New and emerging perspectives for technology assessment: Malevolent creativity and civil security

Alexandros Gazos¹ 2, Octavia Madeira¹ 2, Georg Plattner*, 1 2, Tim Röller¹, Christian Büscher¹ 2

Abstract - The characteristics of new technologies can cause significant harm to society. This does not only apply to unintended consequences, but even more so when the technologies are used for malicious purposes. The latter can be observed in the domain of civil security. Here, negative developments such as social polarization, increasing radicalization, and democratic regression alongside a deteriorating security situation are increasingly associated with technological innovation and malevolent creativity. As society becomes more and more aware of such correlations, technology assessment is called upon to provide orientation knowledge. This requires appropriate approaches and knowledge tools to assess the potential role of technology in relation to radicalization and political extremism.

Neue und künftige Perspektiven für die Technikfolgenabschätzung: Malevolente Kreativität und zivile Sicherheit


Keywords - malevolent creativity, extremism, radicalization, technology assessment, civil security

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.33.2.09

Introduction

The current global decline of democracy (Papada et al. 2023) has raised significant challenges to society. These challenges are partly exacerbated by new technologies, particularly in the realms of communication via technical platforms and artificial intelligence (AI). Social media allows for the radicalization of vulnerable individuals without the need for physical support structures or networks (Calvert 2024; Ware 2023). An example of this is the social media campaign that led thousands of European Muslims to join ISIS (Gates and Podder 2015). The radicalization of some far-right ‘single perpetrators’ is another example (Mølmen and Ravndal 2023). New technologies such as artificial intelligence are currently developing rapidly and could become appealing to extremists. There are already signs of this in the use of deepfakes. The uncertainty they introduce can make people vulnerable to misinformation and radicalize existing views (Niewegowski et al. 2023). Raising awareness of technology-related risks and unintended and undesirable consequences of technology has been a concern of technology...
The challenge for technology assessment (TA) since its inception. So far, however, the focus has tended to be on the opportunistic and (supposedly) benign intentions of scientists, engineers, entrepreneurs, or investors, with little awareness of the societal and environmental consequences (already in Baram 1973). The question of how technology might be deliberately used by malevolent actors to undermine civil security has rarely been addressed in TA research – sometimes as a by-product, sometimes in thematic niches. This usually involves dual-use technologies that can be used for both military and civilian uses (Mahfoud et al. 2018).

In this TATuP Special topic, we want to shed light on new and emerging technologies at the nexus of civil security and malevolent (mis)use. We reflect on the possibilities and limitations of assessing these technologies in terms of their potential for malevolent use, such as extremist radicalization and terrorist violence. The focus is on how malevolent use cases can be anticipated, prevented, or controlled.

Using technology in a function other than that intended by developers and with the intent to harm others is an inherently creative process.

A first area of investigation is ‘civil security’. While technical innovation is generally seen as a desirable deviation, civil security focuses on undesirable deviations. The aim is to prevent the spread of extremist ideas and ideologies as well as violent and terrorist acts. This perspective is based on the premise of preventive action and the logic of suspicion against individuals, groups, and organizations. It analyzes risk factors that increase the likelihood of malicious acts, such as psychological predisposition, social disadvantage, ideological imprint, etc. (Bröckling 2015). However, technology cannot simply be added as a further risk factor since technology is not only ubiquitously available but some technologies depend on, are connected to, or enable other technologies for a potentially unlimited number of applications. Above all, every technological innovation is a deviation from existing norms and knowledge. Under this premise, any technical progress would be generally suspected of being malevolently exploitable (Kusche and Büscher 2021).

Another pillar for future TA research is a branch of research dedicated to the phenomenon of malevolent innovation/actions. Using technology in a function other than that intended by developers and with the intent to harm others is an inherently creative process. Cropley, Kaufman and Cropley (2008) call this “malevolent creativity”. They define it as a form of creativity that “is deemed necessary by some society, group, or individual to fulfill goals they regard as desirable, but has serious negative consequences for some other group, these negative consequences being fully intended by the first group” (Cropley et al. 2008, p. 106). When we talk about ‘malevolent actors,’ we mean those whose actions intentionally harm or negatively affect other groups or individuals (Schmid 2013). The fact that technologies can become the object of malevolent creativity is related to their common openness to unintended uses. Several bodies of literature consequently address the relationship between innovation processes and the intentional use of technology for malevolent ends. The focus of research has been on issues such as the intent of persons or groups (Gill et al. 2013; McLaren 1993), innovation dynamics between extremist or terrorist actors and organizations responsible for civil security (Dolink 2007; Jackson 2001), innovation processes in society at large, and the general availability of technology (Cronin 2020).

All these perspectives provide information about the factors that play a role in the innovation process. However, they do not allow us to say anything about the future use of a particular technology. Every technology is the successful simplification in the medium of causality, i.e., the creation of effects through the rig-
ororous selection of useful causal relationships. If a technology is to be functional, all other cause-effect-relationships, which are still effective, must prove to be of no effect to this system (Luhmann 2005). This applies to mechanical installations as well as algorithms. Innovations also often arise in experiments or niches, i.e., in situations where disruptive influences can be temporarily ignored (de Haan and Rotmans 2011). Moreover, as we know very well today, technology always works most effectively when used without consideration (and often: without responsibility) for all other influences on the natural environment, health, ethics, and society. This is often the case when used for economic and military purposes and, even more so, when used with malevolent intentions, such as terrorist actors indiscriminately harming people or protected goods. We also know very well that the possibilities offered by technologies only reveal themselves at a much later stage of their development. The Tor Project is a prominent example of such changes in use. Originally conceived as a “free haven” (Moore and Rid 2016, p. 16) for military personnel to communicate anonymously on the internet, it has evolved into a decentralized, anonymized, and publicly accessible network. Over time, this led to the emergence of hidden websites and services and eventually to the emergence of a darknet that can be used for nefarious activities (Chertoff 2017).

The psychologist Gibson (2015) addressed this issue by developing the concept of affordances to refer to the possibilities that nature or technical artifacts present to an observer. It was important to him to emphasize the relationship between the object and the observer so that the potentials are not solely attributed to one or the other. Potentials remain unused if they are not discovered. Discovery does not depend on obvious properties. The tinkerers and creative minds keep trying until a useful causal isolation can be realized. What this is used for remains open: “All these benefits and injuries, these safeties and dangers, these positive and negative affordances are properties of things taken with reference to an observer but not properties of the experiences of the observer” (Gibson 2015, p. 137). The concept of technical affordances is of analytical value for the further development of TA concepts because it sensitizes our research to the indeterminacy of technical possibilities in relation to social developments such as political conflict. This raises the following questions for our special topic: What do material and immaterial artifacts allow actors to do? How can we explore unforeseen possibilities in the future use of known and yet unknown technologies by known and unknown actors — possibilities that exceed original intentions? In other words, how can we gain some traction in the face of this factually, temporally, and socially boundless complexity?

How can we explore unforeseen possibilities in the future use of known and yet unknown technologies by known and unknown actors?

Outlook for technology assessment

Given the high degree of ignorance about technological affordances, there is a need for continuous monitoring (see also Stirling 2010). This is especially true in the context of radicalization, extremism, and terrorism. Such monitoring cannot be achieved by individual projects but must be driven by a large number of different observers. In an ongoing research project, we have approached the complexity of this issue step by step, moving from a coarse to a more finely tuned radar through various selections. First, we systematically interviewed experts on the future use of technology by extremist and terrorist actors as part of a two-stage Delphi survey. The results of this survey were used to prioritize detailed analyses, with functionality and availability serving as selective criteria. For example, one particularly relevant technology identified in this analysis is machine learning, which can be used in various areas such as improving translation software, disseminating propaganda, supporting transnational organizations, and creating deepfakes and social bots (Büscher et al. 2022). Second, an expert workshop on surveillance technologies was held to discuss the possibilities and limits of the use of technology and the conditions for intervention, particularly with regard to surveillance technologies for physical and digital locations. The focus was on questions of legal legitimacy, social desirability, and the possibilities and limitations of selective assessment with regard to the information burden and strategy development in policing extremism (Büscher et al. 2022). Third, in our subsequent analysis, we iden-
development and implementation of an approach to preventing malevolent use in the sense of responsible research and innovation and prevention by design. It is crucial for such an approach to take into account technological affordances while utilizing the potential of emerging technologies for the entire spectrum of prevention work, such as measures that focus on improving social conditions, reducing opportunities for misuse, engaging in early intervention, deradicalization, and rehabilitation (Schmid 2020). By networking with other interested groups such as non-governmental organizations (NGOs), mass media, and interest groups, long-term technology monitoring should be able to contribute to the prevention of malevolent use of new technologies.

Contributions in this Special topic

This Special topic addresses the challenges and issues outlined above in various ways. With regard to digital technologies and AI in particular, Tanja Sinozic-Martinez and Jutta Jahnel show how TA can help improve existing cultures of assessing digital technologies. By drawing on Mary Kaldor’s work on security, they demonstrate the contribution TA can make to security studies by focusing specifically on human security and its very own ‘security culture’ in analyzing AI innovation. The article can be understood as a call to action for practitioners and academics in TA to pay more attention to issues of security in AI innovation processes in order to shape a culture that is mindful of the individual, human rights, and societal aspects of innovation. It adds a perspective to this Special topic that considers the predominantly state security or business-oriented approach to AI innovation (both of which also take into account certain types of malevolent creativity) but also shows that there is more at stake for research article analyzes how “disruptive,” “dual-use,” “democratized,” and “diffused” characteristics of a technology can help in assessing its susceptibility for terrorist uses (4D framework). Their two examples of additive manufacturing and unmanned aerial systems show how malevolent creativity is central to the dynamics of terrorist innovation. However, the question of which technology might become the object of malevolent creativity is determined by external realities, such as (market) availability, pricing, or user friendliness. The framework introduced by Veilleux-Lepage and Rassler is an important new way of looking at emerging technologies through the lens of counterterrorism – a perspective that is still sorely lacking in technology assessment. The 4D framework offers great insights not only into how terrorists innovate but also into why terrorists (and by extension extremists) are more likely to use certain technologies maliciously than others. By applying the framework in the assessment of emerging technologies, TA can help prevent malevolently creative actors from using new technologies for their political goals.

In contrast, Dennis Klinkhammer takes a more hands-on-approach to demonstrate the potential of large language models (LLMs) to generate malicious content. Generative AI is becoming increasingly prevalent in society. However, soon after its emergence, more and more voices are pointing out the potential misuse of these tools and urging caution. The author shows how both direct and indirect manipulation can be used to get an LLM to generate malicious content aimed at the target group – possibly even against the security measures implemented. As a result, the production and distribution of misinformation, for example, could become much easier. It is therefore not surprising that this article is also a call for interdisciplinary cooperation to prevent malicious use cases and an important contribution to technology assessment.

It is crucial to take into account technological affordances while utilizing the potential of emerging technologies for the entire spectrum of prevention work.

As a challenge for international humanitarian law, Vasilios Tzoufis and Nikolaos Petropoulos turn their attention to the nexus of autonomous weapons systems and new forms of warfare. The issue of causation and guilt is central to their article: Who is ultimately responsible for any harm caused by an autonomous weapons system? Who will be criminally liable for any potential crime committed by these systems? These questions go to the core of international humanitarian law (IHL), and since autonomous weapons will change the very nature of war as we know it, the authors call on international actors to ensure that these new weapons will respect human rights in the future. Like other authors of the Special topic, Tzoufis and Petropou-
los call for international cooperation in this effort, with international uniform rules and protocols that set clear limits to the use of autonomous weapons systems in regular warfare. In this way, malicious and unlawful uses could be mitigated, at least in the arena of international relations, and, by extension, these international regimes could inform how innovators design their systems to comply with the norms of IHL.

International conventions to address the challenges posed by emerging technologies require national equivalents that take account of national specificities. Petr Machleitd, Jitka Mráčková, and Karel Mráček locate the need for AI regulation in relation to nationally nuanced risk perceptions by comparing (inter)national regulatory frameworks to Czech regulation. To find a holistic and measured approach to preventing and mitigating actual risks of malevolent AI use cases, the authors propose the application of the precautionary principle in regulation and education. The practice of TA could benefit from developing and implementing comprehensive approaches that address several challenges and research areas at the same time. Where these approaches lack clarity and focus, they can be situated in different socio-technical contexts. Situated approaches to holistic concepts can draw attention to commonalities, niches, and conflicting expectations. For example, the authors call for continuous assessment of responsibility and liability, as this would include the regulation of the entire value chain of emerging technologies, from development and innovation practices to use and misuse cases.

By the time technological innovations become relevant to the everyday life of users, their characteristics and thus also the fundamental opportunities for malicious use of technology have already been significantly shaped in previous innovation processes. Today, the innovation landscape and the ability to shape technologies is dominated by the private sector. The article by Niklas Henke presents a qualitative study that explores whether international cooperation in this effort, with international uniform rules and protocols that set clear limits to the use of autonomous weapons systems in regular warfare. In this way, malicious and unlawful uses could be mitigated, at least in the arena of international relations, and, by extension, these international regimes could inform how innovators design their systems to comply with the norms of IHL.

International conventions to address the challenges posed by emerging technologies require national equivalents that take account of national specificities. Petr Machleitd, Jitka Mráčková, and Karel Mráček locate the need for AI regulation in relation to nationally nuanced risk perceptions by comparing (inter)national regulatory frameworks to Czech regulation. To find a holistic and measured approach to preventing and mitigating actual risks of malevolent AI use cases, the authors propose the application of the precautionary principle in regulation and education. The practice of TA could benefit from developing and implementing comprehensive approaches that address several challenges and research areas at the same time. Where these approaches lack clarity and focus, they can be situated in different socio-technical contexts. Situated approaches to holistic concepts can draw attention to commonalities, niches, and conflicting expectations. For example, the authors call for continuous assessment of responsibility and liability, as this would include the regulation of the entire value chain of emerging technologies, from development and innovation practices to use and misuse cases.

By the time technological innovations become relevant to the everyday life of users, their characteristics and thus also the fundamental opportunities for malicious use of technology have already been significantly shaped in previous innovation processes. Today, the innovation landscape and the ability to shape technologies is dominated by the private sector. The article by Niklas Henke presents a qualitative study that explores whether international cooperation in this effort, with international uniform rules and protocols that set clear limits to the use of autonomous weapons systems in regular warfare. In this way, malicious and unlawful uses could be mitigated, at least in the arena of international relations, and, by extension, these international regimes could inform how innovators design their systems to comply with the norms of IHL.

International conventions to address the challenges posed by emerging technologies require national equivalents that take account of national specificities. Petr Machleitd, Jitka Mráčková, and Karel Mráček locate the need for AI regulation in relation to nationally nuanced risk perceptions by comparing (inter)national regulatory frameworks to Czech regulation. To find a holistic and measured approach to preventing and mitigating actual risks of malevolent AI use cases, the authors propose the application of the precautionary principle in regulation and education. The practice of TA could benefit from developing and implementing comprehensive approaches that address several challenges and research areas at the same time. Where these approaches lack clarity and focus, they can be situated in different socio-technical contexts. Situated approaches to holistic concepts can draw attention to commonalities, niches, and conflicting expectations. For example, the authors call for continuous assessment of responsibility and liability, as this would include the regulation of the entire value chain of emerging technologies, from development and innovation practices to use and misuse cases.

By the time technological innovations become relevant to the everyday life of users, their characteristics and thus also the fundamental opportunities for malicious use of technology have already been significantly shaped in previous innovation processes. Today, the innovation landscape and the ability to shape technologies is dominated by the private sector. The article by Niklas Henke presents a qualitative study that explores whether international cooperation in this effort, with international uniform rules and protocols that set clear limits to the use of autonomous weapons systems in regular warfare. In this way, malicious and unlawful uses could be mitigated, at least in the arena of international relations, and, by extension, these international regimes could inform how innovators design their systems to comply with the norms of IHL.

International conventions to address the challenges posed by emerging technologies require national equivalents that take account of national specificities. Petr Machleitd, Jitka Mráčková, and Karel Mráček locate the need for AI regulation in relation to nationally nuanced risk perceptions by comparing (inter)national regulatory frameworks to Czech regulation. To find a holistic and measured approach to preventing and mitigating actual risks of malevolent AI use cases, the authors propose the application of the precautionary principle in regulation and education. The practice of TA could benefit from developing and implementing comprehensive approaches that address several challenges and research areas at the same time. Where these approaches lack clarity and focus, they can be situated in different socio-technical contexts. Situated approaches to holistic concepts can draw attention to commonalities, niches, and conflicting expectations. For example, the authors call for continuous assessment of responsibility and liability, as this would include the regulation of the entire value chain of emerging technologies, from development and innovation practices to use and misuse cases.

By the time technological innovations become relevant to the everyday life of users, their characteristics and thus also the fundamental opportunities for malicious use of technology have already been significantly shaped in previous innovation processes. Today, the innovation landscape and the ability to shape technologies is dominated by the private sector. The article by Niklas Henke presents a qualitative study that explores whether international cooperation in this effort, with international uniform rules and protocols that set clear limits to the use of autonomous weapons systems in regular warfare. In this way, malicious and unlawful uses could be mitigated, at least in the arena of international relations, and, by extension, these international regimes could inform how innovators design their systems to comply with the norms of IHL.

International conventions to address the challenges posed by emerging technologies require national equivalents that take account of national specificities. Petr Machleitd, Jitka Mráčková, and Karel Mráček locate the need for AI regulation in relation to nationally nuanced risk perceptions by comparing (inter)national regulatory frameworks to Czech regulation. To find a holistic and measured approach to preventing and mitigating actual risks of malevolent AI use cases, the authors propose the application of the precautionary principle in regulation and education. The practice of TA could benefit from developing and implementing comprehensive approaches that address several challenges and research areas at the same time. Where these approaches lack clarity and focus, they can be situated in different socio-technical contexts. Situated approaches to holistic concepts can draw attention to commonalities, niches, and conflicting expectations. For example, the authors call for continuous assessment of responsibility and liability, as this would include the regulation of the entire value chain of emerging technologies, from development and innovation practices to use and misuse cases.

By the time technological innovations become relevant to the everyday life of users, their characteristics and thus also the fundamental opportunities for malicious use of technology have already been significantly shaped in previous innovation processes. Today, the innovation landscape and the ability to shape technologies is dominated by the private sector. The article by Niklas Henke presents a qualitative study that explores whether international cooperation in this effort, with international uniform rules and protocols that set clear limits to the use of autonomous weapons systems in regular warfare. In this way, malicious and unlawful uses could be mitigated, at least in the arena of international relations, and, by extension, these international regimes could inform how innovators design their systems to comply with the norms of IHL.
The apparent failure to anticipate the malevolent potential of Web 2.0, which has partly facilitated the decline of democracy seen around the world today, must not be repeated in the age of AI and immersive realities. And while malevolent creativity and use cases can never be completely prevented, it is still important to anticipate potential malevolent uses to keep technology, and by extension societies, as safe as possible.

Much more attention is needed to truly establish civil security and malevolent use of technologies as a central topic of TA, but we hope that with this Special topic the first step has been taken.

Funding • This research was supported by the German Federal Ministry of Education and Research within the framework of the program MOTRA (Grant No. MOTRA-13N15218).

Competing interests • The authors declare no competing interests.

References
Dessecker, Axel; Fecher, Lena; Hirth, Maria-Anna; Hofmann, Rebecca; Muzaqi, Ladvim (2024): Medien- und Technologienutzung durch jahdistische Straftäter*innen. Wiesbaden: MOTRA-Verbund. https://doi.org/10.57671/MOTRA-202401/02
doi.org/10.1016/j.techsoc.2015.01.002
Schmid, Alex (2013): The definition of terrorism. In: Alex Schmid (ed.): The Rout-
Schmid, Alex (2020): Terrorism prevention. Conceptual issues (definitions, typo-
logies and theories). In: Alex Schmid (ed.): Handbook of terrorism preven-
https://www.icct.nl/sites/default/files/2023-01/Chapter-2-Handbook-
pdfhttps://www.motra.info/motra-monitor-2022/, last accessed on
02. 05. 2024.
Schot, Johan, Rip, Arie (1997): The past and future of constructive technology
assessment. In: Technological Forecasting and Social Change 54 (2–3),
pp. 251–268. https://doi.org/10.1016/S0040-1625(96)00180-1
Stilgoe, Jack; Owen, Richard; Macnaghten, Phil (2013): Developing a framework
https://doi.org/10.1016/j.respol.2013.05.008
https://doi.org/10.1038/4681029a
Ware, Jacob (2023): The third generation of online radicalization. Washington,
DC: George Washington University

ALEXANDROS GAZOS
is a member of the scientific staff at ITAS and a PhD student at KIT. He is currently working in the “MOTRA – Technology Monitoring (TM)” project. His research focuses on the resilience and vulnerabilities of critical infrastructures as well as the sociology of technology.

Call for Abstracts

“Beyond short-termism:
Strategies and perspectives for the long-term governance of socio-technical change”

TATuP Special topic in 34/2 (2025). Please send your submission by 12 August 2024

Please read the complete CfA at https://www.tatup.de

Guest editors of this TATuP Special topic:
Dr. Stefania Sardo¹, Dr. Sophie Kuppler¹, and PD Dr. Dirk Scheer¹

¹ Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology