

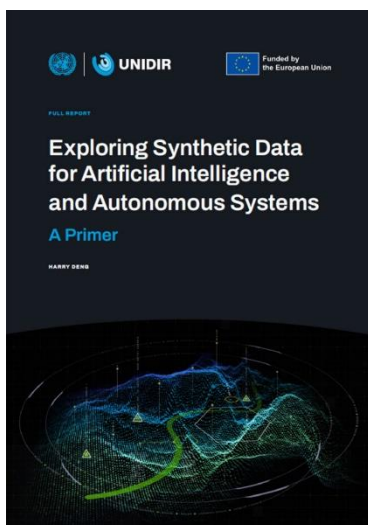
UNIDIR Publications on Lethal Autonomous Weapons and Military Artificial Intelligence

1. Exploring Synthetic Data: for Artificial Intelligence and Autonomous Systems: A Primer



2023

<https://unidir.org/publication/exploring-synthetic-data-for-artificial-intelligence-and-autonomous-systems-a-primer/>



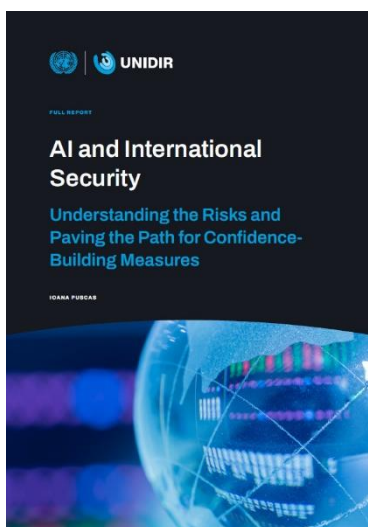
Synthetic data refers to artificially created data that seeks to reproduce the characteristics of real-world datasets in order to have a beneficial effect on training highly complex AI systems. This UNIDIR Primer provides an overview of the main opportunities and limitations of synthetic data in the training of AI systems. While synthetic data can be a proxy for real-world data and help shorten training cycles, among other benefits, there are also significant risks and challenges associated with its use. The Primer explores existing data challenges, both technical and organizational, introduces key technical characteristics and methods of generating synthetic data, and analyzes implications of using synthetic data in the context of international security, including for autonomous systems and in the cyber realm.

2. AI and International Security: Understanding the Risks and Paving the Path for Confidence-Building Measures



2023

<https://unidir.org/publication/ai-and-international-security-understanding-the-risks-and-paving-the-path-for-confidence-building-measures/>



Advances in artificial intelligence (AI) in recent years, combined with the technology's scalability and convergence with other domains, have prompted numerous concerns about the risks of AI to global security, including risks of misuse and escalation. However, policy discussions still lack a comprehensive analysis of the technology's risks and how categories of risks are interrelated. This report provides an overview of the main categories of risks of AI in the context of international peace and security, across domains of use and applications.

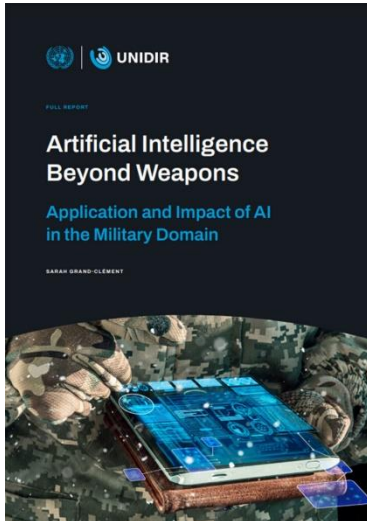
This research concludes phase one of the UNIDIR project on CBMs for AI. It provides a basis for multi-stakeholder engagements to understand the risks and to advance discussions about CBMs, which can help promote a more transparent, safe and responsible environment for the development and use of AI.

3. Artificial Intelligence Beyond Weapons: Application and Impact of AI in the Military Domain



2023

<https://unidir.org/publication/internet-fragmentation-and-cybersecurity-a-primer/>



Within the United Nations, the application of artificial intelligence (AI) in the military domain has, to-date, primarily discussed in the context of the United Nations Group of Governmental Experts (GGE) on emerging technologies in the area of lethal autonomous weapons systems (LAWS). However, the application of AI within the military domain extends beyond the issue of LAWS.

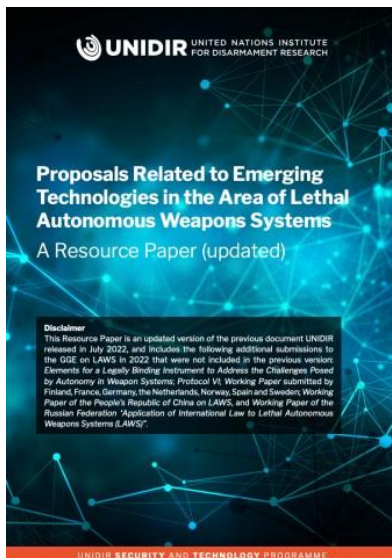
In the midst of discussions and debates around the opportunities and risks of AI for military purposes, as well as the governance and responsible use of these technologies, the United Nations Institute for Disarmament Research (UNIDIR)'s new report aims to increase understanding of the role of AI in the execution of military tasks beyond applications relating to the use of force and the narrow tasks of target selection and target engagement within the targeting process. The report provides an overview of current and near-future AI capabilities relevant to aiding with 18 military tasks. The paper also presents a discussion on the strengths and limitations regarding the application of AI to these military tasks.

4. Proposals Related to Emerging Technologies in the Area of Lethal Autonomous Weapons Systems: A Resource Paper (updated)



2023

<https://unidir.org/publication/proposals-related-emerging-technologies-area-lethal-autonomous-weapons-systems-resource>



This resource paper offers a comparative analysis of the content of the different proposals related to emerging technologies in the area of lethal autonomous weapon systems (LAWS) submitted by States to the Group of Governmental Experts on LAWS up until the end of 2022.

It identifies commonality in views as well as areas that require further discussion in relation to eleven thematic areas covered in the proposals and the Group's discussions.

Also available: [Annex A](#), which includes relevant excerpts from proposals related to emerging technologies in the area of lethal autonomous weapons systems.

5. The 2022 Innovations Dialogue: AI Disruption, Peace and Security (Conference Report)



2023

<https://undir.org/publication/2022-innovations-dialogue-ai-disruption-peace-and-security-conference-report>



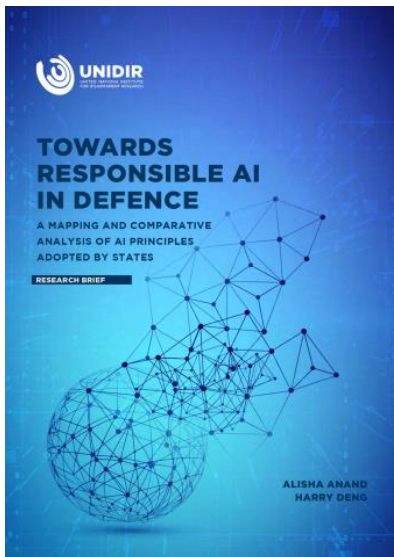
This report provides a summary of the key themes, issues and takeaways that emerged from discussions during the 2022 Innovations Dialogue. Part I of the report seeks to provide a foundational understanding of the concept of AI and its state of play. Part II examines the disruptive impact of AI on international peace and security. In particular, it discusses the risks and benefits of uses of AI in military operations and across domains of warfare as well as the opportunities and challenges of harnessing AI technologies for conflict prevention and peacebuilding. Part III of the report examines the path to Responsible AI. It unpacks the RAI governance approach and discusses how it is and can be operationalized. It also reflects on the value of building an RAI culture.

6. Towards Responsible AI in Defence: A Mapping and Comparative Analysis of AI Principles Adopted by States



2023

<https://www.undir.org/publication/towards-responsible-ai-defence-mapping-and-comparative-analysis-ai-principles-adopted>



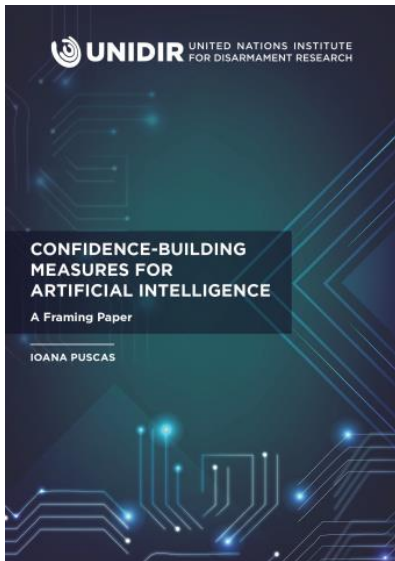
UNIDIR's project Towards Responsible AI in Defence seeks to, first, build a common understanding of the key facets of responsible research, design, development, deployment, and use of AI systems. It will then examine the operationalisation of Responsible AI in the defence sector, including identifying and facilitating the exchange of good practices. This research brief provides an overview of the aims of the project. It also outlines the research methodology for and preliminary findings of the project's first phase: the development of a common taxonomy of principles and a comparative analysis of AI principles adopted by states.

7. Confidence-Building Measures for Artificial Intelligence: A Framing Paper



2022

<https://www.unidir.org/publication/confidence-building-measures-artificial-intelligence-framing-paper>



Confidence-building measures (CBMs) for artificial intelligence (AI) can provide flexible options for the future development and deployment of AI-enabled systems. This framing paper introduces a new UNIDIR project, which aims at developing a possible roadmap for the future elaboration of CBMs for AI. The first phase of the project consists of a risk-mapping analysis, unpacking risks of the technology and assessing how they may translate into risks for international peace and security. The second phase of the project will consider pathways for the elaboration of CBMs in a series of multi-stakeholder engagements.

8. Human-Machine Interfaces in Autonomous Weapon Systems



2022

<https://www.unidir.org/publication/human-machine-interfaces-autonomous-weapon-systems>



Human-machine interfaces (HMIs), the hardware and software components that ensure the communication between machines and human operators, have been touted as a key element of human control over autonomous weapons systems. Strictly speaking, this is correct, as interfaces allow operators to monitor the state and behaviour of a remote system, as well as to manually (re-)establish control when needed, but the optimal use of HMIs relies on a complex mix of factors. These include context of use, ergonomics/human factors design, personnel training, the complexities of decision-making and human-machine interaction in dynamic environments, and the unique challenges related to AI/ML-based learning systems. This report analyses HMIs in autonomous systems and explores key considerations of HMIs use, testing and design. Drawing on these, it considers

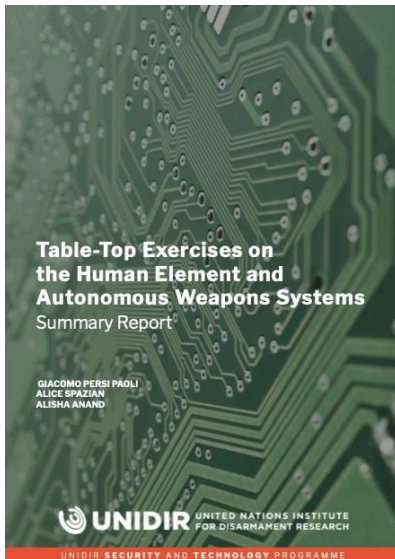
implications for human control and proposes recommendations for the work of the GGE on LAWS.

9. Table-Top Exercises on the Human Element and Autonomous Weapons Systems: Summary Report



2021

<https://www.unidir.org/TTXonLAWS>



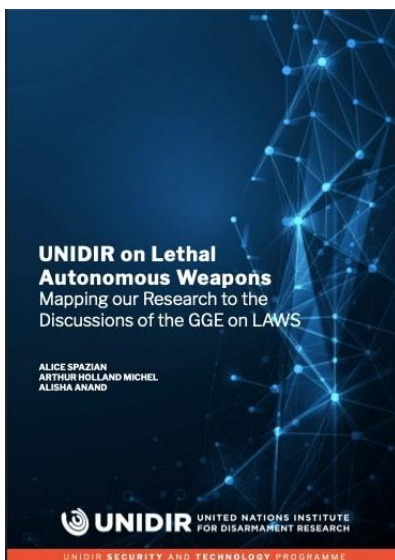
This report summarises the findings of UNIDIR's series of regional table-top exercises conducted between September 2020 and June 2021. The project brought together 198 experts from 75 countries to discuss the technical, military and legal implications of introducing autonomy in various steps of the targeting cycle. By summarising the main findings of this series of exercises, this report aims at creating a common knowledge base to support better-informed negotiations related to a normative and operational framework for Lethal Autonomous Weapons Systems (LAWS).

10. UNIDIR on Lethal Autonomous Weapons: Mapping our Research to the Discussions of the GGE on LAWS



2021

<https://www.unidir.org/mapping-research-GGE-LAWS>



The United Nations Institute for Disarmament Research has been studying the weaponization of increasingly autonomous technologies since 2013. Over this period, the Institute has produced 15 in-depth studies on this issue. These projects have largely focused on considerations that the Group of Government Experts (GGE) on emerging technologies in the area of lethal autonomous weapon systems (LAWS) has identified as having particular relevance to its development of consensus-based recommendations on a potential normative and operational framework for LAWS. To support the GGE's ongoing discussions in 2021 and beyond, this report summarises a range of key contributions that these studies have made in these areas. Each section provides a brief summary of the topic as it has been discussed within the context of the Group, followed by a summary of our relevant

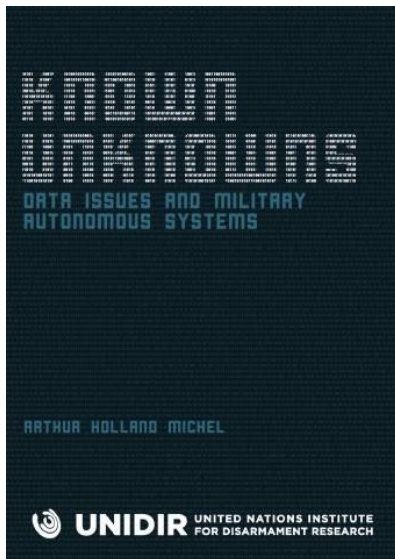
findings sourced to the reports in which those findings are presented.

11. Known Unknowns: Data Issues and Military Autonomous Systems



2021

<https://unidir.org/known-unknowns>



In order to perform as desired, autonomous systems must collect data that are complete, relevant, accurate, and aligned with the data for which the system was developed and tested. But the harsh, dynamic, complex and adversarial nature of conflict environments poses a wide range of obstacles to the collection of such data. As a result, autonomous systems cannot always be expected to achieve the exact same performance in real-world use that they demonstrated in development or testing. And crucially, they will be liable to failures that are both inevitable and impossible to anticipate: "known unknowns."

Data and its vagaries therefore have significant implications for the application of international humanitarian law and other rules of war. This report describes common data issues for autonomous systems and explains how they give rise to

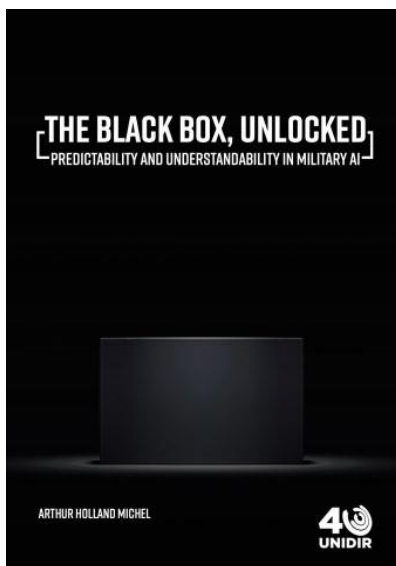
"known unknown" failures. It then explores the legal and operational implications of such failures, and considers a range of potential policy and technical solutions by which they could be addressed.

12. The Black Box, Unlocked: Predictability and Understandability in Military AI



2020

<https://unidir.org/publication/black-box-unlocked>



Predictability and understandability are widely held to be vital characteristics of artificially intelligent systems. Put simply: AI should do what we expect it to do, and it must do so for intelligible reasons. This consideration stands at the heart of the ongoing discussion about lethal autonomous weapon systems and other forms of military AI. But what does it mean for an intelligent system to be "predictable" and "understandable" (or, conversely, unpredictable and unintelligible)? What is the role of predictability and understandability in the development, use, and assessment of military AI? What is the appropriate level of predictability and understandability for AI weapons in any given instance of use? And how can these thresholds be assured?

This study provides a clear, comprehensive introduction to these questions, and proposes a range of avenues for action

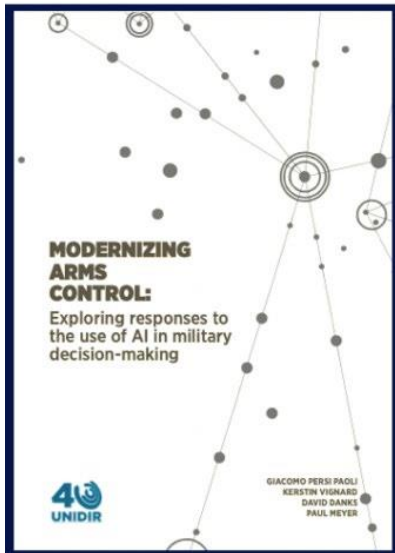
by which they may be addressed.

13. Modernizing Arms Control: Exploring Responses to the Use of AI in Military Decision-Making



2020

<https://undir.org/publication/modernizing-arms-control>



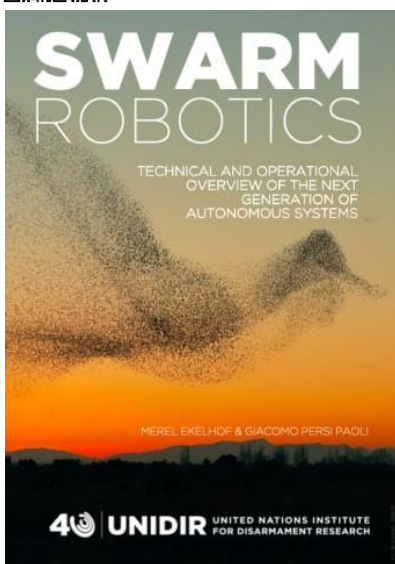
This report provides an initial insight into why the international security community may need to consider regulating artificial intelligence (AI) applications that fall in the digital grey zone between AI-enabled weapon systems (e.g. lethal autonomous weapon systems) and military uses of civilian AI applications (e.g. logistics, transport). It also provides an initial exploration of the familiar tools the community has at its disposal for such regulation.

14. Swarm Robotics: Technical and Operational Overview of the Next Generation of Autonomous Systems



2020

<https://www.undir.org/publication/swarm-robotics-technical-and-operational-overview-next-generation-autonomous-systems>



As the international community continues discussions on lethal autonomous weapons systems (LAWS) in 2020 and 2021 and will focus on the further development and operationalization of the guiding principles, the role of human decision-making will undoubtedly remain one of the core issues. By drawing on near-term technologies, such as swarms, and related command and control models in deliberations about human control and human-machine interaction, the international community can move to develop a more comprehensive understanding of how control may or may not be exercised in military practice – now and in future operations.

An executive summary of this publication is available in English, French, Spanish:



English: <https://undir.org/publication/swarm-robotics-technical-and-operational-overview-next-generation-autonomous-systems>



French: <https://unidir.org/publication/les-essaims-de-robots-dossier-de-recherche>



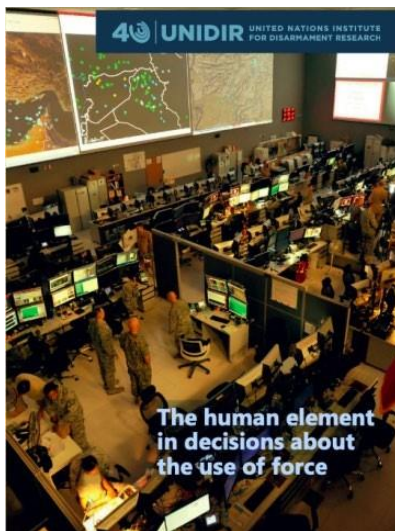
Spanish: <https://unidir.org/publication/robotica-de-enjambre-informe-de-investigacion>

15. The Human Element in Decisions About the Use of Force



2020

<https://unidir.org/publication/human-element-decisions-about-use-force>



Since governments began expert meetings on lethal autonomous weapons systems (LAWS) in the context of the Convention on Certain Conventional Weapons in 2014, maintaining control over emerging technologies in the area of LAWS has been one of the main shared objectives.

States have suggested maintaining human control over weapons, the critical functions of weapons, attacks, the targeting process, and (final) decisions to use force. Although most agree that human control should be more meaningful than the mere possibility of aborting an attack at the final moment, the international community is struggling to determine how the human role in the use of (lethal) force should be defined and implemented.

This infographic offers a framework of the human role in military decision-making – at the strategic, operational and tactical levels – that may guide deeper discussion on the military and legal aspects of human control within the LAWS debate.

The purpose of this infographic is two-fold: 1) it demonstrates how critical decisions about the use of force are taken at various levels and how they may influence one another, 2) it provides some key legal considerations for decision-makers at various stages in the process.

This publication is also available in: [French](#), [Spanish](#):



French: <https://unidir.org/publication/lelement-humain-dans-les-decisions-relatives-lutilisation-de-la-force>



Spanish: <https://unidir.org/publication/el-elemento-humano-en-las-decisiones-sobre-el-uso-de-la-fuerza>

16. The Role of Data in Algorithmic Decision-Making



2019

<https://unidir.org/publication/role-data-algorithmic-decision-making>



The Role of Data
in Algorithmic Decision-Making

A Primer

Lydia Kostopoulos

UNIDIR RESOURCES

The international discussion on military applications of machine learning has focused much more on algorithmic decision-making than on data. Yet without data, an algorithm cannot be trained, and even after training, an algorithm requires data in its processing to make decisions.

This primer explores data in the context of military decision-support tools and increasingly autonomous weapons systems by briefly discussing the links of the data chain (creation, collection, organization and use), potential challenges to data integrity in adversarial environments, and concludes with a few forward-looking questions for policymakers considering military applications of increasingly autonomous systems.

17. Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies



2018

<https://unidir.org/publication/algorithmic-bias-and-weaponization-increasingly-autonomous-technologies>



Algorithmic Bias and the
Weaponization of Increasingly
Autonomous Technologies

A PRIMER

No. 9 UNIDIR RESOURCES

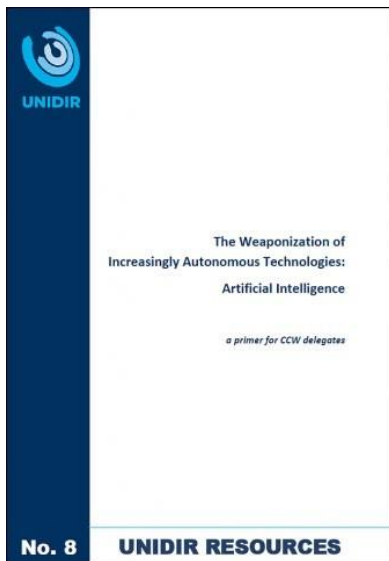
AI-enabled systems depend on algorithms, but those same algorithms are susceptible to bias. Algorithmic biases come in many types, arise for a variety of reasons, and require different standards and techniques for mitigation. This primer characterizes algorithmic biases, explains their potential relevance for decision-making by autonomous weapons systems, and raises key questions about the impacts of and possible responses to these biases.

18. The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence



2018

<https://www.unidir.org/publication/weaponization-increasingly-autonomous-technologies-artificial-intelligence>



The rapidly advancing field of AI and machine learning has significant implications for the role of autonomy in weapon systems. States face the daunting task of trying to understand the legal, policy, ethical, strategic, and other considerations of a technology that is rapidly evolving. This paper is an introductory primer for non-technical audiences on the current state of AI and machine learning, designed to support the international discussions on the weaponization of increasingly autonomous technologies.

19. The Weaponization of Increasingly Autonomous Technologies: Autonomous Weapon Systems and Cyber Operations



2017

<https://www.unidir.org/publication/weaponization-increasingly-autonomous-technologies-autonomous-weapon-systems-and-cyber>



International discussions about autonomous weapons have thus far focused predominantly on conventional weapon systems. These systems are not, however, the only domain in which technological developments in autonomy can have an impact on international security. Rapid advances in machine learning and artificial intelligence also have a significant impact in the field of cyber security, and in particular for offensive operations carried out in cyberspace, so-called "cyber operations". As this paper explains, the interaction of cyber operations and increasingly autonomous physical weapon systems may give rise to new security challenges, as these interactions can multiply complexity and introduce new vulnerabilities.

20. The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches



2017

<https://www.unidir.org/publication/weaponization-increasingly-autonomous-technologies-concerns-characteristics-and>



Agreeing on a working definition of LAWS will be a challenging endeavour, as there are several working definitions already in circulation, and some stakeholders have already stated a preferred policy response. Moreover, each proposed definition attends to a particular set of concerns and characteristics, while omitting others.

One's position on both an appropriate definition and an adequate policy response ultimately depends on what one is concerned about. Different definitions will attend different sets of concerns, as well as privilege different sets of characteristics.

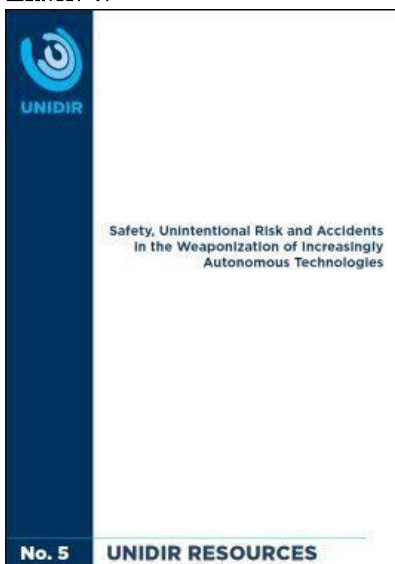
The objective of this primer is to consolidate and give an overview of both concerns and characteristics and illustrate how different definitional approaches attend to these.

21. Safety, Unintentional Risk and Accidents in the Weaponization of Increasingly Autonomous Technologies



2016

<https://unidir.org/publication/safety-unintentional-risk-and-accidents-weaponization-increasingly-autonomous>



Recent international attention on autonomous weapon systems (AWS) has focused on the implications of what amounts to a 'responsibility gap' in machine targeting and attack in war. As important as this is, the full scope for accidents created by the development and deployment of such systems is not captured in this debate. It is necessary to reflect on the potential for AWS to fail in ways that are unanticipated and harmful to humans—a broader set of scenarios that simply those in which international humanitarian law applies.

Of course, any complex, hazardous technology carries 'unintentional' risk, and can have harmful results its designers and operators did not intend. AWS may pose novel, unintended forms of hazard to human life that typical approaches to ensuring responsibility do not effectively

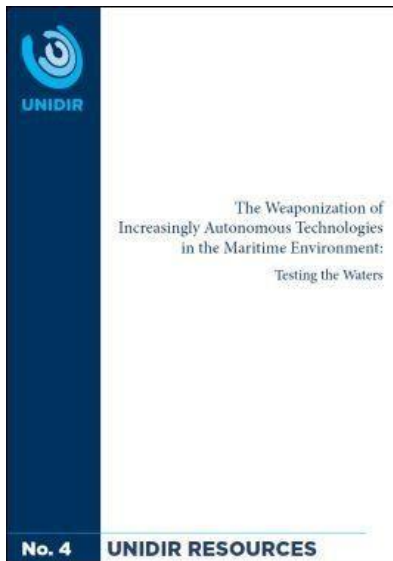
manage because these systems may behave in unpredictable ways that are difficult to prevent. Among other things, this paper suggests human-machine teams would, on their own, be insufficient in ensuring unintended harm from AWS is prevented, something that should bear on discussions about the acceptability of deploying these systems. This is the fifth in a series of UNIDIR papers on the weaponization of increasingly autonomous technologies.

22. The Weaponization of Increasingly Autonomous Technologies in the Maritime Environment: Testing the Waters



2015

<https://unidir.org/publication/weaponization-increasingly-autonomous-technologies-maritime-environment-testing-waters>



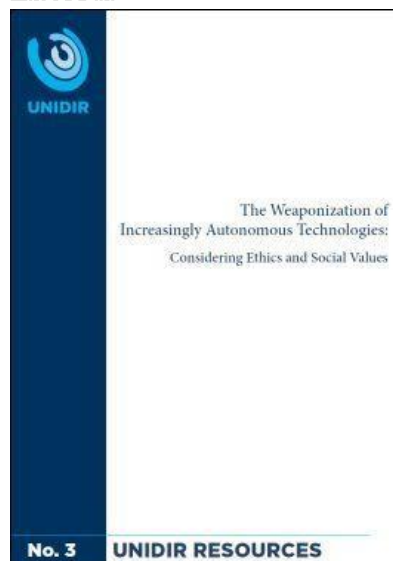
Recent attention among governments, civil society organizations and the media has focused on technical, military, legal and ethical issues of the weaponization of increasingly autonomous technologies. Experts have suggested that fully autonomous weapons are likely to first appear in the relatively “uncluttered” maritime environment. Yet, policy-makers have directed relatively little attention to the specific issues and challenges in this environment that might be different or more acute than on land or in the air. This paper aims to shed light on these issues in order to inform the broader debate on the weaponization of increasingly autonomous technologies. It is the fourth in a series of UNIDIR papers on this theme.

23. The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values



2015

<https://unidir.org/publication/weaponization-increasingly-autonomous-technologies-considering-ethics-and-social-values>



Discussions on the weaponization of increasingly autonomous technologies most often focus on technical aspects of the weapon being considered, potential military missions and legality. This UNIDIR paper highlights some of the ethical and social issues that arise from—and underlie—this discussion. It suggests that far from being extraneous to the policy debate on the weaponization of increasingly autonomous technologies, ethics and social values are close to the core of this discussion. Although legal and technical discussions may produce information about possible technological trajectories, future applications and rules, they will not necessarily produce the insights, wisdom and prudence needed for sound policy that will serve national and international interests. This short paper is aimed at encouraging reflection on different ways that ethics, broadly construed to include social or cultural

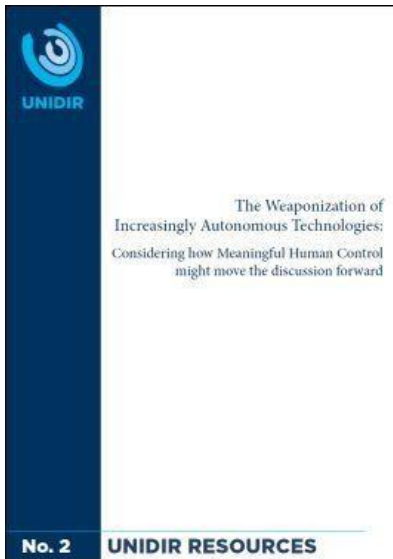
values, might influence consideration of the weaponization of increasingly autonomous technologies.

24. The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward



2014

<https://unidir.org/publication/weaponization-increasingly-autonomous-technologies-considering-how-meaningful-human>



Recent discussions in a variety of intergovernmental, academic and policy forums have considered, among other issues, the objective of maintaining “meaningful human control” over increasingly autonomous weapons. This UNIDIR paper examines what may be understood by “meaningful human control”, its strengths and weaknesses as a framing concept for discussions on autonomy and weapon systems, as well as other conceptual and policy-oriented approaches that address concerns about the weaponization of increasingly autonomous technologies. It is the second in a series of UNIDIR papers on the weaponization of increasing autonomous technologies.

25. Framing Discussions on the Weaponization of Increasingly Autonomous Technologies



2014

<https://unidir.org/publication/framing-discussions-weaponization-increasingly-autonomous-technologies>



There are currently discussions in a variety of national and international fora about autonomy and weapon systems. Yet governments are unsure of what they need to know in order to make responsible policy choices—and not all agree that specific policy is necessary. As these are early days in international, multilateral engagement on this issue, this paper seeks to help frame further dialogue on autonomy and weapon systems in a way that is both concise and relevant to policy-making, by helping direct attention to key issues and the areas of greatest concern.