock enormous potential for humankind, the convergence of AI with other emerging technologies also creates unprecedented vulnerabilities and risks for global security. Think of deep learning systems able to drastically intensify the nature and scope of cyber espionage and cyberattacks within increasingly intelligent and connected cities and laboratories. The same algorithms can rely on emotion analysis to generate deepfakes - highly realistic photos or videos of events

that never occurred - which will enable propaganda, strategic deception and social manipulation to be both more scalable and targeted.

We face a new species of technologies, - those that are increasingly digitized, enable and converge with each other, and that are harnessed in cyberspace. This new species of technologies essentially exploits the data captured from physical and biological systems. Almost all physical and biological matter today can be turned into a digital blueprints or binary code, from the genomes of humans and living organisms, to organs and fingerprints, to DIY drone designs, to nuclear power plant parts, even to brainwaves and nanobots delivering tailored molecules in human bodies.

Converging technologies therefore create networks of digital information that enhance, run, shape and integrate cyberspace with daily ways of living. They slowly “invade” bodies and cities. As they become digitized, converging technologies also become more decentralized, beyond State control, and available to a wider range of actors around the world.

Under the same impulse, cyberspace has become not only a new domain of fierce competition over information, business, and strategic technological operations, but also a new battlefield, in ways that blur the line between war and peace and make each of us a potential target of postmodern conflict. Governance actors are only starting to realize how the manipulation and misuse of converging and connected technologies in cyberspace can threaten the truth, polity, and collective security.

The convergence of AI with emerging technologies can therefore raise new complex security challenges that are neither well understood nor well anticipated globally. There is a crucial need to map out and analyse for whom the convergence of dual-use technologies will generate not only new powerful opportunities to flourish, but also pervasive hybrid threats. This is part of a larger ongoing

“ While these trends may unlock enormous potential for humankind, the convergence of AI with other emerging technologies also creates unprecedented vulnerabilities and risks for global security.

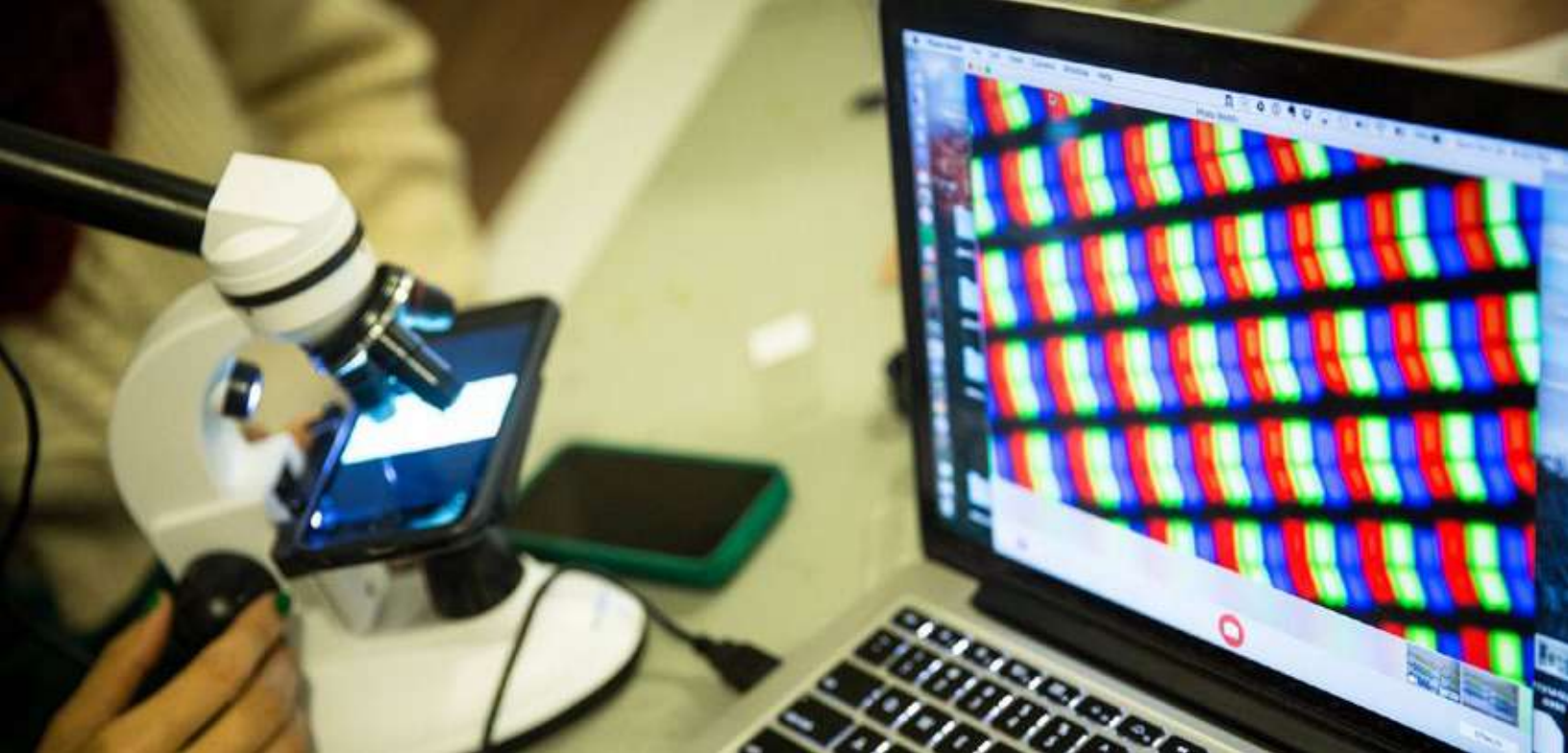
effort to investigate how legacy paradigms of global security and governance are evermore challenged by the combination of powerful dual-use technologies.

Today, converging technologies already impact digital, political, economic, and (bio)-physical security. This affects societal structures and networks with implications for several UN agendas, from prevention, to security, and human rights. Far beyond what was conceived through traditional security and military doctrines, we face new challenges that pertain to human, socio-economic and political security. Competitions and conflicts arise between societies, not armies. What matters is not who wins new territories, but who wins the data, the trust, the hearts and the minds of citizens within a country or polity. In this context, preserving and enhancing societal resilience will become the most important asset for leaders.

AI and converging technologies pose a particular challenge for the multilateral system. The UN was established many decades before the technologies that are driving today’s world. Given its constituency, it is not obviously situated to address the spread of AI-driven technologies beyond the State, into the hands of corporations, groups and individuals. Some major corporations may see little value in bringing multilateral approaches to bear on lucrative, proprietary technologies, while powerful Member States appear more focused on crystallizing their own competitive advantages than in working together to build responsible governance systems in the cyber and bio-tech spheres. Especially in relation to peace and security, Member States may resist attempts to involve the multilateral system in what they see as a fierce competition over powerful new systems of influence. At the same time, there is a clear understanding that the kinds of risks created by AI convergence do not occur neatly within State boundaries: they manifest globally and must be dealt with across and among States.

Against this background, the multilateral system urgently needs to help build a new social contract to ensure that converging technologies, in particular AI, are deployed safely and aligned with the ethical needs of a globalizing world. This new social contract has to create opportunities and incentives for citizens, experts, and in particular, the private sector to commit to a serious, inclusive exercise of foresight to distinguish the technological hype from the hope, the speculation from reality, and in so doing shape responsible technological futures.

The UN can help bridge the gap between the corporate and geopolitical worlds. Among other issues, it will be crucial to consider how corporations should be involved in the future governance of AI and converging technologies at a global scale. The UN should strengthen its engagement with the technology platforms driving innovation in AI and converging technologies. It should also offer a forum for



© David Sun Kong, MIT Media Lab
IT and biotech to design decentralized tools for bio-analysis

“ **What matters is not who wins new territories, but who wins the data, the trust, the hearts and the minds of citizens within a country or polity.**

truly meaningful cooperation between those actors, along with State actors and civil society. This requires trust, but also the prioritization of a real engagement over the comfort of institutional roles.

How can the UN harness internal collective and technological intelligence to offset asymmetries of knowledge and power when shaping technological futures? How can the UN be an engine of trust that can help build a new social contract to ensure that emerging technologies are aligned with the social and ethical needs of a globalizing world?

Above all, this requires humility and vision, inclusion and cooperation. On 12 July 2018, the UN Secretary-General António Guterres established a High-level Panel on Digital Cooperation to foster “the broader public debate on the importance of cooperative and interdisciplinary

approaches to ensure a safe and inclusive digital future for all taking into account relevant human rights norms.”⁴ The mandate of the High-Level Panel on Digital Cooperation is to elicit, in multi-stakeholder discussions, the values that will inspire and the mechanisms that will support new forms of technological cooperation.

For true AI and cyber cooperation, the UN will also need to be a bridge between the interests of nations that are leading in AI and converging technologies and those that are struggling to keep up. In this competition, the UN has a crucial role to play, by providing an inclusive, safe-space where cooperation is rewarded over exclusion, and where innovation in AI and converging technologies are framed as a global public good.

How can the UN better understand the trends and risks associated with AI convergence, anticipating where new crises may emerge? What tools should the multilateral system have to build a common understanding of these risks, and put in place effective prevention strategies? How can Member States begin to overcome what the Secretary-General has called a “trust deficit”⁵ and begin to tackle these issues together?

This report offers a step towards such a common understanding as well as an opportunity to tailor and adapt prevention strategies to the era of AI convergence.



THE PARADIGM OF AI CONVERGENCE

The age of AI convergence builds upon the digital revolution with a global surge in computing power and big data capacities. Such convergence will pervade an array of existing socio-technical systems and become prominent for multiple sectors, from healthcare, transportation, and energy to security and military enterprises.

Think of smart drones, self-driving cars, brain-computer interfaces, and intelligent and connected bio-laboratories. Transformative technological change driven by AI and converging technologies is entering societies at an increasing pace.

AI AND CONVERGING TECHNOLOGIES

The technological story of AI started at the end of World War II.⁶ The birth of AI in the mid-twentieth century coincided with the subjugation by mathematician Alan Turing of the German “Enigma,”⁷ a network of cypher-machines able to encrypt commercial, military, and security information. In the secret rooms of Bletchley Park, Turing’s visionary mind built another electromechanical machine, called the Bombe, which could decipher the combinatorial permutations of Enigma and reveal all daily German Naval traffic.⁸ Turing had the intuition that one day computers would become so powerful that they could think, and so he designed the Turing test of computer intelligence.

The term *artificial intelligence* generally refers to “the use of digital technology to create systems that are capable of performing tasks commonly thought to require intelligence.”⁹ Dr Greg Corrado, Principal Scientist and Director of Augmented Intelligence Research at Google, provides a useful definition: “the art and science of building machines that have the appearance of useful intelligence.”¹⁰ Machine learning, usually considered a subset of AI, describes “the development of digital systems that improve their performance on a given task over time through experience.”¹¹ Deep neural networks

are machine learning architectures based on myriad layers of virtual neurons and designed to mirror the way humans think and learn.¹²

Characterized by extraordinary functional capabilities, AI first captured the imaginations of academic researchers, then industry leaders, and finally of policy-makers around the world.

At its core, what AI does is optimize data analytics. AI can augment human capacity with respect to big data by collecting and cataloguing information faster and more effectively than human cognition alone. In automating those processes, there are opportunities to gain insights about patterns and correlations that would not otherwise be feasible.

In its most innovative form, called deep learning, AI optimizes predictive reasoning by learning how to identify, classify and model high-level abstractions within massive amounts of unstructured or unlabelled data. With this super-computing efficiency – being able to simulate myriads of scenarios in seconds – deep learning offers unmatched investigative opportunities, such as comparing genomes within an entire population, recognizing a certain face out of a crowd, or labelling any location on earth based on millions of pictures. By modelling and predicting complex interdependent systems, AI can help provide decision support for various futures scenarios. For example, work is underway to develop domain ontologies and automated processes for helping emergency managers and first response teams make critical, high-pressure decisions with minimal situational information.

The current technological era is characterized by the convergence of AI with an extraordinary array of other emerging technologies, ranging from cyber and biotechnologies, affective computing and neuro-technologies, to robotic and automated manufacturing.

ARTIFICIAL INTELLIGENCE

AI is a sub-field of computer science created in the 1960s. AI is an umbrella term with no formally agreed upon definition. The narrow form of AI (i.e. the AI currently in existence, not artificial general intelligence or superintelligence) is capable of performing defined tasks such as stock trading, image recognition, medical diagnosis, automated industrial robotics, and remote sensing.

Machine Learning: an iterative process that uses statistical techniques to design software capable of performing specific tasks without having to explicitly program or define rules for each operation. Powered by algorithmic models trained to recognize patterns in collected data, machine learning requires access to exceedingly large amounts of training data to function well.

Deep Learning: also known as neural networks, simulates an array of neurons (called nodes) in order to learn to recognize patterns in digital representations of sounds, images, and other data.

Deep Reinforcement Learning: a type of deep learning that relies on trial-by-error and learns solely from rewards and punishments.

Expert Systems: also referred to as knowledge-based systems, is a complex AI program that uses heuristics to solve decision-making problems with human level competency. The process of building expert systems is called knowledge engineering and has two components: the knowledge base and the reasoning engine. For example, CaDet is an expert system designed to assist primary care physicians in detecting cancer earlier.

Computer Vision: a sub-field of AI that leverages machine learning and vision techniques to train computers to interpret and understand the visual world. Using digital images from cameras, sensors, videos, and deep learning models, machines can accurately identify and classify objects from a single image or sequence of images. Applications of computer vision techniques range from automated industrial inspection and the identification of malign tumors to being able to count the number of people in a crowd.

Natural Language Processing (NLP): aims to bridge the gap between the languages that humans speak and the languages that computers use to operate. By using algorithms that are able to identify key words and phrases in natural language (i.e., unstructured written text), AI applications are able to determine the meaning of text. For example, AI home assistants like Alexa, Cortana, and Siri rely on NLP to understand and perform commands ranging from providing the weather forecast to ordering a pizza.

Affective Computing: aims to bridge the gap between human emotions and computational technology. Current research addresses machine recognition and modeling of human emotional expression, including the invention of new software tools to help people gather, communicate, and express emotional information and better understand the ways emotion impacts health, social interaction, learning, memory, and behavior. Applications of affective computing range from targeted advertisements to assessing the severity of depressive symptoms using wearable technology and personalized humanoid robots for Autism therapy to personalized animated movies.

Neurotechnology: the assembly of methods and instruments that enable a direct connection of technical components with the nervous system. These technical components are electrodes, computers, or intelligent prostheses. They are meant to either record signals from the brain and “translate” them into technical control commands, or to manipulate brain activity by applying electrical or optical stimuli. Applications of neurotechnologies range from brain-computer interfaces and neuro-entertainment such as virtual reality to neuro-privacy and biofeedback.

Robotics: a branch of technology that deals with robots, programmable machines equipped with a suite of sensors and actuators that are usually able to carry out actions autonomously, or semi-autonomously. Examples include drones, industrial robots, surgical robots, etc. Unless equipped with artificial intelligence, robots are not able to learn, analyse, or optimize data.

BIOTECHNOLOGIES

Biotechnology is a broad term that refers to technology that uses living organisms, or uses living organisms as models, to make products.

Genome Editing: the process by which the genomic DNA of living cell is cut and subsequently modified using a nuclease protein. CRISPR, the most widely used and studied genome editing tool, allows scientists to add, delete, or modify multiple genes simultaneously with a high degree of precision.

Gene Drive: a genome editing technology enabled by CRISPR that is capable of passing genes down to successive generations of descendants via sexual reproduction, even if the genes do not necessarily confer a fitness advantage. Since gene drivers could alter the genetic traits of an entire species, they can be powerful tools for sustainable ecosystem management. For example, CRISPR-Cas9 gene drives are being developed as a potential method for suppressing the human malaria vector (i.e., the *Anopheles gambiae*).

Gain-of-Function Research: a research experimentation that aims or is expected to (and/or perhaps, actually does) increase the transmissibility and/or virulence of pathogens. When conducted by responsible scientists, the ultimate objective of GOF research is to better inform public health and preparedness efforts, such as the development of medical countermeasures. However, GOF can also pose risks to biosecurity and biosafety, which resulted in a moratorium on GOF research in the United States from 2014 to 2017.

DNA Origami: Nanotechnologies and cutting-edge biotech are also being co-opted to design and implement new methods of delivery for biological cargos. The advent of DNA origami – the nanoscale fabrication of structures using synthesized DNA molecules– extends the abilities of how drugs or other materials can be delivered into the body to enact a desired change. Recent work showing the ability of DNA origami robots to work together in ways similar to those bacteria employ, highlights the future of biological delivery.

Automated Bio-Manufacturing: using software and hardware automation, biotech companies are increasingly aiming to transform the process behind engineering life. Automated labs or labs in the cloud allow you to perform remote online DNA sequencing, DNA assembly and synthesis as well as genome-editing. The DNA that is synthesized can even be tested in cell lines and bacteria.

CYBERTECHNOLOGIES

Cybertechnology refers to a wide range of computing, communications, and networked devices that operate in cyberspace, from smartphones and computers to the Internet itself and networked devices connected to the Internet.

Internet: the wider network that allows computer networks around the world to communicate with one another. The infrastructure of this global system of interconnected computer networks consists of worldwide fiberoptic cables which are submerged on the ocean floor, satellites, wifi towers, data centers, routers, servers, repeaters, and ultimately, individual computers.

Cloud Computing: the use of a network of remote servers to store, manage, access, and process data rather than a single personal computer or hard drive.

Internet of Things (IoT): refers to the constellation of billions of connected devices that offer exponentially more data points on how smart objects perform within technical systems. Devices that constitute IoT range from smart home devices such as Google Home and Smart Televisions to more inconspicuous internet-connected devices like smart coffee makers and smart trashcans.

Deep Web: 95 per cent of the internet, which is unindexed and invisible to standard browsers. This occurs for various reasons such as: password protected pages; office intranets; and pages no one links to.

Dark Web: Buried in the deep web is the dark web, which hides certain websites by using addresses that are only accessible by using special software. Specialized software such as Tor, originally called The Onion Router because it layers internet traffic, creates user anonymity through encryption, which hides the origin of data. Certain markets (for instance human trafficking) operate primarily within the dark web specifically because the sites are hidden from average internet users and generally use cryptocurrency (another added layer of encryption) as a payment method.

Cyber Immune System: using machine learning techniques to detect cyberattacks based on anomalies detected on normal network traffic in cyberspace, not the traditional way, which is based on signatures. Once trained, the cyber immune system calculates the probability that a certain deviant pattern is malicious and constantly updates its results based on new evidence. Similar to human immune systems, cyber immune systems are vulnerable to autoimmune attacks due to design flaws or bugs in the source code.

Quantum Computing: quantum computers operate on completely different principles to existing computers, which makes them well suited to solving particular mathematical problems such as finding very large prime numbers. Since prime numbers are important in cryptography, it is likely that quantum computers would quickly be able to crack many of the systems that keep online information secure. As such, researchers are already trying to develop technology that is resistant to quantum hacking, making cryptographic systems much more secure than their conventional analogues.

A TIME OF TECHNOLOGICAL TRANSFORMATION

The paradigm of AI convergence embodies *six major shifts* that are leading to powerful opportunities for prevention, but also pervasive hybrid threats to human and political security.

CONVERGENCES

The *first shift* is convergence itself, the combinatorial and converging nature of the current technological revolution. Technologies are becoming complex hybrid systems that are merging and enabling each other, with drastic changes in velocity, scope and system-wide impact.¹³

Harnessing AI within biotech research allows tech-leading nations to optimize the bio-data and capture the value of another country's bio-economy.¹⁴ Emotional analysis enables hyper-personalized campaigns in which key demographics are manipulated to affect voting behaviours at crucial times.¹⁵ Think of self-learning algorithms able to design ever more sophisticated human forgeries based on biometrics analysis and behavioural mimicry.¹⁶

The same intelligent algorithms could be used to model the spread of pathogens and monitor for outbreak in human communities and bodies alike.¹⁷ Such algorithms could also help simulate the effects, on the overall human genome, of editing specific genetic sequences.¹⁸ It is then possible to fathom drone swarms relying on cybernetworks of strategic intelligence to perform precision repair or precision attacks on physical and biotech infrastructures.¹⁹ Technological convergence results in emerging properties that have never been dreamt of, but that are also increasingly difficult to comprehend, foresee, mitigate, and control.

As they become increasingly digitized, converging technologies operate in cyberspace, where the national and international rules of the game have yet to be created. Foucault described the power of modern biology and medicine as an "explosion of numerous and diverse techniques for achieving the subjugation of bodies and the control of populations."²⁰ Cyberspace, as a new operating realm, provides something similar. The new alliance between cyber and biopower will influence the kinds of governance models under which citizens will choose or be forced to live.

“As they become increasingly digitized, converging technologies operate in cyberspace, where the national and international rules of the game have yet to be created.”

DECENTRALIZATION

The *second shift* is characterized by new processes of technological decentralization and democratization, which atomize responsibility among disparate actors for the benefits and risks of new technologies. As they combine, emerging technologies move beyond State control, involving new actors and challenging established expertise and governance models. Gaining access to AI software and scientific findings is relatively easy, in part because the culture of AI research is characterized by a high degree of openness. At conferences over the past few years, for example, AI engineers have shared training data, algorithms and knowledge of how to design software.²¹

Similar dynamics of democratization have transformed the field of biotechnologies. For about \$200, the start-up Odin²² sells gene-editing kits that can design gene-therapies at home, making genetic engineering available to consumers. Biotechnologies have progressed to a point where it is now possible for high school students to learn how to edit the genetic code underlying cells and proteins.²³ As the next generation learns how to turn their own ideas and know-how into new bio-constructs, advances in biotechnologies will accelerate in unprecedented ways. Just like algorithms in software engineering, human cells have become material for intelligent design.

More sobering, current investigations show that malicious hacker communities leverage the deepweb and darkweb to buy, sell, and trade cybermalware, exploits, and stolen data.²⁴ Many security researchers have leveraged this data to understand the activities of such groups with the goal of identifying emerging threats, supporting forensic investigations, and even predicting cyberattacks.²⁵

Ubiquitous access to data, source codes and foundational cyber and biotechnologies means that the responsibility for the misuse of these technologies is distributed among designers, regulators, users and hackers. Increasingly, experts and decision-makers in the public and private sectors will have to face a decisive question: what modalities of governance can be resilient and adaptive enough to the increasing decentralization of converging technologies?

CIVIL-MILITARY FUSION

The *third shift* is spurred by civil-military fusion, the increasing permeability between civil and military contexts, between civilian and weapon technologies. New forms of invasive cyberwar and cognitive-emotional conflicts are waged with powerful implications for collective (bio)-physical, digital, and political security.²⁶

New hybrid threats will arise from the reliance of critical physical and biotech infrastructures on cloud computing and intelligent and connected technologies. AI is increasingly used to map and measure biological functions, parsing through large collections of genomics data stored on cloud platforms.²⁷ Most corporate AI platforms already have access to an individual's online behaviours, relationships, health and emotional states – but, soon enough, they will acquire baseline information about their vital signs, organs and genomes. These integrated intelligent networks could be called the “Internet of Bodies” and the “Internet of Genomes.”

While State-sponsored cyberattacks and cyber espionage are hardly new, the use of automated bio-laboratories, and the growth of the Internet of Genomes, will only exacerbate existing cybersecurity vulnerabilities. Security experts are becoming increasingly concerned about AI technologies being used to automate cyberattacks on medical or biotech supply chains.²⁸ Deceptive algorithms could be used to spread data-poisoning attacks²⁹ with the intent to falsify, erase or steal intelligence, which could affect capacity to detect pathogens in food production, and knowledge about how to treat cancers in subpopulations. Collective bio-intelligence is directly at risk.

Beyond societies' physical and bio-security, another form of national asset will become a target: political and social resilience. Converging technologies in cyberspace let adversaries interfere more directly than ever with a targeted nation's political processes and the minds of its citizens. Conflict now opposes societies and competing ideological systems, not just armies. In a visionary 1993 report, Arquilla and Ronfeldt introduced the concept of “netwar,” in which adversaries aim to “disrupt, damage or modify what a target population knows or thinks it knows about the world around it.”³⁰

“New hybrid threats will arise from the reliance of critical physical and biotech infrastructures on cloud computing and intelligent and connected technologies.”



Unsplash/Giu Vicente

Young child experiences the vast new world of virtual reality

By manipulating sound, images, and content, AI technologies allow the simulation of strategic intelligence, such as sophisticated forgeries from deepfake³¹ videos to deepfake back stories and cover-ups. The use of AI in the information ecosystem will enable propaganda and deception to be both scalable and widespread, but also targeted and efficient. What is at stake is the cognitive and emotional disruption of societies. AI and converging technologies are on the verge of changing the global cyber intelligence system.

POTENTIAL FOR ESCALATION

The *fourth shift* is the death of traditional conflict proportionality in a context of increasing geopolitical competition between nations that are technological leaders. When AI systems are used in cyberoffenses, their autonomy and capacity to improve strategically and to launch increasingly aggressive counter-attacks might lead to breaches of proportionality. Such escalation could, in turn, trigger kinetic conflicts.³² Technological convergence and interdependence will also augment



What is at stake is the cognitive and emotional disruption of societies. AI and converging technologies are on the verge of changing the global cyber intelligence system.



the potential to endanger an array of critical civilian infrastructures, amplifying the risks of escalation and breaches of proportionality. This will be even riskier in an era in which powerful nations are racing to achieve information, economics and security supremacy.³³

GLOBAL COMPETITION

The *fifth shift* entails the severe consequences of global competition over converging technologies³⁴ for tech-taking nations – i.e., those that lack the capacity to develop AI and converging technologies themselves. Instead of greater digital and technological cooperation across borders, increasing competition is likely among major powers, as well as growing inequalities between tech-taking and tech-leading countries. In security terms, this will give rise to “vulnerable links” – States that are too vulnerable to steer governance of converging technologies away from socio-political disruptions and weaponization. These vulnerable States could become liabilities for whole regions.

COGNITIVE-EMOTIONAL CONFLICTS

Sixth and finally, a variety of cognitive-emotional campaigns are underway, enabled by converging technologies and cyber interconnectedness.³⁵ The centre of gravity for at least some conflicts, such as Russia’s intervention in Crimea³⁶ or the Islamic State (IS)’s weaponization of social media, is shifting away from exclusive use of military forces towards targeting critical cyber, bio- and physical civilian infrastructures, political processes and security, and social cohesion of populations. The resilience of a nation and its political institutions ultimately lies in the minds of its citizens, who today are under constant pressure. The most damaging pressure is what Yuval Harari calls the “fear of irrelevance,”³⁷ when citizens, replaced as labour forces by powerful technologies, fear not to be exploited but to become irrelevant to a nation’s political fate and trajectory. Such fear of irrelevance is a fertile ground for a population to become a victim of deception and manipulation by foreign actors skilled in AI and cyber techniques.



PERSON

PERSON

PERSON

PERSON

PERSON

PERSON

SCENARIOS OF AI CONVERGENCE

The prior section offered a narrative description of the major trends in AI convergence. This section explores how those trends could develop in the future. Given the complexity of the issue, there is no linear methodology for anticipating the kinds of change that will take place in the coming period. However, this section relies on well-known foresight methods for constructing scenarios,

including horizon scanning as well as drivers and trends impact analysis (for more in-depth methodological description, see Annex 2). The below scenarios explore how AI and emerging technologies converge to amplify security opportunities and risks in complex socio-technical systems.

THE DECEPTION MACHINE: DEGRADATION OF TRUTH AND TRUST

“Power is in tearing human minds to pieces and putting them together again in new shapes of your own choosing.”

– George Orwell, 1984

The “Deception Machine” scenario revolves around the convergence of AI and affective computing, biometrics, neuro-technologies, and cybertechnologies with systemic implications for political, digital, and human security.

“FAKE INTELLIGENCE”

AI technologies are on the verge of changing the global intelligence system. These technologies will enable the automated generation and simulation of data, media, and strategic intelligence with significant implications for the future of propaganda, deception, and social engineering.

Deep learning algorithms excel at classifying and optimizing data. Yet, increasingly, algorithms can also automatically generate new content such as photographs, video, text, and even basic stories or news articles. Such capability goes from altering facial features and expressions, gait, and biometrics to simulating behaviours on video, in real-time, using commercial webcam equipment. If an individual has a digital footprint that includes, for example, talks and podcasts, deep residual networks are also able to reproduce a synthetic version of that voice.³⁸

Take deepfakes as an example. Sophisticated AI programs can now manipulate sounds, images and videos that create impersonations that are often indistinguishable from the original. Deepfakes rely on deep residual networks that can, with surprising accuracy, read human lips, synthesize speech, and simulate facial expressions and bodily movements. Undergraduates have generated a 3D synthetic version of Tom Hanks’ face, which gave

them the ability to control his facial expressions.³⁹ A team of researchers at Berkeley can make anyone breakdance on video by using deep learning algorithms to simulate breakdance moves.⁴⁰ Today, AI-enabled forgery technology still requires quality tools and relative expertise. In the near future, however, it will be available to amateurs, dark web crusaders, and both State and non-State actors.

Once released outside the lab, such simulations could easily be misused with worrisome implications. On the eve of an election, deepfake videos could falsely portray public officials being involved in criminal or unsavoury behaviours. Public panic could be sowed by videos warning of non-existent epidemics, health safety scandals or widespread cyberattacks. These forged incidents could potentially lead to international political or military escalations.

“FAKE INTELLIGENCE” SCENARIOS

Imagine a few political scenarios around the automated generation of forgeries.

Certain populations around the world accept hyper-targeted disinformation campaigns that favour totalitarian regimes.⁴¹ Digital dictatorships⁴² that rule over official media will have the AI know-how to generate high-quality forgeries to influence and manipulate public opinion.⁴³ Such forgeries, when targeted at an ethnic, religious or cultural group, could provoke violence and discriminations. In Myanmar, a UN report confirmed that Facebook posts have fuelled virulent hate speech

directed at Rohingya Muslims.⁴⁴ As the Myanmar case demonstrates, disruption or tragedy do not always come with fire and smoke, but, sometimes, are virally amplified by online narratives.

With the proliferation of sophisticated deepfake videos, deepfake backstories and cover-ups, even qualified news reporters, decision-makers and diplomats will struggle to parse propaganda and disinformation from real news.⁴⁵

Decision-makers within governments will face pervasive data-manipulation through the use of AI-enabled forgeries.⁴⁶ Opponents of a major military power could generate large quantities of fake intelligence, such as deepfake videos where the military forces in question could be shown murdering innocent civilians in war zones.

Sophisticated forgeries could erode the stability and legitimacy of political and financial institutions at crucial times. Such deceptive tactics could prove difficult to manage during contentious elections in transitioning democracies and could cause market instability.

Fake intelligence could reshape international diplomacy and cooperation. For instance, the company Lyrebird developed an AI-enabled voice imitation algorithm that it says “can not only mimic the speech of a real person but shift its emotional cadence – and do all this with just a tiny snippet of real-world audio.”⁴⁷ In a public demonstration, the company released fake speech generated by deep learning algorithms using voice samples from Presidents Trump and Obama.⁴⁸ The company acknowledges that voice recordings are considered strong pieces of evidence, and that their technology could have significant implications for international diplomatic negotiations. On the eve of an international summit or agreement, forgeries could break down important relationships between States and disrupt the nature of international diplomacy.

Algorithmic power may increasingly be used at scale by an array of decentralized actors to target and manipulate political communication. In this context, there is an urgent need to think proactively, together with the private sector, about what it means to carry out in-depth human rights impact assessments of AI systems. Such assessment could include implications of AI technologies on racial and religious minorities, political opposition parties and activists.

In general, the deployment of AI-enabled forgery technology will drastically alter relationship to evidence and truth across journalism, criminal justice, conflict investigations, political mediation and diplomacy. The capacity of a range of actors to influence public opinion with misleading simulations could have powerful long-term implications for the role of the UN in maintaining peace and security. By eroding the sense of truth and trust between citizens and the State—and indeed among States, and among societies—truly fake intelligence could become deeply corrosive to the global intelligence and governance system.

“ The capacity of a range of actors to influence public opinion with misleading simulations could have powerful long-term implications for the role of the UN in maintaining peace and security.

HYBRID AI-CYBER INFLUENCING

Deep neural networks are increasingly good at self-learning with less supervision – and are therefore both efficient and scalable. By allowing the analysis of individual communication, perception and emotion to be automated, AI systems can increase anonymity and psychological distance in cyberoperations. In the near-future, automated cyberoperations, led by deep learning, will therefore be more effective, finely targeted, difficult to attribute, and likely to exploit evolving vulnerabilities in AI and human systems.⁴⁹ One pervasive security threat will be new forms of hybrid influencing made possible by the automation of social-engineering attacks. Many major cybersecurity incidents rely on social engineering where malicious actors target the social and psychological vulnerabilities of humans within chains of command. The goal is to manipulate command and control organizations to compromise their own safety and security.

The most extreme level of AI-cyber influencing could be the convergence between cyber and kinetic operations.⁵⁰ During combat, for example, a cyberattack could successfully compromise a combination of personal and official channels of communications. Armed forces become the target of psychological manipulation, for instance by being exposed to forged video “evidence” of their adversary’s military superiority. Such hybrid AI-cyber influencing then allows the “supposedly” superior military force to achieve a kinetic advantage.

The consequence of hybrid AI-cyber influencing will be the weaponization of lucrative global digital networks. In this new convergence merging war, tech and politics, the next winning move will be to manipulate information infrastructure and its secrets. We increasingly face geopolitical conflicts in which psychological and algorithmic manipulation are becoming endemic in cyberspace, an ecosystem of nearly four billion minds.⁵¹ Yet, the impacts felt are real in the physical world, from influencing elections, to destabilizing economies and political regimes, to terrorist groups livestreaming their attacks in the streets of Iraq and Syria.

“ In this new convergence merging war, tech and politics, the next winning move will be to manipulate information infrastructure and its secrets.

ELECTIONS, EMOTIONS, AND PSYCHOMETRICS MANIPULATION

By searching for patterns through behavioural data collection – the online footprints left on social media – deep learning algorithms can predict individual personality assessments or digital avatars.⁵² These digital avatars will be used for psychometric micro-targeting to tailor propaganda to specific users. Emotion analysis will also allow the deployment of political bots that can adopt human-like tactics to manipulate users.⁵³ By promoting targeted propaganda, falsehood, confirmation biases, and incendiary content, AI-bots and algorithms will even play a significant role in setting the national and international political agenda.

Swarms of bots, Facebook dark posts, and fake news websites have claimed online territory, with significant global repercussions.⁵⁴ Just consider a few recent events: In the 2016 US presidential elections, a foreign State allegedly conducted a massive campaign that included paid ads, fake social media accounts and polarizing content. Analysts have surmised that Russian operatives relied on social media to spread anger about racial inequalities and to sow confusion among African Americans that was ultimately aimed at discouraging them from voting.⁵⁵

The power of propaganda on social media is certainly difficult to curtail. It requires not only algorithmic detection systems, but also emotional intelligence and linguistic and local knowledge. Across India in summer 2018, manipulative messages on social media sites, including Facebook and WhatsApp, painted certain groups as responsible for child abduction.⁵⁶ The hysteria led to more than 30 deaths and left many injured.

Facebook and WhatsApp are used by over 200 million people in India and the number is expected to double in the next few years.⁵⁷ With an upcoming general election in May 2019, which will determine whether Narendra Modi will have a second consecutive term in office as President, fears are rampant that social media platforms will be used for unrest again.⁵⁸ As a preventive fix, WhatsApp is now limiting users in India to share only up to five online chats.⁵⁹

Increasingly, both engineer and policy communities will work together to try to produce tech fixes and

countermeasures for this post-truth era. The tech sector is creating a whole business for those who can pay to be protected: deep learning to isolate you from disinformation and social nudging.⁶⁰ But for those who cannot pay, tech solutions are quite dire. “Solutions” such as WhatsApp’s preventive fix seem unfair and inadequate for a rapidly changing world.

AFFECTIVE COMPUTING AND NEUROMODULATION

Relying on automated behavioural analysis, AI already enables targeted propaganda to spread more efficiently and at wider scale within the social media ecosystem. For instance, *Cambridge Analytica* deployed intensive and widespread social media data-mining operations to define and curate the psychological profiles of millions of Facebook users.⁶¹ Another company, *Mindstrong*, collects individuals’ passive data (such as how they type, scroll and click) to create digital phenotypes or avatars based on a set of biomarkers that measure their emotional state and cognitive functioning.⁶²

Yet, there are many more strategies in the neuro-biological toolbox. Among them is affective computing, which entails the combination of algorithms with biosensors to create AI systems that can decipher human emotions and predict behavioural and emotional responses.⁶³ Affective computing may one day enable us to identify the emotional triggers that push individuals or population subgroups to violence. This, in turn, would open up the possibility of individuals or organizations (hackers, corporations or government agencies) using such knowledge to spark and manipulate violence.

Emotional analysis and affective computing will enable disinformation campaigns tailored for different subgroups, leading, for instance, to highly effective election interference.⁶⁴ More broadly, the use of converging technologies to amplify cognitive-emotional tensions and vulnerabilities could erode the power of democratic institutions, sow civil violence and enable new forms of dictatorships.

While neuro-technologies are not yet an integral part of daily life, developments in the field aim at deciphering people’s mental processes and modulating the brain functions underlying their intentions, emotions and decisions.⁶⁵ These new forms of neuro-modulations would drastically impact the potential for social engineering, psychological manipulation and other techniques of subversion and deception.⁶⁶

THE RISE OF HACKTIVISM AND CYBER MERCENARIES

“Project Raven” is a story of cyber mercenaries: former US National Security Agency specialists who harnessed cutting-edge cybertheft and spying tools on behalf of

“ **Emotional analysis and affective computing will enable disinformation campaigns tailored for different subgroups, leading, for instance, to highly effective election interference.** ”

a foreign intelligence service that keeps thousands of civil society activists, journalists and political figures under tight digital surveillance.⁶⁷ To do so, these cyber mercenaries relied on an insidious automated hacking application called Karma, which made it possible to hack the iPhones of users all over the globe.⁶⁸ In 2016 and 2017, Karma was used to spy on hundreds of targets in the Middle East and Europe, including senior-level political figures.

The Project Raven story also unveils the role that former American cyber espionage agents play in foreign hacking operations. Increasingly, as cyber, biotech and AI become decentralized, and as tacit knowledge escapes the confines of securitized labs, national and international authorities will have to face a complex question: how to govern the dissemination of dual-use expertise from companies, governments and covert operations to the wild, or from one country to another?

The question about dual-use technologies is not new, but the pervasive digitization and subsequent decentralization of technologies certainly is. From intelligence services, gene-synthesis companies, to top cybersecurity and AI private labs, the talent force that is shaping tomorrow's designs, risks, and opportunities, has much tacit knowledge to offer and transfer.

Secrecy among tech, policy and security communities needs to be urgently rethought. New networks of trust – the new version of the old “guild” – to prevent peer-engineers from being bought for their secrets need to be built. As of right now, the new ecosystem for AI-cyber influencing and potential covert war is shaped by engineers in Silicon Valley, Shenzhen, London, and a small handful of others. The new rules and norms for responsible behaviours in cyberspace await definitions and actions by politicians, lawyers, and diplomats.

The pervasive access to dual-use, converging technologies is also bringing forth a new form of virtual, viral insider threat generated by increasingly powerful “information warriors.” When IS increased its visibility and power through social media, its extremely violent propaganda created another cyberspace phenomenon.⁶⁹

Digital networks turned into a new hypnotic online show about violence and war.

The global hacktivist group Anonymous came to life in response to IS' invasion of social media.⁷⁰ These volunteers, self-appointed custodians of cyberspace, started by systematically hunting for IS Twitter accounts. They shared this information with Twitter, which in turn deactivated several accounts, as well as with governments. With time, Anonymous organized a form of cyber resistance, combining cyber espionage, hacking techniques and automated bots aimed at destroying IS websites, networks and social media presence. Hacktivism unfortunately did not prevent violence on the ground, but it did create a new form of aggression from within homes and cities.

“ **Secrecy among tech, policy and security communities needs to be urgently rethought.** ”

COGNITIVE-EMOTIONAL CONFLICTS

The convergence of AI, affective computing, biometrics, and cybertechnologies in cyberspace empowers new techniques and lowers the entry cost for new actors to use them. This convergence is giving rise to disruptive tools that erode truth, trust and cohesion within societies. Since 2004, at least 27 European and North American countries have allegedly been victims of cyberattacks, disinformation, and financial influence campaigns crafted for destabilization.⁷¹ As the digital scope of these new forms of hybrid threats is not limited to the West, an increasing number of States might be used as “vulnerable links” in a new virtual geography of conflicts.

The near-future will see the rise of cognitive-emotional conflicts: long-term, tech-driven propaganda aimed at generating political and social disruptions, influencing perceptions, and spreading deception.⁷²

While cognitive-emotional conflicts often entail the weaponization of social media, subsequent phases will increasingly target important elements of human and civilian security, including beliefs, discourses, digital systems and infrastructures critical to health, food, political, and economic security. In November 2018, Russia suggested in official statements that the Pentagon is establishing bio-weapons supply chains on the border between Russia and the Republic of Georgia.⁷³ The narrative amplified fears about potential biosecurity threats to populations.

From controversies about the safety of vaccines, baby food or gene-therapies, to disinformation campaigns about the health of financial institutions, attacks like these surface in cognitive-emotional conflict and are extremely large and complex. It will become urgent for governments to create more agile and efficient early warning systems to detect and analyse the sources of forgeries and propaganda. States will also need to continuously map how these new subversive tools influence public discourse and opinion. However, current detection tools might not be effective enough to protect against technological capacities in affective computing, bio- and neuro-technologies that are aimed at manipulating the hearts and minds of (foreign) citizens.

The “Deception Machine” epitomizes the importance of information superiority and its strategic potential for deception and subversion. The supremacy to acquire data and manipulate others’ beliefs, attitudes, emotions and behaviours seems today more important than material power. This could mean first a shift of power towards actors – States, political elites, financial oligarchies and private technological platforms – that have the capital, the data market and the authority to deploy powerful systems for AI convergence.

Alternatively, if innovation in AI technologies contributes to democratizing access to automated simulation of data, propaganda and cyberoperations, these techniques could then become more attractive to less powerful

actors and nations with minor tech capital and declining economies. This democratization trend could further accelerate the proliferation of fake intelligence, AI-cyber hybrid influencing, and the waging of cognitive-emotional wars, leaving the world in a fight of “all against all.”

In turn, the “trust-deficit disorder” will exacerbate public anxiety about the loss of control to an algorithmic revolution, which seems to escape current modes of understanding and accountability. In the context where the UN provides a forum for inter-State security deliberations, some States might strategically use this context of public anxiety and cognitive-emotional dissonance to impose their own competing versions of reality. Is trust in national and global governance at breaking point?

“ The supremacy to acquire data and manipulate others’ beliefs, attitudes, emotions and behaviours seems today more important than material power.

THE INTERNET OF BODIES, GENOMES, AND MINDS

“What will happen to society, politics and daily life when non-conscious but highly intelligent algorithms know us better than we know ourselves?”

– Yuval Noah Harari in *Homo Deus*

This scenario revolves around the convergence of AI, affective computing, genomics and neuro-technologies, with systemic implications for political, socio-economic, civilian and humanitarian security.

NETWORKS OF PRECISION SURVEILLANCE

AI is increasingly used to map and measure biological functions. Most corporate AI platforms already have access to users’ online behaviours, relationships, political and sexual orientations, health and emotional states – but, increasingly, they will acquire baseline information about their biology, including vital signs, brain functions, organs and genomes. The digital representation of characteristic

data could help create the world’s largest precision medicine dataset – or it could render populations more vulnerable to exploitations and intrusions than ever before.

In the near future, biosensors and algorithms will capture and analyse an ever more refined record of humans’ biometrics, emotions and behaviours. AI will watch, track, and evaluate individuals, from the predictive power of one algorithm to the next. Societies may unwittingly give algorithmic networks unprecedented access to bodies, genomes and minds and create possibilities for social and bio-control that surpass Foucauldian concerns. In this scenario, the “Internet of Bodies” is a world in which everyone is under personalized surveillance.⁷⁴

This new form of intrusive computing in individuals' lives has significant implications for self-determination and privacy, especially children's privacy.

For example, the My Friend Cayla smart doll⁷⁵ sends voice and emotional data of the children who play with it to the cloud, which led to a US Federal Trade Commission complaint and a ban on the doll in Germany.⁷⁶ In the US, emotional analysis is already being used in the courtroom to detect remorse in deposition videos. It could soon be part of job interviews to assess candidates' responses and their fitness for a job.⁷⁷

Facebook is perfecting an AI friend that recognizes suicidal thoughts through conversation and emotional analysis.⁷⁸ IBM Frontiers Institute is developing intelligent implants and nanobots to perform, inside the body, exploratory imagery and report on blood-sugar levels. Start-ups Neuralink and Kernel are working on brain-computer interfaces⁷⁹ that will read people's mental processes and influence the brain mechanisms powering their decisions. One company already relies on wireless sensors to analyse workers' brain waves⁸⁰ and monitor their emotional health.

The tech giant Alibaba is deploying millions of cameras equipped with facial recognition across a number of cities.⁸¹ Chinese police forces have debuted AI-augmented glasses to identify and profile individuals in real-time. Government-sponsored databases of faces, genomes, financial and personal information are being created to connect credit ratings, jobs and the social credit scores of citizens, as well as classifications of DNA samples to find related family members. Recently, 5,000 school students allegedly had their photos and saliva samples collected, without informed consent, to feed a database of faces and genomes.⁸² Furthermore, one facial recognition software company, Cloud Walk, is developing AI technology that tracks individuals' movements and behaviour to assess their chances of committing a crime.⁸³

The National Institution for Transforming India, also called NITI Aayog,⁸⁴ is helping the Indian government aggregate private and public data on projects ranging from the optimization of agriculture to healthcare. The Indian government has also mandated compliance with the creation of a country-wide biometrics database as part of Aadhaar's identification profile.⁸⁵ What India intends to do if and when it applies AI technology to such a database is uncertain. What is certain is that national and international governance structures are not well-equipped to handle the concerns over privacy, ownership, and ethics that are already beginning to emerge.

The ability of AI to nudge and control private human behaviour that impacts self-determination could increasingly limit the capacity of the UN to monitor and protect against human rights violations. This capacity is further limited when the private sector and powerful

States own the required data, making them better equipped than the UN with the know-how to understand and design algorithms.

“ The ability of AI to nudge and control private human behaviour that impacts self-determination could increasingly limit the capacity of the UN to monitor and protect against human rights violations.

BIOPOWER AND DEMOCRACY

Similar to nuclear and biological arsenals, AI core platforms may need to be managed by the State to ensure distributed economic prosperity, social well-being, and global security. In China, the State increasingly oversees big data and AI tech platforms, working to align public and private interests. More and more, regulation of private oligopolies might become necessary to maintain minimal stability and mitigate the advent of “tech or AI storms” – constant flows of disruptive innovations that erode social cohesion.

By providing governments with the tools to capture, analyse, and optimize citizens' personal data far more precisely than ever before, AI will serve diverse regime types.⁸⁶ Countries might consider the utility of AI-driven technologies that allow monitoring large swaths of populations and ecosystems. New forms of digital authoritarianism will compete with liberal democracies⁸⁷ over how to most effectively exploit and control technological governance models to ensure that AI convergence contributes to social order with a minimum of distributed progress and benefits.

For decades, long-term economic and social prosperity mainly flourished in nations where citizens were allowed to enhance their skills, take risks, become entrepreneurs, and freely engage in wealth-producing activities. By restraining the agency of their citizens, planned economies did not offer the same sustained path towards financial growth. AI could radically overturn that trend. The commodification of massive streams of citizens' data means that, in the future, governments may be able to not only monitor and control the behaviours of individuals, groups, professions and media communities, but also produce economic value to be redistributed. In brief,

digital authoritarian regimes could provide both growth and security under the form of a repressive social order.

Some countries have started embracing the support that AI technologies could provide to governance. Today, China is building the tenets of a “social credit system,” a large matrix of interconnected databases from which machine learning tools extract a score that can be factored into decisions on jobs, loans, transportation services, medical coverage, and other services.⁸⁸ Personal, behavioural, medical, financial and consumption data of about 20 to 30 million people have now been captured and aggregated to determine digital profiles and rankings.⁸⁹

Slowly, a new form of geopolitics centers around surveillance and control. China excels at diplomacy in this new context, magnifying its influence with trade and investment policies, notably as part of the Belt and Road Initiative,⁹⁰ which includes a “digital silk road”⁹¹ to be deployed through the Asia-Pacific region. By influencing global norms and technical standards, China’s diplomatic efforts, including at the UN, aim to defend a form of cybersovereignty, which normalizes pervasive digital surveillance and control of the flow of information on the internet.

Beijing may plan on exporting its model to other nations. Thailand and Vietnam have adopted a similar approach to the Great Firewall – relying on a combination of legislative and technological tools to regulate the internet domestically. China and the US have exported surveillance and censorship technologies to Ethiopia, Iran, Russia, Zambia and Zimbabwe.⁹²

The consequences of such approaches could be corrosive. The entire history of individuals and populations – in intimate granularity, from online behaviours, dating patterns, medical records, drug consumption, sexually-transmitted diseases – could be leveraged for intimidation and discrimination.

Combined with biometrics and increasingly digitized genomes, algorithmic surveillance can bring ever more order and control. Even the perception of surveillance is enough to keep many in line. The merger and subsequent amplification of genetic and algorithmic governance epitomizes a new alliance between cyber and biopower, which will serve both surveillance capitalism and digital authoritarianism.

“Slowly, a new form of geopolitics centers around surveillance and control.”

In the last chapter of *The Will to Knowledge* entitled ‘Right of Death and Power over Life’, Foucault compares “disciplinary power” with a second form of positive coercion he calls “biopower:”

“The second, formed somewhat later, focused on the species body, the body imbued with the mechanics of life and serving as the basis of the biological processes: propagation, births and mortality, the level of health, life expectancy and longevity, with all the conditions that can cause these to vary. Their supervision was effected through an entire series of interventions and regulatory controls: a biopolitics of the population.”⁹³

Foucault’s words will resonate with the deployment of the Internet of Bodies, Genomes and Minds: “[A] power that exerts a positive influence on life, that endeavours to administer, optimize, and multiply it, subjecting it to precise controls and comprehensive regulations.”⁹⁴

PRIVACY, BIOLOGICAL, AND NEURO-DATA AGENCY

Privacy and security are not new concerns for AI and converging technologies. Yet, as AI and the Internet of Bodies coalesce to optimize precision health, hospitals and cities, these systems operate in the background of daily lives. It is therefore challenging to anticipate and mitigate the risks that AI and converging technologies are creating for privacy and collective security.

A set of factors drastically magnifies these concerns, leading to significant ethical breaches, including: the growth of aggregated public and personal data within the Internet of Bodies; the subsequent weakening of the distinction between what is considered “public” and “personal” data; and the lack of human understanding over how automated intelligent systems conduct data aggregation and analysis, with potential for built-in biases.

The lucrative alliance between AI and data-driven economies has transformed digital networks – social, biometric, and genetic – into the new masters of our digital identities. It is not difficult to envision how some of the large private corporations accelerating today’s AI revolution could use their vast networks and machine learning platforms to commodify continuous streams of extremely intimate data about individuals’ personal lives. Embryonic discussions have raised the idea of shifting the intrinsic model of current digital economies to allow citizens to harness their data as labour.⁹⁵ Yet, these discussions will be met with a hard truth: global aggregates of data are just too valuable.

Business models have created strong incentives for industries, States and agencies to join an economic and technological race to dominate and control the digital landscape. More troubling, this race leads to widespread commodification of personal data, which, combined with



OCHA/N Berger

At one IDP site near Goma, the World Food Programme (WFP) and World Vision are distributing food using the biometric fingerprinting system. Biometric technology provides more accurate demographic data and can be used by emergency aid workers to plan their humanitarian interventions, such as food distributions.

social nudging, is undermining trust and contributing to the deployment of systems that automate pervasive biases rather than correct them.

As the collection of intimate, biological and neural data on every individual accelerates, there is a greater need for each individual to be able to control and have access to a complete set of the type of data captured, analysed and forgotten. In the near future, algorithms and neural devices connected to the internet will make it possible for hackers, corporations or government agencies to track or even influence an individual's mental experience.⁹⁶ Soon enough, there will be a need to decide whether citizens should have the ability and right to keep their neural data private.⁹⁷ Neural data could essentially be treated like organs requiring a consent procedure to be analysed.

In a more distant future, neuro-technologies, such as brain-computer interfaces (BCI), guided by AI predictive reasoning, could help decode individuals' mental processes and intervene in impacting the brain functions behind human intentions, emotions and decisions.⁹⁸ Deep learning will progressively contribute to reproduce and embody, within machines, motor intelligence, language, vision and emotion analysis, including abstract affective concepts and situational awareness across life domains. When used in medical or military contexts, such advances will drastically impact notions of individual agency, the

privacy of thoughts and behaviours, and the integrity of cognitive and emotional processes.

A number of current neurotech applications already have the potential to enhance endurance, sensory, motor and cognitive abilities, with slight enhancement in learning and decision-making.⁹⁹ When mature and democratized, these new forms of enhancement will likely be used in civil and military contexts as indicated by national intelligence research agenda in China and the US.¹⁰⁰ In brief, developments in AI and neurotech could alter societal norms and exacerbate inequalities and discriminations. There is an increasingly urgent need to establish guidelines, both at national and international levels, to accompany the progressive deployment of augmentation technologies in civil and military contexts.

As corporations, governments and others start deploying consumer use of neurotechnological applications, individual identity and agency must be protected as basic human rights. Such trends could even require a reimagining of the frameworks currently in place to monitor and implement international human rights law, and will certainly require the multilateral system to anticipate better and understand the quickly emerging field of "AI-bio-neuro convergence."

“ There is an increasingly urgent need to establish guidelines, both at national and international levels, to accompany the progressive deployment of augmentation technologies in civil and military contexts.

BIO-DATA AND CIVILIAN SECURITY

In early 2019, a 20-year old amateur German hacker accessed and released personal information – photos, phone numbers, credit card numbers – on Twitter of hundreds of German politicians, including Chancellor Angela Merkel.¹⁰¹

In 2018, the Indian government biometrics database, Aadhaar, was the target of multiple cyberattacks that potentially compromised the ID profiles of large swaths of the 1.1 billion registered citizens. The Chandigarh-based Tribune newspaper reported that cybercriminals were monetizing access to the Aadhaar database at a rate of 500 rupees for 10 minutes.¹⁰² Elsewhere in the world in 2018, cybertheft of personal data impacted about 150 million users of the MyFitnessPal application, and around 50 million Facebook users.¹⁰³

Digital identities, the amalgamation of any and all information available online about persons, are vulnerable to the same security threats as physical identification cards. Yet, digital identities go beyond physical identification to include behaviours, social profiles, device information, location, search history, and much more.¹⁰⁴ Governments are increasingly converting to more centralized digital identification mechanisms, which raise serious human right concerns.¹⁰⁵

As early as September 2016, hackers stole the personal data of over 200,000 Malaysian organ donors to create fraudulent identities.¹⁰⁶ While global leaders consider digital identities of the future, for instance in the contexts of migration and displacement, comprehensive strategies to protect personal, biometrics, and genomics data in cyberspace are ill defined.

The complex converging technology landscape requires that humanitarian organizations learn how to systematically assess, understand, and mitigate the risks involved in programme activities that generate data. In recent years, humanitarian organizations collected and generated increasingly more intimate data, usually from persons affected by crises, without a firm understanding of the potential risks they may be inadvertently imposing

upon the same vulnerable populations they aim to protect.¹⁰⁷

The DNA of world leaders is already under high-protection. The US intelligence services, for instance, gather the towels, cups, and other objects that the President has used to prevent anyone from obtaining his genetic material.¹⁰⁸ The fear is that, in an Internet of Bodies and Genomes, politicians, celebrities, leaders of industry could become the targets of precision bio-attacks tailored to their immune systems, microbiomes or other health determinants.

Precision biometrics attacks are not far on the horizon. In 2018, IBM detected an AI malware that can hide a cyberthreat, such as WannaCry, in a video conference application, and launch only when it identifies the face of the target.¹⁰⁹

While State-sponsored cyberattacks and cyber espionage are not new, the proliferation of internet-connected devices, the growth of the Internet of Bodies, and current and future AI technology will only exacerbate existing cybersecurity vulnerabilities. Highly complex, these vulnerabilities are seldom well understood and, therefore, often neglected.

What if skilled hackers were able to automate data-poisoning attacks with the intent to falsify, erase or steal intelligence within genomics data-collections, or perhaps neural imaging databases? These would be the holy grail of population-level datasets, encompassing everything from daily activities to mental and physical health statuses.

Equally troubling, rising tech platforms are often the last line of defence to ensure the security of the massive, precious datasets that fuel e-commerce, and soon, smart cities and much more. The same multinationals that reign over data and its liquidity are also charged with cybersecurity – creating potential conflicts of interest on a global scale.

That tension, of course, is coupled with the fact that the private tech sector is also enabling most of the positive benefits that AI can and will usher in for individuals and societies, from helping to predict natural disasters to finding new warning signs for disease outbreaks. Thinking about how to ensure data liquidity and security will become ever more important as governments aim to reap such benefits.

CYBER COLONIZATION

With extensive control of biological data, whoever gains a monopoly of these powerful resources may well be able to influence the well-being of entire populations and impact innovation in allied countries. Consider how genomics data could be mined in countries that lack robust privacy laws, but analysed in an AI-leading State or corporate platform that would reap most economic

“ **Thinking about how to ensure data liquidity and security will become ever more important as governments aim to reap such benefits.** ”

gain. Forms of “cyber colonization” are increasingly likely, as powerful States are able to harness AI and converging technologies to capture and potentially control the data-value of other countries’ populations, ecosystems, and bio-economies. The ambitions of tech-leading nations are also focused on penetrating and commodifying the data of markets in Southeast Asia, Latin America, Africa, and the Middle East.

In economics and medicine, the implications are massive. Capacities in AI and biotech will allow tech-dominant countries to quietly capture the value of another country’s bio-economy – its genomes, microbiomes and ecosystems – for their own economic growth. Inequality between countries that are tech-leaders and those that are tech-takers will rise if this new form of “cyber colonization” happens without transfer of skills and financial benefit-sharing.

China excels at acquiring markets and new data sets. In large-scale, low cost gene-sequencing, China already surpasses the US.¹¹⁰ The world’s largest genetic research centre, the Beijing Genomics Institute (BGI) based in Shenzhen, holds an estimated 40 million people’s DNA samples.¹¹¹ Increasingly, BGI is gaining access the bio-data of emerging markets in Africa and Brazil where new biological resources might be discovered.¹¹² By providing its sequencing services to health and biotech groups in more than 60 countries, BGI is making a winning bet on the future: reading genomes, on a global scale, to crack illness, famine, evolution – and the secrets of human intelligence.

Another company, UK-based Global Gene Corp¹¹³, has increasingly focused on collecting genomics and personal medical history data from subsets of the Indian population to better serve global precision health.

The consulting firm Frost & Sullivan predicts AI technologies will generate USD 6.7 billion in global revenue from health care by 2021, compared with USD 633.8 million in 2014.¹¹⁴ For instance, liquid biopsies – blood-tests to diagnose cancer – are predicted to become the next commercial gold rush in health care. By 2025, between 100 million and 2 billion human genomes could be sequenced¹¹⁵ – an effort that will exceed the computing challenges of running YouTube and Twitter.

Never before has humankind been equipped to acquire and monitor data about human behaviours, physiology, and ecosystems on such a grand scale. Such capacities will raise ambitions not only to monetize, but also weaponize, ever more bio-intelligence.

Indeed, AI and biotech research are inherently dual-use, and therefore a strategic advantage in a nation’s security arsenal. This knowledge will lead to increasing developments at the forefront of medical countermeasures, including vaccines, antibiotics, and targeted treatments relying on virus-engineering and microbiome research.

Algorithms are becoming a crucial tool in biosecurity to detect potential threats from known and unknown DNA sequences. Applying AI deep learning to genomics datasets could also help geneticists learn how to use genome-editing to “optimize” human health, with potential applications in military enhancements.

With expanding global networks of data, and the integration of civilian and military industries, some superpower States are already pledging to achieve dominance in bio-intelligence. The question is whether other countries will open their eyes to this and try to gain a competitive edge.

“ **Capacities in AI and biotech will allow tech-dominant countries to quietly capture the value of another country’s bio-economy – its genomes, microbiomes and ecosystems – for their own economic growth.** ”

OPTIMIZING HUMANS AND ECOSYSTEMS

Using AI systems to understand how the functioning of genetic systems impact health is of strategic importance for global medical innovation and biodefence. This knowledge will lead to increasing developments at the forefront of medical countermeasures, including vaccines, antibiotics, and targeted treatments relying on virus-engineering and microbiome research.

Applying deep learning to genomics data-sets could help geneticists learn how to use genome-editing (CRISPR) to efficiently engineer living systems, but also



© David Sun Kong, MIT Media Lab

Bio-resources are increasingly being commoditized and optimized for human endeavours.

to treat and even optimize, human health with potential applications in military enhancements. A USD 15 million partnership between a US company, Gingko Bioworks, and the government agency DARPA (Defense Advanced Research Projects Agency) aims to genetically design new probiotics as protection for soldiers against a variety of stomach bugs and illnesses.¹¹⁶

Deep learning may also lead to the identification of “precision maladies,” which are the genetic functions that code for vulnerabilities and interconnections between the immune system and microbiome.¹¹⁷ Using this form of bio-intelligence, malicious actors could engineer pathogens that are tailored to target mechanisms critical in the immune system or the microbiome of specific subpopulations.

In a near future, technological convergence will lead to human augmentation, in particular at the intersection of AI, genome-editing, and neuro-technologies.

Genome-editing already allows for drastic increases in the muscular mass of mammals. In a 2017 technological exploit, doctors in the US restored 80 per cent of the skin

of a seven year old patient afflicted by a rare genetic disease that causes skin to blister and tear off at the slightest touch.¹¹⁸ Research teams in China and the US private sector have already engaged in a competition for performing germline editing while mitigating some of its known damaging side-effects, such as cell mosaicism. In late November 2018, a Chinese scientist took the world by surprise, igniting a storm of condemnation, as he claimed to have used genome-editing to alter embryos for seven couples during fertility treatments, with one pregnancy resulting so far.¹¹⁹ If proved accurate, this first attempt at germline-editing *in vivo* happened without a clear understanding of the potential implications for genomics integrity and human dignity. It represents an inflection point which should be considered a warning by the international ethics and governance community.

Most of these genetic breakthroughs, coupled with increasingly precise predictive deep learning systems, could help optimize human and non-human biology for defensive and offensive purposes. For instance, the use of gene-drives techniques – specific genetic modifications that self-propagate through sexual reproduction – could transform the genetic makeup of mosquitoes or mice, with

the goal of avoiding viruses' transmission to humans.¹²⁰ The same technique, if immature and not well tested, could make a species invasive and disturb ecosystems in largely unforeseen ways.

The alliance between AI, genome-editing and neuro-technologies is likely to generate new distributive inequalities. This is particularly clear when it comes to who is expected to carry the burden of potential risks. Ground-breaking experiments using gene-drives are currently being deployed in Mali, Burkina Faso, and Uganda. Target Malaria, funded by the Bill and Melinda Gates Foundation, is meant to use a gene-drive to drastically reduce the number of mosquitoes that transmit the disease in sub-Saharan Africa, where it kills hundreds of thousands of people a year.¹²¹ While gene-drives technologies sound promising to potentially eradicate endemic vector-borne diseases, even the most prominent engineers in the field have warned of the complex, unintended and long-term consequences of releasing self-propagating genetic payloads into an ecosystem.¹²² In November 2018, the UN Convention on Biological Diversity (CBD) agreed to limit gene-drives but rejected a moratorium.¹²³

In turn, there are also concerns around discrimination to access the empowering effects of converging technologies, such as the proven safe use of AI and

“ The alliance between AI, genome-editing and neuro-technologies is likely to generate new distributive inequalities.

gene-therapy for treating cancers and other afflictions. This persistent problem of inequality in access could be drastically amplified by the rising gap between groups of countries. For instance, the global gene-therapy market for rare disease is expected to surpass USD 3.55 billion by 2026.¹²⁴ Yet, there could be a significant delay in the commercialization of approved therapies in regions such as Latin America, the Asia Pacific, the Middle East, and Africa, as these therapies use cutting-edge genome-editing technologies and prices are high for gene-therapy treatment. While some countries will be lagging behind, a few others will have acquired the capital, data, global talents and intellectual property to develop and use the most optimizing and expensive treatments ever encountered.

INSIDE SMART BUT VULNERABLE CITIES, FACTORIES, AND LABORATORIES

“And then he grasped that this city of machines...that this city...was boundlessly blissful and boundlessly destructive”

– Thea von Harbou, *Metropolis*

This scenario focuses on the convergence of AI and automation, cyber and biotechnologies, robotics and additive manufacturing, with unprecedented implications for human, civilian, socio-economic and physical security.

CYBERATTACKS ON SMART CRITICAL SYSTEMS

With deep learning, hunting for weaknesses in networked infrastructures will be mostly automated.¹²⁵ Automated cyberoperations will be more effective, finely targeted, difficult to attribute, and likely to exploit evolving vulnerabilities in AI systems. The capacity of adversarial algorithms to improve their own strategies and launch increasingly aggressive counter-attacks with each iteration will lead to an expansion and augmentation of

existing cyberattacks, with global damages that may reach USD 6 trillion a year by 2021.¹²⁶

In the 2018 World Economic Forum's Global Risk Perception Survey, the second most frequently cited risk triggered by technological convergence is the combination of cyberattacks with the manipulation and corruption of critical information infrastructure.¹²⁷

For instance, the company DarkTrace¹²⁸ in Cambridge, UK, used the human immune system as a model to design its AI technology to detect and fight back against emerging cyberthreats across corporate networks, cloud and virtualized environments, the IoT, and industrial control systems. By 2023, the value of AI in cybersecurity is projected to increase USD 17 billion.¹²⁹

Similar to a Darwinian experiment, cybersecurity communities will have to rely on the development and survival of the fittest AI.¹³⁰ If AI and cyber capacities keep converging and enabling each other, what is the nature and scope of the attacks that States, the private sector, and civilian populations will have to withstand? Harnessing AI technologies will drastically expand the velocity and reach of cyberattacks¹³¹ as well as their complexity.

Recently, there has been an increase in reports of cyberattacks on critical infrastructures, especially on electric power systems, globally. “This has served to demonstrate just how vulnerable cities, States, and countries have become and the growing importance of achieving global risk agility.”¹³² One example of such an attack occurred in December 2015, when hackers seized control of the Prykarpattyaoblenergo Control Center (PCC) in Western Ukraine, rendering up to 225,000 people without power for up to six hours.¹³³ According to a 2016 Electricity Information Sharing and Analysis Center report, the attackers demonstrated a variety of capabilities, including spear phishing emails, variants of the BlackEnergy 2 malware, and the manipulation of Microsoft Office documents that contained the malware to gain access to the IT networks of the electric companies.¹³⁴

In 2017 alone, States in Europe witnessed cyberattacks targeting national health, telecommunications, energy, and government sectors.¹³⁵ Yet, other nations could perhaps be even more vulnerable to cyberattacks on critical infrastructures if they lack cutting-edge expertise and capacities in cybersecurity. As noted by Landry Signé and Kevin Signé, two experts on cybersecurity in the Global South, “Although Africa is relatively limited in terms of communications infrastructure, due to the high penetration rate of new technologies, it is increasingly a target for cybercriminals, as most African countries still have a low level of commitment to cybersecurity.”¹³⁶

“**Harnessing AI technologies will drastically expand the velocity and reach of cyberattacks as well as their complexity.**”



Unsplash/Clay Banks

In the future, drones will be able to navigate dense environments from forests to urban citiscapes

DRONE SWARMS, 3D-PRINTERS AND URBAN VIOLENCE

Increasingly, the convergence of AI and cloud networked robotics will allow cyberattacks to subvert the functioning of physical systems, from self-driving cars, drone swarms, smart energy systems, surgical robots, to 3D printers and industrial supply chains. New forms of hybrid cyberthreats will likely exploit the boundary with critical physical infrastructures. Often, this will lead to new forms of cognitive-emotional conflicts that aim to create chaos, confusion and uncertainty, not on the battlefield, but inside cities, hospitals and critical institutions.

We recently witnessed the sudden paralysis and the economic burden that the misuse of commercial drones can impose on airports.¹³⁷ Drones are one manifestation of a future “Cambrian explosion”¹³⁸ of robots and smart objects in homes and cities. A State or non-State actor could wage a cyberattack that can subvert networks of robots to steal citizens’ intimate data, manipulate and corrupt strategic information, or disrupt operations in manufacturing plants and hospitals. Similar cyberattacks could automate swarms of micro-drones to spread chemicals and neurotoxins in water and food supply chains. In smart cities, most networked architectures will run on automated work flow and AI-powered data optimization.

The convergence of AI, robotics and additive manufacturing (3D printing) is another evolving domain to watch. AI software will enable automated workflows within decentralized supply chains.¹³⁹ This will enable the production, in the physical world, of digital blueprints of objects, weapons, illicit substances, cells, and fragments of genes. Such decentralized power of invention and creation will serve open innovation labs in Shenzhen,¹⁴⁰ San Francisco,¹⁴¹ and Kumasi,¹⁴² but also nefarious networks that grow in conflict zones and crusade the dark web. For instance, IS took its adversaries by surprise by harnessing off-the-shelf drones the group had modified with 3D-printed parts to drop explosives on unsuspecting forces.¹⁴³ Such a scenario is not, however, limited to the battlefield; it will increasingly concern cities outside conflict zones, too.

Beyond impact on digital security, AI will also endanger the security of physical and biotechnology infrastructures. As additive manufacturing supply chains and biotech laboratories become increasingly automated and run by AI software – such labs already exist¹⁴⁴ – they will be vulnerable to adversarial attacks, in particular data-poisoning, by external AI systems.

A few past experiments have shown how adversarial perturbation – a certain amount of noise within the pixels of a driving sign – could disorient self-driving cars.¹⁴⁵ Instead of a Tesla, the next target of adversarial attacks could be drone swarms or a biotech supply chain. Hackers could inject enough adversarial perturbation

“ **Cyberattacks could automate swarms of micro-drones to spread chemicals and neurotoxins in water and food supply chains.**

into the navigation command of distributed networks of drones. Similar strategies could be used to introduce, into a biotech supply chain, enough structural noise to corrupt data, bio-engineering instructions or slightly change the parameters (temperature) of an experiment.¹⁴⁶ Early warnings might not be visible, but there might be lasting damaging effects on producing antibiotics, vaccines or cancer treatments.

The nature and scale of potential attacks and disruptions are unprecedented, with AI systems that will likely exceed human performance in cybersecurity. Increasingly, deep learning systems will make decisions about what to attack, who to attack, and when to attack, with their targets ranging from consumers, self-driving cars, surgical robots to automated bio-labs.

“ **The nature and scale of potential attacks and disruptions are unprecedented, with AI systems that will likely exceed human performance in cybersecurity.**

INTELLIGENT BIO-MANUFACTURING

The convergence of AI, automation, cloud-computing, and genomics technologies is contributing to the growing intelligence and connectivity of laboratories. Most industrial biological labs today are increasingly automated, equipped with AI analytics software and cloud services.¹⁴⁷ Companies such as Transcriptic¹⁴⁸ and Emerald Therapeutics¹⁴⁹ allow the remote control of up to 50 types of bio-experiments from a computer anywhere in the world. As a result, the biological lab of the future will require less tacit knowledge for use and analysis. In effect, techniques such as genome-editing may become more easily available to those with less or no formal biology training. In this context, it is crucial to assess how converging AI and biotechnologies will drastically impact the tools of biological laboratories, as well as their dual-use potential, to lead both to societal benefits and new vulnerabilities.

While the automation of biotech research may be a boon for the development of novel vaccines and therapeutics by parties that have not traditionally had access to the necessary tools, it also opens the risk of nefarious use to engineer or edit biological agents or neuro-toxins.¹⁵⁰

Using automated labs, it is now possible to assemble benign DNA fragments into gene sequences that can be subsequently modified and synthesized. This form of intelligent design allows an actor, with enough expertise, to synthesize genetic sequences that can become the basis to produce a toxin or a bio-agent. Such a process could enable malicious actors to produce toxins or pathogens, without having to acquire, from a company or a lab, genetic sequences of concern or undergo a series of checks and investigations. Indeed, existing oversight is limited to select agents at the species level, gene sequence orders, and does not include the misuse of genomics data.¹⁵¹

Areas of near-term concern include the automated design of bacteria with multidrug resistance¹⁵² and the modification of commensal bacteria to become super-producers of toxins. Both could be easily spread using drone technologies. Drones, which are easy to buy and control, could significantly break down limits to delivery of bioagents, such as neurotoxins.¹⁵³ Already, drones have been used to fight fires and spray crops with pesticides, both of which require the dissemination of aerosols or liquids, two methods which would be required for the dissemination of biological products.

Within intelligent and connected biolabs, each point in the automated process has cyber and biosecurity vulnerabilities that could be hacked. Security experts are therefore becoming ever more concerned about the risk of adversarial algorithms being used to automate cyberattacks on biotech supply chains.¹⁵⁴ Hackers could corrupt networks of sensors to impact control decisions

on a biotech lab, and damage, destroy or contaminate vital stocks of vaccines, antibiotics, cell or immune-therapies for cancer treatment. Such an event could escalate into biosafety and biosecurity risks.

Deceptive algorithms could also be used to automate data-poisoning attacks¹⁵⁵ with the intent to falsify, erase or steal intelligence within large curation of genomics data. Such data-poisoning would not only affect how to detect and analyse pathogens. It could also corrupt the bio-intelligence collected for decades on complex and personalized diseases, affecting researchers' capacity to target and treat cancers and other afflictions in groups of patients. Building capacity in cyber-biosecurity has become a national and societal asset.

Within intelligent cities and civil infrastructures, new forms of covert data-poisoning, cyber espionage and automated attacks have drastic economic costs as well as significant impact on employment and, in general, societal well-being. Yet, the most damaging impact would be on citizens' trust – trust in governing institutions, emergency data-systems, industrial laboratories, food supply chains, hospitals and critical infrastructures.

“**Building capacity in cyber-biosecurity has become a national and societal asset.**”

AI FOR PREVENTION – A POSITIVE SCENARIO ON AI CONVERGENCE

“I don't try to predict the future, I try to prevent it.”

– Ray Bradbury, author of *Fahrenheit 451*

This scenario explores the idea that AI and other converging technologies could be harnessed to manage complex risks to vulnerable populations, avert outbreaks of crisis, and build resilience against shocks within societies. In other words, AI for prevention.¹⁵⁶

OPTIMIZING AI AND CONVERGING TECHNOLOGIES FOR EMPOWERMENT

Many brilliant technologists and philosophers have challenged the mundane assumption that technology is an apolitical and amoral force.¹⁵⁷ Modern technology

is not about algorithms or self-replicating gene-drives as abstract artefacts used for good or bad. Algorithms and engineered organisms are endowed with a specific structuring function, a certain power and nature, as humans have designed them. This structuring function, the technological design process itself, and the real-world implications these design choices have are increasingly coming under scrutiny.

The automation of predictive intelligence in combination with other emerging technologies can lead to both positive and negative implications for societies. While the previous sections have focused on the potential global

risks generated by AI and converging technologies, this scenario explores predictive intelligence as a powerful tool for prevention across several domains of the Sustainable Development Goals.¹⁵⁸

PREVENTING FAMINE AND EPIDEMICS (SDG 2)

In combination with precision biology, robotics, and decentralized data-capture technologies, AI's capacity for data-optimization and predictive intelligence will be a game-changer to prevent famine, health crises, and epidemics.

Take acute food insecurity. As aforementioned, the World Bank, in collaboration with the UN, Microsoft, and other global partners, is building a Famine Action Mechanism.¹⁵⁹ This early warning system relies on deep learning algorithms to analyse collections of data from satellite imagery to detect when food crises are about to turn into famines. The same tool directly connects agile financing with areas of food insecurity.

Yet, the next step would be to use the Famine Action Mechanism as an intelligent tool to target areas where precision agriculture could prevent food insecurity. Using AI, genomics, and ecosystems data, researchers would determine which crops' genes could be modified to enhance their resilience to an array of climatic events such as drought.¹⁶⁰ Governments, on their end, would be able to understand, in real-time, where the use of genome-edited crops would be the most impactful.¹⁶¹

The same mechanism translates into monitoring epidemic outbreaks. Researchers now use data, including behavioural, demographic, and epidemic disease trends, to create an algorithmic model of disease spread that captures underlying population dynamics and contact patterns between people.¹⁶² In the near future, deep learning will likely lead to a form of "precision prevention," optimizing the allocation of scarce resources and restraining the spread of infectious diseases.

DEMOCRATIZING HEALTHCARE IN THE RURAL SOUTH (SDG 3)

When it comes to health, deep learning will bring ever more exceptional promises. Neural nets are already used in India to diagnose retinopathy and offset the shortage of doctors in rural neighbourhoods.¹⁶³ Equipped with AI analytics software, sensors, and cameras, mobile phones can become a diagnostic tool, increasingly used for digital microscopy, cytometry, immunoassay tests, and vital signs' monitoring. For instance, in a mobile phone equipped with a camera, image recognition systems can detect malaria in a blood smear, or screen for cervix cancer remotely.

A report published in the British Medical Journal in August 2018 demonstrated how AI is already changing healthcare practices, from rural Brazil to sub-Saharan Africa.¹⁶⁴ The expertise of doctors from cutting-edge hospitals in the US is harnessed through IBM Watson by hospitals in Mongolia that crucially lack oncology departments. Remote diagnosis for an increasing array of afflictions is made possible using AI's power to optimize decentralized health data collected through cellphones and other IT devices.¹⁶⁵

A simple but life-changing breakthrough is the use of AI to optimize the delivery of services by community health workers in underserved rural areas. For instance, in Bangladesh, community health workers can use AI to know where and when to be present for birth deliveries, which contributed to a boost in neo-natal survival rates by over 30 per cent.¹⁶⁶ In a similar vein, community health workers in Tanzania now deliver targeted health campaigns to pregnant women and young caregivers using an AI-powered mobile application.¹⁶⁷ Meanwhile, researchers in Brazil rely on machine learning algorithms to forecast the need to resuscitate newborns suffering from birth asphyxia, a condition still endemic in developing countries.¹⁶⁸

Using AI to democratize health is powerful, yet it raises the same questions of privacy, data agency, and security that are raised in the *Internet of Bodies* scenario. Questions also abound about the safety of algorithmic design and the hidden biases that could corrupt health decision-making if training datasets do not carefully mirror the health determinants of local populations.¹⁶⁹ Dynamics of biological data's predation could also be on the rise if AI health services are mainly provided by private companies, which do not allow for transfer of technology, skills, and data to local doctors and hospitals.¹⁷⁰ Helping raise these concerns should be part of the UN's prevention agenda.

“ In the near future, deep learning will likely lead to a form of “precision prevention,” optimizing the allocation of scarce resources and restraining the spread of infectious diseases.

MONITORING MIGRATION FLOWS WITH DRONE SWARMS (SDG 16)

European nations are resorting to swarms of drones equipped with facial recognition, biometrics sensors, and “chemical sniffing” technology to map groups of migrants.¹⁷¹ For instance, the UK has planned to deploy drones, equipped with thermal imaging capabilities, in France to monitor the Channel Tunnel.¹⁷² These drones are likely connected to an earlier €2 billion (USD 2.2 billion) deal for the UK and France to jointly create a “next-generation” drone platform by 2030.¹⁷³

Within a sound policy framework, drone swarms could facilitate rescue operations,¹⁷⁴ prevent trafficking,¹⁷⁵ prepare for integration strategies, and provide aids and goods in “hot spots” on migration roads.¹⁷⁶ Yet, the same predictive tool could also be misused to close borders, and target, detain, or attack those populations already at risk.¹⁷⁷

THE OTHER SIDE OF AI IN WARZONES (SDG 16)

The Syrian conflict is approaching its ninth year. With the number of civilian deaths skyrocketing – approximately 500,000 (and likely more) Syrian civilians have lost their lives since 2011¹⁷⁸ – two Americans, a technologist and an entrepreneur, saw an opportunity to leverage AI and data-capturing technologies to create an early warning system for incoming airstrikes. Five long years after the conflict erupted, they launched Hala Systems,¹⁷⁹ which aims to do what even the international community had failed to accomplish: warn civilians of incoming airstrikes, giving them hope and a chance to run.

In order to effectively warn citizens of impending airstrikes, the team behind Hala Systems needed to create a human network comprised of trusted contacts, recruited teachers, engineers, and even farmers as potential plane spotters. The team supplemented information from the human network with acoustic data, collected from remote sensors hidden in treetops and tall buildings, that helped determine speeds and aircraft models.¹⁸⁰

Within a matter of seconds, Hala’s algorithms compare the data with that of previous air strikes to make predictions. These predictions are then immediately broadcasted across social media sites, triggering a networked alert system of alarms, sirens, and flashing lights.

The result? Civilians were able to disrupt the nature of warfare, even if only in a very small way. Hala Systems’ interoperable platforms – Sentry,¹⁸¹ the early warning system, and Insight,¹⁸² the real-time data analysis portal – work together to warn civilians and predict where warplanes take off, where they will likely hit, where the danger areas are, and whom the planes belong to. Sentry

also allowed first responders to converge at the targeted location to search for survivors buried in the rubble within moments of impact.¹⁸³ Information is power and information can save lives.

Rather than living in darkness as to when their lives, families, and homes could be destroyed in an instant, Hala gave Syrians a glimmer of hope in one of the most volatile conflicts in recent history. Hala has even provided *prima facie* evidence of war crimes related to 75 specific events and disseminated 250 reports on aircraft activity and ceasefire violations to governments, NGOs, and the UN since 2017.¹⁸⁴

According to a preliminary analysis, Hala’s technology solutions resulted in an estimated 20-30 per cent reduction in casualty rates in several areas under heavy bombardment in 2018.¹⁸⁵ In the darkness of a long conflict, Hala Systems are a successful story of human-algorithm collaboration.

INVESTIGATING FORGERIES, ELECTION FRAUD, AND VIOLENT CRIME (SDG 16)

Increasingly, we will witness engineering and policy communities racing to produce tech fixes and countermeasures for the era of post-truth. For instance, the field of media forensics can harness the power of AI technologies to detect visual anomalies with extreme precision.¹⁸⁶ Engineers are working on algorithms that could automatically assess whether an image or video has been forged or tampered with.¹⁸⁷ If successful, media forensics could automatically detect data manipulations, provide detailed information about how these manipulations were performed, and reason for the overall integrity of visual media to facilitate decisions regarding the use of any questionable image or video.

Beyond detecting forgeries and instances of fraud, AI and converging technologies are also being employed in the war against terrorism to scrub the internet of terrorist propaganda. The UK Home Office and ASI Data Science collaborated to develop a technological tool that can automatically detect 94 per cent of IS propaganda with 99.995 per cent accuracy.¹⁸⁸

Predictive algorithms could expose election fraud, crimes, and even terrorists, seeding tools for mediation. Because they can be trained to detect anomalies, AI systems, when combined with facial recognition and precise image recognition, will increasingly play a role in virtual investigations in the context of elections monitoring,¹⁸⁹ predictive policing,¹⁹⁰ and conflict mediation.¹⁹¹

Yet, such algorithms will have to be resilient to adversarial attacks, which could corrupt the anomaly-detection process and blur investigations. Imagine if terrorists added enough adversarial noise to their propaganda

videos. Would the UK's tool be able to look beyond the noise and see the content for what it really is? Technical limitations notwithstanding, the development of countermeasure technologies must be pursued at the same pace as their "evil twins," addressing both ends of the dual-use technology spectrum simultaneously.

FIGHTING THE VIRTUAL WAR ON HATE SPEECH (SDG 16)

In 2016, UN Global Pulse started mining radio content with machine learning to respond to the refugee crisis in Uganda.¹⁹² The experience showed that analysing speech content with natural language processing could bridge an information gap by providing insights on early-warning systems and progress towards the SDGs. Similar reasoning could be applied to monitoring the proliferation of hate speech on social media. This is why UN Global Pulse is also developing machine learning to help flag and reason about signs of hateful speech online and monitoring the effects of extremist violence on hateful speech online.¹⁹³

The same way that swarms of chat bots could nudge and manipulate individuals' inner emotions, desires, and fears, these bots could attempt to police social media in an effort to prevent cyberviolence and bullying. With future progress in affective computing, deep learning could even become a powerful tool to identify early signs of hate speech and violent online behaviour, and to predict their potential to escalate into hate crimes. For now, technical limits and failures abound as both deep learning and emotional analysis are still struggling to understand the contextual, linguistic, symbolic, and behavioural nuances of online human discourse.¹⁹⁴

If we learned anything from the tragedy in Myanmar, it is that media giants such as Facebook need far more than tech fixes to prevent hate speech on their platforms.¹⁹⁵ Human insights and solutions are of strategic importance in order to recognize and understand the impact of hate speech. Hybrid teams of AI systems flagging violence in deluges of data and working with humans, who would conduct the metadata interpretation, could be an intermediary success. Such collaborative intelligence between humans and algorithms would also ensure that any efforts to hunt down hate speech do not restrict free expression and are aligned with international human rights law.¹⁹⁶

When Facebook and Twitter wars amplify discourses that lead to populations suffering atrocities, it is time to pause and ask questions. How have civil society and corporations engaged on the topic of hate speech and disinformation? Do the UN Guiding Principles on Business and Human Rights sufficiently outline the responsibilities of corporate actors in cases of "weaponization" of social media? Are new principles and protocols required to

guide private actors in specifically assessing the human rights impact of AI and cybertechnologies?

The 2018 report of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression outlined the applicable human rights legal framework and proposes standards and processes that companies should adopt to regulate content in accordance with human rights law.¹⁹⁷ Specifically, this report stated that the Guiding Principles on Business and Human Rights applies to both States and companies, enabling "forceful normative responses against undue State restrictions – provided companies play by similar rules."¹⁹⁸

ANTICIPATING AND MEDIATING CONFLICTS (SDG 16)

Peacemaking¹⁹⁹ has traditionally been considered a 'low tech' field in which interpersonal skills, savoir-faire, and emotional intelligence represent the most important attributes of a good mediator. While this remains true, the field is also evolving and searching for tools and skills that may assist the mediators.

Capable of powerful combinatorial computation, AI excels at simulating myriads of quantitative scenarios in seconds. In combination with humans' insights and qualitative analyses, AI could help with dynamic simulation of geopolitical conflicts and political forecasting, as well as alliance- and interest-mapping.²⁰⁰ AI would classify and optimize large amounts of diverse types of data (related to geography, climate, politics, and socio-economic trends) into a form of anticipatory intelligence, that could be harnessed for negotiation, conflict prevention, and peace mediation.

Converging technologies could support mediators in conflict analysis, engagement with conflict parties, inclusivity, and public communications. Already, mediators and their teams use various social media platforms to support analysis, strengthen channels for engaging with conflict parties, foster greater inclusivity and strengthen public communication strategies.²⁰¹

Another example is the Centre for Humanitarian Dialogue's (HD's) use of Liveuamap and geographic information systems (GIS) to track developments on moving frontlines and monitor the implementation of ceasefires during the Syrian conflict, providing mediators with credible information in real-time.²⁰² Liveuamap's software, modelled after Google Maps, uses "AI web crawlers" that filter through relevant stories and reports.²⁰³ The software has, for example, been used in the Ukraine to identify the location and frequency of events related to public protests.

“ With future progress in affective computing, deep learning could even become a powerful tool to identify early signs of hate speech and violent online behaviour, and to predict their potential to escalate into hate crimes.

AI's ability to optimize data will also assist in peacekeeping operations and contribute to the proper use of financial aid. Domestically, it could highlight potential scenarios, filtering news and arguments, and fostering citizen's political literacy. In brief, it could support the management of complex risks and avoiding crises that lead to international interventions.

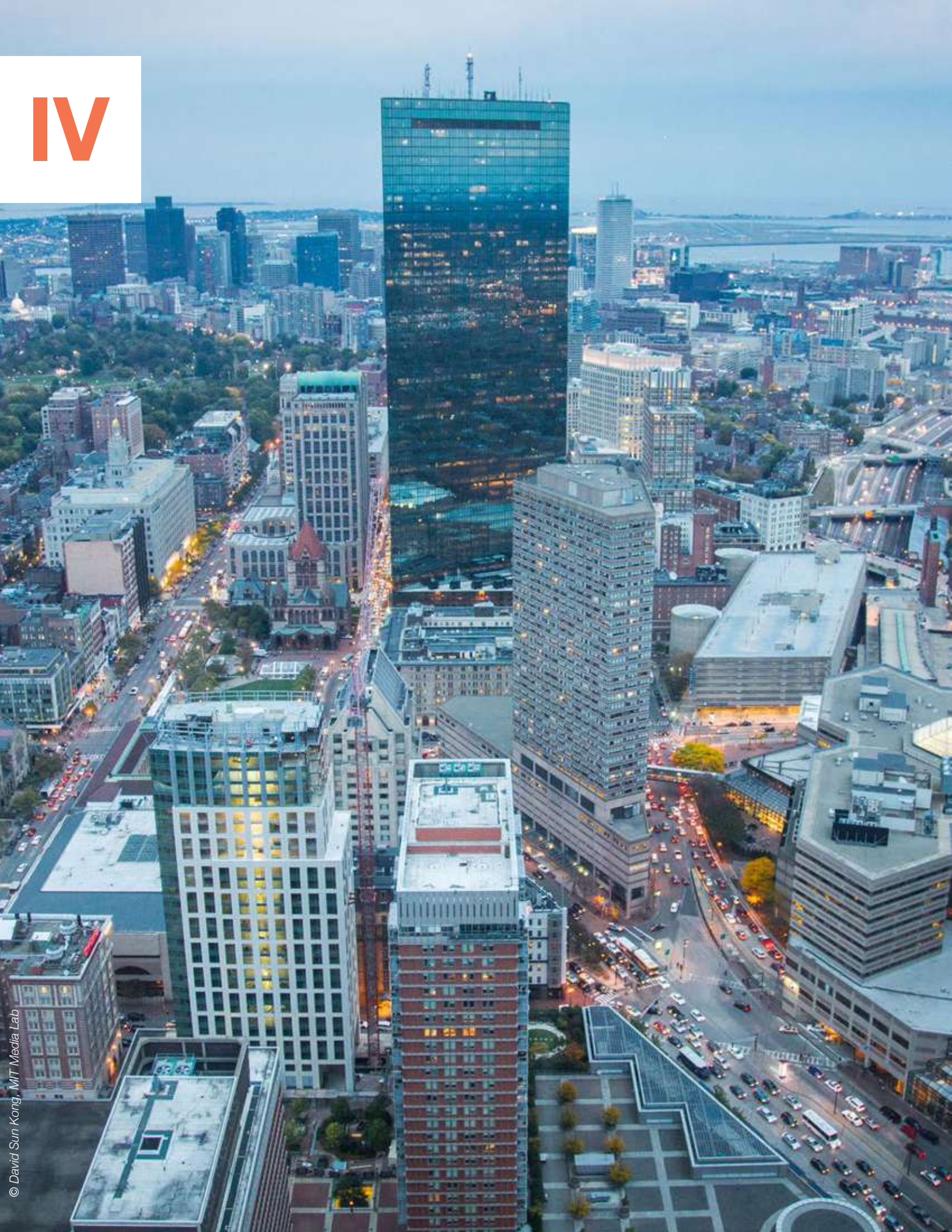
Yet, while the power of data, tools and converging technologies can be harnessed for mediating conflict, mediators will continue to face the same risks as the general public. Swarms of chatbots, trolls, and the proliferation of mis- and disinformation create noise in cyberspace that mediators must be capable of ignoring. In the case of Liveuamap,²⁰⁴ just as mediators can leverage this tool to support their work, it could also be used, by State and non-State actors alike, to identify vulnerable regions, populations, and groups based on the nature, scope, and scale of conflicts mapped.

In combination with human insights, emotional intelligence, and control, AI's capacity for predictive intelligence will provide new technical tools for the prevention agenda. Yet, these tools are not value-neutral, and must be used in accordance with a sound guiding philosophy and a strong ethos of responsible innovation.

The dual-use nature of AI and converging technologies is not equally distributed throughout societies, with some applications promoting human rights for some while adversely affecting the human rights of others. For example, humanitarian organizations use social media platforms such as Facebook, Twitter, and WhatsApp to inform their decision-making processes and operational responses. Yet, in doing so they contribute to the collection of data that can reveal any person's gender, sexuality, religion, location, and anticipated behaviour. Thus, there must be a form of predictive accountability when it comes to the application of AI and converging technologies in conflict prevention efforts.



IV



KEY SCENARIO TAKEAWAYS FOR GLOBAL SECURITY

This section provides a geopolitical lens and an analysis of how the implications and potential risks of AI convergence will amplify rising inequalities between countries that are leading today's technological revolution and those that are lagging behind.

PROTECTING HUMAN WELL-BEING IN GLOBAL SECURITY

As the World Economic Forum's 2019 Global Risk Report emphasizes, threats of all sorts, including technological risks, have a powerful, long-term, and corrosive impact on human security and well-being.²⁰⁵ With 700 million people worldwide suffering from mental health problems, we already face an epidemic of stress and depression.²⁰⁶

Complex technological and social transformations give rise to growing uncertainty about future job security. As a result, larger subsets of the world population see their psychological and emotional well-being receding. Vulnerable groups may experience, with more acute despair, new forms of disempowerment. In turn, disempowerment and the lack of opportunity to play a meaningful role in the innovation journey impacts trust and social cohesion. In decades to come, citizens will need strategies, from lifelong learning to neuro-technological support, to cope with the need for constant, sometimes, radical retraining to adapt to automation.

AI and converging technologies may accelerate, beyond human capacity to adapt, the pace of transformation in major areas of our economies, whether that be in the factory, lab, hospital, law firm, or farm. There is no doubt that AI's capacity for data optimization and predictive

reasoning will serve the data-driven economies of tech-leading countries. Humans and algorithms will engage in collaboration, but also competition for cognitive and creative performance.

In this race, existing and future inequalities will be significant in determining who flourishes and who fails in the converging tech future. In Silicon Valley and the Shenzhen Eldorado, there will be opportunities for empowerment beyond the most advanced AI labs. There, the next generation might benefit from enough exposure to tech know-how in universities and open innovation labs to turn their own data, ideas, and designs into successful applications. In advanced countries like Finland, new educational solutions, such as open-source AI trainings tailored for the non-specialists, aim to more comprehensively include citizens in flourishing innovation ecosystems.²⁰⁷ But what will happen to countries that are not leading the pack in AI and converging technologies?

AUTOMATING TECHNOLOGICAL DIVIDES AND EXCLUSION

AI and converging technologies are already affecting socio-economic trends in many parts of the world, shifting labour markets and potentially creating new stress points for large populations.

In an increasingly globalized world, market interdependence and automation may disrupt the global trade network with disastrous consequences for most vulnerable links. The power of AI to automate supply chains will generate economic growth in regions leading in technological innovation like the US and China, while potentially disrupting the economies of developing and vulnerable countries that rely on low-cost manufacturing, such as El Salvador and Indonesia.²⁰⁸ A 2017 McKinsey Global Institute report found that "30 per cent of 'work activities' could be automated by 2030 and up to 375 million workers worldwide could be affected by emerging technologies."²⁰⁹

With the deployment of additive manufacturing, companies, and perhaps even individuals, will be able to design and develop products at the point-of-need using cloud services, 3D printers and other automated machines.²¹⁰ As the digital power of these new cloud labs and 3D printers could redirect production to innovation hubs in wealthy regions, an increasing number of countries might become vulnerable links in a new economy of intelligent labs and factories. If these countries have

“ AI and converging technologies may accelerate, beyond human capacity to adapt, the pace of transformation in major areas of our economies, whether that be in the factory, lab, hospital, law firm, or farm.

significant swaths of valuable biological resources, they might become vulnerable to data-exploitation or cyber colonization.

The manufacturing industry, finance and retail sectors are already bearing the brunt of the consequences of automation. Narrow automated computing systems may increasingly excel at jobs that require cognitive but repetitive labour in the medical, financial and legal sectors.²¹¹ On Wall Street, for instance, most cognitive functions within the act of trading rely on processing and optimizing data. Except for sporadic flash crashes²¹² – very rapid, deep and volatile loss in security prices – in general, AI's predictive intelligence may become a more efficient trader. When cognitive labour also requires emotional intelligence like in medical diagnosis, centaur teams of humans and AI systems are likely to thrive.

We may find a useful historical analogy for the AI technological storm in the tools and technologies that gave rise to the Industrial Revolution. Disruptive technologies, including steam power, the cotton gin, and more efficient iron production, radically transformed society, upending traditional hierarchies, reshaping economies, and even modifying relationships with the natural world. The benefits reaped from the new manufacturing era were preceded by a period that immiserated much of the workforce and included among its harms, child labour and unsafe working conditions.

In the wake of the manufacturing era, came the rise of industrial capitalism, based on individual rights, private ownership, and free markets. Capitalism and the technologies that enabled it brought with them growing disparities in wealth and opportunity, both among and within countries.

A NEW GEOPOLITICS OF HYBRID RISKS AND VULNERABILITIES

As tech-taker countries line up behind those that are tech-leaders, the resulting rival techno-blocs herald the dawn of a new world order. These new geopolitical dynamics will be shaped by open rivalry over who can control and mitigate the power, failures and externalities of AI and converging technologies.

AI-CYBER ADVERSARIAL ATTACKS: DEVELOPING COUNTRIES AS VULNERABLE LINKS

In 2016, a group of cybercriminals hacked Bangladesh's Central Bank, stealing USD 81 million.²¹³

Cybersecurity vulnerabilities impact low- and high-income countries without discrimination. While higher-income countries such as the UK may have more security resources at their disposal, other countries like Malaysia are not as well equipped to handle large-scale cybersecurity breaches, such as the personal data leak

of hundreds of thousands of organ donors, as noted above.²¹⁴ Keeping pace with rapidly changing security threats will become increasingly more difficult, regardless of the country. Yet it will be the most vulnerable countries, the vulnerable links, that will be impacted the most.

As AI and cyber capabilities expand in developing countries, so too will the attack surface.

A 90 second video of smoke and chaos with armed groups marching over cities and slums went viral on Kenya's social media networks, providing viewers with a dystopian look of what the country should expect if Raila Odinga, Kenya's leading opposition candidate, were to win the elections and establish a violent dictatorship.²¹⁵ And with this fabricated video, unleashed by search-optimizing algorithms on the networks, there was a feeling that Kenyan politics had entered the post-truth era.

The proliferation of digital misinformation and fake political websites during the 2017 Kenyan election unfurled new waves of distrust in a society scarred by successive bouts of post-election violence.²¹⁶ In the past, Kenya might not have been considered a prime target for hybrid AI-cyber influencing. Yet, in many ways, countries that are vulnerable links are the ripest targets for disinformation campaigns. Even though Kenya has one of the highest rates of cyber connectivity in Africa – with four and a half million active Kenyan Facebook accounts and two million Twitter accounts, as of 2015²¹⁷ – the country's history of political corruption and violence renders it vulnerable to the use of social media as an attack vector by both State and non-State actors.

Kenya has a vibrant social media landscape with a wide range of issues and viewpoints represented. As a result, the country has seen a spike in semi-organized "bloggers for hire", who use their influence on platforms such as Twitter and Facebook "to shape public opinion and manipulate the online information landscape."²¹⁸ During the 2017 election, social media allowed these bloggers to contribute to online manipulation and disinformation.²¹⁹ Similar to the disinformation campaigns during the last elections in the US²²⁰ and the UK,²²¹ a number of websites were created using legitimate sounding names to disseminate fake news, thereby undermining the quality of information available online to Kenyan citizens.

Another pervasive anomaly in Kenya's information ecosystem is the potential for the government to exert precision surveillance on citizens via the subversion of

“As AI and cyber capabilities expand in developing countries, so too will the attack surface.

the telecommunication facilities around the country. The UK-based non-profit Privacy International (PI) has reported that Safaricom,²²² Kenya's leading mobile internet provider, allegedly provides data to government authorities, allowing for the interception of both data and content.

As vulnerable States fall victim to AI-cyber influencing, unable to prevent and mitigate data-poisoning and adversarial cyberattacks, they could become fertile operating grounds for cyber mercenaries, terrorist groups, and other actors, increasingly compromising the data-integrity and the robustness of the globalized intelligence system.

CYBER COLONIZATION: DEVELOPING COUNTRIES AS DATA-RESERVOIRS AND TESTBEDS

The ability of AI-driven technologies to influence the livelihood and well-being of large populations is of such immediate and overriding value that it is almost certain to be the theatre for future unconventional conflicts.

There is a very real prospect of a cyber race, in which powerful nations and large technology platforms enter into open competition for collective datasets, as fuel to generate economic, medical and security supremacy across the globe. Forms of cyber colonization are increasingly likely, as powerful States are able to harness AI and converging technologies to capture and potentially control the data-value of other countries' populations, ecosystems, and bio-economies. The ambitions of tech-leading nations are also focused on penetrating and commodifying the data of markets in Southeast Asia, Latin America, Africa, and the Middle East.

It is time to consider the potential for data-predation coming from the convergence of AI and genomics when considering the issue of biopiracy. The Nagoya Protocol is supposed to ensure fair and equitable sharing of benefits arising out of the utilization of genetic resources.

“ The ability of AI-driven technologies to influence the livelihood and well-being of large populations is of such immediate and overriding value that it is almost certain to be the theatre for future unconventional conflicts.

If countries start demanding benefit-sharing in return for the crucial bio-data they provide, they would be able to recover some financial assets after the waves of shocks produced by the impact of automation on the global economy.

Another concern comes from developing countries being used as testbeds for powerful dual-use technologies. It is not rare to see cutting-edge genomics applications being first tested in the ecosystems of countries that are not technically and legally equipped to conduct their own safety and life-cycle assessment.

INCREASING HUMAN SECURITY NEEDS IN A GLOBALIZING WORLD

We are entering an era of hybrid threats and emerging risks triggered by the combination of powerful dual-use technologies. Below are some of the overarching take-aways to consider in further governance analysis (Section V).

The landscape of hybrid security threats is expanding as well as the attack surface. Technologies are becoming hybrid complex systems that are merging the data of our digital, physical and biological lives, with potential for pervasive vulnerabilities and emerging risks. AI and converging technologies are characterized by new processes of decentralization, distributed agency and forms of “atomized responsibility.” Increasingly, these technologies deploy beyond State control, involving new actors such as rogue States and cyber mercenaries, and challenging established expertise and governance models. New forms of invasive cyberwar and cognitive-emotional conflicts will have powerful implications for collective (bio)-physical, digital and political security.

Governance actors, including States and the private sector, will need to adopt techniques of inclusive foresight to be resilient and adaptive enough in the face of hybrid security threats and emerging risks.

States in the Global South are struggling to compete, build and secure capacity in the development and deployment of AI and converging technologies. They risk becoming vulnerable links in a new geography of AI-cyber conflicts. They also may become more vulnerable to dynamics of data-predation and cyber colonization. Such vulnerabilities may fuel and intensify a fierce competition for supremacy in technological convergence, rather than foster digital cooperation. Distributive inequalities will be on the rise between countries that are tech-leaders and tech-takers. Some developing countries may not have the power, influence and foresight tool to shape responsible governance of converging technologies towards social benefits and away from political disruptions and weaponization. These vulnerable States could become a liability for whole regions.

However, States interested in fostering responsible AI convergence could enter into mechanisms of digital cooperation with countries in the Global South to partner around mutually beneficial transfers of data, talent, technologies and security practices.

Technological risks have powerful, long-term and corrosive impacts on human security and well-being.

As job insecurity rises due to complex technological and social transformations, global human psychological and emotional well-being could be receding. Underserved groups in societies may suffer from new forms of disempowerment. In turn, disempowerment and the lack of opportunity to participate in innovation will impact trust and social cohesion.

Governments urgently need to build a new social contract for the era of AI convergence, including by investing in measures that can reinforce networks of social cohesion and resilience. They will need to foster the capacity of individuals, communities and systems to survive, adapt, grow and even transform in the face of political and socio-economic shocks and stresses.

In general, few contemporary analyses capture how AI convergence will impact the security of populations in

low-income and fragile States – those States that struggle to compete and build tech and innovation capacity. There is a crucial need to map and analyse which countries and populations will face pervasive hybrid threats due to the convergence of dual-use technologies. The below matrix (Table 1) offers such a preliminary mapping of emerging threats classified by separate security domains, drawing from the analysis above.

“ Governance actors, including States and the private sector, will need to adopt techniques of inclusive foresight to be resilient and adaptive enough in the face of hybrid security threats and emerging risks.

TABLE 1: Matrix of Emerging Security Threats

RISKS	HUMAN SECURITY	POLITICAL SECURITY	SOCIO-ECONOMIC SECURITY	CYBERSECURITY	BIOSECURITY
AI, AUTOMATION, & AFFECTIVE COMPUTING	<p>Current and future developments in AI and affective computing will exacerbate existing security vulnerabilities, for instance through social engineering</p> <p>Surveillance technologies, emotional manipulation, and behavioural engineering lead to pervasive infringements and violations of human rights (e.g. privacy, data agency, self-determination).</p>	<p>The proliferation of hyper-targeted disinformation campaigns, propaganda, automated forgeries, fake intelligence, emotion manipulation, and other forms of AI-Cyber deception influences public perception and the public's ability to make informed decisions. Affective data, when aggregated, provides insight into personalities, decision-making, and mental states, and could be increasingly useful in politics and diplomacy.</p>	<p>The future of work will be greatly impacted by continued automation of various industries, from manufacturing to data-driven services.</p> <p>There are significant concerns relating to distributive inequalities and equitable access to data-optimization, predictive intelligence, automation and emotional intelligence.</p>	<p>Protection of intimate, sensitive data and digital identities for individuals and populations will become increasingly challenging.</p> <p>The cyberoffense toolkit expands with automated ransomware, adversarial and data-poisoning attacks on critical infrastructures, as well as automated social engineering.</p>	<p>Automated adversarial cyberattacks, data-poisoning, and social engineering will target automated bio-labs and other biotech infrastructures, corrupting or destroying biological data and bio-intelligence.</p>

RISKS	HUMAN SECURITY	POLITICAL SECURITY	SOCIO-ECONOMIC SECURITY	CYBERSECURITY	BIOSECURITY
AI & BIOTECHNOLOGY	By 2025, between 100 million and 2 billion human genomes could be sequenced, rendering vast swaths of the global population vulnerable to hackers, malicious technological applications, and repression of vulnerable populations.	Bio-data capture and dynamics of cyber colonization help tech-leading States acquire increasing geopolitical influence, imposing technological dependence on tech-taker States, and imposing bio-power on foreign populations.	Bio-data capture and dynamics of cyber colonization transfers valuable assets outside of the country of origin and will likely increase inequalities between tech-givers and tech-takers.	More and more intimate data are being collected on individuals and populations from biometrics, health to genomics data; individuals need to control and have access to a complete set of the type of data captured, analyzed, and forgotten.	Using genome-editing to generate “precision maladies” and using germline editing to attempt to direct human evolution lead to unintended vulnerabilities; using gene-drives technologies to modify animal germlines could have various negative impacts on species and environments.
AI & NEUROTECHNOLOGY	Protection of neural data becomes increasingly challenging, undermining individuals’ and populations’ right to autonomy and neural data agency; processes of sharing neural data may need the same protection as organ donation.	Technologies for surveillance, data manipulation, hybrid influencing, and social engineering increasingly focus on a better understanding of neurological processes or use of neuromodulation.	Advancements in the neurosciences and neurotechnologies alter social norms and raise concerns of equitable access and new forms of discrimination.	Individuals’ and populations’ neural data become another target of cyberattacks; Brain-Computer Interfaces just like other internet connected medical devices become target for cyberattacks.	There are growing concerns of precision bio-attacks that may rely on the use of neuro-toxins or that may target the human microbiomes; there are increasing concerns about human neuro-enhancement in civilian and military contexts.
AI & ADDITIVE MANUFACTURING	The combination of drone technologies and additive manufacturing create new security vulnerabilities for individuals and populations; off-the-shelf drones can be modified using 3D-printed parts to drop explosives.		The future of work may be impacted by relocalization of manufacturing into additive supply chains, creating economic loss for countries that have specialized in low-cost manufacturing.	Automated adversarial cyberattacks, data-poisoning, and social engineering will target additive manufacturing infrastructures, corrupting or destroying industrial intelligence.	Automated adversarial cyberattacks and data-poisoning will target 3D-bio-printing corrupting or destroying industrial bio-intelligence.

V



RESPONSIBLE GOVERNANCE IN THE ERA OF AI CONVERGENCE

The scenarios above describe evolving global security threats amplified by the combination of powerful dual-use technologies. The convergence of AI, cyber and biotechnologies raises complex, ambiguous security implications that are poorly understood, leaving the private sector, civil society, States, and the multilateral system with few tools to anticipate and prevent emerging risks.

Politically, legally and ethically, our societies are not properly prepared for the deployment of AI and converging technologies. At national and international levels, we lack a comprehensive understanding of the threats that AI and converging technologies can pose at the individual human level, broader threats to populations, and geopolitical confrontations potentially triggered by the combination of new technological trends. This section focuses on how key actors, including the private sector, governments, and the UN, can contribute to responsible governance in the era of AI convergence.

CORPORATE VISION AND RESPONSIBILITY WITH GLOBAL REACH

Those who have the knowledge to shape the current technological revolution belong primarily to the private sector. Advances in AI, cyber and biotechnologies are driven by private tech platforms that have the cloud and the computing power to commodify and control massive data-sets. Decisions about how AI convergence will shape the future are taken by a new corporate order, not made of diplomats and lawyers, but of CEOs with global reach.

Major technological platforms have long integrated the potential opportunities that technological convergence and multifaceted data (personal, biological and neural data) will bring to designing powerful forms of general AI. Better tools for data-optimization and predictive intelligence are just one step on the path towards “solving” human intelligence all together. Competition over the economic gain produced by AI’s convergence with bio- and neuro-technologies is the reason why Silicon Valley and China are engaged in a race.

Google and Amazon benefit from their powerful cloud platforms, which are required for leading the largest genomics studies conducted in the US and abroad. About USD 13 billion²²³ of Facebook shares have recently been invested in the Chan-Zuckerberg Initiative,²²⁴ which pursues the ambition to develop a “human cells’ digital

atlas” to cure, prevent and manage all diseases in our children’s lifetime. China’s WuXiNextCODE,²²⁵ is one of the largest genomic data platforms using machine learning to better diagnose rare diseases and cancer as well as design tailored, improved therapeutics.

Partnering with the private sector and academic research labs will be instrumental in understanding and anticipating the implications of AI convergence. As they face the complex political implications of their creations, some private sector actors are exhibiting growing interest in collaborating with the UN to foster normative guidance and align technological convergence with the public interest.

In 2018, six AI-leading companies published ethical codes of conduct (compared in Table 2), showing their interest in self-regulation.²²⁶ If these principles truly shape algorithmic design and materialize into risk assessment and governance practices, they could form a solid basis for responsible innovation in AI convergence. The question remains if responsible governance can crystallize exclusively as an internal ethos without the collaboration and oversight of policy communities.

SEVERAL INSIGHTS EMERGE FROM THIS COMPARATIVE ANALYSIS:

Transparency versus Secrecy: The first is an aspiration to promote transparency and openness when conducting AI research and deployment. Such an effort is certainly paved with ambiguities. Research communities are increasingly harnessing converging technologies, which had until now been traditionally siloed in corporate labs. Many actors are therefore making decisions and actions in extended supply chains that determine the deployment of converging technologies. In the field of AI, characterized by ubiquitous access to data and source codes, the responsibility for the misuse of algorithms is easily “atomized,” distributed among peer designers, regulators, users and hackers. It is therefore increasingly difficult to limit access to knowledge and tools involved in dual-use research of concern.

Corporate leaders, entrepreneurs and engineers will have to learn how to balance the powerful tension between expectations of transparency – an ethos of openness – and the need for secrecy and security in dual-use research and innovation. This is not only true for AI, but increasingly for its convergence with other technologies. This tension is drastically magnified by incidents that have already occurred in the field of biotechnologies, such as the recent synthesis of horsepox and studies of avian

TABLE 2: Matrix of AI Industry Guiding Principles

 = principle is targeted;
  = principle is not targeted

PRIVATE COMPANIES & INSTITUTIONAL ORGANIZATIONS						
Values & Principles	GOOGLE	DEEPMIND	IBM	MICROSOFT	ACCENTURE	PARTNERSHIP ON AI
Societal Benefit						
Limit Harmful Applications						
Transparency						
Openness						
Accountability						
Responsibility						
Inclusion						
Value Alignment						
Fairness						
Explainability						
Reliability						
Safety						
User Data Rights						
Privacy by Design						
Security by Design						
Safety by Design						
Trust by Design						

PRIVATE COMPANIES & INSTITUTIONAL ORGANIZATIONS

Values & Principles	GOOGLE	DEEPMIND	IBM	MICROSOFT	ACCENTURE	PARTNERSHIP ON AI
No Weaponization						
Avoid Algorithmic Bias						
No Violation of Human Rights						
Strong Cybersecurity						
Human Control						
Non-Subversion						
Engage the Public & Stakeholders						
Ethical, Social, & Legal Implications						
Scientific Excellence						
Rigorous & Evidenced-Based						

flu strains H5N1 and H7N9. In 2018, scientists at the University of Alberta published the process required to synthesize a smallpox relative from scratch, using publicly available DNA fragments.²²⁷ And, in 2011 and 2012, two scientific publications became the centre of a global controversy for disclosing which mutations can convert a deadly strain of influenza in birds (H5N1) to one that could spread among mammals.²²⁸

In the future, with more advances in deep learning for functional genomics, one may not even need much tacit knowledge of bio-experimentation to obtain an agent capable of causing harm. Moreover, the increasing automation of biotech in “intelligent and connected labs”²²⁹ provides new avenues to avoid existing national and international biosecurity oversight.

The same reasoning applies to AI and cybertechnologies. Socially responsible principles recently published by companies such as IBM and Microsoft emphasize that AI systems should be easy for users to understand how these systems work. Yet, in this effort toward transparency, AI research labs and private companies will need to think

seriously about adequate safeguards when disseminating training data, source codes of concern, or information that could be harnessed to devise malware and adversarial attacks. This became clear when the research institute OpenAI²³⁰ decided not to publish the tacit knowledge, training data and full algorithmic model of a new AI-powered text generator. The AI model is powerful enough to create, from scratch, texts, arguments and opinionated book reviews that could be misinterpreted as coming from a human writer.

Transparency and understandability are in direct tension with the traditional approach to intangible technology transfer in dual-use research. These intangible transfers of technology matters beyond the debates of non-proliferation experts – the transfers take place in civilian research, outside of secured labs, in open innovation ecosystems, and are increasingly difficult to control. Such tacit knowledge also plays an important role in the collaboration between private sector actors and governments when assessing security threats, preparedness and modalities of governance.

“ AI research labs and private companies will need to think seriously about adequate safeguards when disseminating training data, source codes of concern, or information that could be harnessed to devise malware and adversarial attacks.

Predictive Accountability versus Liability: As unveiled in Section III scenarios, the development and deployment of converging technologies are fraught with ethical, legal and social quandaries that call for foresight, for a form of collective and anticipatory wisdom. Whose duty is it in global socio-technical systems to foresee the unintended consequences of dual-use technologies, and who possesses the required expertise for preventing harm?

In the UN Disarmament Agenda, *Securing our Common Future, and the Strategy on New Technologies*,²³¹ UN Secretary-General Guterres has called for partnerships across the scientific, engineering and multilateral communities to foster responsible technological innovation and use of scientific knowledge. This is based upon an understanding that technological convergence is already having a direct impact on global risks, without the understanding and foresight needed to anticipate and respond to such risks.

Founders and engineers of AI and tech companies have a unique visionary capacity and expertise to anticipate emerging threats in their domains.²³² Their knowledge is instrumental when trying to delineate the drivers, nature, potential impact, likelihood and velocity of emerging technological risks. They can help policymakers identify a set of examples of plausible futures that provide a valuable point of reference when assessing current strategies or formulating new ones.

Engineers and entrepreneurs can help policymakers discuss governance models that rely on “predictive accountability” for converging technologies. Under “predictive accountability,” inventors and producers of dual-use technologies would conduct foresight to anticipate potential scenarios of technological misuses. Equipped with such predictive analysis, they would be able to improve technological design to account for and mitigate unintended consequences. These efforts

should include human rights impact assessment from the beginning, to assess converging technologies’ risks and benefits, and define guiding principles for ethical and safe design.

Competition versus Shared Prosperity: Most corporate guiding principles insist on social and shared benefits. This prompts broader ethical questions from the sidelines of the global conversation on AI convergence and policy: how are advances in AI and converging technologies affecting existing socio-economic inequalities across our burdened planet?

Corporate vision and business strategies that ignore social ills and contingencies will tend to crystallize algorithmic and genetic divides and, in turn, limit opportunities to bridge them. Increasingly, we may need different epistemic, cultural and political approaches to adequately assess emerging technologies’ risks and their broader social impacts. Without adequate reflection, the risk is that ground-breaking scientific discoveries and converging technological applications will outpace societies’ capacity for control. A “techstorm” is essentially a period of overzealous technological innovation that can have destructive consequences.

Techstorms are enabled by rhetoric that hypes immediate benefits while downplaying risks. This premise allows technology creators, such as those in the AI, biotech industry and the military-industrial complex, to practice a form of “permission-less” innovation, evade scrutiny and quickly entrench themselves in society. Security and defence research priorities as drivers of the techstorm are in line with what is happening in the current competition in AI and biotechnologies. This paradigm of fast-paced or permission-less innovation, driven by economic competition and security imperatives, incorporates little input from citizens beyond consumers’ market preferences.

“ Whose duty is it in global socio-technical systems to foresee the unintended consequences of dual-use technologies, and who possesses the required expertise for preventing harm?

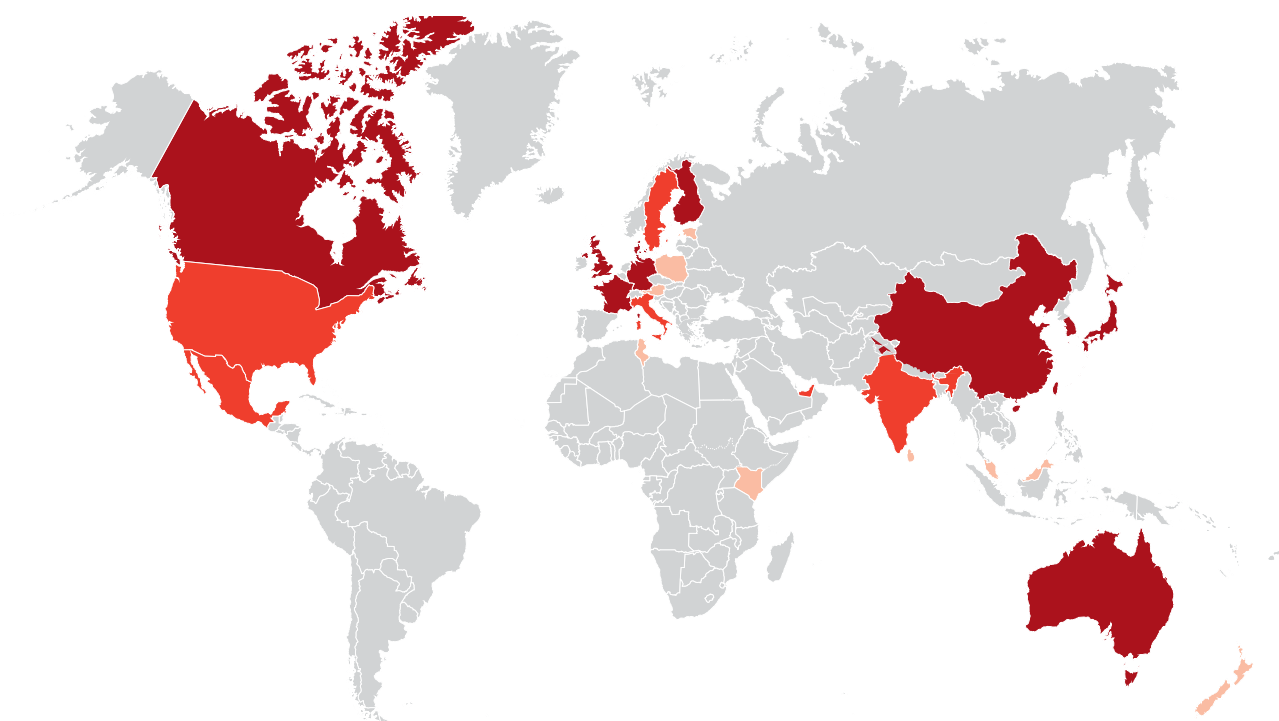
FRAGILE STATES: AI CONVERGENCE DEVELOPMENT AND EXCLUSION

In the future, UN Member States will need to develop a common understanding of technological convergence to be able to design proper oversight in collaboration with strategic actors in the private sector and civil society. States lagging behind in AI convergence are the most at risk and the least likely to have any foresight capacity.

Pressured in a race to develop AI talents, research and industrialization pathways, States are not equal in their ability to understand and anticipate evolving security risks, including in their political and socio-economic dimensions.

Most national AI strategies have been developed by tech-leading States, leaving large unprepared regions in this new geopolitical landscape of converging technologies. These trends are apparent in the below Map, which shows where national AI strategies have been developed.

Global Heatmap of National AI Strategies



KEY

The key of the map should show the difference between countries that already have full-fledged national strategies, the countries with guidance documents (but no national strategy), and those countries that are currently in the development process of national strategies

The map was inspired by a CIFAR December 2018 report (see figure 1 on page 8), but has been updated to include more recent events

- = Funded and implemented strategy
- = Guiding document; not implemented
- = Working on a strategy

COUNTRIES WITH NATIONAL AI STRATEGIES

IN DEVELOPMENT	GUIDING DOCUMENT	FUNDED
<ul style="list-style-type: none"> • Austria • Estonia • Kenya • Malaysia • New Zealand • Poland • Sri Lanka • Tunisia 	<ul style="list-style-type: none"> • India • Italy • Mexico • Sweden • United Arab Emirates • United States 	<ul style="list-style-type: none"> • Australia • Canada • China • Denmark • European Union • Finland • France • Germany • Japan • Singapore • South Korea • Taiwan • United Kingdom

“UN Member States will need to develop a common understanding of technological convergence to be able to design proper oversight in collaboration with strategic actors in the private sector and civil society.

Moreover, most AI national strategies are treating issues related to political security and socio-economic cohesion as a secondary policy priority.²³³ UN Member States are struggling to foresee the broader, transformative landscape of AI and converging technologies.

When it comes to the dynamics of tech exclusion, an urgent priority may be to create incentive structures to encourage more founders in the private sector to take on some of the real-world problems faced by developing countries and to do so with ethics at their heart. Another imperative is to create, in lower-income

countries, opportunities for academics and entrepreneurs to create channels of learning and collaboration for the next generation to gain literacy in AI and converging technologies.

Yet, there are also opportunities for developing countries with respect to AI and converging technologies. Data-optimization and predictive intelligence, in combination with biotechnologies, will enable innovation in sectors as vast as climate change, precision health and agriculture, personalized education, and additive manufacturing, all activities that might be more transformative for the Global South than the Global North.

The capacity of AI to fight corruption and improve government efficiency, service delivery and public administration is too often forgotten. Cooperation between lower-income countries and international financial or development institutions could harness predictive intelligence for public good. This, however, comes with a significant caveat: the private sector leading AI convergence would have to help lower-income countries avoid the array of hybrid threats described in the scenarios above, and facilitate understanding of how complex algorithms operate and impact already vulnerable societies.

The ultimate goal of AI-cyber development would be to promote endogenous responsible innovation and strengthen economic and social resilience.



UN Photo/Eskinder Debebe
The Secretary-General's High Level Panel on Digital Cooperation

“ **The ultimate goal of AI-cyber development would be to promote endogenous responsible innovation and strengthen economic and social resilience.** ”

TECH FRONTIER AND DIGITAL COOPERATION AT THE UN

To address the enormous potential benefits and threats triggered by emerging technologies, the UN Secretary-General has launched a series of initiatives on how the UN system may integrate these technologies to support its mandates. The primary goal of the Secretary-General's Strategy on New Technologies²³⁴ is to determine how the UN can use these technologies to accelerate the achievement of the 2030 Sustainable Development Agenda²³⁵ and facilitate their alignment with the values enshrined in the UN Charter.²³⁶ In support of this endeavour, the Secretary-General identified five principles to guide UN engagement with new technologies, which revolve around adherence to global values, fostering inclusion and transparency, and building on existing capabilities and mandates by working in partnership with multi-stakeholders.

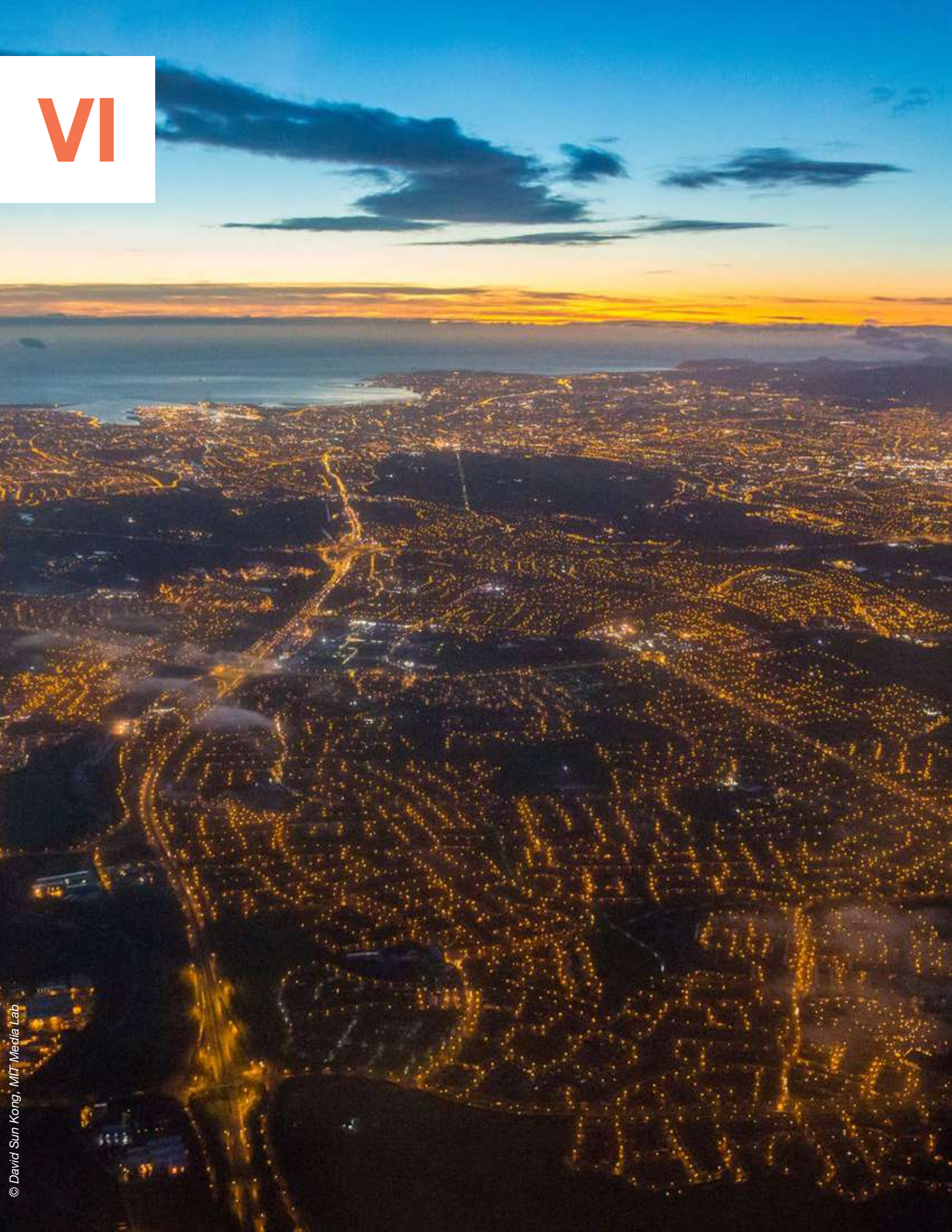
The Strategy identifies four areas in which the UN may enhance its external and internal engagement on emerging technologies:

- Deepening the UN's internal capacities and exposure to new technologies;
- Increasing understanding, advocacy and dialogue;
- Supporting consensus on normative frameworks;
- Enhancing the UN's support to government capacity development

The Strategy is intended to be a “living document” – that is, it is supposed to be adapted as the UN learns and acts on it, and as external factors and needs shift. This will be challenging. Such a complex task will demand access to expertise that is inclusive of diverse technical, ethical, regional, gender and age-group perspectives.

In order to successfully implement the Strategy, the Secretary-General established a High-Level Panel on Digital Cooperation²³⁷ to identify principles, values and mechanisms for working cooperatively across sectors, disciplines, and borders when addressing the challenges of the digital age. The High-Level Panel is expected to release a final report in 2019 with concrete recommendations for how the UN should move forward in such a complex environment.

VI



PREVENTION IN THE ERA OF AI CONVERGENCE

Evolving security risks will certainly require the multilateral system to anticipate and better understand the rapidly emerging field of AI convergence.

How can the UN help develop a capacity of foresight and oversight that promotes responsible innovation in AI and converging technologies – empowering those in need while preventing vulnerabilities to turn into pervasive hybrid threats? How can corporate leaders be encouraged or required to design responsible applications of AI and converging technologies needed to remedy global problems of epidemics, climate change and endemic conflicts – while giving affected communities a role in, or even control of, their technological futures? How can a diversity of stakeholders collaborate to anticipate the risks involved and the sources of inequalities and disempowerment that hide in dual-use technologies?

Responsible innovation provides a cooperative approach to manage the governance tensions between the private sector and policymakers. “Specifically, it expands the notion of ‘market failure’ – that grants governments the authority to intervene in free markets over areas in which the price system fails – to include ‘public value failure’ or situations in which the innovation system fails to deliver outcomes consistent with widely held public values – such as safety, privacy and choice.”²³⁸

To achieve this goal, responsible innovation relies at its core on what experts call anticipatory governance or forms of “predictive accountability.” When facing uncertain and complex technological outcomes, anticipation implies building individual and collective abilities for adaptation, resilience, and preparedness. Such abilities require understanding plausible scenarios related to AI convergence futures. Equipped with foresight analysis, inventors, producers and regulators

of dual-use technologies are in better positions to improve technological design to account for and mitigate unintended consequences.

To what extent are national and international governing bodies equipped with the knowledge, foresight, and analytical tools to foster multilateral discussions about responsible innovation?

The UN is uniquely positioned to develop an ethical and inclusive foresight function for the anticipatory governance of AI and converging technologies. The UN can assist Member States, private sector actors and other stakeholders define their interests, identify strategies for cooperation and be more responsive to benefits, but also to emerging risks and human rights abuses triggered by the combination of dual-use technologies. Inclusive foresight can help anticipate, learn, adapt, be resilient, and ultimately prevent technological, governance and ethical failures. In the end, this is about prevention in the era of AI convergence.

“**Inclusive foresight can help anticipate, learn, adapt, be resilient, and ultimately prevent technological, governance and ethical failures.**”

ANTICIPATING GLOBAL POLICY ISSUES IN AI CONVERGENCE

The scenario about the “Internet of Bodies, Genomes and Minds” unveils the extent to which AI convergence promises to drastically impact human well-being, human rights and the human side of global security risks.

We are starting to understand how future generations will be increasingly challenged when it comes to knowing

what it means to be human. At the same time, we find ourselves often powerless or, at least, poorly equipped to comprehend and anticipate how AI and converging technologies will commodify our secrets in novel ways, mining thoughts, aspirations, words, habits, bodies, genomes and emotions as fuel to engender always more profit. As eloquently said by UN Assistant Secretary-

General, Fabrizio Hochschild, “what we understand far less is what all these [technological] changes mean for us socially, politically and psychologically; what they will mean for the relationship between citizen and State, for the conduct of conflict, for our economies, for our psyche and for our human rights.”²³⁹

Urgent efforts are needed using inclusive foresight to better prepare societies for the emerging threats that target truth and trust, political and social cohesion. With the private sector and civil society, diplomats and policymakers will also need better foresight to define, shape and anticipate what norms and oversight models will protect human safety and security in an era of technological convergence. What model of anticipatory and responsible governance can prevent emerging risks when dual-use technologies combine?

If the UN is to collaborate effectively with the private sector, Member States and civil society, on global anticipatory

and responsible governance of AI and converging technologies, it will have to do so across mandates. As shown in this report’ scenarios, AI convergence will have complex, ambiguous implications for the different prevention and peace and security domains under the purview of the UN.

In each sector, an array of entities within the UN system could play a role that is sorely needed at the international level: 1) technological foresight, which is inclusive of diverse countries’ challenges; 2) negotiating adequate normative frameworks; and 3) monitoring and coordinating standards and oversight.

The following matrix defines the contours of an evolving set of emerging challenges and risks that connects with the below categories of global policy issues within the prevention and peace and security mandates.

TABLE 3: Emerging Global Policy Challenges

<div>PEACE & SECURITY AGENDA</div>	<p>Techniques of deception, such as: political influencing, simulation of data, targeted propaganda and election interferences; corrosive degradation of trust;</p> <p>The rise of cognitive-emotional conflicts;</p> <p>Conflict prevention, mediation and resolution is increasingly complex; preventive diplomacy is challenged;</p> <p>Pervasive data breaches through automated adversarial cyberattacks increased human insecurity;</p> <p>Hybrid AI-Cyber influencing leading to kinetic conflicts;</p> <p>Automated adversarial cyberattacks and data-poisoning with implications for how to build, control and secure biotech, manufacturing and physical systems;</p> <p>The weaponization of decentralized networks of information and intangible knowledge transfers;</p> <p>The rise of hacktivism and cyber mercenaries;</p> <p>Combinations of dual-use technologies create potential for precision attacks on populations within cities such as optimized drone swarms;</p> <p>Fragile States and developing countries as vulnerable links in a whole region.</p>
--	---

<p>DEMOCRACY, HUMAN WELL-BEING, & HUMAN RIGHTS AGENDA</p>	<p>Pervasive, intrusive, large-scale surveillance of individuals and populations;</p> <p>The rise of digital dictatorship and surveillance capitalism;</p> <p>Increased infringements of human rights, such as pervasive threats to privacy, self-determination, and biological and neural data agency;</p> <p>Social engineering, emotion manipulation and neuromodulation of individuals and subsets of populations;</p> <p>Social cohesion, human psychological and emotional well-being is receding;</p> <p>The augmentation of human populations and ecosystems.</p>
<p>DEVELOPMENT AGENDA & SOCIO-ECONOMIC ASPECTS</p>	<p>The automation of work provokes major disruption in labour markets and new stress points for large populations;</p> <p>Data-predation, cyber colonization and technological dependence;</p> <p>Destruction of industrial manufacturing and bio-intelligence;</p> <p>Automating exclusion and technological, genetic and algorithmic divides;</p> <p>Developing countries as data-reservoirs and testbeds.</p>

The mapping shows that foresight will be needed for global policy challenges across various UN sector of actions.

Among these policy issues some fall clearly into traditional approaches to peace and security, which belong to disarmament and non-proliferation mandates where foresight approaches could modernize legacy toolkits.

Yet, others are murky, cross-boundary issues, which could be left unaddressed with corrosive implications for individuals and broader populations. Such cross-boundary policy issues are complex, ambiguous and emerging from the merger of our digital, physical and biological lives: 1) pervasive, intrusive, large-scale surveillance of individuals and populations; 2) threats to data agency and self-determination; 3) data-poisoning with implications for how to build and control physical systems and security; 4) social engineering and emotion manipulation of individuals and subsets of populations; 5) political influencing/deception, election interferences and cognitive-emotional conflicts; 6) the weaponization of decentralized networks of information and intangible knowledge transfers; 7) data-predation, cyber colonization and technological dependence; 8) the augmentation of human populations and ecosystems.

A majority of these new threats could be labelled or addressed under political, socio-economic and human security risks. They concern both, prevention and peace and security mandates, but without matching previous definitions of global policy issues by the UN. In this context, inclusive foresight will be crucial to develop a form of technical, normative and political preparedness based on anticipating potential dual-use and misuse of AI and converging technologies.

Anticipating Political, Socio-economic and Human security risks: The ability of AI-driven technologies to influence large populations is of such immediate and overriding value that it is almost certain to be the theatre for future conflicts. Achieving the UN prevention agenda will require providing sharp and inclusive horizon scanning to anticipate the nature and scope of emerging security risks that will threaten not only nations, but also individuals and vulnerable populations. Such foresight will become increasingly critical as AI converges with other technologies that are beyond State control and more accessible to a wider range of actors around the world.

New forms of social and bio-control could in fact require a reimagining of the framework currently in place to monitor and implement the Universal Declaration of Human Rights and will certainly require the multilateral system to better anticipate and understand this quickly emerging field.²⁴⁰

“ Inclusive foresight will be crucial to develop a form of technical, normative and political preparedness based on anticipating potential dual-use and misuse of AI and converging technologies.

Promoting and Protecting Human Rights: Inclusive foresight and normative monitoring and coordination will be particularly crucial in the promotion and protection of human rights.

Given the powerful, sometimes corrosive, implications that AI may have for self-determination, privacy and other individual freedoms, UN entities will need to collaborate to monitor and guarantee coherence across multiple normative efforts spurred by national, regional and private actors.

The alliance of AI, genome-editing and neurotechnologies is also likely to change societal norms, raise issues of equitable access and generate new forms of discrimination.

Preventing Distributive Inequalities and Their Impact on Human Security: While promises are unprecedented to read genomes, crack illnesses, famine, and evolution on a global scale, for large subsets of the global population, inequalities and suffering remain. The first inequality which will arise will be about who is expected to bear the burden of technological risks.²⁴¹ It is not rare to see cutting-edge genomics applications, such as gene-drives, being first tested in the ecosystems of countries that are not technically and legally equipped to conduct their own safety and life-cycle assessment.²⁴² The same way, countries that are lagging behind in AI convergence are likely to become vulnerable links in future cyberwars.

Yet, another kind of inequality is lurking. Most populations in the Global South are in no position to claim the right to envision and shape how converging technologies will impact their futures. The gap keeps widening between frugal innovation in the South and the wildest dreams of visionary minds in societies of abundance who design new ways to optimize humans and ecosystems. There is no equality of rights when it comes to imagining, anticipating and choosing the designs and powers of converging technologies. Those have often been conceptualized in vibrant, wealthy universities or by successful corporations with global reach.

MULTI-STAKEHOLDER COOPERATION ON FORESIGHT: TWO USE-CASES

To achieve strategic foresight, the UN needs to modernize its models of governance cooperation by increasingly relying on agile but inclusive networks. They should include entrepreneurs, scientists, social-scientists, policymakers, legislators and diplomats, including from the Global South. Such networks can facilitate a deep understanding of how complex algorithmic and genetic innovation will impact societies.

USE-CASE 1: TECHNOLOGICAL PREPAREDNESS FOR VULNERABLE STATES AND THE GLOBAL SOUTH

To respond to emerging threats and build technological preparedness, governance actors, including States and the private sector in collaboration with the UN, will need to adopt oversight models that allow for resilience and adaptation in the face of hybrid security threats.

It will be crucial for these governance actors to compare and monitor how well oversight models perform at local, national, regional, and international levels. Such a complex monitoring effort would benefit from developing strategic and cohesive regional dialogue to discuss principles of responsible innovation for the deployment of dual-use, converging technologies.

Importantly, foresight exercises have to provide a more in-depth understanding of how AI and converging technologies can impact multiple security domains for States of the Global South, which may be struggling to compete, build and secure capacity in the development and deployment of AI convergence. Equipped with such a predictive analysis, entrepreneurs and experts in the North and the South would be in position to improve technological design to account for and mitigate unintended consequences.

Nascent efforts around foresight and predictive accountability could take increasingly agile forms and thrive on alliances between States such as North-South, and South-South collaborations. For instance, within a network of trust, entrepreneurs, policymakers and UN officials could use red-teaming exercises²⁴³ to anticipate potential vulnerabilities as well as safety and security best practices relevant to the convergence of AI, cyber and biotechnologies.

This knowledge-sharing process has to break silos and promote learning across technologies (AI-Cyber-Bio-Neurotech-Quantum) and security domains. Decades of managing ethics and dual-use research of concern in biotech can certainly inform the field of AI and cybertechnologies where algorithms and malware present similarities with self-replicating organisms. Cultures of responsibility should be shared across converging technologies.

Red-teaming could also turn into “sparring” exercises where corporate and government actors learn to collaborate and build a trusted space where to test AI and cybersystems to report fatal anomalies and discuss optimized defence capabilities.²⁴⁴ Such sparring exercises could be crucial for States that are currently vulnerable links in the new geopolitics of converging technologies.

States that are willing to lead in the responsible governance of AI, cyber and biotechnologies could develop ad-hoc mechanisms of inter-State cooperation, a “safe space” where to discuss predictive accountability, adherence to emerging norms, as well as legitimate and illegitimate behaviours when deploying dual-use technologies.

To be even more successful, foresight efforts should include human security and human-rights impact assessment²⁴⁵ from the beginning, to assess converging technologies’ risks and benefits, and define guiding principles for ethical and safe design. Human rights impact assessments provide a step-by-step evaluation of the impact of practices or technologies deployed in a given context on aspects such as privacy, data agency and self-determination.

Collaborations should take place between teams that specialize in foresight and predictive accountability with teams that systematically monitor for ethical breaches and violations of human rights in the field or for vulnerable populations. Building relationships between UN entities, UN Innovation labs and human rights’ labs inside AI companies would constitute an interesting collaborative network to foster principles of predictive accountability.

The UN has a significant and unique role to play in ensuring that wider participation, from more vulnerable States or underserved groups, is enabled and supported within strategic foresight. New forms of adaptive social engagements with entrepreneurs, engineers, civil society and, in particular, the next generation, could help renew interest in applying the norms and values of the UN Charter to the era of AI convergence.

USE-CASE 2: POLITICAL PREPAREDNESS -AI, CYBERTECHNOLOGIES AND PREVENTIVE DIPLOMACY

The combination of AI, affective computing, biometrics and cybertechnologies provide non-State actors with more tools to interfere directly with a State’s political processes and the minds of its citizens. With new forms of fake intelligence and smart terror, information, politics and war are increasingly entangled.

In this context, the UN is needed more than ever, as a community of experts, diplomats and policymakers with knowledge and understanding of local, regional and international politics, and as a convening power for entrepreneurs, civil societies, and States, including the most vulnerable ones.

The UN can work with social media companies and AI engineers to help them anticipate and mitigate how technological convergence will be harnessed to manipulate populations and political systems. Entrepreneurs and engineers are currently the main actors shaping what implications AI and converging technologies will have for the conduct of conflict, social cohesion, and human rights. Far from being just neutral information providers, through the use of their determinative algorithms, social media platforms play a substantial role in enabling the waging of virtual conflicts and their real-world results.

The UN can also support Member States in anticipating and understanding the implications that technological convergence can have in a new geography of virtual conflicts. Such conflict space, with its potential for deception and subversion, will increasingly matter and bear significant consequences to both, national security and individual citizens’ security.

In this context, reinforcing the resilience of societies to deceptive campaigns is even more complex and deserves urgent attention in conflict studies. One strategic element to remember is that deception thrives on existing vulnerabilities, from economic exclusion, lack of social cohesion, political discord, disengagement or polarization, to cultural, religious and ethnic dissensions.

Experts in preventive diplomacy have developed comprehensive expertise in analysing conflict dynamics, building sensitive political relationships in fractured countries, and conducting “framework diplomacy” with allies to create a safer space for crisis management. These experts are skilled to analyse the internal drivers of recent and current conflicts, to identify likely threats to peace and to anticipate how they may evolve if left unaddressed.

This expertise and collaborative practices can be leveraged to better anticipate the nature, depth and scope of social and political vulnerabilities that technological deception may target.

“ The UN can also support Member States in anticipating and understanding the implications that technological convergence can have in a new geography of virtual conflicts.

Foresight to Prevent Deception: One strategic option might be to build collaborative teams of experts in preventive diplomacy, conflict resolution and AI-Cybertechnologies. Such a braintrust could conduct a combination of political and technological foresight to 1) analyse emerging tensions, anomalies and divisions in fractured societies; 2) anticipate subsequent scenarios of deception; and 3) plan for strategies to rapidly and effectively counter deception, in particular at crucial times such as election processes.

These types of collaborations could serve as a form of early-warning system to help social media and AI technological platforms better understand, detect and mitigate data-manipulation, forgeries, social engineering, tailored propaganda and disinformation. Such an effort would also help raise general awareness of potential vulnerabilities before and during election times. It could help enter into mediation process with more tactical insights on information warfare, which could be critical in strategic engagement with the parties at stake. It would also benefit larger networks of diplomats and policymakers, journalists and citizens.

Anticipating both, the internal and external tensions that influence internationalized civil wars, and understanding how, in this context, outside powers could harness AI and cyberdeception, is likely to be a recurrent prevention challenge for regional players and multilateral organisations in the years ahead.

Metrics for Threats Prioritization: Experts in preventive diplomacy and AI-Cyber technologies could also work together on prioritizing what form of deception would constitute a minor interference, and what other form would turn into major threats to sow violence or contribute to the growing intractability of existing conflicts. Collaborative teams could do so by developing a metric of damages that acknowledges a range of target objectives and assess if these objectives have led to actual disruptions.

Countering Disinformation: AI-driven tools exist to counter disinformation by filtering fake news, reinforcing known facts, detecting nefarious content, eliminating trollbots, and verifying the authenticity of audio and video content.²⁴⁶ Yet, for now, technical limits and failures abound in these counter-disinformation systems. The main reason is that deep learning algorithms fail to understand contextual, linguistic, symbolic, and behavioural nuances of human online discourse.

If the tragedy in Myanmar²⁴⁷ told us anything, it is that media giants such as Facebook need far more than tech fixes to prevent or even curb hate speech on their platforms. Expertise in languages, cultures, history and politics are of strategic importance in order to recognize and understand the impact of hate speech. Collaboration between AI systems flagging violence in deluges of data and experts doing the meta-interpretation could be an intermediary success. This form of collaborative intelligence between prevention expertise and private tech corporations would also ensure that any efforts to restrict hate speech do not stifle free expression and are aligned with international human rights law.

The few above examples of collaboration between preventive diplomacy and the AI-Cyber tech sector would have larger benefits to foster novel expertise in “AI for prevention.” For instance, this collaborative intelligence could help turn machine learning into a much better tool for preventive diplomacy. Two examples come to mind: using deep learning to recognize patterns for populations’ displacement in satellite images; or to monitor open data for crisis detection and even maybe crisis prevention.

In the end, fostering political preparedness against new forms of deception requires expertise in prevention to understand factors of social resilience. Increasingly, social resilience in one country could be impacted by the long-term unintended implications of technologies developed and controlled very far away by wealthy corporations.

A delicate but necessary alliance will need to be formed between the two worlds of tech companies and preventive diplomacy.

“ A delicate but necessary alliance will need to be formed between the two worlds of tech companies and preventive diplomacy.

A GLOBAL FORESIGHT OBSERVATORY FOR AI CONVERGENCE

The UN is in a unique position to initiate and guide an inclusive foresight effort on how AI and converging technologies can be designed and governed so that they meet the ethical needs of a globalizing world. Such a foresight observatory – focusing on converging technologies such as genome-editing, AI and affective computing, cyber and neurotechnologies – would provide the UN with decisive insights to promote peace and security, the respect of human rights, human well-being and responsible, sustainable development across different regions. A global foresight observatory would serve the prevention agenda in its cross-pillar, comprehensive nature.

The Observatory would equip the UN to: (i) articulate tailored, robust scenarios from which innovative strategies can emerge; (ii) map and involve key stakeholders that reflect the unique ways in which technologies are converging; and (iii) develop coherent, responsible approaches to leverage innovation and technology for prevention.

In simple terms, the Observatory would consist of:

A “FORESIGHT FUSION CELL”

This “fusion cell” would consist of a small team of technology and policy experts skilled in foresight methodologies. The team will provide an “interface”

that can integrate foresight into the strategic work of the Secretariat. In addition to producing original foresight and red-teaming on an in-house basis, the “fusion cell” could serve an aggregating function based on a hybrid organizational approach: 1) bridging existing foresight activities taking place in various entities across the UN system; 2) converting findings and insights into actionable policy recommendations that are integrated into the strategic planning of the UN.

A constellation of multistakeholder foresight groups, including key actors from private sector, academia, civil society and UN agencies, could be stood up to conduct foresight on a topical or regional basis. The “Foresight Fusion Cell” would initiate, support and oversee the work of these multistakeholder constellations.

The growing network of UN Innovation Labs could conduct inclusive foresight exercises in collaboration with democratized innovation ecosystems. Such a bottom-up approach would open the foresight process to teams of young technologists and innovators from start-ups and grassroots open innovation labs across the world.

The three above layers of actors in the Global Foresight Observatory could conduct a combination of regional and topical foresight, focusing on how AI convergence will impact societies from regional economic trends to global human security and prevention.

INCLUSIVENESS AND EQUALITY IN ANTICIPATION

The UN has a significant and unique role to play in ensuring that wider participation, from a critical mass of underserved groups and vulnerable States, is enabled and supported by strategic foresight. This promise of “equality in anticipation” at global scale has never really been achieved.

Never have we faced a form of technological convergence – whose design is in the hands of a few, and who are mostly born in societies of abundance. Yet, this set of technologies is powerful enough to shape multifaceted aspects of daily lives. This asymmetry of knowledge and power raises significant challenges for global cooperation.

Only a diversity of knowledge and experience will help foster diligent technical design, anticipate ethical failures, and minimize the risks of unintended harms. An example of how AI could be detrimental to underserved,

vulnerable populations is the use of biased algorithms by humanitarian organizations. “By embedding the logic of the powerful [with algorithms] to determine what happens to people at the periphery, humanitarian AI becomes a neo-colonial mechanism that acts in lieu of direct control.”²⁴⁸ In essence, biased algorithms will undermine the supposed neutrality of humanitarian organizations, leaving those whom they are trying to help potentially more vulnerable.

This is exactly why there is an urgent need to think and talk about digital cooperation. This strategic process should engage, not only the best and the brightest, but also the broadest range of engineers, academics, and civil society activists among others to figure out together what kind of world we want to live in and how AI can help us achieve the positive technical and ethical outcomes that will lead to that future.

Access to the knowledge and education required for anticipating the role of an emerging technology like AI, genome-editing or gene-drives is still a luxury. How to foster a global, inclusive and complex model of cooperation and governance in AI and converging technologies?

THE UN AND DEMOCRATIZED INNOVATION ECOSYSTEMS

As the UN builds its own innovation labs – for instance, UNTIL²⁴⁹ in Helsinki – the institution could also play a role in supporting and collaborating with the global networks of democratized innovation ecosystems that currently thrive and flourish in San Francisco, Shenzhen, Yucatan, Kumasi to Mumbai. In Kumasi's Lab, young innovators build drones that can deliver vaccines to isolated health centres.²⁵⁰ In Mexico's Lab, *Interspecifics*, inventors have designed algorithms that can recognize signals and interactions between micro-organisms inside biological cultures.²⁵¹ In Shenzhen's Open Innovation Lab,²⁵² mashups of self-made engineers, inventors and artists are working on projects ranging from drone swarming for diagnosing diseases on crops to precision algorithms for companion robots.

These democratized innovation ecosystems are ready to learn, adapt and harness AI and converging technologies for their own needs. They made a simple, winning bet: hacking the future.

Imagine a democratized innovation ecosystem where AI engineers perform mentorship for the next generation of algorithms' designers, where a young woman can be certified for new skills that add value to computation, where underserved users can experiment with new technological designs. Such a vision could be scaled up with the support of the UN and in collaboration with schools, universities and the mentorship of tech engineers and companies. For instance, the university of Helsinki together with a start-up is testing AI learning models tailored to large publics of the Finnish population.²⁵³ Such models could be studied and adapted to other national contexts and underserved needs.

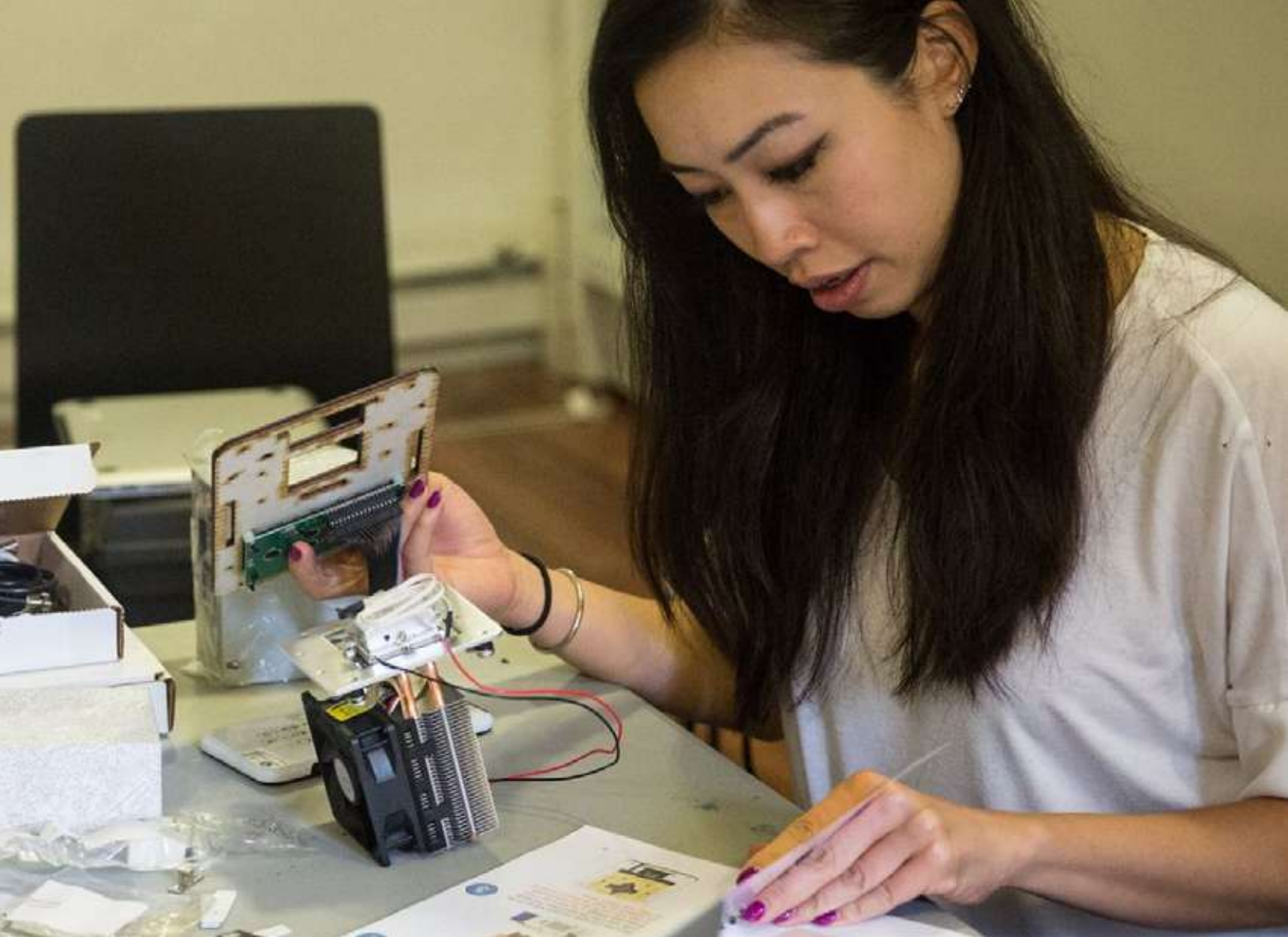
Democratized ecosystems, which build on open source approaches to new technologies, have developed unprecedented active local innovation hubs with incentives to better tailor tech applications to real social issues. States or regions can truly deploy, implement and flourish with new technologies only if they promote a normative environment that enables technological transfer to the grassroots level, delivering valuable products and services not only to the wealthy and powerful but also to social and economic peripheries.²⁵⁴



The global innovation ecosystem could be slowly transformed if community labs continue to export their democratized approach to AI and converging technologies. The future of AI convergence could be defined, invented and implemented by bottom-up networks of inventors and engineers, rather than large corporate platforms alone.

In this context, the UN and its Member States, the World Bank and international foundations, should discuss how to empower and oversee democratized innovation ecosystems, the “grassroots,” in their effort to design and deploy AI and converging technologies for solving local social problems.

To foster inclusiveness, which is a core value and priority defined by the High-level Panel on Digital Cooperation,



© David Kong, MIT Media Lab

Young entrepreneurs build new tools for biological analysis and health prevention in “hotzones,” bringing the lab to the city and the jungle

the UN could help support mentorship and certification programs, as well as fellowship exchanges between democratized ecosystems, private companies, start-ups and the UN Innovation Labs. The ultimate goal would be to foster inclusive forums of engagement in which to share lessons learned, exchanges of innovation and oversight practices for AI convergence.

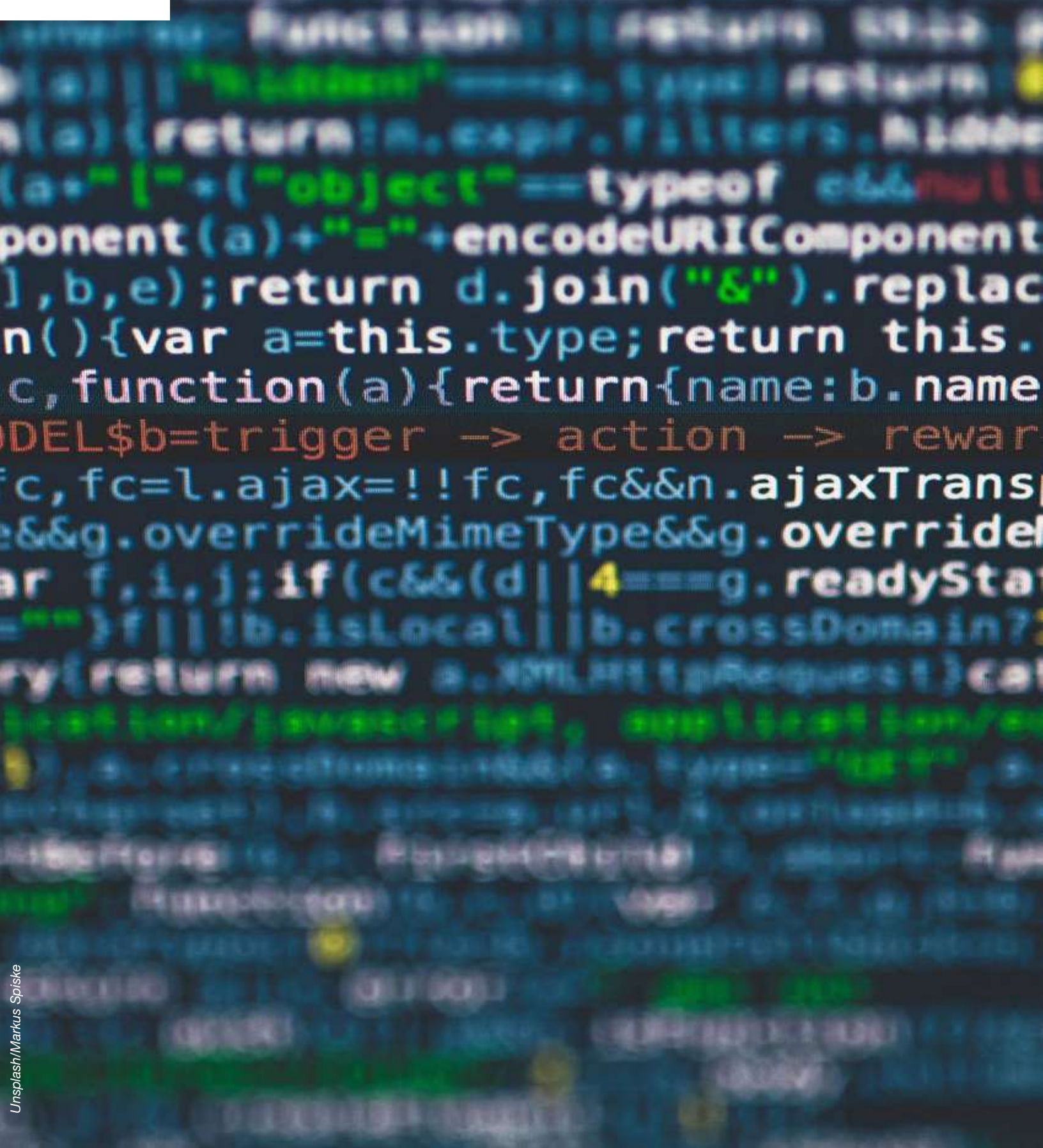
Just like any automated AI and biotech labs today, democratized innovation ecosystems could be targeted to produce malicious uses of AI and converging technologies. Yet, to avoid regulatory sanctions, most community labs have agreed to collaborate with security experts and develop codes of conduct and ethos of responsible innovation to be shared and implemented between peers. Such an effort gives us insights into the kind of responsible practices that could ratify knowledge-

sharing between entrepreneurs, engineers and security communities.

Adaptive regulatory frameworks could ensure safe and responsible citizen participation in AI and converging technologies. The way forward is to create a dialogue through which regulators can help citizens embed tailored governance mechanisms into their endeavours.

A more equal and inclusive world will not appear by chance. It will rest on the empowerment of those who can imagine globally beneficial intelligent designs. As shown in the Scenario “AI for Prevention,” there is a need for an inclusive, “cosmopolitan” conversation to anticipate and shape not only the risks but also AI’s promises.

VII



CONCLUSION

Today's technological era is generating a new geopolitics of hybrid emerging risks and vulnerabilities produced by the convergence of powerful dual-use technologies.

First, the combined convergence and decentralization of AI and other emerging technologies is not only disrupting war and conflicts, but politics, social cohesion and human well-being.

As unveiled in the scenario "The Deception Machine," the combination of AI, affective computing, and cyberdeception has turned micro-targeted disinformation campaigns into a global epidemic. Emotional manipulation, confirmation biases and sophisticated forgeries can swiftly mobilize millions and lead to "digital tragedies" like the one that unleashed "click by click" in Myanmar. Moreover, the learning potential of AI, penetrating through historical, social and personal data patterns, makes it possible to simulate "information" (facts, evidences, and narratives) from scratch. The survival of the global strategic, diplomatic and political intelligence system is at stake.

The "Internet of Bodies, Genomes and Minds" scenario explores how, in the near-future, an individual's identity will exist within a vast array of interconnected databases of faces, vital signs, genomes, emotions, and thoughts. The rise of personal algorithmic avatar may be the next target for hackers and cyber mercenaries, drastically increasing human insecurity at individual and population levels.

The Internet of Bodies also provides authoritarian regimes with new techniques to impose precision surveillance and implement biopower, managing, nudging and controlling the behaviours of entire populations. The capacity to mine and optimize ever more biological data from populations and ecosystems allows tech-leading nations and large technological platforms to subjugate other more fragile States into dynamics of cyber-colonization. The resulting effects are value exfiltration, technological dependence and increased distributive inequalities. As cutting-edge AI and genetic technologies bring new therapies in societies of abundance, other populations will have to be satisfied with more frugal innovation.

Within smart cities and automated (bio)-manufacturing supply chains, AI and cybertechnologies will offer both potential for cyberoffence and defence. New forms of terrorist violence have emerged with the weaponization of social media. Cyberattacks will increasingly target smart cities, and critical infrastructures as means of production become digitized. Combinations of algorithms, robots

and drones offer new intelligent and connected armies. New forms of invasive cyberwar and cognitive-emotional conflicts will have powerful implications for (bio-)physical, digital, socio-economic and political security.

In its dual-use nature, the current technological era also delivers opportunities for predictive reasoning within large-scale prevention operations. With enough technical and political preparedness, AI convergence could be harnessed to prevent human trafficking, reduce civilian casualties, anticipate and mediate conflicts. With the collaboration of in-depth human and political expertise, algorithms will help combat hate speech and investigate forgeries, election fraud and violent crimes. Some of the most promising uses of AI and converging technologies will materialize to optimize human health, preventing famine and epidemics, with tools in precision biotech and agriculture. If humans can learn how to anticipate and mitigate unintended consequences, future lives could be empowered on our burdened planet.

Governance actors, including States and the private sector, will need to adopt techniques of inclusive foresight to be resilient and adaptive enough in the face of hybrid security threats and emerging risks.

A second implication is that States in the Global South will be the first vulnerable targets in this new geopolitical landscape of virtual conflicts and cyber colonization.

Urgent support to develop foresight and responsible innovation is needed for those States that are struggling to compete, build and secure capacity in the development and deployment of AI and converging technologies. They risk becoming vulnerable links or "ungoverned cyberspaces" and may turn more susceptible to dynamics of data-predation and value exfiltration. Such vulnerabilities may fuel and intensify a fierce competition for supremacy in technological convergence rather than foster digital cooperation. Developing countries may lack the power, influence and foresight tools to shape responsible governance of converging technologies towards social benefits and away from political disruptions and weaponization. These fragile States could become a liability for a whole region.

Preventing such growing inequalities will rest on incentives coming from the multilateral system. States interested in fostering responsible AI convergence could enter into mechanisms of digital cooperation with countries in the Global South to partner around mutually beneficial transfers of data, talent, technologies and security practices.

“ States interested in fostering responsible AI convergence could enter into mechanisms of digital cooperation with countries in the Global South to partner around mutually beneficial transfers of data, talent, technologies and security practices.

III A third implication concerns the human side of global security risks. Technological risks could have a powerful, long-term and corrosive impact on human security and well-being.

In contexts where uncertainty about future job security is rising, when facing complex technological and social transformations, larger subsets of the world population may face fears of becoming useless and irrelevant classes. Global human psychological and emotional well-being could be receding. Underserved groups in societies may suffer from new forms of disempowerment. In turn, disempowerment and the lack of opportunity to participate in innovation will impact trust and social cohesion. Without proper citizen engagement and a means to contain the power of minority interests, technological development will proceed unhindered, for better or, quite possibly, worse. The sheer speed of change will assuredly result not just in people who surrender their lives to intelligent and connected machines, but in societal disruptions that will be difficult to mitigate.

It is time for governments to build a new social contract for the era of AI convergence. For instance, they could invest in measures that can reinforce networks of social cohesion and resilience. Such an effort would enhance the capacity of individuals, communities and systems to survive, learn, adapt and even transform in the face of political and socio-economic shocks and stresses.

In the end, the most important message from this report is the need to harness the full force of the UN Charter and the multilateral system to shape technological progress according to a diversity of values and life experiences.

This report recommends that the UN should form a Global Foresight Observatory for the convergence of AI and other emerging technologies. The UN is in a unique position to initiate and guide an inclusive foresight effort on how AI and converging technologies can be designed and governed to generate ethical and social benefits as well as greater shared prosperity in a globalized world.

Such a foresight observatory – focusing on converging technologies such as genome-editing, AI and affective computing, cyber and neurotechnologies – would provide the UN with decisive insights to promote peace and security, the respect of human rights, human well-being and responsible, sustainable development across different regions. A global foresight observatory would serve the prevention agenda in its cross-pillar, comprehensive nature.

It will rest on the UN to create the required incentives and mechanisms for involving Member States, the private sector, civil society and citizens in a foresight and responsible innovation journey. Yet, this report's governance diagnosis unveils three incentives. First, “vulnerable links” and “ungoverned spaces” in the development of AI and converging technologies create security weaknesses and potential liability for all. Second, the matrix of emerging risks (Section IV) unveils substantial threats that could compromise global human security in the near future. Finally, the prospect of a technological race that outpaces normative discussions is not conducive to responsible innovation.

This report provides a mapping of the global policy issues related to the prevention and peace and security agendas. They coalesce around: 1) anticipating political, socio-economic and human security risks; 2) promoting and protecting human rights; and 3) preventing distributive inequalities and their impact on human security.

This report also provides two use-cases of multi-stakeholder cooperation on foresight for technological and political preparedness: 1) technological preparedness for vulnerable States and the Global South; and 2) political preparedness, preventive diplomacy, AI and cybertechnologies.

Finally, one crucial role for the Global Foresight Observatory would be to foster inclusiveness and equality in anticipation for a diversity of populations and Member States. In this context, the report focuses on how the UN and its growing networks of innovation labs could collaborate and empower democratized ecosystems – citizen science and open innovation labs across the globe – that have engaged into efforts to promote public scientific and digital literacy. This Observatory could partner with the private sector by connecting with companies' labs specialized in examining safety, ethical and human rights' issues. In this “safe space” where foresight and responsible innovation plays a strategic role, research institutes and civil society's groups would provide further expertise.

The success of a Global Foresight Observatory – led by the UN – will depend on the humility of those who thought they could fully master AI, and the empowerment of others who can imagine globally beneficial intelligent designs. The Observatory will offer a chance to foster an inclusive, “cosmopolitan” effort to anticipate and shape AI's promises and risks.

ANNEX 1: MECHANISMS WITHIN THE GLOBAL FORESIGHT OBSERVATORY

GLOBAL FORESIGHT OBSERVATORY	
MECHANISMS	
MULTI-STAKEHOLDER COOPERATION ON TECHNOLOGICAL PREPAREDNESS	<p>Red-teaming exercises: Nascent efforts around foresight and predictive accountability could take increasingly agile forms and thrive on alliances between States such as North-South, and South-South collaborations. For instance, within a network of trust, entrepreneurs, policymakers and UN officials could use red-teaming exercises to anticipate potential vulnerabilities as well as safety and security best practices relevant to the convergence of AI, cyber and biotechnologies.</p>
	<p>Sparring exercises: Red-teaming could also turn into “sparring” exercises where corporate and government actors learn to collaborate and build a trusted space where to test AI and cybersystems to report fatal anomalies and discuss optimized defence capabilities. Such sparring exercises could be crucial for States that are currently vulnerable links in the new geopolitics of converging technologies.</p>
	<p>Safe-space for predictive accountability: States that are willing to lead in the responsible governance of AI, cyber and biotechnologies could develop ad-hoc mechanisms of inter-State cooperation, a “safe space” where to discuss predictive accountability, adherence to emerging norms, as well as legitimate and illegitimate behaviours when deploying dual-use technologies.</p>
	<p>Human rights impact assessment: To be even more successful, foresight efforts should include human security and human-rights impact assessment from the beginning, to assess converging technologies’ risks and benefits, and define guiding principles for ethical and safe design. Human rights impact assessments provide a step-by-step evaluation of the impact of practices or technologies deployed in a given context on aspects such as privacy, data agency and self-determination. Collaborations should take place between teams that specialize in foresight and predictive accountability with teams that systematically monitor for ethical breaches and violations of human rights in the field or for vulnerable populations. Building relationships between UN entities, UN Innovation labs and human rights’ labs inside AI companies would constitute an interesting collaborative network to foster principles of predictive accountability.</p>

GLOBAL FORESIGHT OBSERVATORY

MECHANISMS

MULTI-STAKEHOLDER COOPERATION ON POLITICAL PREPAREDNESS

Foresight to prevent deception in conflicts: One strategic option might to build collaborative teams of experts in preventive diplomacy, conflict resolution and AI-Cyber technologies. Such braintrust could conduct a combination of political and technological foresight to 1) analyse emerging tensions, anomalies and divisions in fractured societies; 2) anticipate subsequent scenarios of deception; and 3) plan for strategies to rapidly and effectively counter deception, in particular at crucial times such as election processes.

Metrics for Threats Prioritization: Experts in preventive diplomacy and AI-Cyber technologies could also work together on prioritizing what form of deception would constitute a minor interference, and what other form would turn into major threats to sow violence or contribute to the growing intractability of existing conflicts. Collaborative teams could do so by developing a metric of damages that acknowledges a range of target objectives and assess if these objectives have led to actual disruptions.

Countering disinformation: AI-driven tools exist to counter disinformation by filtering fake news, reinforcing known facts, detecting nefarious content, eliminating troll bots, and verifying the authenticity of audio and video content. Yet, for now, technical limits and failures abound in these counter-disinformation systems. The main reason is that deep learning algorithms fail to understand contextual, linguistic, symbolic, and behavioural nuances of human online discourse.

INCLUSIVENESS AND EQUALITY IN ANTICIPATION

Decentralized Foresight Exercises with UN Innovation Labs and Citizen Science/ Open Innovation Ecosystems; Opportunities for enhancing scientific and digital literacy through mentorship, certification programs and a culture of experimentation.

ANNEX 2: METHODOLOGY FOR SCENARIOS OF AI CONVERGENCE

Section III relies on the analysis of “scenarios” —plausible stories and vignettes about the future designed to tease out the present assumptions of technologists and policymakers in order to confront them with potential outcomes of their designs and decisions. Scenario planning does not attempt to predict what will happen, but identifies a set of examples of possible futures that provide a valuable point of reference when assessing current strategies or formulating new ones. A scenario-writing exercise requires insightful and rigorous thinking about plausible futures, which may be based on horizon-scanning and a systemic analysis of socio-technological drivers and trends. Drivers and trend impact analysis is a forecasting process that analyses the drivers, nature, potential impact, likelihood, and velocity of an emerging issue.

HORIZON SCANNING

Horizon scanning is a foresight method for dealing with the complexity of technological convergence. This report employs the methodology produced by the UK Government Office for Science, which is considered

as best practice when undertaking a detailed horizon scanning exercise. The first step identified the potential threats of AI convergence by gathering intelligence through desk research. Then, findings are organized into a matrix of pervasive hybrid threats to be refined through an analysis of drivers and trends impact.

DRIVERS & TRENDS IMPACT ANALYSIS

Sometimes referred to as cross-impact analysis, drivers and trend analysis exercises aim to gain more insight into the interactions between the threats identified during the horizon scanning phase.

Using the output produced during the horizon scanning exercise, we appraised the possibility of each threat occurring by means of a scale from 1 (very low) to 5 (highly probable). Engaging drivers and trend analysis exercises such as the table below allowed us to think through and construct plausible scenarios of AI and technological convergence.

PERVERSIVE HYBRID SECURITY THREATS		
DIGITAL SECURITY	Automated forgeries	
	Social engineering	
	Data Poisoning	
	Cyber Espionage	
	CyberTheft	
BIO-PHYSICAL SECURITY	Automated adversarial attacks	
	Damaging critical infrastructures	
	Corrupting biotech supply chains	
	Harnessing drone swarms and robots	
	Weaponizing additive manufacturing	

PERVASIVE HYBRID SECURITY THREATS		
POLITICAL SECURITY	Fake intelligence	
	Election interference	
	Hybrid influencing	
	Cognitive-emotional conflicts	
	Hactivism	
SOCIO-ECONOMIC SECURITY	Destruction of bio-intelligence and medical innovation	
	Data colonization and value exfiltration	
	Technological dependence	

		KEY
1	Low probability of occurrence	
2		
3		
4		
5	High probability of occurrence	

SCENARIO-BUILDING

Though complex systems – such as the convergence of AI, cyber and biotechnologies – often defy efforts to anticipate future outcomes and events, in part because of the sheer number of interactions, hybrid foresight methods that engage complex system theory help manage uncertainty.²⁵⁵ Our approach to drivers and trend analysis, as well as our scenario-writing, is framed by the research of Guston and Sarewitz²⁵⁶ on anticipatory governance and the work of Stirling and Smith²⁵⁷ on analyses aimed to understand and manage complexity. Exploring plausible future scenarios through the lens of complex systems theory provides approaches to prepare for the global human security challenges of tomorrow.

One of the most well-known foresight techniques, scenario-building, is the development of plausible narratives of the future that “explore how the world would change if certain trends were to strengthen or diminish, or various events were to occur.”²⁵⁸ Upon the completion of the drivers and trends impact analysis, we began to identify a set of possible futures which provide a valuable point of reference for evaluating current policy strategies and formulating new ones. Specifically, we identified four clusters of potential challenges posed by AI convergence: 1) The Deception Machine; 2) the Internet of Bodies, Genomes and Minds; 3) Our Smart, but vulnerable Cities; and 4) AI for Prevention.

REFERENCES

- 1 For more information on Affectiva's *Brain Power System* project for autism, see: McManus A. 2017. "Brain Power Launches Anticipated Emotion-Enabled Autism System for Smart Glasses." *Affectiva*, 7 November. <https://blog.affectiva.com/brain-power-launches-much-anticipated-emotion-enabled-autism-system-for-smart-glasses>. For more information on Affectiva's *SDK* project for suicide prevention, see: "SDK on the Spot: Suicide Prevention Project with Emotion Recognition." *Affectiva*, 14 August 2017. <https://blog.affectiva.com/sdk-on-the-spot-suicide-prevention-project-with-emotion-recognition> (*hyperlinks checked as of 29 March 2019*)
- 2 For more information on Peppy Pals, see: McManus A. 2017. "SDK on the Sport: Peppy Pals Education Apps Teaches Children SEL/EQ Skills." *Affectiva*, 23 March. <https://blog.affectiva.com/sdk-on-the-spot-peppy-pals-educational-apps-teaches-children-sel/eq-skills>
- 3 Parliament of Finland Committee for the Future. 2018. *Societal Transformation 2018-2037*, 20 Regimes, 100 Radical Technologies. p 3
- 4 UN Secretary General's High-Level Panel on Digital Cooperation. 2018. "United Nations Secretary-General Appoints High-Level Panel on Digital Cooperation;" 12 July. https://www.un.org/en/pdfs/HLP-on-Digital-Cooperation_Press-Release.pdf
- 5 UN Secretary General's Address to the General Assembly, New York, 25 September 2018. <https://www.un.org/sg/en/content/sg/statement/2018-09-25/secretary-generals-address-general-assembly-delivered-trilingual>
- 6 See: Press G. 2016. "A Very Short History of Artificial Intelligence." *Forbes*; 30 December. <https://www.forbes.com/sites/gilpress/2016/12/30/a-very-short-history-of-artificial-intelligence-ai/#5f91e9826fba>; and the first publication mention the term "artificial intelligence": McCarthy J., et al. 1955. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence." *Dartmouth College*; 31 August. <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>.
- 7 See Section 4 of: Callahan K. 2013. "The Impact of Allied Cryptographers on World War II: Cryptanalysis of the Japanese and German Cipher Machines." *Semantic Scholar*; 14 December: pp. 7-14. <https://pdfs.semanticscholar.org/c7cf/0c41932d61457dd943dc4dffca2c8bb92e95.pdf>.
- 8 Mathiseon S.A. 2012. "Bletchley Park: where government started computing." *The Guardian*; 15 May. <https://www.theguardian.com/government-computing-network/2012/may/15/bletchley-park-codebreaking-colossus-government-computing>
- 9 See: Brundage et al. 2018. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.* University of Oxford. February: p. 9. <https://arxiv.org/pdf/1802.07228.pdf>; and, Pauwels E. and Denton SW. 2018. "Searching for Privacy in the Internet of Bodies." *The Wilson Quarterly*; May. <https://wilsonquarterly.com/quarterly/living-with-artificial-intelligence/searching-for-privacy-in-the-internet-of-bodies/>
- 10 Definition presented by Dr. Greg Corrado, Principal Scientist and Director of Augmented Intelligence Research, Google, during an event titled "Governing Artificial Intelligence" co-hosted by the International Institute of Peace and the UN University held on June 22, 2018. For more information about this event and video recordings of Dr. Corrado's keynote speech, see: <https://www.ipinst.org/2018/06/governing-artificial-intelligence#15>
- 11 Brundage et al. 2018. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.* University of Oxford. February: p. 9. <https://arxiv.org/pdf/1802.07228.pdf>
- 12 For more information on this topic, please see Dr. Greg Corrado's keynote speech at the *Governing Artificial Intelligence* event co-hosted by the International Institute of Peace and the UN University held on June 22, 2018: <https://www.ipinst.org/2018/06/governing-artificial-intelligence#15>
- 13 Schwab K. 2016. "The Fourth Industrial Revolution: what it means, how to respond." *World Economic Forum*; 14 January. <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>
- 14 Pauwels E. 2018. "China pins its hopes on beating US in race for bio-intelligence supremacy." *South China Morning Post*; 27 November. <https://www.scmp.com/news/china/science/article/2174815/china-pins-its-hopes-beating-us-race-bio-intelligence-supremacy>
- 15 See: Hern A. 2018. "Cambridge Analytica: how did it turn clicks into votes?" *The Guardian*; 6 May. <https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie>.

- 16 Turgeman A. 2018. "Machine Learning and Behavioral Biometrics: A Match Made in Heaven." *Forbes*; 18 January. <https://www.forbes.com/sites/forbestechcouncil/2018/01/18/machine-learning-and-behavioral-biometrics-a-match-made-in-heaven/#3b5a038e3306>
- 17 McKay S. and Bali S. 2018. "Disruptive Technologies: Keeping an AI on Disease Surveillance." *Global Health Now*, Johns Hopkins Bloomberg School of Public Health; 25 April. <https://www.globalhealthnow.org/2018-04/disruptive-technologies-keeping-ai-disease-surveillance>
- 18 Maxmen A. 2018. "Machine learning spots natural selection at work in human genome." *Nature*; 1 November. <https://www.nature.com/articles/d41586-018-07225-z>
- 19 Foucault described the power of modern biology and medicine as an "explosion of numerous and diverse techniques for achieving the subjugation of bodies and the control of populations." See, also Wrona K. 2017. "Cyber warfare in smart environments." In *Defence Future Technologies: What We See on the Horizon*, Schweizerische Eidgenossenschaft: pp. 93-96. https://deftech.ch/What-We-See-On-The-Horizon/armasuisseW%2BT_Defence-Future-Technologies-What-We-See-On-The-Horizon-2017_HD.pdf
- 20 Foucault M. 1998. *The History of Sexuality*. Volume 1. London: Penguin, p. 140. Also see: Adams R. 2017. "Michel Foucault: Biopolitics and Biopower." *Critical Legal Thinking*; 10 May. <http://criticallegalthinking.com/2017/05/10/michel-foucault-biopolitics-biopower/>
- 21 Hutson M. 2018. "Missing data hinder replication of artificial intelligence studies." *Science*; 15 February. https://www.sciencemag.org/news/2018/02/missing-data-hinder-replication-artificial-intelligence-studies?r3f_986=https://www.google.com/
- 22 See The Odin's website, available at: <http://www.the-odin.com/diy-crispr-kit/>
- 23 Locklear M. 2017. "These Kids Are Learning CRISPR at Summer Camp." *Motherboard*; 27 July. https://motherboard.vice.com/en_us/article/kzavja/these-kids-are-learning-crispr-at-summer-camp
- 24 Curran D. 2018. "My terrifying deep dive into one of Russia's largest hacking forums." *The Guardian*; 24 July. <https://www.theguardian.com/commentisfree/2018/jul/24/darknet-dark-web-hacking-forum-internet-safety>
- 25 Robertson J., et al. 2016. "Darknet Mining and Game Theory for Enhanced Cyber Threat Intelligence." *Cyber Defense Review*; Fall. https://cyberdefensereview.army.mil/Portals/6/Documents/CDR%20Journal%20Articles/Darknet_Mining_and_Game_Theory_Robertson_et_al.pdf?ver=2018-08-01-090210-620
- 26 Brundage et al. 2018. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. University of Oxford. February. <https://arxiv.org/pdf/1802.07228.pdf>
- 27 Dunlap G. and Pauwels E. 2017. "The Intelligent and Connected Bio-Labs of the Future: Promise and Peril in the Fourth Industrial Revolution," *Wilson Center Policy Briefs*; October. https://www.wilsoncenter.org/sites/default/files/dunlap_pauwels_intelligent_connected_biolabs_of_future.pdf
- 28 Heinbockel WJ., Landerman ER., and Serrao GJ. 2017. "Supply Chain Attacks and Resiliency Mitigations." *MITRE*; October. <https://www.mitre.org/sites/default/files/publications/pr-18-0854-supply-chain-cyber-resiliency-mitigations.pdf>. Also see: King M. and Rosen J. 2018. "The Real Challenges of Artificial Intelligence: Automating Cyber Attacks." *The Woodrow Wilson Center for International Scholars*, CTRL Forward; 28 November. <https://www.wilsoncenter.org/blog-post/the-real-challenges-artificial-intelligence-automating-cyber-attacks>
- 29 Chan-Hon-Tong A. 2018. "An Algorithm for Generating Invisible Data Poisoning Using Adversarial Noise That Breaks Image Classification Deep Learning." *Machine Learning and Knowledge Extraction*, 1: pp. 192-204. https://www.researchgate.net/publication/328907709_An_Algorithm_for_Generating_Invisible_Data_Poisoning_Using_Adversarial_Noise_That_Breaks_Image_Classification_Deep_Learning
- 30 Arquilla J. and Ronfeldt D. 1993. "Cyberwar is Coming!" *Comparative Strategy*, Vol 12, N°2, p. 144. <https://www.tandfonline.com/doi/abs/10.1080/01495939308402915>
- 31 Villasenor J. 2019. "Artificial intelligence, deep fakes, and the uncertain future of truth." Brookings Institute; 14 February. <https://www.brookings.edu/blog/techtank/2019/02/14/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/>
- 32 Taddeo M. 2018. "How to Deter in Cyberspace." *Strategic Analysis June-July 2018*. University of Oxford. https://www.researchgate.net/publication/326033611_How_to_Deter_in_Cyberspace

- 33 Horowitz MC., et al. *Strategic Competition in an Era of Artificial Intelligence*. CNAS; July. https://s3.amazonaws.com/files.cnas.org/documents/CNAS-Strategic-Competition-in-an-Era-of-AI-July-2018_v2.pdf?mtime=20180716122000
- 34 Parkins S. 2018. "The battle for digital supremacy." *The Economist*; 15 March. <https://www.economist.com/leaders/2018/03/15/the-battle-for-digital-supremacy>. Also see: Turner J. 2018. "The new battlefield: the race to integrate cyber and electronic warfare." *Army Technology*; 6 June. <https://www.army-technology.com/features/new-battlefield-race-integrate-cyber-electronic-warfare/>
- 35 Wells II L. 2017. "Cognitive-Emotional Conflict: Adversary Will and Social Resilience." *Prism: National Defense University*; No. 2. https://ccco.ndu.edu/Portals/96/Documents/prism/prism_7-2/2-Cognitive-Emotional_Conflict.pdf?ver=2017-12-21-110638-877
- 36 Treverton GF., et al. Addressing Hybrid Threats. Swedish Defence University: Arkitektkopia AB, Bromma; 2018. <https://www.hybridcoe.fi/wp-content/uploads/2018/05/Treverton-AddressingHybridThreats.pdf>
- 37 Harari YN. 2018. "Why Technology Favors Tyranny." *The Atlantic*; October. <https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330/>
- 38 Arik SO, et al. 2018. "Neural Voice Cloning with a Few Examples." *32nd Conference on Neural Information Processing Systems: Montreal, Canada*. <https://arxiv.org/pdf/1802.06006.pdf>
- 39 Rothman J. 2018. "In the Age of AI, Is Seeing Still Believing?" *The New Yorker*; 12 November. <https://www.newyorker.com/magazine/2018/11/12/in-the-age-of-ai-is-seeing-still-believing>
- 40 O'Neill M. 2018. "What a 'virtual stuntman' dance and perform martial arts in stunning machine learning breakthrough that could make animations more realistic." *The Daily Mail*, 11 April. <https://www.dailymail.co.uk/sciencetech/article-5604585/Watch-virtual-stuntman-break-dance-perform-martial-arts-machine-learning-breakthrough.html>. Also see this YouTube video published by US Sciencetech: https://www.youtube.com/watch?v=jHP7N_-RYGE.
- 41 See: Harari YN. 2018. "Why Technology Favors Tyranny." *The Atlantic*; October. <https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330/>. Also see: A Joint Report by Access Now, Civil Liberties Union for Europe, and European Digital Rights. 2018. *Informing the 'Disinformation' Debate*. 18 October. https://edri.org/files/online_disinformation.pdf.
- 42 Freedom House. 2018. *The Rise of Digital Authoritarianism: Fake news, data collection and the challenge to democracy*. <https://freedomhouse.org/article/rise-digital-authoritarianism-fake-news-data-collection-and-challenge-democracy>
- 43 Bradshaw S. and Howard PN. 2018. *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation*. Computational Propaganda Research Project. <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf>
- 44 Human Rights Council. 2018. *Report of the independent international fact-finding mission on Myanmar*: p. 14. https://www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_64.pdf
- 45 Bergmann M. and Kenney C. 2017. "War by Other Means." *Center for American Progress*; 6 June. <https://www.americanprogress.org/issues/security/reports/2017/06/06/433345/war-by-other-means/>
- 46 Chesney R. and Citron DK. 2018. "Disinformation on Steroids." *Council on Foreign Relations*; 16 October. <https://www.cfr.org/report/deep-fake-disinformation-steroids>
- 47 Lomas N. 2017. "Lyrebird is a voice mimic for the fake news era." *TechCrunch*. <https://techcrunch.com/2017/04/25/lyrebird-is-a-voice-mimic-for-the-fake-news-era/>
- 48 See Lyrebird, Vocal Avatar API on their website, available at: <https://lyrebird.ai/vocal-avatar-api>
- 49 Mayer M. 2018. "Artificial Intelligence and Cyber Power from a Strategic Perspective." *IFS Insights*; April. https://brage.bibsys.no/xmlui/bitstream/handle/11250/2497514/IFS%20Insights_4_2018_Mayer.pdf
- 50 Applegate SD. 2013. "The Dawn of Kinetic Cyber." *5th International Conference on Cyber Conflict*; K. Podins, J. Stinissen, M. Maybaum (Eds.). <https://ieeexplore.ieee.org/document/6568376>
- 51 Villasenor J. 2018. "Artificial Intelligence and the future of geopolitics." *Brookings Institute*; 14 November. <https://www.brookings.edu/blog/techtank/2018/11/14/artificial-intelligence-and-the-future-of-geopolitics/>. Also see: Gavrilovic A. 2018. "[WebDebate summary] Algorithmic diplomacy: Better geopolitical analysis? Concerns about human rights?" *DiploFoundation*; 12 June. <https://www.diplomacy.edu/blog/webdebate-summary-algorithmic-diplomacy-better-geopolitical-analysis-concerns-about-human>

- 52 Wei H., et al. 2016. "Beyond the Words: Predicting User Personality from Heterogeneous Information." *Microsoft, Microsoft Research, Tsinghua University*; February. https://www.microsoft.com/en-us/research/wp-content/uploads/2017/01/WSDM_personality.pdf
- 53 Woolley SC and Howard PN. 2017. "Computational Propaganda Worldwide: Executive Summary." *Computational Propaganda Research Project*. <https://comprop.oii.ox.ac.uk/research/working-papers/computational-propaganda-worldwide-executive-summary/>
- 54 Woolley SC and Guilbeault DR. 2017. "Computational Propaganda in the United States of America: Manufacturing Consensus Online." *Computational Propaganda Research Project*; No. 5. <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-USA.pdf>
- 55 Shane S and Frenkel S. 2018. "Russian 2016 Influence Operation Targeted African-Americans on Social Media." *The New York Times*; 17 December. <https://www.nytimes.com/2018/12/17/us/politics/russia-2016-influence-campaign.html>
- 56 Iwanek K. 2018. "WhatsApp, Fake News? The Internet and Risks of Misinformation in India." *The Diplomat*; 30 July. <https://thediplomat.com/2018/07/whatsapp-fake-news-the-internet-and-risks-of-misinformation-in-india/>
- 57 Rai S. 2018. "How Facebook uses 'WhatsApp phones' to tap next emerging market." *The Economic Times*; 3 December. <https://economictimes.indiatimes.com/news/international/world-news/how-facebook-uses-whatsapp-phones-to-tap-next-emerging-market/articleshow/66914406.cms>
- 58 Robinson G., et al. 2019. "Bots and the ballot box: Is Facebook prepared for Asia's elections?" *Nikkei Asian Review*; 30 January. <https://asia.nikkei.com/Spotlight/Cover-Story/Bots-and-the-ballot-box-Is-Facebook-prepared-for-Asia-s-elections>
- 59 2018. "Soon WhatsApp Will Only Let You Forward 5 Messages At Once, As It Tries To Wrestle With Rumours." *India Times*; 20 July. <https://www.indiatimes.com/technology/news/whatsapp-will-only-let-you-forward-5-messages-in-india-as-a-way-to-restrict-spread-of-rumours-349697.html>
- 60 Elvy SA. 2017. "Paying for Privacy and the Personal Data Economy." *Columbia Law Review* 117(6): pp. 1369-1460. https://columbialawreview.org/wp-content/uploads/2017/10/Elvy_Paying-for-Privacy-and-the-Personal-Data-Economy.pdf
- 61 Granville K. 2018. "Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens." *The New York Times*; 19 March. <https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html>
- 62 For more information about Mindstrong Health's Digital Biomarkers, see Mindstrong's website available at: <https://mindstronghealth.com/science/>
- 63 Kleber S. 2018. "3 Ways AI is Getting More Emotional." *Harvard Business Review*; 31 July. <https://hbr.org/2018/07/3-ways-ai-is-getting-more-emotional>
- 64 Marwick A. and Lewis R. 2017. *Media Manipulation and Disinformation Online*. Data & Society. https://centerformediajustice.org/wp-content/uploads/2017/07/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf. Also see: Polonski V. 2017. "Artificial Intelligence Has the Power to Destroy or Save Democracy." *Council on Foreign Relations*; 7 August. <https://www.cfr.org/blog/artificial-intelligence-has-power-destroy-or-save-democracy>
- 65 Ienca M. and Andorno R. 2017. "Towards new human rights in the age of neuroscience and neurotechnology." *Life Sciences, Society and Policy* 13(5). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5447561/>
- 66 Ibid.
- 67 Bing C. and Schectman J. 2019. "Inside the UAE's Secret Hacking Team of American Mercenaries." *Reuters*; 30 January. <https://www.reuters.com/investigates/special-report/usa-spying-raven/>
- 68 Ibid.
- 69 Ward A. 2018. "ISIS's Use of Social Media Still Poses a Threat to Stability in the Middle East and Africa." *RAND*; December. <https://www.rand.org/blog/2018/12/isis-use-of-social-media-still-poses-a-threat-to-stability.html>. Also see: Alfifi M., et al. 2018. "Measuring the Impact of ISIS Social Media Strategy." http://snap.stanford.edu/mis2/files/MIS2_paper_23.pdf
- 70 Hern A. 2015. "Anonymous 'at war' with ISIS, hacktivist group confirms." *The Guardian*; 17 November. <https://www.theguardian.com/technology/2015/nov/17/anonymous-war-isis-hacktivist-group-confirms>

- 71 Dorell O. 2017. "Alleged Russian political meddling documented in 27 countries since 2004." *USA Today*; 7 September. <https://www.usatoday.com/story/news/world/2017/09/07/alleged-russian-political-meddling-documented-27-countries-since-2004/619056001/>
- 72 Wells II L. 2017. "Cognitive-Emotional Conflict: Adversary Will and Social Resilience." *Prism: National Defense University*; No. 2. https://cco.ndu.edu/Portals/96/Documents/prism/prism_7-2/2-Cognitive-Emotional_Conflict.pdf?ver=2017-12-21-110638-877
- 73 Lentzos F. 2018. "The Russian disinformation attack that poses a biological danger." *The Bulletin of the Atomic Scientists*; 19 November. <https://thebulletin.org/2018/11/the-russian-disinformation-attack-that-poses-a-biological-danger/>
- 74 Denton SW., Pauwels E., et al. 2018. "There's Nowhere to Hide: Artificial Intelligence and Privacy in the Fourth Industrial Revolution," *Wilson Center Policy Report* (March): pp. 10-11. https://www.wilsoncenter.org/sites/default/files/ai_and_privacy.pdf
- 75 For more information about the My Friend Cayla Doll, see their website available at: <https://www.myfriendcayla.com/meet-cayla-c8hw>
- 76 Nienaber M. 2017. "Germany bans talking doll Cayla, citing security risk." *Reuters*; 17 February. <https://www.reuters.com/article/us-germany-cyber-dolls/germany-bans-talking-doll-cayla-citing-security-risk-idUSKBN15W20Q>
- 77 Zetlin M. 2018. "AI is now analyzing candidates' facial expressions during video job interviews." *Inc.*; 28 February. <https://www.inc.com/minda-zetlin/ai-is-now-analyzing-candidates-facial-expressions-during-video-job-interviews.html>
- 78 Novet J. 2018. "Facebook is using A.I. to help predict when users may be suicidal." *CNBC*; 21 February. <https://www.cnn.com/2018/02/21/how-facebook-uses-ai-for-suicide-prevention.html>
- 79 Yuste R., et al. 2017. "Four ethical priorities for neurotechnologies and AI." *Nature*, 551(7679); 8 November. <https://www.nature.com/news/four-ethical-priorities-for-neurotechnologies-and-ai-1.22960>
- 80 Chen S. 2018. "'Forget the Facebook leak': China is mining data directly from workers' brains on an industrial scale." *South China Morning Post*; 29 April. <https://www.scmp.com/news/china/society/article/2143899/forget-facebook-leak-china-mining-data-directly-workers-brains>
- 81 Xiansheng H. 2017. "City brain and comprehensive urban cognition." *Alibaba Cloud blog*; 30 September. https://www.alibabacloud.com/blog/interview-with-idst-deputy-managing-director-hua-xiansheng-city-brain--comprehensive-urban-cognition_221544
- 82 Fan W., Khan N., and Lin L. 2017. "China snares innocent and guilty alike to build world's biggest DNA database." *The Wall Street Journal*; 26 December. <https://www.wsj.com/articles/china-snares-innocent-and-guilty-alike-to-build-worlds-biggest-dna-database-1514310353>
- 83 For more information about CloudWalk Technology, see the company's website at <http://www.cloudwalk.cn/>
- 84 For more information about this organization, see NITI Aayog's website: <http://niti.gov.in/content/overview>
- 85 Roy S. 2018. "Aadhaar: India's Flawed Biometric Database." *The Diplomat*; 6 March. <https://thediplomat.com/2018/03/aadhaar-indias-flawed-biometric-database/>. Also see the Aadhaar website: <https://uidai.gov.in/your-aadhaar/about-aadhaar.html>
- 86 Sherman J. and Morgus R. 2018. "Authoritarians Are Exporting Surveillance Tech, and With It Their Vision of the Internet." *Council on Foreign Relations*; 5 December. <https://www.cfr.org/blog/authoritarians-are-exporting-surveillance-tech-and-it-their-vision-internet>
- 87 Shahbaz A. 2018. "Fake news, data collection, and the challenge to democracy." *Freedom on the Net 2018-The Rise of Digital Authoritarianism*, Freedom House. <https://freedomhouse.org/report/freedom-net/freedom-net-2018/rise-digital-authoritarianism>
- 88 Xinhua. 2018. "Across China: Credit system provides support for social development." *XinhuaNet*; 3 July. http://www.xinhuanet.com/english/2018-07/03/c_137298542.htm. Raicu I. 2017. "China's Social Credit Score: An Ethics Case Study." *Markkula Center for Applied Ethics*; 16 May. <https://www.scu.edu/ethics/focus-areas/internet-ethics/resources/chinas-social-credit-score/>
- 89 Mozur P. 2018. "Inside China's Dystopian Dreams: AI, Shame and Lots of Cameras." *The New York Times*; 8 July. <https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html>
- 90 For more information about the Belt and Road Initiative, see the World Bank's overview available at: <https://www.worldbank.org/en/topic/regional-integration/brief/belt-and-road-initiative>

- 91 2018. "China talks of building a 'digital silk road.'" *The Economist*; 31 March. <https://www.economist.com/china/2018/05/31/china-talks-of-building-a-digital-silk-road>
- 92 Weber V. 2017. "Why China's Internet Censorship Model Will Prevail Over Russia's." *Council on Foreign Relations*; 12 December. <https://www.cfr.org/blog/why-chinas-internet-censorship-model-will-prevail-over-russias>
- 93 Foucault M. *The History of Sexuality*. An Introduction, Vol. 1. Trans. Robert Hurley. 1976. New York: Vintage Books, 1990: p. 138-139.
- 94 Adams R. 2017. "Michel Foucault: Biopolitics and Biopower." *Critical Legal Thinking*; 10 May. <http://criticallegalthinking.com/2017/05/10/michel-foucault-biopolitics-biopower/#fnref-22546-11>
- 95 Ibarra IA., et al. 2018. "Should we treat data as labor? Let's open up the discussion." *Brookings Institution*; 21 February. <https://www.brookings.edu/blog/techtank/2018/02/21/should-we-treat-data-as-labor-lets-open-up-the-discussion/>
- 96 Two companies, Kernel and Neuralink, aim to improve and expand human cognition. For more information about, Kernel, visit their website available at: <https://kernel.co/>. Neuralink, founded by Elon Musk, has a website, but provide no information on the company. Neuralink's website is available at: <https://www.neuralink.com/>
- 97 Ienca M. 2017. "Brain Machine Interfaces, Artificial Intelligence and Neurorights." *IEEE*. <https://brain.ieee.org/newsletter/2017-issue-3/brain-machine-interfaces-artificial-intelligence-neurorights/>
- 98 Coleman J. 2018. "Brain Computer Interfaces with Artificial Intelligence and Reinforcement Learning." *Medium*; 4 May. <https://medium.com/@askwhy/brain-computer-interface-with-artificial-intelligence-and-reinforcement-learning-9c94b0454209>
- 99 Potomac Institute for Policy Studies. 2014. *Neurotechnology: Enhancing the Human Brain and Reshaping Society*. <http://www.potomacinstitute.org/images/stories/publications/22JanNeurotechEnhancementReport.pdf>
- 100 Pandya J. 2019. "The Dual-Use Dilemma of Artificial Intelligence." *Forbes*; 7 January. <https://www.forbes.com/sites/cognitiveworld/2019/01/07/the-dual-use-dilemma-of-artificial-intelligence/#3a55f2b16cf0>
- 101 Chappell B. 2019. "Police Say Hacking Suspect, 20, Confessed to Posting German Leaders' Private Data." *NPR*; 8 January. <https://www.npr.org/2019/01/08/683272309/hacking-suspect-20-confesses-to-posting-private-data-of-hundreds-of-german-leade>
- 102 Khaira R. 2018. "Rs 500, 10 minutes, and you have access to billion Aadhaar details." *The Tribune*; 5 January. <https://www.tribuneindia.com/news/nation/rs-500-10-minutes-and-you-have-access-to-billion-aadhaar-details/523361.html>
- 103 Axel. 2018. "Enough Is Enough: 2018 Has Seen 600 Too Many Data Breaches." *Medium*; 24 July. <https://medium.com/@AxelUnlimited/enough-is-enough-2018-has-seen-600-too-many-data-breaches-9e3e5cd8ff78>
- 104 Turgeman A. 2018. "Demystifying Digital Identity: What It Is, What It Isn't and What It Can Be." *Forbes*; 15 November. <https://www.forbes.com/sites/forbestechcouncil/2018/11/15/demystifying-digital-identity-what-it-is-what-it-isnt-and-what-it-can-be/#4dd5d6ed2af1>
- 105 For example, Aadhaar is India's digital identity portal, which encompasses both physical and digital identification. For more information about Aadhaar, see: <https://uidai.gov.in/my-aadhaar/about-your-aadhaar.html>
- 106 Ananthalakshmi A. 2018. "Personal details of over 200,000 Malaysian organ donors leaked online: report." *Reuters*; 23 January. <https://www.reuters.com/article/us-malaysia-cybercrime/personal-details-of-over-200000-malaysian-organ-donors-leaked-online-report-idUSKBN1FD07B>
- 107 Privacy International. 2018. *The Humanitarian Metadata Problem: "Doing No Harm" in the Digital Era*. <https://privacyinternational.org/sites/default/files/2018-12/The%20Humanitarian%20Metadata%20Problem%20-%20Doing%20No%20Harm%20in%20the%20Digital%20Era.pdf>
- 108 Hessel A., et al. 2012. "Hacking the President's DNA." *The Atlantic*; November. <https://www.theatlantic.com/magazine/archive/2012/11/hacking-the-presidents-dna/309147/>
- 109 Osborne C. 2018. "Deeplocker: When malware turns artificial intelligence into a weapon." *ZDNet*; 8 August. <https://www.zdnet.com/article/deeplocker-when-malware-turns-artificial-intelligence-into-a-weapon/>. For a more in-depth explanation of DeepLocker, see: Kirat D., Jang J., and Stoecklin MP. 2018. "DeepLocker: Concealing Targeted Attacks with AI Locksmithing." *Black Hat USA 2018*. <https://i.blackhat.com/us-18/Thu-August-9/us-18-Kirat-DeepLocker-Concealing-Targeted-Attacks-with-AI-Locksmithing.pdf>

- 110 Pauwels E. 2018. "China pins its hopes on beating US in race for bio-intelligence supremacy." *South China Morning Post*; 26 November. <https://www.scmp.com/news/china/science/article/2174815/china-pins-its-hopes-beating-us-race-bio-intelligence-supremacy>. Also see: Robbins R. and Sheridan K. 2018. "A Chinese company unveils a powerful new sequencer. But can it compete in the U.S.?" *STAT*; 25 October. <https://www.statnews.com/2018/10/25/a-chinese-company-unveils-a-powerful-new-sequencer-but-can-it-compete-in-the-u-s/>
- 111 See: Human Rights Watch. 2017. "China: Minority Region Collects DNA from Millions." 17 December. <https://www.hrw.org/news/2017/12/13/china-minority-region-collects-dna-millions>. Also see: Specter M. 2014. "The Gene Factory." *The New Yorker*; 6 January. <https://www.newyorker.com/magazine/2014/01/06/the-gene-factory>
- 112 Yuan Sun I., Jayaram K., and Kassiri O. 2017. "Dance of the lions and dragons: How are African and China engaging, and how will the partnership evolve?" *McKinsey Institute*; June. <https://www.mckinsey.com/~media/McKinsey/Featured%20Insights/Middle%20East%20and%20Africa/The%20closest%20look%20yet%20at%20Chinese%20economic%20engagement%20in%20Africa/Dance-of-the-lions-and-dragons.aspx>
- 113 See Global Gene Corp's website, available at: <https://globalgenecorp.com>
- 114 Valente F. 2018. "Frost & Sullivan Experts Announce Global Smart Cities to Raise a Market of Over \$2 Trillion by 2025." *Frost & Sullivan*; 4 April. <https://www2.frost.com/news/press-releases/frost-sullivan-experts-announce-global-smart-cities-raise-market-over-2-trillion-2025/>
- 115 Hayden EC. 2015. "Genome researchers raise alarm over big data." *Nature*; 7 July. <https://www.nature.com/news/genome-researchers-raise-alarm-over-big-data-1.17912>
- 116 Cross R. 2018. "Ginkgo Bioworks launches biosecurity initiative." *C&en* 96(28); 8 July. <https://cen.acs.org/business/Ginkgo-Bioworks-launches-biosecurity-initiative/96/i28>
- 117 Ghassemi M., et al. 2018. "Opportunities in Machine Learning for Healthcare." *arXiv*; 5 June. <https://arxiv.org/pdf/1806.00388.pdf>
- 118 Servick K. 2017. "A boy with a rare disease gets new skin, thanks to gene-corrected stem cells." *Science Magazine*; 8 November. https://www.sciencemag.org/news/2017/11/boy-rare-disease-gets-new-skin-thanks-gene-corrected-stem-cells?r3f_986=https://www.google.com/
- 119 Associated Press. 2018. "Chinese researcher claims birth of first gene-editing babies – twin girls." *STAT*; 25 November. <https://www.statnews.com/2018/11/25/china-first-gene-edited-babies-born/>
- 120 See: Kirkpatrick J., et al. 2018. *Editing Biosecurity: Needs and Strategies for Governing Genome Editing*. George Mason University: Arlington, VA. <https://static1.squarespace.com/static/58eaf565197aea9c401e0497/t/5c05209d70a6ad22a0abb98b/1543839915478/Editing-Bio-+Report-Final-.pdf>. Also see the following working paper from the same project: Esvelt K. "Gene Drive Technology: The Thing to Fear is Fear Itself." *Editing Biosecurity Issue Brief No. 4*. Arlington, VA: George Mason University; December 2018. http://mars.gmu.edu/bitstream/handle/1920/11337/FINAL_11.21.18_Esvelt_IB.pdf?sequence=1&isAllowed=y
- 121 For more information about *Target Malaria*, see the website available at: <https://targetmalaria.org>
- 122 Collins JP. 2018. "Gene drives in our future: challenges of and opportunities for using a self-sustaining technology in pest and vector management." *BMC Proceedings* 12(Supp 8); 19 July. <https://www.ncbi.nlm.nih.gov/pubmed/30079101>
- 123 Callaway E. 2018. "UN treaty agrees to limit gene drives but rejects a moratorium." *Nature*; 29 November. <https://www.nature.com/articles/d41586-018-07600-w>
- 124 2019. "Global Gene Therapy for Rare Disease Market to Surpass US\$3.55 Billion by 2026." *Market Watch*; 30 January. <https://www.marketwatch.com/press-release/global-gene-therapy-for-rare-disease-market-to-surpass-us-355-billion-by-2026-2019-01-30>
- 125 In the 2018 World Economic Forum's Global Risk Perception Survey, the second most frequently cited risk triggered by technological convergence is the combination of cyberattacks with the manipulation and corruption of critical information infrastructure. *The Global Risks Report 2018: 13th Edition*. World Economic Forum, Geneva: Switzerland. http://www3.weforum.org/docs/WEF_GRR18_Report.pdf
- 126 Eubanks N. 2017. "The Trust Cost of Cybercrime for Business." *Forbes*; 13 July. <https://www.forbes.com/sites/theyec/2017/07/13/the-true-cost-of-cybercrime-for-businesses/#38953fec4947>

- 127 In the 2018 World Economic Forum's Global Risk Perception Survey, the second most frequently cited risk triggered by technological convergence is the combination of cyberattacks with the manipulation and corruption of critical information infrastructure. *The Global Risks Report 2018: 13th Edition*. World Economic Forum, Geneva: Switzerland. http://www3.weforum.org/docs/WEF_GRR18_Report.pdf
- 128 See DarkTrace's website: <https://www.darktrace.com/technology/>. "The Threat Visualizer is Darktrace's real-time, 3D threat notification interface. As well as displaying threat alerts, the Threat Visualizer provides a graphical overview of the day-to-day activity of your network(s), which is easy to use, and accessible for both security specialists and business executives. Using cutting-edge visualization techniques, the Threat Visualizer user interface automatically alerts analysts to significant incidents and threats within their environments, enabling analysts to proactively investigate specific areas of the infrastructure."
- 129 P&S Market Research. 2017. *Artificial Intelligence (AI) in Cyber Security Market*. <https://www.psmarketresearch.com/market-analysis/artificial-intelligence-in-cyber-security-market>
- 130 Dvorsky G. 2017. "Hackers Have Already Started to Weaponize Artificial Intelligence," *Gizmodo* 11 September. <https://gizmodo.com/hackers-have-already-started-to-weaponize-artificial-in-1797688425>
- 131 See: Brundage et al. 2018. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. University of Oxford. February: p. 44. <https://arxiv.org/pdf/1802.07228.pdf>; and, Press G. 2018. "The AI Cybersecurity Arms-Race: The Bad Guys Are Way Ahead." *Forbes*; 26 April. <https://www.forbes.com/sites/gilpress/2018/04/26/the-ai-cybersecurity-arms-race-the-bad-guys-are-way-ahead/#3fc3ac28148e>
- 132 Wagner D. 2016. "Expert Commentary: The Growing Threat of Cyber-Attacks on Critical Infrastructure." *IRMI*; 2016. <https://www.irmi.com/articles/expert-commentary/cyber-attack-critical-infrastructure>
- 133 E-ISAC. 2016. *Analysis of the Cyber Attack on the Ukrainian Power Grid*, Defense Use Case; 18 March. https://www.nerc.com/pa/CI/ESISAC/Documents/E-ISAC_SANS_Ukraine_DUC_18Mar2016.pdf
- 134 Ibid: p. 4-5.
- 135 United Kingdom Parliament. 2018. Cyber Security of the UK's *Critical National Infrastructure: Third Report of Session 2017-19*: p. 3. <https://publications.parliament.uk/pa/jt201719/jtselect/jtnatsec/1708/1708.pdf>
- 136 Signe L. and Signe K. 2018. "Global cybercrimes and weak cybersecurity threaten businesses in Africa." Brookings Institute; 20 May. <https://www.brookings.edu/blog/africa-in-focus/2018/05/30/global-cybercrimes-and-weak-cybersecurity-threaten-businesses-in-africa/>
- 137 2019. "Heathrow airport drone sighting halts departures." *BBC News*; 8 January. <https://www.bbc.com/news/uk-46803713>. Also see: 2019. "Gatwick and Heathrow buying anti-drone equipment." *BBC News*; 4 January. <https://www.bbc.com/news/uk-46754489>
- 138 Fox D. 2016. "What sparked the Cambrian explosion?" *Nature*; 16 February. <https://www.nature.com/news/what-sparked-the-cambrian-explosion-1.19379>
- 139 Firestone K. 2018. "2018 Manufacturing Outlook and 3D-Printing's Impact." *manufacturing.net*; 16 January. <https://www.manufacturing.net/blog/2018/01/2018-manufacturing-outlook-and-3d-printings-impact>
- 140 See Shenzhen Open Innovation Lab's website, available at: <https://www.szoil.org>
- 141 See Accenture's website, available at: <https://www.accenture.com/us-en/service-open-to-disrupt>
- 142 For more information about Kumasi Hive, see: Ahiataku S. 2018. "Engaging Young Innovators in Kumasi with Innovation Clinic." *Global Lab Network*; 3 March. <https://glabghana.wordpress.com>
- 143 Watson B. 2017. "The Drones of ISIS." *Defense One*; 12 January. <https://www.defenseone.com/technology/2017/01/drones-isis/134542/>
- 144 Dunlap G. and Pauwels E. 2017. "The Intelligent and Connected Bio-Labs of the Future: Promise and Peril in the Fourth Industrial Revolution," *Wilson Center Policy Briefs* (October): pp. 4-6. https://www.wilsoncenter.org/sites/default/files/dunlap_pauwels_intelligent_connected_biolabs_of_future.pdf; Also see: Gwynne P. and Heebner G. 2018. "Laboratory Technology Trends: Lab Automation and Robotics, the Brave New World of 24/7 Research." *Science*; 18 January. <http://www.sciencemag.org/site/products/robotfinal.xhtml>

- 145 Refer to Dr. Seán ÓhÉigeartaigh's panel discussion at the *Governing Artificial Intelligence* expert meeting held on June 22, 2018 at the UN University. See: Session I: Does the AI race threaten international peace and security? "[Learning to Live with Artificial Intelligence: A Virtuous Circle or a Vicious One?](https://www.ipinst.org/2018/06/governing-artificial-intelligence#21)" *International Peace Institute*; 22 June 2018. <https://www.ipinst.org/2018/06/governing-artificial-intelligence#21>
- 146 Wrona K. 2017. "Cyber warfare in smart environments." In *Defence Future Technologies: What We See on the Horizon*, Schweizerische Eidgenossenschaft: pp. 93-96. https://deftech.ch/What-We-See-On-The-Horizon/armasuisseW%2BT_Defence-Future-Technologies-What-We-See-On-The-Horizon-2017_HD.pdf
- 147 Scudellari M. 2017. "Software Startups Aim to Automate Bio Labs." *IEEE Spectrum*; 27 April. <https://spectrum.ieee.org/the-human-os/biomedical/devices/software-startups-aim-to-automate-bio-labs>
- 148 See Transcryptic's website, available at: <https://www.transcriptic.com>
- 149 See Emerald Cloud Lab's website, available at: <https://www.emeraldcloudlab.com>
- 150 Dunlap G. and Pauwels E. 2017. "The Intelligent and Connected Bio-Labs of the Future: Promise and Peril in the Fourth Industrial Revolution," *Wilson Center Policy Briefs* (October): p. 6. https://www.wilsoncenter.org/sites/default/files/dunlap_pauwels_intelligent_connected_bioblabs_of_future.pdf
- 151 Kirkpatrick J., et al. 2018. *Editing Biosecurity: Needs and Strategies for Governing Genome Editing*. George Mason University: Arlington, VA. <https://static1.squarespace.com/static/58eaf565197aea9c401e0497/t/5c05209d70a6ad22a0abb98b/1543839915478/Editing-Bio-+Report-Final-.pdf>
- 152 Yang YT., et al. 2016. "Design and Use of Low Cost, Automated Morbidostat Evolution of Bacteria Under Antibiotic Drug Selection." *Journal of Visualized Experiments*; July. https://www.researchgate.net/publication/303893986_Design_and_Use_of_a_Low_Cost_Automated_Morbidostat_for_Adaptive_Evolution_of_Bacteria_Under_Antibiotic_Drug_Selection
- 153 Global Biodefense Staff. 2017. "Countering the Threat of UAV-Delivered Chemical and Biological Weapons." *Global Biodefense*; 7 November. <https://globalbiodefense.com/2017/11/07/counter-weaponized-drones-us-needs-joint-public-private-solutions/>
- 154 Heinbockel WJ., Landerman ER., and Serrao GJ. 2017. "Supply Chain Attacks and Resiliency Mitigations." *MITRE*; October. <https://www.mitre.org/sites/default/files/publications/pr-18-0854-supply-chain-cyber-resiliency-mitigations.pdf>
- 155 Chan-Hon-Tong A. 2018. "An Algorithm for Generating Invisible Data Poisoning Using Adversarial Noise That Breaks Image Classification Deep Learning." *Machine Learning and Knowledge Extraction*, 1: pp. 192-204. https://www.researchgate.net/publication/328907709_An_Algorithm_for_Generating_Invisible_Data_Poisoning_Using_Adversarial_Noise_That_Breaks_Image_Classification_Deep_Learning
- 156 Special thanks to Robert Kirkpatrick, Director of UN Global Pulse, for his constant inspiration when it comes to harnessing AI for prevention.
- 157 See: Danaher J. 2017. "Is Technology Value-Neutral? New Technologies and Collective Action." *Institute for Ethics and Emerging Technologies*; 18 December. <https://ieet.org/index.php/IEET2/more/Danaher20171218>. For information regarding how to design value-sensitive AI, see: Umbrello S. 2019. « Beneficial Artificial Intelligence Coordination by Means of a Value Sensitive Design Approach." *Big Data and Cognitive Computing*; 3(5). <https://philarchive.org/archive/UMBBAC>
- 158 UN. 2017. "Looking to the future, UN to consider how artificial intelligence could help achieve economic growth and reduce inequalities." <https://www.un.org/sustainabledevelopment/blog/2017/10/looking-to-future-un-to-consider-how-artificial-intelligence-could-help-achieve-economic-growth-and-reduce-inequalities/2017/>. For more information about the UN Sustainable Development Goals (SDGs), see the UN website, available at: <https://sustainabledevelopment.un.org/?menu=1300>
- 159 See: The World Bank. 2018. "Famine Action Mechanism (FAM).", available at: <http://www.worldbank.org/en/programs/famine-early-action-mechanism>
- 160 See: World Economic Forum. 2018. *Harnessing Artificial Intelligence for the Earth*; January. http://www3.weforum.org/docs/Harnessing_Artificial_Intelligence_for_the_Earth_report_2018.pdf. Also see: Cho R. 2018. "Artificial Intelligence – A Game Changer for Climate Change and the Environment." *State of the Planet*. Earth Institute, Columbia University; 5 June. <https://blogs.ei.columbia.edu/2018/06/05/artificial-intelligence-climate-environment/>
- 161 Sablich E. 2019. "How artificial intelligence will affect the future of energy and climate." *Brookings Institute*; 10 January. <https://www.brookings.edu/research/how-artificial-intelligence-will-affect-the-future-of-energy-and-climate/>
- 162 Blumenthal A. 2108. "Artificial Intelligence to fight the spread of infectious diseases." *Phys.org*; 20 February. <https://phys.org/news/2018-02-artificial-intelligence-infectious-diseases.html#jCp>

- 163 See: Raju M., et al. « Development of a Deep Learning Algorithm for Automatic Diagnosis of Diabetic Retinopathy." *Studies in Health Technology Informatics*; 245: pp. 559-563. <http://ebooks.iospress.nl/publication/48210>. Also see: Comstock J. 2016. "Google researchers use deep learning to detect diabetic retinopathy with upwards of 90 percent accuracy." *Mobile Health News*; 29 November. <https://www.mobihealthnews.com/content/google-researchers-use-deep-learning-detect-diabetic-retinopathy-upwards-90-percent-accuracy>
- 164 Wahl B., et al. 2018. "Artificial Intelligence (AI) and global health: how can AI contribute to health in resource-poor settings?" *BMJ Global Health*; 27 July. <https://gh.bmj.com/content/bmjgh/3/4/e000798.full.pdf>
- 165 Murgia M. 2017. "How smartphones are transforming healthcare." *Financial Times*; 12 January. <https://www.ft.com/content/1efb95ba-d852-11e6-944b-e7eb37a6aa8e>
- 166 Brunskill E. and Lesh D. 2010. "Routing for Rural Health: Optimizing Community Health Worker Visit Schedules." *AAAI*. <http://www.aaai.org/ocs/index.php/SSS/SSS10/paper/download/1139/1369>
- 167 UNICEF. 2017. *Annual Report, Tanzania*. https://www.unicef.org/about/annualreport/files/Tanzania_2017_COAR.pdf
- 168 Reis MAM., et al. 2004 "Fuzzy expert system in the prediction of neonatal resuscitation." *Brazilian Journal of Medical and Biological Research*; 37(5): pp. 755-764. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-879X2004000500018&lng=en&tlng=en
- 169 Danks D. and London AJ. 2017. "Algorithmic Bias in Autonomous Systems." *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. <https://www.cmu.edu/dietrich/philosophy/docs/london/IJCAI17-AlgorithmicBias-Distrib.pdf>
- 170 Tanner A. 2017. *Our Bodies, Our Data*. Beacon Press; 10 January. <https://www.penguinrandomhouse.com/books/539847/our-bodies-our-data-by-adam-tanner/9780807033340/>
- 171 Prakash A. 2017. "European Drones Monitor Migrants as Policies Firm Up." *Robotics Business Review*; 17 May. <https://www.roboticsbusinessreview.com/unmanned/european-drones-monitor-migrants-policies-firm/>
- 172 Dearden L. 2016. "Drones deployed to keep migrants and refugees out of Channel Tunnel amid warnings of post-Brexit surge." *The Independent*; 28 June. <https://www.independent.co.uk/news/world/europe/drones-deployed-channel-tunnel-eurotunnel-stop-migrants-refugees-warnings-post-brexit-surge-a7107186.html>
- 173 Ibid.
- 174 Lomonaco V., et al. 2018. "Intelligent Drone Swarm for Search and Rescue Operations at Sea." *Workshop on AI for Good, Neural Information Processing Systems*; December. <https://hal.archives-ouvertes.fr/hal-01951515/document>
- 175 Velasquez-Manoff M. 2016. "We Really Can Stop Poaching. And It Starts with Drones." *Wired*; 6 July. <https://www.wired.com/2016/07/we-really-can-stop-poaching-and-it-starts-with-drones/>
- 176 George A. 2013. "Forget roads – drones are the future of goods transport." *New Scientist*; 4 September. <https://www.newscientist.com/article/mg21929334-900-forget-roads-drones-are-the-future-of-goods-transport/>
- 177 Lachow I. 2017. "The upside and downside of swarming drones." *Bulletin of Atomic Scientists* 73(2): pp. 96-101. <https://www.tandfonline.com/doi/pdf/10.1080/00963402.2017.1290879>
- 178 Gold D. 2018. "Saving Lives with Tech Amid Syria's Endless Civil War." *Wired*; 16 August. <https://www.wired.com/story/syria-civil-war-hala-sentry/>
- 179 See Hala Systems' website, available at: <https://halasystems.com>
- 180 Loveluck L. 2018. "The secret app that gives Syrian civilians minutes to escape airstrikes." *The Washington Post*; 18 August. https://www.washingtonpost.com/world/the-secret-app-that-gives-syrian-civilians-minutes-to-escape-airstrikes/2018/08/17/e91e66be-9cbf-11e8-b55e-5002300ef004_story.html?noredirect=on&utm_term=.ba902d1fbbfb
- 181 See Hala Systems' reader on the Sentry Civilian Early Warning System, available at: <https://drive.google.com/file/d/1fGDOV-r512OTnZef-NJWqUlar4gaM9P4/view>
- 182 See Hala Systems' reader on the Insight Data Analysis Portal, available at: https://drive.google.com/file/d/17PTjBKzf31Q7XHRmn0BRUQruYss_UyoJ/view
- 183 See this video testimony from a first responder Sentry user in Syria, available at: <https://drive.google.com/file/d/1rBSb10CWvXOhu-BI5RfooT1oY5Y08TDA/view>

- 184 See Hala Systems Inc. 2019. https://drive.google.com/file/d/1KDGJnt8_wC_ymIK_GfjG3LfTogD22nQA/view
- 185 See Hala Systems Inc. 2019. https://drive.google.com/file/d/1KDGJnt8_wC_ymIK_GfjG3LfTogD22nQA/view
- 186 Hatmaker T. 2018. "DARPA is funding new tech that can identify manipulated videos and 'deepfakes'." *TechCrunch*. <https://techcrunch.com/2018/04/30/deepfakes-fake-videos-darpa-sri-international-media-forensics/>
- 187 Bolles RJ., et al. 2018. "Spotting Audio-Visual Inconsistencies (SAVI) in Manipulated Video." *University of Amsterdam*. <https://staff.fnwi.uva.nl/t.e.j.mensink/publications/bolles17cvprwmf.pdf>
- 188 United Kingdom Home Office. 2018. "New technology revealed to help fight terrorist content online." 13 February. <https://www.gov.uk/government/news/new-technology-revealed-to-help-fight-terrorist-content-online>
- 189 BBC News. 2018. "Election interference to be sniffed out by early-alert system." 17 July. <https://www.bbc.com/news/technology-44820416>
- 190 Perry WL., et al. "Predictive Policing: The Rose of Crime forecasting in Law Enforcement Operations." *RAND: Safety and Justice Program*. Supported by the National Institute of Justice. https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.sum.pdf
- 191 Habtemariam E., et al. 2007. "Artificial Intelligence for Conflict Management." *DBLP*. https://www.researchgate.net/publication/220487240_Artificial_Intelligence_for_Conflict_Management
- 192 UN Global Pulse. 2017. *Using Machine Learning to Analyse Radio Content in Uganda*. http://unglobalpulse.org/sites/default/files/Radio%20Analysis%20Report_Preview%20%283%29.pdf
- 193 Olteanu A., et al. 2018. "The Effect of Extremist Violence on Hateful Speech Online." *Association for the Advancement of Artificial Intelligence*. http://chato.cl/papers/olteanu_castillo_boy_varshney_2018_hate_speech_online_islamic_terrorism_islamophobia.pdf
- 194 Picard RW. "Towards Machines with Emotional Intelligence." *MIT Media Library*. <https://affect.media.mit.edu/pdfs/07.picard-El-chapter.pdf>
- 195 Stecklow S. 2018. "Why Facebook is losing the war on hate speech in Myanmar." *Reuters*; 15 August. <https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/>
- 196 Tarantola A. 2017. "How artificial intelligence can be corrupted to repress free speech." *Engadget*; 20 January. <https://www.engadget.com/2017/01/20/artificial-intelligence-can-repress-free-speech/>
- 197 Human Rights Council. 2018. *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. 38th Session, UN General Assembly; 6 April. <https://www.ohchr.org/en/issues/freedomofopinion/pages/opinionindex.aspx>
- 198 Human Rights Council. 2018. *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. 38th Session, UN General Assembly; 6 April: p. 14. <https://www.ohchr.org/en/issues/freedomofopinion/pages/opinionindex.aspx>
- 199 Johns Hopkins University, School of Advanced International Studies. *Peacemaking*. <https://www.sais-jhu.edu/content/peacemaking>
- 200 Olsher DJ. 2015. "New Artificial Intelligence Tools for Deep Conflict Resolution and Humanitarian Response." *Procedia Engineering* 107: pp. 282-292. https://ac.els-cdn.com/S187770581501036X/1-s2.0-S187770581501036X-main.pdf?_tid=c0dc5171-8cdb-4fce-b6b1-64285ce658b5&acdnat=1549733620_ab62abadf20e185648d1ebc0c1d86b2a
- 201 *Report on Digital Technologies and Mediation in Armed Conflict*. Referencing: D. Lanz and A.Eleiba (2018). The Good, the bad and the Ugly: Social Media and Peace Mediation. Swisspeace. <https://www.swisspeace.ch/assets/publications/downloads/Policy-Briefs/aa3fc8830f/Social-Media-and-Peace-Mediation-Policy-Brief-12-2018.pdf>
- 202 Centre for Humanitarian Dialogue. 2018. *Digital Technologies and Mediation in Armed Conflict*: pp. 18-19. <https://www.hdcentre.org/wp-content/uploads/2018/12/MPS-8-Peacemaking-and-New-Technologies.pdf>
- 203 Centre for Humanitarian Dialogue. 2018. *Digital Technologies and Mediation in Armed Conflict*: p. 19. <https://www.hdcentre.org/wp-content/uploads/2018/12/MPS-8-Peacemaking-and-New-Technologies.pdf>

- 204 See Liveuamap's website, available at: <https://usa.liveuamap.com/>
- 205 *The Global Risks Report 2019: 14th Edition*. World Economic Forum, Geneva: Switzerland. http://www3.weforum.org/docs/WEF_Global_Risks_Report_2019.pdf
- 206 World Health Organization. 2018. *Depression and Other Common Mental Disorders: Global Health Estimates*. <https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf?sequence=1>
- 207 Moon M. 2018. "Finnish university's online AI course is open to everyone." *endgadget*; 20 May. <https://www.engadget.com/2018/05/20/finland-helsinki-university-ai-course/>
- 208 UNCTAD. 2018. *Technology and Innovation Report*. https://unctad.org/en/PublicationsLibrary/tir2018_en.pdf
- 209 McKinsey Global Institute. 2017. *Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation*; December. <https://www.mckinsey.com/~media/mckinsey/featured%20insights/future%20of%20organizations/what%20the%20future%20of%20work%20will%20mean%20for%20jobs%20skills%20and%20wages/mgi-jobs-lost-jobs-gained-report-december-6-2017.ashx>
- 210 Reichental A. 2018. "The Future of #DPrinting." *Forbes*; 23 January. <https://www.forbes.com/sites/forbestechcouncil/2018/01/23/the-future-of-3-d-printing/#74f0be865f65>
- 211 Rotman D. 2013. "How Technology is Destroying Jobs." *MIT Technology Review*; 12 June. <https://www.technologyreview.com/s/515926/how-technology-is-destroying-jobs/>
- 212 Harwell D. 2018. "A down day on the markets? Analysts say blame the machines." *The Washington Post*; 6 February. https://www.washingtonpost.com/news/the-switch/wp/2018/02/06/algorithms-just-made-a-couple-crazy-trading-days-that-much-crazier/?utm_term=.0e3be1e7419b
- 213 Gotalakrishnan R. and Mogato M. 2016. "Bangladesh Bank official's computer was hacked to carry out \$81 million heist: diplomat." *Reuters*; 19 May. <https://www.reuters.com/article/us-cyber-heist-philippines/bangladesh-bank-officials-computer-was-hacked-to-carry-out-81-million-heist-diplomat-idUSKCN0YA0CH>
- 214 Ananthalakshmi A. 2018. "Personal details of over 200,000 Malaysian organ donors leaked online: report." *Reuters*; 23 January. <https://www.reuters.com/article/us-malaysia-cybercrime/personal-details-of-over-200000-malaysian-organ-donors-leaked-online-report-idUSKBN1FD07B>
- 215 The attack ad was posted to Facebook on 10 July 2017 by a page named "The Real Raila". <https://www.facebook.com/TheRealRaila/videos/vb.246999515773477/291055968034498/?type=2&theater>
- 216 Nyabola N. 2017. "Texts, Lies, and Videotape." *Foreign Policy*; 1 August. <https://foreignpolicy.com/2017/08/01/texts-lies-and-videotape-kenya-election-fake-news/>
- 217 KNA. 2015. "Daily Facebook Users in Kenya reach 2.2 million." *Nairobi News*; 12 September. <https://nairobionews.nation.co.ke/news/daily-facebook-users-in-kenya-reach-2-2-million/>. Also see: Jäntti P. 2015. "The Usage of Social Media Among Young Adults Living in Nairobi." A Master's Thesis. University of Jyväskylä; March. <https://jyx.jyu.fi/bitstream/handle/123456789/45684/1/URN%3ANBN%3Afi%3Aju-201504211638.pdf>
- 218 Freedom House. 2018. *Freedom on the Net 2018*. Kenya Report. <https://freedomhouse.org/report/freedom-net/2018/kenya>
- 219 Ibid.
- 220 Howard PN., et al. 2018. *The IRA, Social Media and Political Polarization in the United States, 2012-2018*. Computational Propaganda Research Project. <https://comprop.oii.ox.ac.uk/research/ira-political-polarization/>
- 221 House of Commons: Digital, Culture, Media and Sport Committee. 2018. *Disinformation and 'fake news': Interim Report Government Response to the Committee's Fifth Report of Session 2017-19*. <https://publications.parliament.uk/pa/cm201719/cmselect/cmcmds/1630/1630.pdf>
- 222 Marczak B., et al. 2018. "Hide and Seek: Tracking NSO Group's Pegasus Spyware to Operations in 45 Countries." *The Citizens Lab*; 18 September. <https://citizenlab.ca/2018/09/hide-and-peek-tracking-nso-groups-pegasus-spyware-to-operations-in-45-countries/>. Also see: Freedom House. 2018. *Freedom on the Net 2018*. Kenya Report. <https://freedomhouse.org/report/freedom-net/2018/kenya>

- 223 Farr C. 2018. "Mark Zuckerberg is selling up to USD 13 billion of Facebook stock to fund an ambitious project to end disease: Here's an early look inside." *CNBC*; 15 September. <https://www.cnn.com/2018/09/14/chan-zuckerberg-initiative-what-is-it-doing-so-far.html>
- 224 For more information about the Chan-Zuckerberg Initiative, see their website, available at: <https://chanzuckerberg.com/>
- 225 For more information about WuXiNextCode, see their website, available at: <https://www.wuxinextcode.com/genomic-insights/google-for-genomic-data/>
- 226 Dutton T., et al. *Building an AI World: Report on National and Regional AIU Strategies*. CIFAR. https://www.cifar.ca/docs/default-source/ai-society/buildinganaiworld_eng.pdf?sfvrsn=fb18d129_4
- 227 Noyce RS., Lederman S., and Evans DH. 2018. "Construction of an infectious horsepox virus vaccine from chemically synthesized DNA fragments." *Plos One*; 19 January. <https://journals.plos.org/plosone/article/comment?id=10.1371/annotation/495105f7-8b0a-4b7e-a617-c89c26d5b308>
- 228 For more information about H5Ni research and ethics, see: Resnik DB. 2013. "H5Ni Avian Flu Research and the Ethics of Knowledge." *Hastings Center Report* 43(2): pp. 22-33. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3953619/>
- 229 Dunlap G. and Pauwels E. 2017. "The Intelligent and Connected Bio-Labs of the Future: Promise and Peril in the Fourth Industrial Revolution," *Wilson Center Policy Briefs* (October): p. 4-6. https://www.wilsoncenter.org/sites/default/files/dunlap_pauwels_intelligent_connected_bioblabs_of_future.pdf. Also see: Gwynne P. and Heebner G. 2018. "Laboratory Technology Trends: Lab Automation and Robotics, the Brave New World of 24/7 Research." *Science*; 18 January. <http://www.sciencemag.org/site/products/robotfinal.xhtml>
- 230 See OpenAI's website, available at: <https://openai.com/>
- 231 UNODA. 2018. *Securing Our Common Future: An Agenda for Disarmament*. Office for Disarmament Affairs. <https://unoda-epub.s3.amazonaws.com/i/index.html?book=sg-disarmament-agenda.epub>
- 232 Robertson J., et al. 2016. "Darknet Mining and Game Theory for Enhanced Cyber Threat Intelligence." *Cyber Defense Review*; Fall. https://cyberdefensereview.army.mil/Portals/6/Documents/CDR%20Journal%20Articles/Darknet_Mining_and_Game_Theory_Robertson_et_al.pdf?ver=2018-08-01-090210-620
- 233 Treverton GF., et al. Addressing Hybrid Threats. Swedish Defence University: Arkitektkopia AB, Bromma; 2018. <https://www.hybridcoe.fi/wp-content/uploads/2018/05/Treverton-AddressingHybridThreats.pdf>
- 234 UN Secretary-General. 2018. *Strategy on New Technologies*. <http://www.un.org/en/newtechnologies/images/pdf/SGs-Strategy-on-New-Technologies.pdf>
- 235 For more information about the 2030 Sustainable Development Agenda, see: <https://www.un.org/sustainabledevelopment/development-agenda/>
- 236 See the Charter of the UN, available at: <http://www.un.org/en/charter-united-nations/>
- 237 For more information about the High-Level Panel on Digital Cooperation, see: <http://www.un.org/en/digital-cooperation-panel/>
- 238 Valdivia WD., and Guston DH. 2015. "Responsible innovation: A primer for policymakers." *Brookings Institution*; May. https://www.brookings.edu/wp-content/uploads/2016/06/Valdivia-Guston_Responsible-Innovation_v9.pdf
- 239 Hochschild F. 2018. "The Secretary-General's Strategy on New Technologies." *UN Chronicle*; December. <https://unchronicle.un.org/article/secretary-general-s-strategy-new-technologies>
- 240 Risse M. 2018. "Human Rights and Artificial Intelligence: An Urgently Needed Agenda." *Carr Center for Human Rights Policy*, Harvard University; May. https://carrcenter.hks.harvard.edu/files/cchr/files/humanrightssai_designed.pdf
- 241 Metzl J. and Pauwels E. 2018. "Is America's national security Facebook and Google's problem?" *TechCrunch*. <https://techcrunch.com/2018/04/15/is-americas-national-security-facebook-and-googles-problem/>. "...it would be self-defeating for American policymakers to not at least partly consider America's tech giants in the context of the important role they play in America's national security."
- 242 Glover B., et al. 2018. "Strengthening regulatory capacity for gene drives in Africa: leveraging NEPAD's experience in establishing regulatory systems for medicines and GM crops in Africa." *BMC Proceedings* 12(Suppl 8). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6069767/>

- 243 For an introduction to the idea of red-teaming, see: *Red Teaming and Alternative Analysis*, available at: <https://redteamjournal.com/red-teaming-and-alternative-analysis/>
- 244 See: Deloitte. *Red Teaming*. https://www2.deloitte.com/content/dam/Deloitte/de/Documents/risk/181016_Deloitte_Risk_Advisory_Red_Team_EN.pdf
- 245 Reisman D., et al. 2018. *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*. AI Now Institute; April. <https://ainowinstitute.org/aiareport2018.pdf>
- 246 Snow J. 2017. "Can AI Win the War Against Fake News?" *MIT Technology Review*; 13 December. <https://www.technologyreview.com/s/609717/can-ai-win-the-war-against-fake-news/>
- 247 See: Douek E. 2018. "Facebook's Role in the Genocide in Myanmar: New Reporting Complicates the Narrative." *Lawfare*; 22 October. <https://www.lawfareblog.com/facebooks-role-genocide-myanmar-new-reporting-complicates-narrative>. Also see: Wong JC. 2019. "Overreacting to failure': Facebook's new Myanmar strategy baffles local activists." *The Guardian*; 7 February. <https://www.theguardian.com/technology/2019/feb/07/facebook-myanmar-genocide-violence-hate-speech>
- 248 McQuillan D. 2018. "AI will be used by humanitarian organisations – this could deepen neocolonial tendencies." *The Conversation*; 23 April. <https://theconversation.com/ai-will-be-used-by-humanitarian-organisations-this-could-deepen-neocolonial-tendencies-92547>
- 249 See UN Technology Innovation Labs' website, available at: <https://until.un.org/>
- 250 See the Kumasi Hive website, available at: <https://www.globalinnovationexchange.org/organizations/kumasi-hive>
- 251 See Interspecifics' website, available at: <http://www.interspecifics.cc/-/about/>
- 252 See Shenzhen Open Innovation Lab's website, available at: <https://www.szoil.org/>
- 253 Delcker J. 2019. "Finland's grand AI experiment." *Politico*; 2 January. <https://www.politico.eu/article/finland-one-percent-ai-artificial-intelligence-courses-learning-training/>
- 254 Li D. and Pauwels E. 2018. "AI & Global Governance: AI for Mass Flourishing," *UN Centre for Policy Research*, 15 October. <https://cpr.unu.edu/ai-global-governance-ai-for-mass-flourishing.html>
- 255 For an overview of how complex systems approaches can better deal with uncertainty, see: Sargut G., and RG McGrath. 2011. "Learning to Live with Uncertainty." *Harvard Business Review*; September. <https://hbr.org/2011/09/learning-to-live-with-complexity>. Also see the OECD's systems approach to managing complexity in the public sector, available at: OECD. 2017. *Working with Change: Systems approaches to public sector challenges*. Preliminary Version: pp. 12-19. <https://www.oecd.org/media/oecdorg/satellitesites/opsi/contents/files/SystemsApproachesDraft.pdf>
- 256 Guston, D., and D. Sarawitz. 2002. Real-time technology assessment. *Technology in Society* 23 (4): 93–109.
- 257 Smith A, Stirling A (2007) Moving outside or inside? Objectification and reflexivity in the governance of socio-technical systems. *J Environ Polic Plann*, 9(3-4):351–373
- 258 UNDP. 2014. *Foresight: The Manual*. Global Centre for Public Service Excellence: p. 25. https://www.undp.org/content/dam/undp/library/capacity-development/English/Singapore%20Centre/GCPSE_ForesightManual_online.pdf



**UNITED NATIONS
UNIVERSITY**
Centre for Policy Research

cpr.unu.edu
@UNUCPR

767 Third Avenue, Suite 35B
New York, NY 10017
USA