Managing the risks and rewards of emerging technologies: International cooperation, national policy and the role of the individual

Regina Surber was invited to take part on the panel on 'AI: Civilian, Transdisciplinary, International Perspectives' at a side event to the UN GGE (United Nations Group of Governmental Experts) discussion on Lethal Autonomous Weapons Systems (LAWS), on 27 March 2019 at the United Nations in Geneva. The side event's goal was to highlight the importance of interdisciplinary and civilian participation in discussions surrounding LAWS and ethical AI development.

1. The position of the ICT for Peace Foundation (ICT4Peace)

For the last 2.5 years, ICT4Peace has been doing research on how emerging technologies – especially mathematical models for AI – impact individual human beings and society. In the context of this work, the foundation has focused on the analysis of the social and ethical impact of autonomous technology and LAWS in-depth.[1] ICT4Peace is also a regular observer of the UN debates on the topic since 2016.

ICT4Peace's standpoint on the peace and security implications of emerging technologies in general, and AI and LAWS in particular, goes beyond the existing mandate of the GGE in two ways:

(1) The CCW (Convention on Certain Conventional Weapons) is a framework underpinned by IHL (International Humanitarian Law), which narrows the GGE debate's focus on weapons and their use to situations of *armed conflict*.[2] However, emerging technologies may not only have an effect on the individual and society during war scenarios, but also during peace-time, e.g. autonomous technology can and is integrated into weapons used during law enforcement operations.[3] Further, emerging technologies raise broader social and human rights concerns relating to (data) privacy, bias and fairness, justice, and even existential risks for humanity (peace-time threats).[4] These concerns are prevalent independent of armed conflict.

(2) The UN GGE's debate on LAWS focuses on peace and security implications of emerging technologies and LAWS for traditional territorial state sovereignty. However, the challenges arising from emerging technologies do not fit within our traditional concept of borders and state sovereignty and do not only affect the state as a collective construct. These

[1] See e.g., Surber, Regina, 2018: Artificial Intelligence: Autonomous Technology (AT), Lethal Autonomous Weapons Systems (LAWS) and Peace-Time Threats, Geneva: ICT4Peace Foundation, available at: https://ict4peace.org/wp-content/uploads/2018/02/2018_RSurber_AI-AT-LAWS-Peace-Time-Threats_final.pdf; Weekes, Barbara, 2018, Digital Human Security 2020, Geneva: ICT4Peace Foundation, available at: https://ict4peace.org/wp-content/uploads/2018/12/Digital-Human-Security-Final-DSmlogos.pdf.

[2] Art. 1 and 2 CCW.

[3] See e.g. Opall-Rome, Barbara, 2016, Introducing: Israeli 12-Kilo Killer Robot, DefenseNews.com, May 8, 2016, available at: https://www.defensenews.com/global/mideast-africa/2016/05/08/introducing-israeli-12-kilo-killer-robot/ (accessed on February 4, 2018); Hurst, Luke, 2015, Indian Police Buy Pepper Spraying Drones To Control 'Unruly Mobs', Newsweek.com, April 7, 2015, available at: http://www.newsweek.com/pepper-spraying-drones-control-unruly-mobs-say-police-india-320189 (accessed on February 4, 2018). The 'Mozzy Wildlife Darting Copter' is promoted for wildlife capture, Desert Wolf: Leaders in Technology and Innovation, available at: http://www.desert-wolf.com/dw/products/unmanned-aerial-systems/mozzy-wildlife-darting-copter.html (accessed on February 4, 2018).

[4] See Weekes, Barbara, 2018.

challenges arising from emerging technologies are also inherently local and citizen-based, precisely because they affect an individual's data security, privacy, autonomy, or the (truth or falsehood of) information available. Therefore, ICT4Peace is interested in bringing individual human beings back into the epicenter of security concerns,[5] an urgency also highlighted by Sweden's Foreign Minister Margot Wallström at a recent arms control conference.[6]

2. Peace-time threats[7]

Peace-time threats include the effects of emerging technologies on society that are subtler than LAWS, potentially permanent, and, therefore, very transformative. Those effects raise questions about (a) the self-understanding of the human being, (b), the role and make-up of social regulation, and (c) the perception society has of the individual.

Peace-time threats are structural, manifold, (still and potentially ever-) evolving, and, hence, require an immensely broad observational focus in order to be identified. Further, they require a holistic understanding of the interplay of emerging technologies. As the core of emerging technologies are technologically highly complex, and as they are developed at a very rapid pace, there exist exceptionally broad and currently unsolvable uncertainties about the trajectories of their future development, which in turn makes it difficult to delineate a clear risk environment. However, the beginning of certain social transformations resulting from emerging technologies can arguably already be observed.

2.1. Information: the blurring of the truth

We live in a world where almost everyone has access to certain pieces of information. Those can be manipulated to offer exactly the piece of information that one individual, or a group of individuals, want or need to hear. The world has already witnessed incidents of mass information manipulation campaigns, targeting national elections and political parties, thereby undermining democratic processes.[8] In addition to *general* mass manipulation through widely spread disinformation, *individualized*[9] mis- or disinformation can also create an interesting landscape of perception: when people have access to different individualized news, a common reference point for knowledge is lost. Truth becomes something (even more) subjective and fluid. Further, what is true, nowadays often depends on 'likes'. Therefore, quantitative support and not qualitative substance, seems to be the arbiter of truth. As a consequence, the borders between reality and artificial creation with regards to knowledge through individual research are blurring. This raises questions such as 'how might this affect social cohesion?', 'are we still

[5] 'Digital Human Security' - Ibid.
[6] Statement by Margot Wallström, Capturing Technology – Rethinking Arms Control, Berlin, 16 March 2019.
[7] For further examples of peace-time threats, see Surber, Regina, 2018, 16-18.
[8] See e.g., Hern, A., 2018, Cambridge Analytice: how did it turn clicks into votes?, The Guardian, available at: https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie (accessed on 23 April 2019).
[9] Cambridge Analytica has made lucrative use of those technological developments, see e.g. Hall, Jessica, 2017, Meet the weaponized propaganda that knows you better than yourself, Extremetech.com, March 1, 2017, accessible at: https://www.extremetech.com/extreme/245014-meet-sneaky-facebook-powered-propaganda-ai-might-just-know-better-know (accessed on February 15, 2018).

knowingly shaping our (e.g. democratic) environment?', or 'do we need a human right to true information?'

2.2. Human data and AI

We live in a world where the individual human is arguably fading into irrelevance behind the vast economic and political possibilities of his/her data. Data can be willingly leveraged for economic and political interests, or for humanitarian purposes, e.g. when states try to attract tech companies that invest in AI by offering them access to their citizens' data.[10] Or, when a "great power" trains its AI algorithms in developing countries to diversify its datasets.[11] Or, when refugees receive humanitarian aid only when giving away biometric data.[12] Also, data can unwillingly increase existing global inequalities, especially through insensitive choices in training data for AI applications in the medical sector. In the Global South, medical data is often scarce and 'bad'.[13] Hence, citizens from those resource-poor environments are generally excluded from clinical trials and from developments of AI systems for health care.[14] As differences in disease incidence between different ethnic groups or 'races' are scientifically well-established,[15] those AI health applications might not fit for a population subset underrepresented in the training data. Consequently, both conscious data geopolitics as well as missing consideration of existing inequalities when designing new technologies can lead to the exploitation of vulnerable communities and, thereby, enhance global inequality – something that the international community wants to reduce (SDG 10).

2.3. Life-enhancement technologies: from augmenting to invading

Life- or human-enhancement technologies (LETs or HETs respectively) may represent an a priori more 'physical' way of transformation. LETs/HETs aim to improve human physical, psychological or intellectual capabilities, and rely on a range of emerging technologies such as genetic modification or body implants. In principle, they could extend capacity beyond the typical range of human experience, e.g. not only restore missing eye-sight to normal, but make us see for miles. This rapidly advancing scientific field raises pressing social questions, e.g. 'what if

[10] Moody, Glyn, 2017, Detailed medical records of 61 million Italian citizens to be given to IBM for its 'cognitive computing' system Watson, Privacy News Online, available at: https://www.privateinternetaccess.com/blog/2017/05/detailed-medical-records-61-million-italian-citizens-given-ibm-cognitive-computing-system-watson/ (accessed on 23 April 2019).

[11] Council on Foreign Relations, 2018, Exporting Repression? China's Artificial Intelligence Push into Africa, available at: https://www.cfr.org/blog/exporting-repression-chinas-artificial-intelligence-push-africa (accessed on 23 April 2019).

[12] Indrajit, Sneha, 2017, The Cybersecurity Risks of Using Biometric Data to Issue Refugee Aid, The Henry M. Jackson School of International Studies, University of Washington, available at: https://jsis.washington.edu/news/cybersecurity-risks-using-biometric-data-issue-refugee-aid/ (accessed on 23 April 2019).

[13] Mate KS, Bennett B, Mphatswe W, Barker P, Rollins N., 2009, Challenges for routine health system data manage- ment in a large public programme to prevent mother-to- child HIV transmission in South Africa. PLoS One. 4(5): e5483; Carrell, D. S., Schoen, R. E., Leffler, D. A., et al., 2017, Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings, Journal of the American Medical Informatics Association 24(5), 986-991: 988-989; Fraser, Hamish S. F. et al., 2010, Implementing medical information systems in developing countries: what works and what doesn't, American Medical Informatics Association (AMIA) Symposium 2010, 232-236, available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041413/pdf/amia-2010_sympproc_0232.pdf (accessed on 6 March 2019).

[14] Wahl, Brian, Cossy-Gatner, Aline, Germann, Stefan, and Schwalbe, Nina R, 2018, Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settinge?, BMJ Global Health, available at: https://gh.bmj.com/content/bmjgh/3/4/e000798.full.pdf (accessed on 20 February 2019).

[15] Coakley, Meghan, et al., 2012, Dialogues on Diversifying Clinic Trials: Successful Strategies for Engaging Women and Minorities in Clinical Trials, Journal of Women's Health 21(7): 713-716; see also e.g. Basu, D., Lopez, I., Kulkarni, A., and Sellin, J. H., 2005, Impact of race and ethnicity on inflammatory bowel disease, American Journal of Gastroenterology 100(10): 2254-2261.

LETs/HETs become mandatory, e.g. for police officers?', 'what if they recreate or augment inequality, because only 1% of society can afford them?'[16], or 'how autonomous is an individual who is 'modified' by deep-brain stimulation?'[17]

Besides tremendous ethical pressure to discuss those questions and many more, LETs/HETs also have a further aspect: In the future, possibilities to not only modify, but enhance our physical bodies and our cognitive functions might be more and more technological instead of biological – we might have body implants the size of micrometers. [18] In other words, technology might move from augmenting the human to invading the human body, with further implications when considering the IoT (Internet of Things). This might raise issues with regards to hacking, and may require new methods to secure the physical integrity of the human being.

2.4. Health: a changing definition?

Advances in biomedicine and biotechnology might eventually lead to ever earlier diagnostics: through implanted monitoring devices, we might be capable of constantly controlling our bodily processes and notice slightest deviation from a pre-set 'healthy norm'. Further, as those controlling devices are individually-tailored and potentially implanted, health management might slowly move into the private sphere[19] and more within the sphere of (perceived?) responsibility of individuals without any in-depth medical knowledge. Through constant individual supervision of bodily changes, the understanding of what is 'healthy' and what is (potentially) 'ill' might not depend on our individual physical and sensory feeling and awareness, but on our health monitoring devices. 'Feeling healthy' or 'feeling ill' might fall behind 'monitored health' and 'diagnosed disease'. Besides raising the question of whether a 'healthy' human is a human that can *feel* healthy or ill, biomedical research and developments seem to imply a *new understanding* of the term '*health*' (and 'illness'). Consequently, biomedical research might change our understanding of what it means to be healthy. As biomedical research and developments evolve at a highly rapid pace, we risk that the changed health landscape they produce sets a (new) definition of the (healthy) human being *without* us having time to reflect upon this question, let alone guide research towards our *chosen* understanding of what is 'human'.

3. A three-fold strategy for future policy development

3.1. The danger of convergence: the underexamined interplay of emerging technologies and dual-use applications

Discussing emerging technologies in the area of LAWS, as the UN GGE's mandate requires, is a highly important task. However, it is crucial for the international community to understand that it is not only Machine Learning (ML) or Deep Learning (DL) and robotics, and that it is not

[16] Whitman et al., 2018, What Americans Think of Human-Enhancement Technologies, Scientific American Blog Network, available at: https://blogs.scientificamerican.com/observations/what-americans-think-of-human-enhancement-technologies/ (accessed on 23 April 2019).
[17] Maslen, H., Pugh, J., and Savulescu, J., 2015, The Ethics of Deep Brain Stimulation for the Treatment of Anorexia Nervosa, Neuroethics 8(3), 215-230.
[18] Prof. Simone Schürle, Biomedicine, Personal Statement, 11 April 2019.
[19] Ibid.

only LAWS, that challenge international peace and security. Other emerging technologies, such as biomedicine, biotechnology, additive manufacturing, quantum computing or micro- and nanotechnology (a) also offer new ways of using traditional weapons, (b) enhance traditional weapons' lethality, accuracy, reach, and speed, or (c) may be used to create new weapons. Different emerging technologies may converge into a new weapons landscape, which requires a breaking-up of the traditional weapons 'silos' of nuclear weapons, cyber-weapons/-attacks, biological weapons, or, more recently, LAWS. Unfortunately, there currently exists a lack of a holistic understanding of emerging technologies, as well as a lack of understanding of the interplay of emerging technologies and the resulting security risks of potential dual-use applications.

For example, AI and robotics are drivers for autonomous weapons. But AI and robotics also make access and production of pathogens– bacteria and viruses for example – much easier because they can automate steps in the design process of a pathogen. Therefore, they can influence the production and proliferation of biological (and chemical) weapons. What is more, pathogens could potentially be deployed using autonomous drones, created through, e.g., 3D printing (additive manufacturing).[20] Further, autonomous intelligent agents are of great interest in the cyber domain. ML algorithms now offer the means to handle the incredible processing speed and the enormous amount of data used in cyber-operations, which the human cannot handle. In addition, they offer the flexibility that is needed to navigate within the fast-changing cyber environment, because they have the capability to learn and adapt. This makes cyber-operations cheaper, easier, and hence, more militarily lucrative.[21] What is more, quantum computing might change approaches to data security, because it offers novel ways to break encryption. This could have a game-changing effect for cyber-operations.[22] Cyber operations can be (and already are) used to sabotage nuclear weapons systems. Command and control-, alert- or launch systems of nuclear weapons could be targeted through cyber-attacks, and this could lead to accidental nuclear conflicts. This can have a 'game-changing' effect on the perceived value of nuclear weapons.[23]

There is a need to understand how emerging technologies converge into new weapons systems and weapons enhancements, which also leads to interconnection of 'classical' weapons categories. Separately analyzing and regulating different and currently pre-set weapons categories might not prove to be effective (anymore).

It could be advisable to create permanent international scientific expert groups for different weapons areas or technological sectors, that can continuously inform diplomatic debates, and that also regularly exchange on how their technological fields are converging.

[20] Brockmann, K., Bauer, S., Boulanin, V., and Lentzos, F., 2019, New Developments in Biotechnology, Stockholm International Peace Research Institute (SIPRI), in: Capturing Technology. Rethinking Arms Control, Conference Reader, 25-32.
[21] King, M., and Rosen, J., 2018, The Real Challenges of Artificial Intelligence: Automating Cyber Attacks, Wilson Center, available at: https://www.wilsoncenter.org/blog-post/the-real-challenges-artificial-intelligence-automating-cyber-attacks (accessed on 23 April 2019).
[22] Usas, A., 2018, The quantum computing cyber storm is coming, CSOOnline, available at: https://www.csoonline.com/article/3287979/the-quantum-computing-cyber-storm-is-coming.html (accessed on 23 April 2019).
[23] Van der meer, S., Cyber Warfare and Nuclear Weapons: Game-changing Consequences?, in: Meier, O., and Suh, E. (eds.), 2016, Reviving Nuclear Disarmament, Paths Towards a Joint Enterprise, Working Paper of the Research Division 'International Security', German Institute for International and Security Affairs, 37-38.

3.2. The role of government: re-claiming the regulation and safeguarding of basic rights and ethical principles in a digital world

National governments need to understand that the 'digital world' is an infrastructure like any other – if not the most important one. Currently, major tech companies are starting to create ethical principles (privacy, data security, transparency).[24] Those are principles that strive to safeguard basic rights, like the right to privacy or the right to physical integrity. Those rights are often guaranteed by national constitutions. Classically, if a new development or law risked to limit or violate a basic right, it needed to pass through parliament. However, now, with regards to potential risks of basic rights by emerging technology applications, the tech sector is taking on the task of deciding on the legality of limits and potential violations of those basic rights – and not governments. What is more, those ethical guidelines set up by representatives of the tech industry, are necessarily inspired by competitive thinking, and are developed under time pressure of global business. Whether or not this is 'ethical washing', i.e. marketing, or real added value, remains an open question. Generally, it is highly important not to 'abuse' ethical considerations and principles as a means to an end, but as an end in themselves.

Based on those observations, it would be advisable to create forums and mechanisms for increased dialogue between governments and the tech industry in order for governments to catch up on technological advances, and to develop appropriate policies to meet new social and political needs. It would also be constructive to create continuous polity-technology interfaces, e.g. through state departments for technology, that would generate the knowledge and understanding that governments need in the digital age.

3.3. Adapting education to the digital age: A bottom-up approach

As emerging technologies – as arguably any other technology – are dual-use, criminalizing them will also limit their tremendous potential for good. Hence, bans or prohibitions are not a practicable long-term strategy. As long as individuals (or a state) feels insecure, or has the potential need to harm another, dual-use tools will be used for this end. Consequently, we need to strive towards an altering of the human (or state) wish to harm. This goal requires understanding and a tremendous level of individual awareness of the new technological environment we live in, the social and ethical implications of new technologies, as well as awareness of individual responsibility for those implications. As this required transformation is located at the individual level, ICT4Peace calls for a bottom-up, educational approach.

Steps to raise awareness about those issues could be a promotion of responsible technological research, e.g. via fixed ethical guidelines for different technological fields, and/or a promotion of value-added design. Value-added design offers an approach to treat values, in addition to safety, as design specifications. For example, systems can be designed to maximize users' privacy. Determining the liability and responsibility as a design specification can sensitize

[24] See e.g. Artificial Intelligence Principles at Google: https://ai.google/principles/ (accessed on 23 April 2019), at Microsoft: https://www.microsoft.com/en-us/ai/our-approach-to-ai (accessed on 23 April 2019), or at IBM: https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf (accessed on 23 April 2019).

engineers to the risks and societal impacts of the technologies they develop.[25] Further, it seems highly advisable to sensitize young researchers about ethical questions and social implications of their own research. Education must offer a toolkit about how to approach ethical questions relating to technological research and developments, in order for graduates to have the competence to answer these questions in their later day-to-day work. The sensitization of students regarding ethics, social questions and individual responsibility must, arguably, be included even at an earlier age prior to university. The reasons are two-fold. First, a 'confrontation' with ethical reasoning at an undergraduate or graduate age might be 'too late'. Young adolescents choose an academic field, such as one of the MINT subjects, often also because those fields are so clearly delineated from philosophy and social sciences. Hence, the importance of ethical reasoning must be taught at an earlier age, so that it becomes *natural* to also study MINT subjects through an ethical lens. And secondly, as technological tools start to increasingly shape our environment without our input, very early stage reflection on individual human power, responsibility, and control is necessary.

Children and young adults have to learn through updated and technologically savvy educational programs that the way our society is built today is based on ideas and developments that we as humans have developed over hundreds of years. And they have to learn that those ideas and developments can be influenced and changed – by humans.

Regina Surber 23 April 2019

[25] Wallach, W., and Marchant, G., 2019, Toward the Agile and Comprehensive International Governance of AI and Robotics, Proceedings of the IEEE 107(3), 505-508.