TATuP Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis
Journal for Technology Assessment in Theory and Practice

oekom science

RESEARCH ARTICLE

TA for human security: Aligning security cultures with human security in AI innovation

Tanja Sinozic-Martinez*,[1], Jutta Jahnel[2]

Abstract • This research article addresses the growing concerns about digital technologies and artificial intelligence and their impact on human security. It explores the inadequacies of current digital technology regulation in protecting fundamental human rights. The authors present a typology of three "security cultures for digital technology" based on Mary Kaldor's work on human security, linking international relations and critical security studies with technology assessment (TA). The following cultures are distinguished: dual use, cybersecurity, and TA. The article concludes with a call for collaborative efforts among policy makers, industry, and civil society to prioritize a human-centered approach and global cooperation mechanisms and to strengthen the TA culture in order to address AI innovation without compromising human rights.

***TA für menschliche Sicherheit**: Ausrichtung der Sicherheitskulturen auf die menschliche Sicherheit bei KI-Innovationen*

Zusammenfassung • *Dieser Forschungsartikel befasst sich mit den zunehmenden Bedenken hinsichtlich digitaler Technologien und künstlicher Intelligenz sowie deren Auswirkungen auf die menschliche Sicherheit. Er untersucht die Unzulänglichkeiten der derzeitigen Regulierung digitaler Technologien beim Schutz grundlegender Menschenrechte. Die Autorinnen stellen eine Typologie von drei „Sicherheitskulturen für die digitale Technologie" vor, die auf Mary Kaldors Arbeit zu menschlicher Sicherheit basiert und die Disziplin internationale Beziehungen und Critical Security Studies mit Technikfolgenabschätzung (TA) verbindet. Folgende Kulturen werden unterschieden: Dual Use, Cybersecurity und TA. Der Beitrag schließt mit einem Aufruf zur Zusammenarbeit zwischen politischen Entscheidungsträger*innen, Branchenführer*innen und der Zivilgesellschaft, um einen menschenzentrierten Ansatz und globale Kooperationsmechanismen in den Vordergrund zu stellen und die TA-Kultur zu stärken, damit KI-Innovationen ohne Beeinträchtigung der Menschenrechte angegangen werden können.*

Keywords • *security cultures, artificial intelligence, digital humanism, fundamental human rights, digital ethics*

This article is part of the Special topic "Malevolent creativity and civil security: The ambivalence of emergent technologies," *edited by A. Gazos, O. Madeira, G. Plattner, T. Röller, and C. Büscher. https://doi.org/10.14512/tatup.33.2.08*

Introduction

As of the current moment, there exists a state of concern surrounding the latest digital technology field, namely artificial intelligence (AI). This has led to a convergence of world leaders and prominent figures in the tech industry coming together at the 'AI Safety Summit' in the United Kingdom to deliberate on the profound risks associated with AI. In addition, the European Union (EU) is in the process of finalizing its AI Act, civil society is mobilizing to include foundational models in AI regulation, and the United States government is issuing an order aimed at regulating large language models (LLMs). Although an escalation in human security risks posed by increasing loss of privacy, addiction economy, and algorithmic bias are widely recognized, scientific evidence shows that digital technology regulation is unsuccessful in the protection of fundamental human rights. Global AI threats are 'new' and 'different' from those posed by nuclear, biological and chemical technologies, and dominant

* *Corresponding author: tanja.sinozic-martinez@wien.gv.at*
[1] *Magistratsdirektion, Bereichsleitung für Wissenschaft, Stadt Wien, Wien, AT*
[2] *Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology, Karlsruhe, DE*

https://doi.org/10.14512/tatup.33.2.16
Received: 05. 01. 2024; revised version accepted: 08. 04. 2024;
published online: 28. 06. 2024 (peer review)

security regimes are 'old' ways in dealing with them. This paper tries to make sense of these 'new' differences by developing a typology of three 'digital technology security cultures' to better understand the challenges of addressing the threats posed by AI-supported digital technology innovations to humans. We apply the terms 'security cultures' directly from Mary Kaldor's work on human security (Kaldor 2007). In developing the three cultural types, we link the discipline of international relations (IR) and its sub-field of critical security studies with the field of technology assessment (TA) for an exploration of the security dimensions of regulating digital technology for the purposes of human security and the protection of humanitarian space.

Kaldor uses the terms security culture to denote "a pattern of behavior, which includes norms and standards which correspond to a particular interpretation of security that is associated with a form of political authority or a set of power relations. A security culture comprises different combinations of ideas, rules, people, tools, tactics and infrastructure, linked to different types of political authority that come together to address or engage in large-scale violence" (Kaldor 2018, p. 2). For the purpose of this paper, security is used to draw attention to the violation of human rights such as privacy, freedom from discrimination, safety, freedom from cruel treatment, equality, democracy, social security, work protection, and education, through the design, manufacture and use of technological innovations using the example of AI-supported digital technologies. The objective is to frame AI innovation as part of an ecosystem for maintaining progress towards preserving and enhancing human life and wellbeing, while drawing attention to the structures and processes in society which ensure such progress. The structures and processes are here associated with a 'TA security culture' which is also the research gap that this theoretical contribution intends to fill. Our method relied upon a literature review of AI threats, security cultures and TA, with the aim of continuing existing discussions on the role of culture in addressing the ambivalence of emerging technologies. Culture is the unit of analysis, rather than technologies or sectors or products. By viewing TA through an IR lens, the dimension of global governance for human rights protection in the face of technological change can be more deeply investigated.

The following section develops the concept of digital technology security cultures, drawing the main inspiration from Mary Kaldor's work on security and linking it to the most recent steps in digital technology evolution, namely AI capabilities. Thereafter, the dual use security culture, which is historically based on international cooperation in the regulation of nuclear, biological and chemical weapons is described. A discussion on the cybersecurity security culture follows, which we differentiate from the dual use paradigm in its greater reliance on private firms and private investors. Thereafter, the TA security culture is explained, which we associated with a social and humanistic focus on individual rights protection in the face of digital innovation, but as lacking the governance frameworks and private sector support of the other two cultures. The final section concludes.

Digital technology security cultures

AI threats to humanity are currently a 'hot topic' in the literature (Werthner et al. 2024). The concentration of power in the hands of a few large firms, rapidly increasing capabilities to spread misinformation at scale and the challenges these processes create for democracy are counterpoints to the benefits of AI for progress in areas such as science and medicine. Since at least the 1960s scientific evidence has shown that science and technology does not cause societal ills such as bias and inequality, but that technology evolves through socio-technical and techno-economic processes and cultures, embedding itself within pathways which already exist. To counteract societal deficiencies such as environmental degradation and inequality, and to redirect technology away from being grafted on these existing deficiencies of socio-economic systems, it is helpful to look at the ways in which the current technological paradigm interacts with them. The phrase 'digital security cultures' refers to the various methods and strategies employed to safeguard people from digital threats. This is essential because the ongoing development of digital technology poses risks not only in military contexts but also in non-military settings, and these risks impact human safety and well-being. Kaldor in her description of the evolution of the concept of 'security' (defined more broadly than technology threats) following Buzan and Hansen (Kaldor 2018, p. 13; Buzan and Hansen 2009) state that although the term emerged to define threats of war and defense, more recently include societal cohesion and non-military threats and vulnerabilities. The term security can be used to denote objectives (such as safety), or practices and artefacts in use (such as scanners at airports) (Kaldor 2018, pp. 13–14). The UN Development Programme (UNDP 1994) report describes seven types of risks to security: economic, food, health, environmental, personal, community and political. Security cultures, following Kaldor, are non-national, they include strategy, focus on functions, and are defined by specific ways of exercising power (Kaldor 2018). A central feature of this conceptualization of cultures is transition and change. Kaldor (2018, p. 24) draws attention to the usefulness of looking at mechanisms through which cultures are reproduced, such as the market, events, public debates, media representations, professional career and training structures, to find out how to reverse what is happening that is worsening human security.

‚Digital technology' is used to mean a dynamically evolving bundle of products, services, knowledge, skills and resources originating in the telecommunications industries in the 1960s and 70s and associated with capabilities of high data processing power (Freeman 2009). The information and communication technologies (ICT) paradigm is characterized by changes in production, design, consumption, speed and ease of communication and networking, and interactions between people. Recent emphasized changes have been the use of robotics, biotechnology, and the Internet of Things. Further focus is placed on emerging products and services enabled through artificial intelligence and the related changes which occur in governance, power struc-

https://doi.org/10.14512/tatup.33.2.16

tures and human identity, drawing attention to threats to humanity and civility (Nida-Rümelin and Weidenfeld 2022; Werthner et al. 2022).

The risk of technology negatively affecting humanity is of course not a new concern. However, while propaganda, misinformation, precise drone strikes, addictive product features, and biased algorithms are old issues, they present impacts which go beyond regular technological advancements of automation, even though they share some similarities. A central concern is the extent and degree to which they conflict with and gradually erode the social, civil and human rights institutions which have taken a very long time to build (such as for example individual property rights, the right to privacy, and social security). In the previous four technological revolutions (Perez 2002) not as many institutions protecting human rights and social welfare existed as today. The year 2023 marks 75 years of the Universal Declaration of Human Rights. Children were not protected during the industrial revolution from doing hard labor, but after it. The social welfare state in the UK is a post-WWII phenomenon. The digital shifts that violate human rights are highly incremental (Frischmann and Selinger 2018), much more subtle than in previous technological revolutions, and therefore more difficult to reverse.

Dual use security culture

The perhaps oldest technology security culture has its roots in the Cold War era (although the concept of dual use is traceable back to the 17th century (McLeish and Nightingale 2007), and initially focused on the international governance of nuclear, chemical and biological weapons (Molas-Gallart and Robinson 1997) and can be called the 'dual use technology security culture'. The term 'dual use' was originally used to refer to technologies that can be applied for both military and civilian purposes (Alic 1994), but have evolved to include the concepts peaceful versus non-peaceful, legitimate versus illegitimate, and benevolent versus malevolent (Rath et al. 2014) technology use. An important institutionalization of the dual use technology security culture was established for nuclear weapons in the Treaty on the Non-Proliferation of Nuclear Weapons (NPT) (effective in 1970) and for chemicals with the Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction (CWC) (effective in 1997), emerging out of the Cold War era. The aim of CWC was to regulate "the development, production, acquisition, stockpiling, retention, transfer, and use of chemical weapons" (Tuerlings and Robinson n.d., p. 4) at the international level. The CWC, unlike the NPT which did not depend on a civilian industry, had a unique impact on private industry, namely the chemicals industry (Tuerlings and Robinson n.d.). The main challenge, as for digital technologies which are also, like chemicals, primarily a civilian industry, is in defeating the development of illegal weapons and the malevolent use of civilian products while maintaining their scientific (McLeish and Nightingale 2007), technical and civilian usefulness.

The dual use technology security culture for nuclear and chemical technologies was created by various actors from the civil sector, public policy makers, private actors, governments, NGOs, and, for chemicals in particular, the private sector. It took a period of 50–70 years to create agreements and treaties. Non-state actors and their networks played a crucial role in shaping this culture. The absence of a central agency, or global technology governance at the international level, and private industry backing, makes processes very slow. Further, McGrew (1992, p. 318) states that "international regime and institutions of global management tend to lack the authoritative means to ensure compliance with their decisions".

As technological change has taken place, the dual use culture has extended beyond traditional military applications and the potential threat of over-regulation to science (McLeish and Nightingale 2007), to include ICTs and neurotechnologies (Mahfoud et al. 2018). Armed drones have become prevalent in ongoing wars, as has the use of mobile phones as triggering devices (Kaldor 2018, p. 28), the use of social media for mobilization and spreading fear (Kaldor 2018, p. 28) and continuous surveillance is enabled by the use of virtually any ICT product and system. Especially content-generation techniques evolved as a powerful weapon for disinformation and deep fakes, where one can no longer rely on what we are seeing or hearing. These hybrid threats can disrupt public order, attack communication infrastructures and weaken democratic structures (Chesney and Citron 2019). Compared to nuclear weapons, the production of biological weapons is easy and inexpensive, and information on how to produce them is readily available (Selgelid 2009). Similarly, digital technologies rapidly become easier to misuse. This is in big contrast to nuclear science, where discoveries on weapons and illegitimate use were usually classified (Selgelid 2009). Following the three general laws of Isaac Asimov invented in the field of robot ethics the priority of tackling AI technologies should be based on the precondition that there will be no harmful effects for human beings even in the case of a malevolent misuse (Clarke 1994). Effective dual use governance requires collaboration among states, private sector, international organizations and civil society.

Cybersecurity security culture

‚Cybersecurity' is variably used to raise awareness to threats to nation states and individuals which are dependent on, or are enabled by, innovation in ICTs. The here so-called 'cybersecurity security culture' has originated in and is shaped by the US through a collaboration between the US government and US-based ICTs firms in Silicon Valley (Cavelty 2007). Included in pushing it forward are public and private sectors. It relies on technical fixes to 'cyber-threats', which are reliant upon innovation in the ICTs sector and include products and systems to re-

 https://doi.org/10.14512/tatup.33.2.16

duce harm from for example worms, viruses, hacking, malware and ransomware directed at information system infrastructures at the national level, as well as personal computers.

It is analyzed, amongst other disciplines and fields, in the IR literature in the sub-field of security studies. The cyber-threat narrative, which is shaped by public communications of the threats to national security from and to digital technology-based products and systems, originated in the US in the Reagan administration, and later on the military through the concept of information warfare, and was first institutionalized with the Computer Security Act in 1987. The US military-funded innovation response to the national threat narrative has since 1991 been focused on the development of products and systems such as surveillance, reconnaissance systems and long-range precision weapons systems (Cavelty 2007). The high public investment in data collection and weapons technologies continues a long historical tradition in military and public spending on private sector innovation (Mazzucato 2013). Private industry specializing in these products has become highly profitable (Penney et al. 2018). Cavelty (2007, p. 4) says that "cyber-threats have clearly not materialized as a 'real' national security threat". She argues that instead of focusing on the terms cyber and security, it is necessary to look at the narratives which produce threat indicators, who produces the knowledge and how, and how this shapes measures and policies (Dwyer et al. 2022, p. 9).

In addition to public investments to the private sectors for technological innovation for military purposes to mitigate the threats described in the 'threat narrative' of the US, a big part of the cybersecurity culture is the protection of individual computers from crimes which depend upon and attack digital infrastructures such as hacking, malware, and denial of service and crimes which occur in other domains as well such as financial fraud, phishing, pharming and extortion). AI technologies could also enhance cybercrime threats by malevolent impersonations or identity theft through deep fakes (Ciancaglini 2020). The main challenge for human rights protection is difficult to achieve under the leverage of the cybersecurity security culture because both the infrastructures and the solutions for protecting the infrastructures, systems and its components are reliant on the private sector (Bay 2016; Penney et al. 2018). Lindstrom (2012) says that although the dominant cybersecurity narrative focuses on risks to national security on a macro scale level, they are not necessarily the biggest risks, those are accrued by individuals on a micro scale level. Human rights and social protection are being compromised by digital technology firms and products, such as for example in the case of labor protection. The global 'laissez-faire' approach to the regulation of digital innovation has been criticized for being insufficient in protecting human rights, marginalizing civil society in this narrative. Cavelty, referring to the words of Coles-Kemp in a collaborative piece published in 2002, states that what is needed in security culture is "a positive change that benefits the security of people" (Dwyer et al. 2022, p. 18). This is where technology security cultures can benefit from the integration of the human rights focus of TA and related fields such as AI ethics, digital humanism and responsible innovation (RI).

Technology assessment security culture

By focusing on individual security through the strengthening of civil society and global collaboration (Kaldor 2007), the third technology security culture can be called the 'TA security culture'. The dominant narrative is about the threats to fundamental human rights such as privacy, autonomy, health and well-being perpetuated by technological innovation in specific products and systems. The TA security culture is created by communities of TA scholars and practitioners, scientists from disciplines such as AI ethics, philosophy, sociology, Science and Technology Studies (STS) and science communication, the public sector and civil servants. Apart from scientific research, activities include public engagement of science and technology, science communication, and policy advice. An important step in formalizing this culture in policy circles has been the inclusion of academics specializing in the ethics of AI and TA in ethics councils such as the German Ethics Council. The emerging social movement of digital humanism (Nida-Rümelin and Weidenfeld 2022; Werthner et al. 2019) is inclusive of different actors such as urban planning and administration, cultural institutions (such as museums) public decision-makers, private industry, scientists from disciplines such as computer science, philosophy, ethics, TA, sociology, political science, and business. Early work on the concept of digital humanism is associated with the Vienna Manifesto on Digital Humanism (Werthner et al. 2019), a document outlining core principles for digital progress in line with the protection and strengthening of fundamental human rights. The objective of this culture (which includes different scientific communities such as AI ethics, RI, STS and sociology of technology, among others) is to maintain human rights in the digital realm, rather than only criminalizing their violation (borrowing Kaldor's words in her description of human security and humanitarian space). The TA culture can be traced back to the Office of Technology Assessment and the Ethical Legal and Social Aspects method of technology assessment, and more recently the digital rights, RI, AI ethics and STS communities, and providing reflections on the compromising of human rights, ethics and social institutions with the deployment and use of AI-enabled products.

The TA security culture depends upon government support and the public financing of research projects, social movements, education, designated organizations and research institutions, as well as the private sector in maintaining an ethical and human right centered approach to innovation. The threat narrative of this culture are violations of human rights at the micro, meso and macro levels. Risks at the micro level are associated with safety issues of deep learning technologies and data processing. This category includes prejudice against disabled people (Venkit et al. 2022), violation of privacy and intellectual property rights

https://doi.org/10.14512/tatup.33.2.16

infringement during training, and risks of using LLMs without understanding how they generate digital content (Taecharungroj 2023; Daws 2020). In addition, threats with regard to illegal criminal activities such as the generation of non-consensual pornography or abuse of malevolent impersonations play an important role. Besides single persons, also firms and organizations could be targeted (Ciancaglini 2020). The third risk category summarizes societal and systemic threats on the macro scale, for example political security issues posed by generative AI due to its high potential for disinformation and manipulating public opinion (Chesney and Citron 2019). AI technologies could be misused for propaganda, manipulation of elections or for stoking social unrest, political polarization or radicalization with the aim to undermine democracy. Finally, some experts are warning against an upcoming enhanced human intelligence or superintelligence that could not be controlled by humans and may pose an existential risk for humanity (Bostrom 2014).

The TA security culture does not yet have strong support from national governments or from private industry. Its main supporters are academia, research institutions directed at civil society, human rights organizations, and some parts of the public sector. As the EU AI Act comes into force in 2026, it can be expected that firms will engage more productively with standards such as Value-Based Engineering (Spiekermann and Winkler 2022) integrating ethical concerns with their innovation processes. However, the limited support from large firm oligopolies is a critical vulnerability. Nevertheless, there are echoes of this culture in employee movements in some of these large firms (for example, Google), and this can increase as tools are developed and firm ecosystems are shaped in the direction of the protection of human rights in the face of rapid AI advancements.

Conclusion: Navigating the nexus of security cultures, human rights, and AI

In analyzing the complex intersections between security cultures, human rights, and AI our paper has identified critical issues in framing digital technologies as inclusive of threats to fundamental human rights, similarly to nuclear, biological and chemical technologies, and therefore a relevant and 'new' security issue. As AI-enabled products and systems rapidly diffuse, it becomes evident that safeguarding human rights necessitates a human security focused approach to technological innovation.

Our examination reveals three technology security cultures: dual use, cybersecurity, and TA. Each culture reflects a unique set of norms, practices, and actors influencing the governance of AI-enabled digital technology. The dual use culture, rooted in Cold War dynamics, emphasizes international cooperation to regulate high-risk technologies. In contrast, the cybersecurity culture, pioneered by the US, relies on technical fixes and private sector engagement to counter digital threats. The TA culture, emerging from academic and ethical domains, places human rights at the forefront of technological development.

Each security culture is accompanied by its own threat narrative. The dual use culture contends with the complex task of governing dual-use technologies and preventing their malevolent use predominantly at the macro level. The cybersecurity culture grapples with the dual role of private sector entities in both protecting and potentially compromising individuals and firms. The TA culture focuses on safeguarding rights at different levels but faces challenges in garnering support from governments and large firms. To reverse what is happening in the compromising of human rights from rapid datafication and diffusion of ICTs and AI-enabled products and systems, policy makers, industry leaders and civil society must collaborate to fill the gap of a human-centered approach in the two dominant technology security cultures. Governments need to, address structural issues contributing to human rights violations, and firms need to embed ethical considerations into their AI innovation processes as far as possible. Global cooperation mechanisms, similar to those established for the dual use of nuclear, chemical and biological technologies, need to be negotiated and implemented for digital technologies as Bostrom says "to shift world politics into a more harmonious register" (Bostrom 2014, p. 259). Civil society has a pivotal role in shaping technology governance, and initiatives that amplify the voices of advocacy groups, ethical scholars, and human rights organizations, need to be strengthened to provide a counterbalance to the influence of powerful oligopolies. Inclusion in AI development, which incorporates diverse perspectives from marginalized groups needs to be encourage to minimize algorithmic biases and improve the adaptability of AI to diverse cultural contexts. The strategic analysis presented in this paper underscored the centrality of human security and the importance of strengthening the TA security culture. Focused on social, environmental and ethical objectives, it offers a pathway to navigate the challenges posed by AI, support the beneficial and useful aspects of AI innovation, without compromising fundamental human rights.

Funding • This work has received no external funding.
Competing interests • The authors declare no competing interests.

References

Alic, John (1994): The dual use of technology. Concepts and policies. In: Technology in Society 16 (2), pp. 155–172. https://doi.org/10.1016/0160-791X(94)90027-2

Bay, Morten (2016): What is cybersecurity? In search of an encompassing definition for the post-Snowden era. In: French Journal for Media Research 6, pp. 1–28. Available online at https://frenchjournalformediaresearch.com:443/lodel-1.0/main/index.php?id=988, last accessed on 07.06.2024.

Bostrom, Nick (2014): Superintelligence. Paths, dangers, strategies. Oxford: Oxford University Press.

Buzan, Barry; Hansen, Lene (2009): The evolution of international security studies. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511817762

Cavelty, Myriam (2007): Cyber-security and threat politics. US efforts to secure the information age. Abingdon: Routledge.

Chesney, Robert; Citron, Danielle (2019): Deep fakes. A looming challenge for privacy, democracy, and national security. In: California Law Review 107, pp. 1753–1819. https://doi.org/10.2139/ssrn.3213954

Ciancaglini, Vincenzo et al. (2020): Malicious uses and abuses of artificial intelligence. Available online at https://www.europol.europa.eu/cms/sites/default/files/documents/malicious_uses_and_abuses_of_artificial_intelligence_europol.pdf, last accessed on 25. 04. 2024.

Clarke, Roger (1994): Asimov's laws of robotics. Implications for information technology. In: Computer 27 (1), pp. 57–66. https://doi.org/10.1109/2.248881

Daws, Ryan (2020): Medical chatbot using OpenAI's GPT-3 told a fake patient to kill themselves. In: AI News, 28. 10. 2020. Available online at https://www.artificialintelligence-news.com/2020/10/28/medical-chatbot-openai-gpt3-patient-kill-themselves/, last accessed on 25. 04. 2024.

Dwyer, Andrew; Stevens, Clare; Muller, Lilly; Cavelty, Myriam; Coles-Kemp, Lizzie; Thornton, Pip (2022): What can a critical cybersecurity do? In: International Political Sociology 16 (3), pp. 1–26. https://doi.org/10.1093/ips/olac013

Freeman, Christopher (2009): The ICT paradigm. In: Robin Mansell et al. (eds.): The Oxford Handbook of Information and Communication Technologies. Oxford: Oxford University Press, pp. 34–54. https://doi.org/10.1093/oxfordhb/9780199548798.003.0002

Frischmann, Brett; Selinger, Evan (2018): Re-engineering humanity. Cambridge, UK: Cambridge University Press.

Kaldor, Mary (2007): Human security. Cambridge, UK: Polity.

Kaldor, Mary (2018): Global security cultures. Cambridge, UK: Polity.

Lindstrom, Gustav (2012): Meeting the cyber security challenge. In: GCSP Geneva Papers Research Series 7. Available online at https://www.files.ethz.ch/isn/147788/7-2012.pdf, last accessed on 25. 04. 2024.

Mahfoud, Tara; Aicardi, Christine; Datta, Saheli; Rose, Nikolas (2018): The limits of dual use. In: Issues in Science and Technology 34 (4), pp. 73–78. Available online at https://issues.org/the-limits-of-dual-use/, last accessed on 25. 04. 2024.

Mazzucato, Mariana (2013): The entrepreneurial state. Debunking public vs. private sector myths. London: Anthem Press.

McGrew, Anthony (1992): Global politics in a transitional era. In: Anthony McGrew and Paul Lewis (eds.): Global politics. Globalization and the Nation-State. Cambridge, UK: Polity, pp. 312–330.

McLeish, Caitriona; Nightingale, Paul (2007): Biosecurity, bioterrorism and the governance of science. The increasing convergence of science and security policy. In: Research Policy 36 (10), pp. 1635–1654. https://doi.org/10.1016/j.respol.2007.10.003

Molas-Gallart, Jordi; Robinson, Julian (1997): Assessment of dual technologies in the context of European security and defence. Report for the Scientific and Technological Options Assessment (STOA). Luxembourg: European Parliament.

Nida-Rümelin, Julian; Weidenfeld, Nathalie (2022): Digital humanism. For a humane transformation of democracy, economy and culture in the digital age. Cham: Springer. https://doi.org/10.1007/978-3-031-12482-2

Penney, Jon; McKune, Sarah; Gill, Lex; Deibert, Ronald (2018): Advancing human rights-by-design in the dual-use technology industry. In: Columbia Journal of International Affairs 71 (2), pp. 103–110. Available online at https://ssrn.com/abstract=3218975, last accessed on 07. 05. 2024.

Perez, Carlota (2003): Technological revolutions and financial capital. The dynamics of bubbles and golden ages. Cheltenham: Edward Elgar.

Rath, Johannes; Ischi, Monique; Perkins, Dana (2014): Evolution of different dual-use concepts in international and national law and its implications on research ethics and governance. In: Science and Engineering Ethics 20 (3), pp. 769–790. https://doi.org/10.1007/s11948-014-9519-y

Selgelid, Michael (2009): Governance of dual-use research. An ethical dilemma. In: Bulletin of the World Health Organization 87 (9), pp. 720–723. https://doi.org/10.2471/BLT.08.051383

Spiekermann, Sarah; Winkler, Till (2022): Value-based engineering with IEEE 7000. IEEE Technology and Society Magazine 41 (3), pp. 71–80. https://doi.org/10.1109/MTS.2022.3197116

Taecharungroj, Viriya (2023): "What can ChatGPT do?" Analyzing early reactions to the innovative AI chatbot on Twitter. In: Big Data and Cognitive Computing 7 (1), p. 35. https://doi.org/10.3390/bdcc7010035

Tuerlings, Emmanuelle; Robinson, Julian (n.d.): The trilateral network associated with the Chemical Weapons Convention. Case study for the UN vision project on global public policy networks. Sussex: University of Sussex. Available online at https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6142cc246bd45db237f409d982a965f958fbbd1c, last accessed: 25. 04. 2024.

UNDP – United Nations Development Programme (1994): Human development report 1994. New York, NY: Oxford University Press. Available online at https://hdr.undp.org/system/files/documents/hdr1994encompletenostats.pdf, last accessed: 25. 04. 2024.

Venkit, Pranav; Srinath, Mukund; Wilson, Shomir (2022): A study of implicit bias in pretrained language models against people with disabilities. In: Nicoletta Calzolari et al. (eds.): Proceedings of the 29th International Conference on Computational Linguistics. Gyeongju: COLING, pp. 1324–1332. Available online at https://aclanthology.org/2022.coling-1.113.pdf, last accessed: 25. 04. 2024.

Werthner, Hannes et al. (2019): Vienna manifesto on digital humanism. Available online at https://caiml.org/dighum/dighum-manifesto/Vienna_Manifesto_on_Digital_Humanism_EN.pdf, last accessed: 25. 04. 2024.

Werthner, Hannes et al. (eds.) (2024): Introduction to digital humanism. A Textbook. Cham: Springer. https://doi.org/10.1007/978-3-031-45304-5

DR. TANJA SINOZIC-MARTINEZ
is expert in digital humanism at the City of Vienna director's office since 2024. While doing the research for this article, she was academy scientist at ITA, Vienna, and guest researcher at ITAS, Karlsruhe. Her educational background is in economics (LSE), and science and technology policy studies (SPRU, Sussex). Her work focuses on humanism in the innovation and regulation of AI.

DR. JUTTA JAHNEL
studied food chemistry and gained experiences at the interface between analytics, technology and law. Since 2010 she works as a researcher at ITAS, Karlsruhe. Her recent work focuses on risk governance of artificial intelligence. She led a project to advise the EU Parliament on how to deal with deepfakes in the new AI regulation.

https://doi.org/10.14512/tatup.33.2.16