**ARTIFICIAL INTELLIGENCE, WARFARE, AND EVIDENCE**

**IN INTERNATIONAL CONTEXTS**

**A Select Annotated Bibliography**

Patrick McGee

Archival Technologies Lab, City University of New York

For the Creation and Use Working Group, InterPARES Trust AI

March 2025

## **Table of Contents**

[Table of Contents 1](#_rie2adpf4y44)

[Introduction 2](#_939llzvsfra9)

[Methodology 3](#_145i7879qezg)

[Annotated Bibliography 4](#_i65zs7u3j7aw)

[Books and Special Issues 4](#_h9vzj5lx5yt0)

[Book Chapters 17](#_4dn8kdr3iqys)

[Peer Reviewed Articles 24](#_nsw35hyrcnzb)

[Non-peer Reviewed Articles and Journalism 66](#_r9ytwx7tb789)

[Reports and Protocols 76](#_u1k40g4dh683)

[Conference Papers, Proceedings, and Workshops 105](#_xmx791e63ul7)

[Dissertations and Theses: 110](#_3wr4ri7qex9m)

[Projects and Campaigns 115](#_yrvqs5wep3h9)

## 

## 

## **Introduction**

With *+972 Magazine’s* exposé of Israel’s use of artificial intelligence (AI) for target production in Gaza, and the citation of the report in South Africa’s International Court of Justice (ICJ) case against the State of Israel, the issue of the adjudication of artificial intelligence by international courts has become a global reality. This annotated bibliography collects peer-reviewed journal articles, news reports, resolutions, research briefings, social media and other sources relating to the issues of artificial intelligence, warfare, and evidence in international contexts. These writings are collected as a preliminary step towards further research relating artificial intelligence to authenticity in evidence before international courts.

Many of the pieces surveyed here seek to understand artificial intelligence / machine learning, and autonomous weapon systems (AWS) in particular, as they relate to the norms of international humanitarian law (IHL). For this reason, there are more often than not ambiguities where one might seek precedent. There has been little use of open-source data as evidence in international criminal courts, and much of the writing culminates in speculation. Many of the writers claim that laws must shift in order to judge cases of war crimes involving AI or AWS, for their adjudication still falls under the vague concepts of *just war,* as determined by international bodies after World War II. There is a great need for scholarship on how material evidence of military use of artificial intelligence is processed and used as evidence, as the scholarship regarding legal precedent and code develops. Archival and information theorists should rise to this challenge, bringing the collective knowledge developed through the InterPARES program to bear on the military use of artificial intelligence. The black box of this technology obscures human agency, but as civilian casualties mount around the world, and belligerent nation-states exercise asymmetric power against the colonized and dispossessed, we must work to preserve authentic proof of culpability and complicity.

It is hoped that this bibliography may serve as a basis for further study. As we face the onslaught of imperial violence around the world, we must ensure the logic of this violence does not go uncontested. Gaining familiarity with contemporary articulations of international humanitarian law and its relationship to artificial intelligence is one step toward a vital international information ethics.

## **Methodology**

This annotated bibliography developed from a literature search using the following tools and search terms. Upon finding pieces relating to the subject, the author traced footnotes and references to provide comprehensive oversight of the subject and discourse. These led to some of the most fruitful discoveries. The annotations consist of the authors’ abstracts when available, or a brief excerpt from the piece representative of its content. In respect to the government statements in the *Reports & Protocols* section, the original documents are too short to provide an annotation.

*Search engines used:*

Google and Google Scholar

CUNY Queens College Library OneSearch

*Search term combinations:*

artificial intelligence + evidence + international + warfare

Artificial intelligence + evidence + international courts + warfare

artificial intelligence + evidence + international law

Artificial intelligence + authenticity + international law + warfare

Military artificial intelligence + evidence + international + warfare

*Journals:*

Journal of International Criminal Justice

*Books:*

This section contains any bookwith two or more chapters devoted to the subject. Individual chapters are compiled in the following section. While there was no date range set, the relatively new nature of the subject means that all of the books have been published in the 21st century.

Recourse was also made to UNIDIR and UNODA for resource discovery. These bodies of the United Nations are dedicated to disarmament research and affairs. Related publications are included in “Reports & Protocols,” and UNIDIR’s focus area *Artificial Intelligence* is included in “Projects & Campaigns.”

## **Annotated Bibliography**

## 

## **Books and Special Issues**

**Arkin, R. C. (2009). *Governing Lethal Behavior in Autonomous Robots*. Boca Raton: CRC Press.**

Expounding on the results of the author’s work with the US Army Research Office, DARPA, the Office of Naval Research, and various defense industry contractors, Governing Lethal Behavior in Autonomous Robots explores how to produce an "artificial conscience" in a new class of robots, *humane-oids*, which are robots that can potentially perform more ethically than humans in the battlefield. The author examines the philosophical basis, motivation, theory, and design recommendations for the implementation of an ethical control and reasoning system in autonomous robot systems, taking into account the Laws of War and Rules of Engagement.

The book presents robot architectural design recommendations for

* *Post facto* suppression of unethical behavior,
* Behavioral design that incorporates ethical constraints from the onset,
* The use of affective functions as an adaptive component in the event of unethical action, and
* A mechanism that identifies and advises operators regarding their ultimate responsibility for the deployment of autonomous systems.

It also examines why soldiers fail in battle regarding ethical decisions; discusses the opinions of the public, researchers, policymakers, and military personnel on the use of lethality by autonomous systems; provides examples that illustrate autonomous systems’ ethical use of force; and includes relevant Laws of War.

Helping ensure that warfare is conducted justly with the advent of autonomous robots, this book shows that the first steps toward creating robots that not only conform to international law but outperform human soldiers in their ethical capacity are within reach in the future. It supplies the motivation, philosophy, formalisms, representational requirements, architectural design criteria, recommendations, and test scenarios to design and construct an autonomous robotic system capable of ethically using lethal force.

**Bode, I., & Huelss, H. (2022). *Autonomous weapons systems and international norms*. McGill-Queen's Press-MQUP.**

Autonomous weapons systems seem to be on the path to becoming accepted technologies of warfare. The weaponization of artificial intelligence raises questions about whether human beings will maintain control of the use of force. The notion of meaningful human control has become a focus of international debate on lethal autonomous weapons systems among members of the United Nations: many states have diverging ideas about various complex forms of human-machine interaction and the point at which human control stops being meaningful.

In Autonomous Weapons Systems and International Norms Ingvild Bode and Hendrik Huelss present an innovative study of how testing, developing, and using weapons systems with autonomous features shapes ethical and legal norms, and how standards manifest and change in practice. Autonomous weapons systems are not a matter for the distant future - some autonomous features, such as in air defence systems, have been in use for decades. They have already incrementally changed use-of-force norms by setting emerging standards for what counts as meaningful human control. As UN discussions drag on with minimal progress, the trend towards autonomizing weapons systems continues.

A thought-provoking and urgent book, Autonomous Weapons Systems and International Norms provides an in-depth analysis of the normative repercussions of weaponizing artificial intelligence.

**Caron, J. (2019). *Contemporary technologies and the morality of warfare: The war of the machines*. Routledge.**

Supported by genuine historical cases, this book argues that certain new technologies in warfare can not only be justified within the current framework of the just war theory, but that their use is mandatory from a moral perspective.

Technological developments raise questions about the manner in which wars ought to be fought. The growing use of drones, capacity-increasing technologies, and cyberattacks are perceived by many as posing great challenges to Just War Theory. Instead of seeing these technologies as inherently unethical, this book adopts a different perspective by arguing that they are morally necessary since they can limit the potential violations of the moral rules of war and ensure that militaries better respect their obligation to protect their members. Caron’s research offers insights into how and under what conditions autonomous or semi-autonomous robots, artificial intelligence, cyberwarfare, and capacity increasing technologies can be considered as legitimate weapons.

This book will be of interest to students, members of the armed forces, and scholars studying Politics, International Relations, Security Studies, Ethics, and Just War Theory.

**Cristiano, F., et al. (Eds). (2023). *Artificial Intelligence and International Conflict in Cyberspace (1st ed.).* Routledge.** [**https://doi.org/10.4324/9781003284093**](https://doi.org/10.4324/9781003284093)

This edited volume explores how artificial intelligence (AI) is transforming international conflict in cyberspace.

Over the past three decades, cyberspace developed into a crucial frontier and issue of international conflict. However, scholarly work on the relationship between AI and conflict in cyberspace has been produced along somewhat rigid disciplinary boundaries and an even more rigid sociotechnical divide – wherein technical and social scholarship are seldomly brought into a conversation. This is the first volume to address these themes through a comprehensive and cross-disciplinary approach. With the intent of exploring the question ‘what is at stake with the use of automation in international conflict in cyberspace through AI?’, the chapters in the volume focus on three broad themes, namely: (1) technical and operational, (2) strategic and geopolitical and (3) normative and legal. These also constitute the three parts in which the chapters of this volume are organised, although these thematic sections should not be considered as an analytical or a disciplinary demarcation.

This book will be of much interest to students of cyber-conflict, AI, security studies and International Relations.

**Dubberley, S., Koenig, A. & Murray, D. (eds). (2019). *Digital witness: Using open source information for human rights investigation, documentation, and accountability.* Oxford Academic.** [**https://doi.org/10.1093/law/9780198836063.001.0001**](https://doi.org/10.1093/law/9780198836063.001.0001)

From videos of rights violations, to satellite images of environmental degradation, to eyewitness accounts disseminated on social media, human rights practitioners have access to more data today than ever before. To say that mobile technologies, social media, and increased connectivity are having a significant impact on human rights practice would be an understatement. Modern technology – and the enhanced access it provides to information about abuse – has the potential to revolutionise human rights reporting and documentation, as well as the pursuit of legal accountability. However, these new methods for information gathering and dissemination have also created significant challenges for investigators and researchers. For example, videos and photographs depicting alleged human rights violations or war crimes are often captured on the mobile phones of victims or political sympathisers. The capture and dissemination of content often happens haphazardly, and for a variety of motivations, including raising awareness of the plight of those who have been most affected, or for advocacy purposes with the goal of mobilising international public opinion. For this content to be of use to investigators it must be discovered, verified, and authenticated. Discovery, verification, and authentication have, therefore, become critical skills for human rights organisations and human rights lawyers. This book is the first to cover the history, ethics, methods, and best-practice associated with open source research. It is intended to equip the next generation of lawyers, journalists, sociologists, data scientists, other human rights activists, and researchers with the cutting-edge skills needed to work in an increasingly digitized, and information-saturated environment.

**Fernández-Sánchez, P. A. (Eds.). (2022, May 9). *The Limitations of the Law of Armed Conflicts: New Means and Methods of Warfare.* Leiden, The Netherlands: Brill | Nijhoff.** [**https://doi.org/10.1163/9789004468863**](https://doi.org/10.1163/9789004468863)

In the law of armed conflicts, one of the elements that has changed the most has been the means and methods of warfare. Yet there are few legal answers for the many questions these changes pose. This volume, therefore, seeks to identify the limitations of current international law on this double plane, the means and methods of combat, and to offer insights about how to address them. Topics include the use of nuclear energy, which without being a weapon, can have the same effect as one, chemical and biological weapons, autonomous artificial intelligence weapons, and biobots.

Similarly, fake news, the hostile use of cyberspace, lawfare, the use of big data, terrorism as a combat method, premeditated poisoning, sexual humiliation, the impact of such news on the armed forces and the reorganization needed to face the new scenarios are all situations not contemplated in classical law and which require new legal and operational responses.

**Gaeta, P. & Bo, M. (Eds.). (2023, November). Autonomous Weapon Systems and War Crimes. Special Issue of the *Journal of International Criminal Justice, 21(5).*** [**https://academic.oup.com/jicj/issue/21/5**](https://academic.oup.com/jicj/issue/21/5)

This special issue of the *Journal of International Criminal Justice* is part of the research output of a 4-year project, funded by the Swiss National Science Foundation, entitled *Lethal Autonomous Weapon Systems and War Crimes: Who is to Bear Criminal Responsibility?*

# **Geiß, R., Lahmann, H. (Eds). (2024). *Research Handbook on Warfare and Artificial Intelligence.* Edward Elgar.**

The Research Handbook on Warfare and Artificial Intelligence provides a multi-disciplinary exploration of the urgent issues emerging from the increasing use of AI-supported technologies in military operations. Bringing together scholarship from leading experts in the fields of technology and security from across the globe, it sheds light on the wide spectrum of existing and prospective cases of AI in armed conflict.

**Gow, J., Dijxhoorn, Ernst., Kerr, R., & Verdirame, G. (Eds.). (2019). *Routledge handbook of war, law and technology* (1st ed.). Routledge.**

This volume provides an authoritative, cutting-edge resource on the characteristics of both technological and social change in warfare in the twenty-first century, and the challenges such change presents to international law. The character of contemporary warfare has recently undergone significant transformation in several important respects: the nature of the actors, the changing technological capabilities available to them, and the sites and spaces in which war is fought. These changes have augmented the phenomenon of non-obvious warfare, making understanding warfare one of the key challenges. Such developments have been accompanied by significant flux and uncertainty in the international legal sphere. This handbook brings together a unique blend of expertise, combining scholars and practitioners in science and technology, international law, strategy and policy, in order properly to understand and identify the chief characteristics and features of a range of innovative developments, means and processes in the context of obvious and non-obvious warfare. The handbook has six thematic sections: Law, war and technology Cyber warfare Autonomy, robotics and drones Synthetic biology New frontiers International perspectives. This interdisciplinary blend and the novel, rich and insightful contribution that it makes across various fields will make this volume a crucial research tool and guide for practitioners, scholars and students of war studies, security studies, technology and design, ethics, international relations and international law.

**Hoijtink, M. & Leese, M. (2019). *Technology and Agency in International Relations*. Routledge.**

This book responds to a gap in the literature in International Relations (IR) by integrating technology more systematically into analyses of global politics.

Technology facilitates, accelerates, automates, and exercises capabilities that are greater than human abilities. And yet, within IR, the role of technology often remains under-studied. Building on insights from science and technology studies (STS), assemblage theory and new materialism, this volume asks how international politics are made possible, knowable, and durable by and through technology. The contributors provide empirically rich and pertinent accounts of a variety of technologies relevant to the discipline, including drones, algorithms, satellite imagery, border management databases, and blockchains.

Problematizing various technologically mediated issues, such as secrecy, violence, and questions of how authority and evidence become constituted in international contexts, this book will be of interest to scholars in IR, in particular those who work in the subfields of (critical) security studies, International Political Economy, and Global Governance.

# **Hong, I. (2018). International Digital Forensic Investigation at the ICC. In *Handling and exchanging electronic evidence across Europe.* Springer International Publishing.**

The International Criminal Court (“ICC” or “the Court”) investigates and tries individuals charged with crimes of concern to the international community: genocide, war crimes and crimes against humanity. With regard to digital forensic investigations, the ICC has been confronted with various challenges especially derived from the nature of the crimes it handles and the fact that its procedure is distinct from national criminal procedure. This paper provides an introduction to the activities and challenges of digital forensics in international criminal investigations, and draws attention to requirements for more international cooperation, awareness improvement, standard establishment and the need for a joint effort at solving technical issues.

**Johnson, J. (2021). *Artificial intelligence and the future of warfare: The USA, China, and strategic stability*. Manchester University Press.**

*Artificial intelligence and the future of warfare* sketches a clear and sobering picture of the potential impact of *artificial intelligence (*AI) on the digitized battlefield, broadening our understanding of critical questions facing decision-makers today.

**Koenig, A., et al. (2021). New Technologies and the Investigation of International Crimes. Special issue of the *Journal of International Criminal Justice, 21*(5)*.*** [***https://academic.oup.com/jicj/issue/19/1***](https://academic.oup.com/jicj/issue/19/1)

The latest Special Issue of the Journal of International Criminal Justice explores the increasing reliance on the use of digital technologies in the investigation and prosecution of international crimes.

**Liljefors, M., Noll, G., Steuer, D. (2019). *War and Algorithm*. Rowman & Littlefield.**

New military technologies are animated by fantasies of perfect knowledge, lawfulness, and vision that contrast sharply with the very real limits of human understanding, law, and vision. Thus, various kinds of violent acts are proliferating while their precise nature remains unclear. Especially man–machine ensembles, guided by algorithms, are operating in ways that challenge conceptual understanding.

*War and Algorithm* looks at the increasing power of algorithms in these emerging forms of warfare from the perspectives of critical theory, philosophy, legal studies, and visual studies. The contributions in this volume grapple with the challenges posed by algorithmic warfare and trace the roots of new forms of war in the technological practices and forms of representation of the digital age. Together, these contributions provide a first step toward understanding—and resisting—our emerging world of war.

**Lushenko, P., Bose, S. & Maley, W., (eds). (2021). *Drones and Global Order*. Routledge.**

This book explores the implications of drone warfare for the legitimacy of global order.

The literature on drone warfare has evolved from studying the proliferation of drones, to measuring their effectiveness, to exploring their legal, moral, and ethical impacts. These "three waves" of scholarship do not, however, address the implications of drone warfare for global order. This book fills the gap by contributing to a "fourth wave" of literature concerned with the trade-offs imposed by drone warfare for global order. The book draws on the "English School" of International Relations Theory, which is premised on the existence of a society of states bounded by common norms, values, and institutions, to argue that drone warfare imposes contradictions on the structural and normative pillars of global order. These consist of the structure of international society and diffusion of military capabilities, as well as the sovereign equality of states and laws of armed conflict. The book presents a typology of contradictions imposed by drone warfare within and across these axes that threaten the legitimacy of global order. This framework also suggests a confounding consequence of drone warfare that scholars have not hitherto explored rigorously: drone warfare can sometimes strengthen global order. The volume concludes by proposing a research agenda to reconcile the complex and often counter-intuitive impacts of drone warfare for global order.

This book will be of considerable interest to students of security studies, global governance, and International Relations.

**Mecacci, G., Amoroso, D., Siebert, L.C., Abbink, D., etc. (Eds). (2024). *Research Handbook on Meaningful Human Control of Artificial Intelligence Systems*. Edward Elgar.**

This prescient Research Handbook analyses the ethical development of Artificial Intelligence systems through the prism of meaningful human control. It encapsulates a multitude of disciplinary lenses including technical, philosophical and legal, making a crucial contribution to the ongoing discourse about control and responsibility in the field of AI.

**Nissan, E., & Martino, A. A. (2004). *Applied Artificial Intelligence, 18*(3-4).** [**https://www.tandfonline.com/toc/uaai20/18/3-4?nav=tocList**](https://www.tandfonline.com/toc/uaai20/18/3-4?nav=tocList)

Thanks to the hospitality of Prof. Robert Trappl, we have recently been able to publish, in *Cybernetics and Systems*, a double special issue on the novel field that applies artificial intelligence to legal evidence. This field is more specific than “Artificial Intelligence and Law,” and an area that has been developing over the last three decades. Moreover, it is different from artificial intelligence methods as applied to the forensic sciences.

We are now able to introduce the reader to this other special issue, this time in the journal *Applied Artificial Intelligence*. The introductory article has the aim of providing the readers with an overview of the literature that already exists (and which we are pleased to have been the ones to edit, for the most part). The reader is tutored as well throughout major issues and problems that affect the area.

**Nissan, E., & Martino, A. A. (2003). *Cybernetics and Systems, 34*(4-5).** [**https://www.tandfonline.com/toc/ucbs20/34/4-5**](https://www.tandfonline.com/toc/ucbs20/34/4-5)

Artificial intelligence for law is a thriving area of research. It has been in existence for three decades and is not to be confused with more prosaic, less ambitious applications of information technology in the legal professions. Surprisingly enough, legal evidence hasn't traditionally been much of a presence, let alone a conspicuous subdomain, within "Al and Law" research, for all of legal evidence looming large in legal scholarship. to say nothing of forensics, itself subdivided into various disciplinary compartments. Nevertheless, among legal evidence theorists, especially at American schools of law, there has long been a controversy about forensic statistics. Papers relevant to this controversy can be found in the companion double special issue, of which we are the editors, which was published in 2001 in Artificial Intelligence and Law (Vol. 9, Nos. 2-3).

*Cybernetics and Systems* kindly agreed to publish the present col- lection of papers on Al and legal evidence as two consecutive double issues, nos. 4-5 and 6-7, in the current volume. This editorial also applies to the next double issue.

**Nissan, E., & Martino, A. A. (2003). *Cybernetics and Systems, 34*(6-7).** [**https://www.tandfonline.com/toc/ucbs20/34/6-7**](https://www.tandfonline.com/toc/ucbs20/34/6-7)

The physical journal issue you are holding, a double issue (Nos. 6-7), is Part II, following another double issue (Nos. 4-5). We guest editors are delighted, of course, that Prof. Trappl, the editor of *Cybernetics and Systems*, has shown such appreciation to the novelty of our subject, to devote a multiple issue of the journal to hosting the collection of papers we have put together. Parts I and II together constitute the centerpiece of an effort of many years, which resulted in publishing thematic issues in this and other journals. Together these represent a critical mass of research, intended to bootstrap artificial intelligence (AI) for legal evidence as an emergent field within the already well-established field of artificial intelligence and law. In particular, refer to our guest-edited thematic issues “Formal Approaches to Legal Evidence” (*Artificial Intelligence and Law,* Vol. 9, Nos. 2-3, Sept. 2001); “Software, Formal Models, and Artificial Intelligence for Legal Evidence” (*Computing and Informatics,* Vol. 20, No. 6, 2001); as well as “The Construction of Judicial Proof: A Challenge for Artificial Intelligence Modelling” (*Applied Artificial Intelligence*, in press).

**Payne, K. (2021). *I, warbot: The dawn of artificially intelligent conflict*. Oxford University Press.**

Artificial Intelligence is going to war. Intelligent weapon systems are here today, and many more are on the way tomorrow. Already, they're reshaping conflict--from the chaos of battle, with pilotless drones, robot tanks and unmanned submersibles, to the headquarters far from the action, where generals and politicians use technology to weigh up what to do. AI changes how we fight, and even how likely it is that we will.

In battle, warbots will be faster, more agile and more deadly than today's crewed weapons. New tactics and concepts will emerge, with spoofing and swarming to fool and overwhelm enemies. Strategies are changing too. When will an intelligent machine escalate, and how can it be deterred? Can robots predict the future? And what happens to the 'art of war' as machines themselves become creative?

Autonomous warfare makes many people uneasy. An international campaign against 'killer robots' hopes to ban AI from conflict. But the genie is out--AI weapons are too useful for states to outlaw. Still, crafting sensible rules for warbots is possible. This fascinating book shows how it might be done.

**Raska, M., & Bitzinger, R.A. (Eds.). (2023). *The AI Wave in Defence Innovation: Assessing Military Artificial Intelligence Strategies, Capabilities, and Trajectories (1st ed.).* Routledge.** [**https://doi.org/10.4324/9781003218326**](https://doi.org/10.4324/9781003218326)

An international and interdisciplinary perspective on the adoption and governance of artificial intelligence (AI) and machine learning (ML) in defence and military innovation by major and middle powers.

Advancements in AI and ML pose pressing questions related to evolving conceptions of military power, compliance with international humanitarian law, peace promotion, strategic stability, arms control, future operational environments, and technology races. To navigate the breadth of this AI and international security agenda, the contributors to this book include experts on AI, technology governance, and defence innovation to assess military AI strategic perspectives from major and middle AI powers alike. These include views of how the United States, China, Japan, South Korea, the European Union, and Russia see AI/ML as a technology with the potential to reshape military affairs and power structures in the broader international system. This diverse set of views aims to help elucidate key similarities and differences between AI powers in the evolving strategic context.

A valuable read for scholars of security studies, public policy, and STS studies with an interest in the impacts of AI and ML technologies.

**Roach, S.C. & Eckert, A.E. (2020). *Moral Responsibility in Twenty-First Century Warfare: Just War Theory and the Ethical Challenges of Autonomous Weapons Systems*. State University of New York Press.**

*Moral Responsibility in Twenty-First-Century Warfare* explores the complex relationship between just war theory and the ethics of autonomous weapons systems (AWS). One of the challenges facing ethicists of war, particularly just war theorists, is that AWS is an applicative concept that seems, in many ways, to lie beyond the human(ist) scope of the just war theory tradition. The book examines the various ethical gaps between just war theory and the legal and moral status of AWS, addresses the limits of both traditional and revisionist just war theory, and proposes ways of bridging some of these gaps. It adopts a dualistic notion of moral responsibility-or differing, related notions of moral responsibility and legitimate authority-to study the conflicts and contradictions of legitimizing the autonomous weapons that are designed to secure peace and neutralize the effects of violence. Focusing on the changing conditions and dynamics of accountability, responsibility, autonomy, and rights in twenty-first-century warfare, the volume sheds light on the effects of violence and the future ethics of modern warfare.

**Roy, K. (Ed.). (2024). *Artificial Intelligence, Ethics and the Future of Warfare: Global Perspectives*. Taylor & Francis.**

This volume examines how the adoption of AI technologies is likely to impact strategic and operational planning, and the possible future tactical scenarios for conventional, unconventional, cyber, space and nuclear force structures. In addition to developments in the USA, Britain, Russia and China, the volume also explores how different Asian and European countries are actively integrating AI into their military readiness. It studies the effect of AI and related technologies in training regimens and command structures. The book also covers the ethical and legal aspects of AI augmented warfare.

The volume will be of great interest to scholars, students and researchers of military and strategic studies, defence studies, artificial intelligence and ethics.

**Schraagen, J. M. (2024). *Responsible Use of AI in Military Systems.* CRC Press LLC.**

Artificial Intelligence (AI) is widely used in society today. The (mis)use of biased data sets in machine learning applications is well‑known, resulting in discrimination and exclusion of citizens. Another example is the use of non‑transparent algorithms that can’t explain themselves to users, resulting in the AI not being trusted and therefore not being used when it might be beneficial to use it.

*Responsible Use of AI in Military Systems* lays out what is required to develop and use AI in military systems in a responsible manner. Current developments in the emerging field of Responsible AI as applied to military systems in general (not merely weapons systems) are discussed. The book takes a broad and transdisciplinary scope by including contributions from the fields of philosophy, law, human factors, AI, systems engineering, and policy development.

Divided into five sections, Section I covers various practical models and approaches to implementing military AI responsibly; Section II focuses on liability and accountability of individuals and states; Section III deals with human control in human‑AI military teams; Section IV addresses policy aspects such as multilateral security negotiations; and Section V focuses on ‘autonomy’ and ‘meaningful human control’ in weapons systems.

Key Features:

* Takes a broad transdisciplinary approach to responsible AI
* Examines military systems in the broad sense of the word
* Focuses on the practical development and use of responsible AI
* Presents a coherent set of chapters, as all authors spent two days discussing each other’s work

This book provides the reader with a broad overview of all relevant aspects involved with the responsible development, deployment and use of AI in military systems. It stresses both the advantages of AI as well as the potential downsides of including AI in military systems.

**Tangredi, S. J., & Galdorisi, G. (Eds). (2021). *AI at war: How big data, artificial intelligence, and machine learning are changing naval warfare*. Naval Institute Press.**

Artificial intelligence (AI) may be the most beneficial technological development of the twenty-first century, but it is often misunderstood outside of specialists in the field. *AI at War* provides a balanced and practical understanding of this dawning new technology, explaining the importance of machine learning, human-machine interfaces, and big data analysis, components that are often omitted or misunderstood. While AI has many potential applications, Tangredi and Galdorisi have brought together more than thirty experts to focus on those elements relating to national security, making clear the importance and the potential of AI in defending the nation and in warfighting. Contributors include Robert Work, former Deputy Secretary of Defense; Admiral James Stavridis, former Commander, Supreme Allied Commander Europe; Admiral Michael Rogers, former Director of the National Security Agency; and Admiral Scott Swift, former Commander of the U.S. Pacific Fleet; as well as scientists and operators who share their theoretical and experiential knowledge of this “game-changing” new field.

**Wu, M. (2022). *Intelligent warfare: Prospects of military development in the age of AI*. Taylor & Francis.**

This book examines the future trend toward "intelligent" warfare considering the global environment, the history of warfare, and scientific and technological advancement. It develops a comprehensive set of theoretical frameworks, application concepts, and evaluation criteria for military intelligence.

The volume is packed with theoretical highlights and vivid examples, including the tracking of Osama bin Laden, the decapitation strike against Qasem Soleimani, the remote assassination of Iranian nuclear scientists, the drone war in the Nagorno–Karabakh conflict, modern equipment deployed in the Palestinian–Israeli conflict, and the war between social media groups. In addition, the author envisions a possible future for "intelligent" wars in which adversarial parties engage in combat through virtual and unmanned systems. This nature may help avoid the brutality and high death toll associated with traditional warfare.

The book explores the possibility of future civilized warfare. It will be of interest to researchers, academics, and students in the fields of politics, military intelligence, and military technology, and to those who are interested in intelligent warfare in general.

[**Završnik**](https://link.springer.com/book/10.1007/978-3-031-19149-7#author-1-0)**, A.,** [**Simončič**](https://link.springer.com/book/10.1007/978-3-031-19149-7#author-1-1)**, K. (Eds.). (2023). *Artificial intelligence, social harms and human rights*. Palgrave MacMillan.**

T​his book critically explores how and to what extent artificial intelligence (AI) can infringe human rights and/or lead to socially harmful consequences and how to avoid these. The European Union has outlined how it will use big data, machine learning, and AI to tackle a number of inherently social problems, including poverty, climate change, social inequality and criminality. The contributors of this book argue that the developments in AI must take place in an appropriate legal and ethical framework and they make recommendations to ensure that harm and human rights violations are avoided. The book is split into two parts: the first addresses human rights violations and harms that may occur in relation to AI in different domains (e.g. border control, surveillance, facial recognition) and the second part offers recommendations to address these issues. It draws on interdisciplinary research and speaks to policy-makers and criminologists, sociologists, scholars in STS studies, security studies scholars and legal scholars.

**Zwitter, A., Gstrein, O (Eds). (2023). *Handbook on the Politics and Governance of Big Data and Artificial Intelligence*. Edward Elgar Publishing.**

Drawing on the theoretical debates, practical applications, and sectoral approaches in the field, this ground-breaking Handbook unpacks the political and regulatory developments in AI and big data governance. Covering the political implications of big data and AI on international relations, as well as emerging initiatives for legal regulation, it provides an accessible overview of ongoing data science discourses in politics, law and governance.

## **Book Chapters**

**Gibson, J. (2021). Death by data: drones, kill lists and algorithms. In *Remote Warfare* (A. McKay, A. Watson, M. Karlshøj-Pedersen, Eds.). E-International Relations.** [**https://www.e-ir.info/publication/remote-warfare-interdisciplinary-perspectives/**](https://www.e-ir.info/publication/remote-warfare-interdisciplinary-perspectives/)**.**

In 2018 Google employees made headlines when they openly protested the company’s involvement in Project Maven – a controversial US programme aimed at integrating artificial intelligence into military operations. Google argued it was simply helping automate analysis of drone footage. Employees signed an open letter to CEO Sundar Pichai arguing Google ‘should not be in the business of war’ (BBC 2018). For many communities in places like Pakistan and Yemen, computers are already making life and death decisions. Massive amounts of signals intelligence are being run through algorithms that make decisions as to who is ‘suspicious’ and who ‘isn’t.’ For populations with a drone flying overhead, those decisions can be deadly. Nobody knows the damage America’s covert drone war can wreak better than Faisal bin ali Jaber. Faisal’s brother-in-law, Salem, was killed by a drone just days after he preached against al-Qaeda in 2012 (Jaber 2016). The strike was likely a ‘signature’ strike, one taken based on a suspicious ‘pattern of behaviour.’ This chapter will examine the case of Faisal bin ali Jaber and address just some of the troubling questions that arise as big data and remote warfare converge. Can targeting based on metadata ever be compliant with international humanitarian law (IHL) and its principle of ‘distinction’ and just what are the ‘feasible’ precautions the US must take to ensure it is?

**Hughes, J.G. (2020). The Law of Armed Conflict Issues Created by Programming Automatic Target Recognition Systems Using Deep Learning Methods. In: Gill, T., Geiß, R., Krieger, H., Paulussen, C. (eds) *Yearbook of International Humanitarian Law, Volume 21* (2018). T.M.C. Asser Press, The Hague.** [**https://doi.org/10.1007/978-94-6265-343-6\_4**](https://doi.org/10.1007/978-94-6265-343-6_4)**.**

Deep learning is a method of machine learning which has advanced several headline-grabbing technologies, from self-driving cars to systems recognising mental health issues in medical data. Due to these successes, its capabilities in image and target recognition is currently being researched for use in armed conflicts. However, this programming method contains inherent limitations, including an inability for the resultant algorithms to comprehend context and the near impossibility for humans to understand the decision-making process of the algorithms. This can lead to the appearance that the algorithms are functioning as intended even when they are not. This chapter examines these problems, amongst others, with regard to the potential use of deep learning to programme automatic target recognition systems, which may be used in an autonomous weapon system during an armed conflict. This chapter evaluates how the limitations of deep learning affect the ability of these systems to perform target recognition in compliance with the law of armed conflict. Ultimately, this chapter concludes that whilst there are some very narrow circumstances where these algorithms could be used in compliance with targeting rules, there are significant risks of unlawful targets being selected. Further, these algorithms impair the exercise of legal duties by autonomous weapon system operators, commanders, and weapons reviewers. As such, this chapter concludes that deep learning-generated algorithms should not be used for target recognition by fully-autonomous weapon systems in armed conflicts, unless they can be made in such a way as to understand the context of targeting decisions and be explainable.

**Koenig, A. & Freeman, L. (2024). Links in the chain: How the Berkeley Protocol is strengthening digital investigation standards in international justice. In *Open Source Investigations in the Age of Google* (H. Wilson, O. Samuel, D. Plesch, Eds.). World Scientific.** [**https://humanrights.berkeley.edu/publications/links-in-the-chain-how-the-berkeley-protocol-is-strengthening-digital-investigation-standards-in-international-justice/**](https://humanrights.berkeley.edu/publications/links-in-the-chain-how-the-berkeley-protocol-is-strengthening-digital-investigation-standards-in-international-justice/)

The international justice community comprises a complex ecosystem of actors with differing interests and end goals — from humanitarian efforts aimed at predicting and preventing mass atrocities, to advocacy efforts designed to pressure governments to respond to violations of international law, to legal efforts targeted at ensuring judicial accountability for the perpetrators of international crimes. Increasingly, however, these end goals can only be achieved if the diverse actors work together. Cooperation and coordination between first responders, civil society groups, intergovernmental organizations and judicial institutions are essential to ensuring that those most responsible for the world’s gravest crimes are identified and held to account. Depending on the degree and quality of cooperation, the diversity of actors involved in international justice and security can be beneficial or detrimental. In this chapter, we discuss these issues, illustrating them with examples from active war zones, and highlight how the further development of international standards for open source investigations is essential to ensure these efforts positively impact human security and justice worldwide. We also introduce the [Berkeley Protocol on Digital Open Source Investigations](https://humanrights.berkeley.edu/publications/berkeley-protocol-on-digital-open-source-investigations/), a recently-launched UN manual providing common standards and ethical principles for conducting digital investigations.

**Koenig, A., Freeman, L. (2023). ‘Nor is it neutral’: New technologies and the International Criminal Court. In *The International Criminal Court in its third decade* (C. Stahn, Ed., R.B. da Silva, Asst. Ed.). Brill | Nijhoff.** [**https://humanrights.berkeley.edu/publications/nor-is-it-neutral-new-technologies-and-the-international-criminal-court/**](https://humanrights.berkeley.edu/publications/nor-is-it-neutral-new-technologies-and-the-international-criminal-court/)

In this essay, we focus on the activities of the International Criminal Court, with an emphasis on the Office of the Prosecutor to outline three ways in which digital technologies have been and are increasingly impacting investigations and prosecutions. We also consider trends that we anticipate will have a growing impact on the Court’s operations. Specifically, we assess 1) the benefits digital technologies have provided (and are increasingly providing) with regards to conducting international criminal investigations; 2) the limitations of and challenges to their use; and 3) the ways in which digital technologies are enabling crimes within the jurisdiction of the Court.

**Lewis, D. A. (2022). On ‘Responsible AI’ in War: Exploring Preconditions for Respecting International Law in Armed Conflict. In: *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives* (S. Voeneky, P. Kellmeyer, O. Mueller, & W. Burgard, Eds.). Cambridge University Press.**

The law scholar Dustin Lewis explores the requirements of international law with regard to the employments of AI-related tools and techniques in armed conflict. The scope of this chapter is not limited to Lethal Autonomous Weapons (AWS) but also encompasses other AI-related tools and techniques related to warfighting, detention, and humanitarian services. After providing an overview of international law applicable to armed conflict, the author outlines some preconditions necessary to respect international law. According to Lewis, current international law essentially presupposes humans – and not artificial, non-humans – as legal agents. From that premise, the author argues that any employment of AI-related tools or techniques in an armed conflict needs to be susceptible to being administered, discerned, attributed, understood, and assessed by human agents.

**Lewis,** [**D.A. (2021).**](https://ccdcoe.org/uploads/2021/05/Autonomous_Cyber_Capabilities_210525.pdf#page=121) **Preconditions for Applying International Law to Autonomous Cyber Capabilities. In *Autonomous Cyber Capabilities under International Law* (R. Liivoja, A. Väljataga, Eds.). NATO CCDCOE.** [**https://ccdcoe.org/uploads/2021/05/Autonomous\_Cyber\_Capabilities\_210525.pdf#page=121**](https://ccdcoe.org/uploads/2021/05/Autonomous_Cyber_Capabilities_210525.pdf#page=121)

In this chapter, I seek to set out some of the preconditions arguably necessary to apply international law to employments—by a State, an international organization (IO) or a natural person — of autonomous cyber capabilities. Through this thought experiment, I aim in part to help detect preconditions arguably necessary to facilitate compliance with international law or incurrence of responsibility for violations of international law that may arise in respect of such employments.

**Miotto, N. (2024, July). The Vienna Document 2011 and Military Applications of Artificial Intelligence. In *OSCE Insights*(pp. 1-14). Nomos Verlagsgesellschaft mbH & Co. KG. DOI:** [**10.5771/9783748945857-01**](http://dx.doi.org/10.5771/9783748945857-01)

The development and deployment of military applications of artiﬁcial intelligence (AI) is raising concerns about their negative implications for international security. Misperception, unintended escalation, and proliferation are some of the key potential risks stemming from military uses of AI. This article argues that states within and outside the OSCE region should draw on the OSCE Vienna Document 2011 to develop conﬁdence- and security-building measures (CSBMs) applicable to the military uses of AI. Such CSBMs could help foster dialogue and co-operation by increasing transparency and predictability concerning military applications of AI.

**Margulies, P. (2016). Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts. In *Research Handbook on Remote Warfare* (J.D. Ohlin, Ed.). Edward Elgar Press. Forthcoming, Roger Williams University Legal Studies Paper No. 166, Available at SSRN:** [**https://ssrn.com/abstract=2734900**](https://ssrn.com/abstract=2734900)

Autonomous weapons systems, in which a computer makes targeting decisions without specific human authorization, pose challenges for international humanitarian law (IHL). The most salient challenge is accountability for autonomous IHL violations. An autonomous weapons system (AWS) that violates IHL cannot be a defendant in a war crimes trial or a subject of military discipline. Moreover, accountability for IHL violations would be ill-served by human combatants who shrugged off their own role in an AWS’s IHL violations, lamely claiming to be “outside the loop” of the computer’s autonomous decisions. To fill the AWS accountability gap, this paper relies on the doctrine of command responsibility. A human in command should have responsibility for autonomous decisions, just as a commander is currently held responsible for an unreasonable failure to prevent a subordinate’s IHL violations. Holding commanders responsible for an AWS is a logical refinement of current law, since it imposes liability on an individual with power and access to information who benefits most concretely from the AWS’s capabilities in war-fighting.

Accountability requires what I call dynamic diligence, a three-pronged approach entailing a flexible human/machine interface, periodic assessment, and parameters tailored to IHL compliance. The model features a dedicated AWS command, staffed by officers familiar with the capabilities of autonomous weapons. A dynamic human/machine interface will not require human authorization or real-time monitoring of targeting, but must enable humans to override an AWS’s decisions. Dynamic assessment should include regular reviews of the AWS’s learning process, to ensure that an AWS in the field does not learn behavior that violates IHL. Dynamic parameters encourage interpretability of AWS targeting decisions: a substantive, verbal explanation, rather than hidden layers of computer calculations, will facilitate effective review. Constrained by dynamic diligence, a commander can harness an AWS’s freedom from flawed human emotions, while furnishing “meaningful human control” to ensure fidelity to IHL principles.

**Nunn, N. J. (2021). Creating legal frameworks to afford human accountability for AI decisions in war. In *Futures of International Criminal Justice* (E. Palmer, E. Bikundo, S.H. Rimmer, M. Clark, Eds.). Routledge.**

Human control over weapons is being phased out. In its place, we are seeing the development of artificial intelligence (AI), which will take over many of a weapon’s functions. Seek, engage, and destroy functions are being programmed into lethal fully autonomous weapon systems (LAWS) to remove the human from the arena of armed conflict. AI allowing superior detection, faster analysis and response times, and lack of physically limiting factors such as exhaustion, hunger, and stress makes LAWS a desirable option for military operations, despite raising ethical concerns. These developments create several legal issues, with the most pressing being how to assign accountability. This chapter will focus on the issue of human accountability under international criminal law (ICL) for breaches of law enacted by LAWS. It will look at how existing frameworks are not sufficient to regulate accountability for their use and will suggest a new framework for responsibility that extends existing law to safely regulate AI in battle.

**Shepitko, V. Y., & Shepitko, M. V. (2024). International standards of war crimes investigation with usage of digital data and technologies. In *Impact of martial law in Ukraine on the further development of jurisprudence*. Baltija Publishing. DOI:**[**10.30525/978-9934-26-440-5-16**](http://dx.doi.org/10.30525/978-9934-26-440-5-16)

During the Russian aggression against Ukraine, criminologists and judicial experts face new tasks related to providing scientific recommendations and documenting, recording and proving the facts of war crimes, proposing and applying separate methods of investigation of the crime of aggression, the crime of genocide, crimes against humanity, as well as researching evidence of war crimes. There are questions about the need to use optical imaging systems, 3D scanners, drones, mobile DNA laboratories and the implementation of artificial intelligence in the activities of law enforcement agencies and expert institutions. The investigation of war and other crimes of an international nature is connected with the need to create and operate joint international investigative teams, obtain information in the warzone, the dangers of conducting investigative (search) actions during war, the presence of misinformation and countermeasures from the aggressor state.

**Weber, J. (2010). Armchair Warfare ‘on Terrorism'. On Robots, Targeted Assassinations and Strategic Violations of International Law. In *Thinking Machines and the Philosophy of Computer Science: Concepts and Principles* (Jordi Vallverdú, Ed.). IGI Global.**

In the 21st century, militaries are no competing for military dominance through specific superior weapon systems but through networking these systems via information and communication technologies. The ‘Revolution in Military Affairs’ (RMA) relies on network centric warfare, ‘precision’ weaponry and ‘intelligent’ systems such as uninhabited, modular, globally connected robot systems. While some Western forces (and the U.S. Central Intelligence Service C.I.A.) claim that robots help to avoid the death of one’s soldiers (respectively agents), NGOs point out the increase of killed civilians. In my paper, I discuss the deployment of uninhabited combat aerial vehicles (UCAV) in Western ‘wars on terror’ and their political and techno-ethical consequences. The question arises whether the new military philosophy, network centric (armchair) warfare, targeted assassinations and robot technology work towards the weakening of international humanitarian law.

**Weber, J. (2009). Robotic warfare, human rights & the rhetorics of ethical machines. In *Ethics and robotics* (R. Capurro & M. Nagenborg, Eds.)*.* IOS Press.**

Killing with robots is no more a future scenario but became a reality in the first decade of the 21st century. The U.S. and Israel forces are using uninhabited combat aerial vehicles (UCAVs) in their so-called wars on terror, especially for targeted killing missions in Iraq, Pakistan, Afghanistan as well as in Lebanon and the Palestinian occupied territories (for example in Israel's recent war on Gaza). In the last years, the number of UCAV air attacks is rising significantly as well as the number of killed civilians. Nevertheless, the automation of warfare is envisioned by the US government and military for 2032 at the latest and military robots are increasingly used in civilian contexts. In the face of these developments, discussions on robotic warfare as well as security technology from a science and technology studies and technoethical perspective are highly needed. Important questions are how robotic warfare and security applications may find their way into society on a broad scale and whether this might lead to a new global arms race, violation of the international law of warfare, an increasing endangerment of civilians transporting racist and sexist implications, and the blurring of boundaries between military, police and civil society.

**Wilson, C. (2020). Artificial Intelligence and Warfare. In *21st Century Prometheus* (M. Martellini & R. Trapp, Eds.)*.* Springer.** [**https://doi.org/10.1007/978-3-030-28285-1\_7**](https://doi.org/10.1007/978-3-030-28285-1_7)

New developments in Artificial Intelligence may overshadow the traditional foundation of nuclear deterrence by reducing the survivability of missile launch platforms. AI has the capability to gather and quickly synthesize information from thousands of sources to produce highly accurate estimates of locations for missile-submarines, or land-based mobile launchers. Thus, mobile nuclear assets for deterrence may become easier targets, which reduces the threat of successful retaliation after a deliberate first-strike. Other traditional forms of defense may also be rendered obsolete. While traditional defenses are designed to protect against larger, stronger weapons, AI has the new capability to operate autonomous weapons at the miniaturized level. Miniature devices can fly toward a target undetected and then swarm together at the last minute to deliver a powerful, coordinated blow through autonomous synchronization. Many critics encourage policy makers to prevent AI weapons from ever killing humans as part of a military mission without first receiving explicit permission from a human operator. However, other experts warn that AI may someday evolve beyond submission to their human controllers, and may independently and dynamically set up their own new missions and goals which may conflict with the original designers. AI is currently an area of intense research for China and Russia. Moral objections to AI by contractor-technicians in the US may slow new development by the DOD. However, US commercial development of AI is moving forward rapidly, and can be seen in new products such as self-driving cars. AI may be considered a dual-use technology, similar to strong encryption, which makes legal controls to limit uses and prevent export somewhat difficult.

**Yan, G. (2020). The impact of Artificial Intelligence on hybrid warfare. In *Robotics, Autonomous Systems and Contemporary International Security* (A. Rossiter, Ed.). Routledge.**

## Through a brief survey of the typical definitions of hybrid warfare (HW), this article illustrates the five salient features of HW: synergy, ambiguity, asymmetry, innovative disruption and battle over psychology; then based on a HW model proposed by Erik Reichborn-Kjennerud and Patrick Cullen, the article discusses the impact of Artificial Intelligence on the five instruments of power – military, political, economic, civil and informational (MPECI), and analyses the changes and continuities of HW in the age of Artificial Intelligence.

## 

## **Peer Reviewed Articles**

**Abouzari, M., Barzegar, M. R., & Naderi, Z. (2023). Feasibility of AI-powered war weapons criminal liability and the challenge of impunity in the International Criminal Court. *ModernTechnologies Law*, *4*(8), 119-134. doi: 10.22133/mtlj.2023.389496.1184**

The issue of delegating the authority to kill humans to artificial intelligence in wartime is one of the most challenging emerging disciplines, as it raises numerous legal issues. This is the first time that legal knowledge has encountered a phenomenon that, despite being created by humans, is independent of humans in thinking and decision-making, and it may be possible after multiple stages that even the creator of the system cannot comprehend or control its operation. Therefore, this article discusses the obstacles that the International Criminal Court must overcome to apply direct criminal responsibility and the doctrine of commander's responsibility in cases where artificial intelligence weapons have led to the killing of civilians and war crimes, as well as the steps that should be taken to overcome these obstacles. Following a documentary approach and a descriptive-analytical analysis, this article concludes that the current provisions of the statute regarding artificial intelligence need to be revised due to the emergence of a concept called the gap in responsibility, which is caused by the unique nature of artificial intelligence. To adapt to the new conditions, the international community must require manufacturers to provide maximum transparency and prohibit producing certain types of artificial intelligence. Regarding the possible killings caused by the use of artificial intelligence weapons with the ability to self-learn after production, the court statute must explicitly and exclusively specify the rules of criminal responsibility so that the existence of AI weapons does not lead to impunity.

**Acquaviva, G. (2022). Autonomous weapons systems controlled by Artificial Intelligence: a conceptual roadmap for international criminal responsibility. *The Military Law and the Law of War Review*, *60*(1), 89-121.** [**http://dx.doi.org/10.2139/ssrn.4070447**](https://dx.doi.org/10.2139/ssrn.4070447)**.**

The article explores a conceptual roadmap for considering autonomous weapons systems (AWS) not just in the framework of international humanitarian law (IHL), but also in terms of individual responsibility under international criminal law (ICL). After setting the stage with some preliminary and terminological considerations, the author considers the different perspectives through which IHL and ICL look at Artificial Intelligence as applied to weapons systems. Within this framework, the article then hones in on certain essential aspects of the contemporary understanding of ICL – a system aimed at protecting certain interests (or goods) based on individual agency as demonstrated by its approach to modes of responsibility, and sentencing. On the basis of these reflections, three main theoretical possibilities are explored: that humans remain the only subject capable of being considered criminally responsible for harm to interests protected by ICL; that AWS come to be considered ‘persons’ subject to (a form of) individual criminal responsibility; or, a hybrid new system combining aspects of the two. The author considers that, while the second option appears implausible and unrealistic under the present circumstances, the challenges faced by the other two options remain daunting, posing potentially irresolvable issues within the current ICL framework.

**Asaro, P. (2012). On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. *International review of the Red Cross*, *94*(886), 687-709. DOI:** [**https://doi.org/10.1017/S1816383112000768**](https://doi.org/10.1017/S1816383112000768)**.**

This article considers the recent literature concerned with establishing an international prohibition on autonomous weapon systems. It seeks to address concerns expressed by some scholars that such a ban might be problematic for various reasons. It argues in favour of a theoretical foundation for such a ban based on human rights and humanitarian principles that are not only moral, but also legal ones. In particular, an implicit requirement for human judgement can be found in international humanitarian law governing armed conflict. Indeed, this requirement is implicit in the principles of distinction, proportionality, and military necessity that are found in international treaties, such as the 1949 Geneva Conventions, and firmly established in international customary law. Similar principles are also implicit in international human rights law, which ensures certain human rights for all people, regardless of national origins or local laws, at all times. I argue that the human rights to life and due process, and the limited conditions under which they can be overridden, imply a specific duty with respect to a broad range of automated and autonomous technologies. In particular, there is a duty upon individuals and states in peacetime, as well as combatants, military organizations, and states in armed conflict situations, not to delegate to a machine or automated process the authority or capability to initiate the use of lethal force independently of human determinations of its moral and legal legitimacy in each and every case. I argue that it would be beneficial to establish this duty as an international norm, and express this with a treaty, before the emergence of a broad range of automated and autonomous weapons systems begin to appear that are likely to pose grave threats to the basic rights of individuals.

**Asghar, M. S., Naz, H., & Saqib, K. M. (2023). Of Rules and Robots: The Dawn of Modern Warfare and the Looming Challenge of Lethal Autonomous Robots (LARs). *Journal of Law & Social Studies (JLSS)*, *5*(3), 432-439. DOI:**[**10.52279/jlss.05.03.432439**](http://dx.doi.org/10.52279/jlss.05.03.432439)

Due to our endless quest to be faster and stronger military technology has progressed drastically during the last few decades. People are aware of certain technological developments, such as guns and jets, which have revolutionized the modern warfare. But a lot of technological advancements happened under the radar, in the private test sites or in the military labs, making majority of citizens oblivious of technological advents. Robotics is one of the military technologies which has essentially escaped public notice to date. The development of Artificial Intelligence (AI) based robots capable to exert lethal force, called lethal autonomous robots (LARs), has significant legal and ethical implications. Their creation has already ignited a heated discourse on these allusions and two polarized sides could easily be discerned. Aim of the paper is to present a brief analysis of the regulations and rules of International Humanitarian Law (IHL) that are pertinent to comprehend the discourse on the legality of lethal autonomous robots (LARs).

**Barber, I. A. (2020). Autonomous Weapons Systems & Accountability: Rethinking Criminal Responsibility for War Crimes at the ICC. *SOAS Law Journal*, *7, 5.*** [**https://vlex.co.uk/vid/847575247**](https://vlex.co.uk/vid/847575247)**.**

The ongoing development and future deployment of Autonomous Weapons Systems (AWS) has spurred intense debate within the international community. Scholars and global policymakers have recently begun to consider the various legal and logistical ramifications of such weapons systems. However, there is currently no adequate legal framework in place to attribute individual criminal responsibility for the conduct of AWS in armed conflict. The purpose of this research is to explore the problematic nature of attributing individual criminal liability for the actions of AWS, and subsequently evaluate proposals to provide accountability when an AWS commits a war crime. An examination of these proposed accountability schemes, the relevant jurisprudence, and the existing scholarship suggests that legal mechanisms could be established to effectively attribute criminal responsibility under international criminal law. Accordingly, this paper ultimately advocates for an amendment to the Rome Statute of the International Criminal Court (ICC) to establish individual criminal liability for human agents who deploy AWS that perpetrate war crimes. Support for this amendment is based primarily on the ICC's unique construction as the world's first permanent international criminal tribunal and the success of previous efforts to amend the Rome Statute.

**Bo, Marta. (2021, May). Autonomous Weapons and the Responsibility Gap in light of the *Mens Rea* of the War Crime of Attacking Civilians in the ICC Statute, *Journal of International Criminal Justice, Volume 19*(2), 275–299.** [**https://doi.org/10.1093/jicj/mqab005**](https://doi.org/10.1093/jicj/mqab005)

Within the broader context of the problems raised by the interaction between humans and machines in weaponry and targeting, this paper deals with the specific issue of the *mens rea* required to establish responsibility for the war crime of indiscriminate attacks, in the context of attacks performed with semi-autonomous weapons or with the support of artificial intelligence (AI) in targeting decision-making. The author presents the difficulties that are determined by the interaction between humans and machines, and highlights that an interpretation that would allow for risk-taking mental elements such as *dolus eventualis* and recklessness in the framework of the war crime of attacking civilians would be better able to capture the criminality of the conduct of the person who knowingly accepts the risk of killing civilians as part of an AI-powered attack where the result of hitting the civilian target is one of the possible outcomes. However, the article indicates that this construction can be employed only in specific circumstances, since in most scenarios even these lowered *mens rea* requirements would not be met. In most human-machine teaming scenarios, lower types of intent such as *dolus eventualis* would still be insufficient for the ascription of criminal responsibility for such indiscriminate attacks against civilians. This is because of the specific risks posed by the integration of autonomy in the targeting process and the resulting changes to the cognitive environment in which human agents operate, which significantly affect specific components of *mens rea* standards.

**Bo, Marta. (2023, November). Criminal Responsibility by Omission for Failures to Stop Autonomous Weapon Systems, *Journal of International Criminal Justice* *21*(5), 1057–1075.** [**https://doi.org/10.1093/jicj/mqad029**](https://doi.org/10.1093/jicj/mqad029)

This article deals with establishing the criminal responsibility, through the model of commission by omission, of autonomous weapon systems (AWS) users in situations where failures to suspend AWS-driven attacks have caused a war crime. The author tackles the question of whether an omission to stop such an AWS may amount to the actus reus of war crimes of unlawful attacks and does so by establishing how the doctrine of commission by omission can be applied on the basis of the grave breaches regime in the First Additional Protocol to the Geneva Conventions and the Rome Statute of the International Criminal Court. In deconstructing the status of commission by omission under both these legal frameworks, this article analyses whether the substantive conditions of commission by omission, namely, the legal duty to act and the capacity to act, are met. The author suggests that ‘human control’, manifested in the ability to supervise, intervene and stop an AWS-driven attack, should be considered a necessary pre-condition for the imputation of criminal responsibility in at least some expected scenarios of AWS use. In the absence of such human control, there would be no accountability for unlawful attacks, including indiscriminate attacks, caused by AWS, which would lead to impunity for such crimes. On the one hand, the attribution of responsibility by omission has, therefore, crucial implications for closing the ‘responsibility gap’ within this context. On the other hand, based on the analysis of ‘control’ as the key principle for criminal responsibility by omission, the author argues that an additional treaty obligation should be adopted to ensure human control over AWS and preserve accountability for potential unlawful attacks.

**Borgesano, N. (2021). Lethal Machines’ ‘Acts’: The Use of Artificial Intelligence and the Principles of International Law. *Humanitäres Völkerrecht: Journal of International Law of Peace and Armed Conflict*, *4*(3/4), 174–187.** [**https://www.jstor.org/stable/48646323**](https://www.jstor.org/stable/48646323)**.**

In the most flourishing era in terms of artificial intelligence advancements and achievements, an ever-increasing uncertainty regarding the possibility to ensure human rights protection at war requires in-depth analysis and considerations. Notably, there are three overlapping areas that presuppose the usage of artificial intelligence: physical robotic systems, cyber weapons, and decision-support systems. The article focuses on the employment of robotic systems that use physical force and are considered ‘deadly’, given the inherent capability of ‘acting’ and putting human life in peril, thus causing an immediate concern from a humanitarian, legal, and ethical perspective. The article introduces a functional definition of Lethal Autonomous Weapon Systems and examines the existence of international liability in relation to the aggressive ‘act’ of a machine. After evaluating the ‘acts’ of those machines under the main principles of international humanitarian law and international human rights law, the article concludes that, even though some problems may be overcome, ethical issues pose the greatest risk and suggest the necessity of a permanent human-centred approach.

**Chin, W. (2019). Technology, war and the state: past, present and future. *International Affairs*, *95*(4), 765-783.** [**https://doi.org/10.1093/ia/iiz106**](https://doi.org/10.1093/ia/iiz106)

War made the state, and the state made war, but does this statement hold true today? Will it apply in the future? The consensus is that the absence of major war within the western world, post 1945, did cause the war–state relationship to change, but each became significantly less important to the other. This article argues that the relationship was closer and deeper than has been assumed. It proposes that the peculiar strategic conditions created by the nuclear age caused states to wage a ritualistic style of war, in which demonstration rather than the physical application of violence became increasingly important. Within this setting, the state drove the process of technological innovation in defence to its limits in an effort to demonstrate its military superiority. This massive peacetime investment in defence technology exerted a huge impact on the character of war, which led to new strategic forms. However, most importantly, the diffusion of military technology also affected the wider economy and society, leading to a form of internal power transition within states. The author speculates on how these elemental forces will play out in the future, what will happen to war and the state, and whether we will reach a point where war leads to the unmaking of the state.

**Deeks, A., Lubell, N., & Murray, D. (2018, November 16). Machine Learning, Artificial Intelligence, and the Use of Force by States. *Journal of National Security Law & Policy 10,* *Virginia Public Law and Legal Theory Research Paper No. 2018-63*.** [**https://ssrn.com/abstract=3285879**](https://ssrn.com/abstract=3285879)**.**

Big data technology and machine learning techniques play a growing role across all areas of modern society. Machine learning provides the ability to predict likely future outcomes, to calculate risks between competing choices, to make sense of vast amounts of data at speed, and to draw insights from data that would be otherwise invisible to human analysts. Despite the significant attention given to machine learning generally in academic writing and public discourse, however, there has been little analysis of how it may affect war-making decisions, and even less analysis from an international law perspective. The advantages that flow from machine learning algorithms mean that it is inevitable that governments will begin to employ them to help officials decide whether, when, and how to resort to force internationally. In some cases, these algorithms may lead to more accurate and defensible uses of force than we see today; in other cases, states may intentionally abuse these algorithms to engage in acts of aggression, or unintentionally misuse algorithms in ways that lead them to make inferior decisions relating to force.

This essay’s goal is to draw attention to current and near future developments that may have profound implications for international law, and to present a blueprint for the necessary analysis. More specifically, this article seeks to identify the most likely ways in which states will begin to employ machine learning algorithms to guide their decisions about when and how to use force, to identify legal challenges raised by use of force-related algorithms, and to recommend prophylactic measures for states as they begin to employ these tools.

**Deeks, A. (2018, December 22). Predicting Enemies. *Virginia Law Review 104*:8, 1529.** [**https://virginialawreview.org/articles/predicting-enemies/**](https://virginialawreview.org/articles/predicting-enemies/)

Actors in our criminal justice system increasingly rely on computer algorithms to help them predict how dangerous certain people and certain physical locations are. These predictive algorithms have spawned controversies because their operations are often opaque and some algorithms use biased data. Yet these same types of predictive algorithms inevitably will migrate into the national security sphere as the military tries to predict who and where its enemies are. Because military operations face fewer legal strictures and more limited oversight than criminal justice processes do, the military might expect—and hope—that its use of predictive algorithms will remain both unfettered and unseen.

This Article shows why that is a flawed approach, descriptively and normatively. First, in the post-September 11 era, any military operations associated with detention or targeting will draw intense scrutiny. Anticipating that scrutiny, the military should learn from the legal and policy challenges that criminal justice actors have faced in managing the transparency, reliability, and lawful use of predictive algorithms. Second, the military should clearly identify the laws and policies that govern its use of predictive algorithms. Doing so would avoid exacerbating the “double black box” problem of conducting operations that are already difficult to legally oversee and contest, using algorithms whose predictions are often difficult to explain. Instead, being transparent about how, when, why, and on what legal basis the military is using predictive algorithms will improve the quality of military decision-making and enhance public support for a new generation of national security tools.

**D'Evereux, V. (2024). Israeli Military Artificial Intelligence, Its Possible Use in the War in Gaza. *Univerzita Obrany. Ustav Strategickych Studii. Obrana a Strategie, 24(1)*, 125-142.** [**https://doi.org/10.3849/1802-7199.24.2024.01.125-142**](https://doi.org/10.3849/1802-7199.24.2024.01.125-142)**.**

This paper is focused on examining selected Israeli technologies operating based on artificial intelligence and the possible use of these technologies by the Israeli Defense Forces in the war in Gaza. These technologies include the Besorah system, an AI technology for identifying and recommending targets suitable for aerial bombardment, as well as other forms of attacks carried out from a distance. The SMASH system is an automatic targeting and firing system for small arms. The Goshawk system is a fully autonomous UAV used for aerial protection. The IRIS robot can be used for the initial investigation of tunnels ahead of assaults. The alleged use of these technologies in the war in Gaza cannot be verified by independent sources, mainly due to the ongoing conflict. The paper strives to present a general theoretical assessment of these technologies and weapons from the perspective of international law and armed conflicts.

**Duke, Joshua E. (2021). “Autonomous Robotics and the Laws of War: Methods and Consequences of Regulating Artificial Intelligence in Warfare.” *Security and Intelligence* *6* (2).** [**https://doi.org/10.18278/gsis.6.2.6**](https://doi.org/10.18278/gsis.6.2.6)**.**

This article addresses the question of what impact International Humanitarian Law (IHL), the Laws of Armed Conflict (LOAC), and the international community can or should have on the international development/deployment of autonomous and semi-autonomous weapon systems, and how the international community can achieve a significant impact with emerging national or cooperative international regulations or laws with regards to the developing relationship between robotics and warfare, without hindering technological developments in other areas of human life. The author, using primarily case studies related to weapon autonomy and robotics in warfare, tests the following theory: Technological advancements related to the development and implementation of autonomous and semi-autonomous weapons in warfare have the potential to be directly impacted by IHL and the LOAC, by using a reactive approach guided by historical underlying principles related to other technologies and the moral spirit of existing laws in order to proactively regulate the field. In testing the theory, the author shows the differences in lasting and effective technological impact of reactive versus proactive international actions. The case studies highlight the effectiveness of reactive international action, while framing the underlying issues of the past in the context of modern autonomous weaponry developments. The article highlights the record of weapon systems with autonomous functions and discusses fully autonomous lethal weapon systems’ inherent inability to comply with international human rights laws.

**Emery, J. R. (2020). “Probabilities Towards Death: Bugsplat, Algorithmic Assassinations, and Ethical Due Care.” *Critical Military Studies* 8 (2): 179–197.** [**https://www.tandfonline.com/doi/full/10.1080/23337486.2020.1809251**](https://www.tandfonline.com/doi/full/10.1080/23337486.2020.1809251)

This article explores the principle of due care in war and the myth that improved battlefield technology makes Western warfare inherently more ethical. The discursive construction – which I term virtuous chaoplexic militarism – of the US as ethical by virtue of its utilization of technologically advanced modes of killing, seeks to dissolve the ethico-political dilemmas of war into quantifiable problems to-be-solved. This article illustrates this dissolution by outlining the transformation within US military decision-making from an ethics of practical judgement to a computational techno-ethics. To do this, I evaluate two concrete cases of US algorithms of militarism. The first case traces the rise of collateral damage estimation algorithms, colloquially known as bugsplat. I examine how bugsplat is programmed, its fundamental design flaws, and its practical exploitation by commanders to erroneously tick the box of ethical due care. The second case explores the SKYNET machine-learning algorithm that was designed to construct ‘legitimate targets’ for US drone strikes via heterogeneous correlations of SIM card metadata. While drone strikes are widely praised for their capacity to individualize targeting, the algorithmic process of SKYNET ultimately erodes the individual subjectivity that is foundational for ethics of war through data constructions of ‘terroristness.’As both cases demonstrate, the ultimate goal of this virtuous chaoplexic militarism is to render the ethico-political dilemmas of killing quantifiable, predictable, and solvable. There exists an urgent need to interrogate socio-technical interactions in the military setting; and specifically, the degree to which practical judgement has been outsourced to a morally problematic computational techno-ethics.

**Ekelhof, M.A.C. (2018). Lifting the Fog of Targeting: “Autonomous Weapons” and Human Control through the Lens of Military Targeting. *Naval War College Review,* *71*(3).** [**https://digital-commons.usnwc.edu/nwc-review/vol71/iss3/6**](https://digital-commons.usnwc.edu/nwc-review/vol71/iss3/6)**.**

Military targeting practices should be the core of any analysis that seeks a better understanding both of the concept of meaningful human control within a context of increasingly autonomous weapons, and of what human-machine relationship we require.

**Freeman, L. (2018). Digital evidence and war crimes prosecutions: The impact of digital technologies on international criminal investigations and trials. *Fordham International Law Journal, 41*(2), pp. 283-336.** [**https://humanrights.berkeley.edu/publications/digital-evidence-and-war-crimes-prosecutions-the-impact-of-digital-technologies-on-international-criminal-investigations-and-trials/**](https://humanrights.berkeley.edu/publications/digital-evidence-and-war-crimes-prosecutions-the-impact-of-digital-technologies-on-international-criminal-investigations-and-trials/)**.**

As technology develops, new tools are continually being introduced that alter the nature and availability of courtroom evidence. The proliferation, connectivity, and capabilities of camera embedded and internet-enabled mobile devices, which record far more information about people’s activities and communications than ever before, are transforming the way criminal investigators and prosecutors collect, evaluate, and present evidence at trial. This is particularly true in international criminal trials, where prosecutors must present a voluminous and varied body of evidence to prove multiple charges related to complex conflicts. It is the prosecutor’s job to present evidence in a way that assists the fact-finder in evaluating its significance and understanding how it fits into the greater narrative. Advanced digital devices can now capture far more information about a situation than a witness can perceive, and innovative presentation tools allow lawyers to augment and strengthen the evidence in their cases by adding supplementary data and creating compelling visuals. In cases involving war crimes, crimes against humanity, and genocide, a large quantity and diversity of evidence is necessary to explain the context of the conflict and to prove the requisite elements of crimes and modes of liability. By examining the evidence and presentation techniques used in recent cases before international criminal courts, this article illustrates how war crimes prosecutions are evolving to meet the challenges and advantages of modern times. Part II explains the applicable law and describes how the use of emerging types of evidence in international criminal cases has expanded and been refined over the years. Part III analyzes three exceptional, yet emblematic cases from 2016, which call attention to an important trend that is predictive of the future use of digital evidence in war crimes prosecutions. Part IV discusses cases on the horizon and what these technological developments mean for members of the international justice community.

**Freeman, L. (2022). Digitally Disappeared: The Struggle to Preserve Social Media Evidence of Mass Atrocities. *Georgetown Journal of International Affairs* *23*(1), 105-113.** [**https://dx.doi.org/10.1353/gia.2022.0017**](https://dx.doi.org/10.1353/gia.2022.0017)**.**

Videos, images, and posts on social media — often referred to as “user-generated content”— can serve as valuable evidence of international crimes, but only if they are identified by investigators in time, forensically preserved, and made available to prosecuting authorities. This article examines the impact of content moderation, retention, and disclosure policies on the investigation of war crimes, crimes against humanity, and genocide. It highlights tensions between certain human rights, such as freedom of speech, freedom of expression, access to information, and the right to privacy, as well as competing interests between counterterrorism professionals, human rights advocates, and international justice practitioners. Ultimately, this article considers whether data protection rules like the “right to be forgotten” can be reconciled with broader principles of international justice, like the core precept that there are horrors in this world that we must “never forget.”

**Freeman, L. (2021). Hacked and leaked: Legal issues arising from the use of unlawfully obtained digital evidence in international criminal cases. *UCLA Journal of International Law and Foreign Affairs, 25*(2)*.*** [**https://escholarship.org/uc/item/5b87861x**](https://escholarship.org/uc/item/5b87861x)**.**

Digital open source investigations — the use of publicly available information on the internet for intelligence, leads, or evidence — are becoming an increasingly critical part of international criminal investigations. While the definition of open source information is simple, there are several categories of information that fall into a gray area between private and public — in particular, the growing amount of illegally hacked and leaked information on the web. Online leaks, whether the result of hacking or whistleblowing, fit the definition of open source information. Yet, there is something inherently different about information in the public domain that was not intended to be public. The dissemination of incriminating information unlawfully obtained by a third party creates a complex situation in which, on one hand, the illegal method of acquisition should not be rewarded, while at the same time, the illegal acts that are exposed in the documents should not go unpunished. The public interest can cut both ways. What are the rules and practical implications of using this information in criminal investigations or, more importantly, criminal trials? By examining specific hacks and leaks, describing their relevance to international criminal cases, and identifying the applicable evidentiary rules, this Article explores the challenges to admitting hacked and leaked digital documents into evidence.

**Freeman, L. (2019). Law in conflict: The technological transformation of warfare and its consequences for the International Criminal Court. *NYU Journal of International Law and Politics 50*, 807.** [**https://humanrights.berkeley.edu/publications/law-in-conflict-the-technological-transformation-of-warfare-and-its-consequences-for-the-international-criminal-court/**](https://humanrights.berkeley.edu/publications/law-in-conflict-the-technological-transformation-of-warfare-and-its-consequences-for-the-international-criminal-court/)

Modern armed conflicts and military strategies have undergone dramatic shifts as a result of new technologies, and the next generation of innovations will have profound consequences for how wars are fought, where they are fought, and who fights them. This, in turn, will inevitably have a pronounced influence on the development of the laws of war and the justice mechanisms mandated with enforcing those laws. Therefore, as new strategies and dynamics of war emerge related to the use of new technologies, war crimes investigators and prosecutors must adapt in order to meet the goals of establishing the truth, protecting the historical record, and holding individuals accountable for grave violations of international law. This article examines the characteristics of contemporary armed conflicts related to the use of new technologies and asks how this technological transformation of warfare will affect the ability of the global community to reach the goals of international justice. First, it examines the development of the use of technology-what technologies are used, how they are used, and what impact they have on armed conflicts and military affairs generally. Second, it identifies the complex legal issues arising from the use of new technologies and advocates for needed revisions to the definitions of crimes and modes of liability. Finally, it looks at the impact of the technological transformation of warfare on the fact-finding process for international criminal investigators and recommends a modified approach to evidence and the rules of procedure. In sum, this article takes a big picture approach to examining a current revolution based on the complex interplay of technology, law, and investigations in armed conflicts, and uses this understanding to chart a new way forward for the International Criminal Court.

**Garcia, D. (2016). Future arms, technologies, and international law: Preventive security governance. *European Journal of International Security*, *1*(1), 94-111.** [**https://doi.org/10.1017/eis.2015.7**](https://doi.org/10.1017/eis.2015.7)

This article presents an initial discussion of the political and legal challenges associated with weaponised technologies in three interconnected areas that may impinge upon the ability to protect civilian populations during peace and war and imperil international security: armed unmanned combat aerial vehicles (commonly known as drones); autonomous weapons systems (known as ‘killer robots’); and the potential militarisation of cyberspace, or its use as a weapon, and the operation of drones and killer robots in the cyber domain. Supporting the argument that the world is ‘facing new methods of warfare’ and that international security governance and law are not keeping up, the article provides an overview and interpretation of three technologies in connection with aspects of five branches of law: state responsibility, use of force, international humanitarian law, human rights law, and law of the commons. I argue therefore that ‘preventive security governance’ could be a strategy to curtail uncertainty in the preservation of stability and international order. I define ‘preventive security governance’ as the codification of specific or new global norms, arising from existing international law that will clarify expectations and universally agreed behaviour on a given issue-area. This is essential for a peaceful future for humanity and for international order and stability.

**Garcia, D. (2018). Lethal Artificial Intelligence and Change: The Future of International Peace and Security. *International Studies Review* *20*(2), pp. 334–341,** [**https://doi.org/10.1093/isr/viy029**](https://doi.org/10.1093/isr/viy029)**.**

The development of artificial intelligence and its uses for lethal purposes in war will fundamentally change the nature of warfare as well as law-enforcement and thus pose fundamental problems for the stability of the international system. To cope with such changes, states should adopt preventive security governance frameworks based upon the precautionary principle of international law, and upon previous cases where prevention brought stability to all countries. Such new global governance frameworks should be innovative as current models will not suffice. The World Economic Forum has advanced that the two areas that will bring most benefits but also biggest dangers to the future are robotics and artificial intelligence. Additionally, they are also the areas in most urgent need for innovative global governance.

Leading scientists working on artificial intelligence have argued that the militarization and use of lethal artificial intelligence would be highly destabilizing. Here I examine twenty-two existing treaties that acted under a “preventive framework” to establish new regimes of prohibition or control of weapons systems that had been deemed to be destabilizing. These treaties achieved one or all of three goals: prevented further militarization, made weaponization unlawful, and stopped proliferation with cooperative frameworks of transparency and common rules. As a result of my findings, it is clear that there is a significant emerging norm in regards to all weapons systems: the utilization of disarmament and arms regulations as a tool and mechanism to protect civilians. The development of lethal autonomous weapons systems would severely jeopardize this emerging norm. I show under what conditions lethal autonomous weapons systems will be disruptive for peace and security and show alternative governance structures based upon international law with robust precautionary frameworks.

**Gevers, A. (2014). Is Johnny Five Alive or Did It Short Circuit: Can and Should an Artificially Intelligent Machine Be Held Accountable in War Or Is It Merely a Weapon. *Rutgers Journal of Law & Public Policy*, *12*, 384.** [**https://rutgerspolicyjournal.org/jlpp/wp-content/uploads/sites/26/2017/02/Gevers.pdf**](https://rutgerspolicyjournal.org/jlpp/wp-content/uploads/sites/26/2017/02/Gevers.pdf)**.**

Imagine the scene: in the not too distant future you, as a U.S. Army Captain, are on a mission in the snowy mountain topography of some far off land. Accompanying you are several soldiers, one of which, Sergeant Johnny Five, sits prone behind a large caliber machine gun as you overlook a village. You have been advised that enemy combatants are the sole residents of the village below and you are instructed to eliminate the targets. You sit behind a spotting scope and direct Johnny to engage five targets dressed in distinct white camouflage and clear the village. Johnny confirms he sees the targets. You give a resounding “Execute” order. Johnny then, being a well programmed, trained, and armored artificially intelligent robot, proceeds down to the village. Johnny opens fire upon reaching a distance of thirty meters from the combatants. Johnny confirms the targets are down. You go to confirm his assessment through the spotting scope but realize that Johnny has not stopped firing and is now clearing the village of all its residents – residents who now lay in the blood-soaked snow – and are, quite clearly, women and children non-combatants. You watch speechless in horror, stunned at this effortless slaughter, unable to give an order over the radio.

**Gill, A. S. (2019). Artificial Intelligence and International Security: The Long View. *Ethics & International Affairs*, *33*(2), 169–179.** [**https://doi.org/10.1017/S0892679419000145**](https://doi.org/10.1017/S0892679419000145)**.**

How will emerging autonomous and intelligent systems affect the international landscape of power and coercion two decades from now? Will the world see a new set of artificial intelligence (AI) hegemons just as it saw a handful of nuclear powers for most of the twentieth century? Will autonomous weapon systems make conflict more likely or will states find ways to control proliferation and build deterrence, as they have done (fitfully) with nuclear weapons? And importantly, will multilateral forums find ways to engage the technology holders, states as well as industry, in norm setting and other forms of controlling the competition? The answers to these questions lie not only in the scope and spread of military applications of AI technologies but also in how pervasive their civilian applications will be. Just as civil nuclear energy and peaceful uses of outer space have cut into and often shaped discussions on nuclear weapons and missiles, the burgeoning uses of AI in consumer products and services, health, education, and public infrastructure will shape views on norm setting and arms control. New mechanisms for trust and confidence-building measures might be needed not only between China and the United States—the top competitors in comprehensive national strength today—but also among a larger group of AI players, including Canada, France, Germany, India, Israel, Japan, Russia, South Korea, and the United Kingdom.

**Goldfarb, A., & Lindsay, J. R. (2021). Prediction and judgment: Why artificial intelligence increases the importance of humans in war. *International Security*, *46*(3), 7-50.** [**https://doi.org/10.1162/isec\_a\_00425**](https://doi.org/10.1162/isec_a_00425)**.**

Recent scholarship on artificial intelligence (AI) and international security focuses on the political and ethical consequences of replacing human warriors with machines. Yet AI is not a simple substitute for human decision-making. The advances in commercial machine learning that are reducing the costs of statistical prediction are simultaneously increasing the value of data (which enable prediction) and judgment (which determines why prediction matters). But these key complements—quality data and clear judgment—may not be present, or present to the same degree, in the uncertain and conflictual business of war. This has two important strategic implications. First, military organizations that adopt AI will tend to become more complex to accommodate the challenges of data and judgment across a variety of decision-making tasks. Second, data and judgment will tend to become attractive targets in strategic competition. As a result, conflicts involving AI complements are likely to unfold very differently than visions of AI substitution would suggest. Rather than rapid robotic wars and decisive shifts in military power, AI-enabled conflict will likely involve significant uncertainty, organizational friction, and chronic controversy. Greater military reliance on AI will therefore make the human element in war even more important, not less.

**Greipl, A. R. (2023). Artificial intelligence in urban warfare: opportunities to enhance the protection of civilians?. *The Military Law and the Law of War Review*, *61*(2), 191-211.** [**https://doi.org/10.4337/mllwr.2023.02.03**](https://doi.org/10.4337/mllwr.2023.02.03)

As the hostilities in Ukrainian cities remind us once again, urban warfare persistently causes immense suffering and devastating consequences for civilians’ lives and livelihoods. At the same time, the events reveal the challenges urban warfare represents for militaries. The urban environment is one of the most complex environments within which to conduct military operations. To address the challenges this environment poses to militaries, several technologically advanced States are investing in the development of artificial intelligence to enhance a range of their military activities. The ways States have thus far prioritized the development of artificial intelligence systems, however, evidence that investments to improve militaries’ ability to mitigate civilian harm during urban warfare remain rather neglected. Hence, this article aims to demonstrate that despite the risks related to artificial intelligence applications, this technology has great potential for enhancing militaries’ ability to mitigate civilian harm further. But this requires governments to invest more in its development and use to that end. Recognizing artificial intelligence systems’ potential to reduce the military challenge of protecting civilians from harm during urban warfare is a pressing need, considering that conflicts increasingly occur in urban environments where the risks for civilians’ lives and their livelihoods grow exponentially.

**Grimm, P.W., Grossman, M.R., and Cormack, G.V. (2021). Artificial intelligence as evidence. *Northwestern Journal of Technology and Intellectual Property, 19(9).*** [**https://scholarlycommons.law.northwestern.edu/njtip/vol19/iss1/2**](https://scholarlycommons.law.northwestern.edu/njtip/vol19/iss1/2)**.**

This article explores issues that govern the admissibility of Artificial Intelligence (“AI”) applications in civil and criminal cases, from the perspective of a federal trial judge and two computer scientists, one of whom also is an experienced attorney. It provides a detailed yet intelligible discussion of what AI is and how it works, a history of its development, and a description of the wide variety of functions that it is designed to accomplish, stressing that AI applications are ubiquitous, both in the private and public sectors. Applications today include: health care, education, employment-related decision-making, finance, law enforcement, and the legal profession. The article underscores the importance of determining the validity of an AI application (i.e., how accurately the AI measures, classifies, or predicts what it is designed to), as well as its reliability (i.e., the consistency with which the AI produces accurate results when applied to the same or substantially similar circumstances), in deciding whether it should be admitted into evidence in civil and criminal cases. The article further discusses factors that can affect the validity and reliability of AI evidence, including bias of various types, “function creep,” lack of transparency and explainability, and the sufficiency of the objective testing of AI applications before they are released for public use. The article next provides an in-depth discussion of the evidentiary principles that govern whether AI evidence should be admitted in court cases, a topic which, at present, is not the subject of comprehensive analysis in decisional law. The focus of this discussion is on providing a step-by-step analysis of the most important issues, and the factors that affect decisions on whether to admit AI evidence. Finally, the article concludes with a discussion of practical suggestions intended to assist lawyers and judges as they are called upon to introduce, object to, or decide on whether to admit AI evidence.

**Grut, C. (2013). The Challenge of Autonomous Lethal Robotics to International Humanitarian Law. *Journal of Conflict & Security Law*, *18*(1), 5–23.** [**https://www.jstor.org/stable/26296246**](https://www.jstor.org/stable/26296246)

The concept of a truly autonomous weapons system—a system which is capable of operating itself, independently from human oversight—sounds more like science fiction than science fact. However, the reality is that weapons development is increasingly moving in this direction. Despite reassurances that humans will always be 'in the loop', significant amounts of autonomy have been given to certain weapons systems already. Such weapons present unique regulatory problems, arising not so much from their nature as weapons, but from their replacement of the human role in war and killing. This article considers the implications of increasing weapon autonomy for the humanitarian law principles of distinction and proportionality, and the concept of accountability for breaches of international humanitarian law.

**Hamilton, R. J. (2022). Future-Proofing US Law for War Crimes Investigations in the Digital Era. *Georgia Law Review* *57*(4), Article 2.** [**https://digitalcommons.law.uga.edu/glr/vol57/iss4/2**](https://digitalcommons.law.uga.edu/glr/vol57/iss4/2)

This is the first Article to look at the legal landscape facing international investigators seeking access to digital evidence regarding genocide, war crimes, crimes against humanity, and aggression. It analyzes Republic of Gambia v. Facebook(Meta), the first case to seek digital evidence from a U.S. social media company for an international proceeding on genocide. And it draws on material gleaned from background interviews with international investigators seeking digital evidence held by U.S. social media companies in relation to atrocities in Myanmar and Ukraine. This reveals two key problems facing international investigators. First, and in contrast to their counterparts in domestic criminal investigations, the Stored Communications Act provides no pathway through which international investigators can overcome the prohibition on disclosure of private digital evidence. Second, the ability of international investigators to access quasi-public digital evidence, and/or digital evidence that was public but has been removed by a social media company, is at the discretion of the social media company. A significant risk emerging from this arrangement is that evidence disclosure decisions are not made in a consistent and principled manner, but are instead driven by the self-interest of a few U.S. corporations, creating disparate outcomes across victim groups.

The Article recommends two, non-exclusive, reforms that could be undertaken in the short term to advance principled disclosure decisions for accountability, while ensuring privacy protections and data security. It also urges U.S. social media companies to develop and publish their own interim guidelines on how they make evidence disclosure decisions, with a presumption in favor of disclosing removed public and quasi-public evidence needed for the pursuit of accountability for the international crimes of genocide, war crimes, crimes against humanity, and aggression. The Article concludes by pointing to the need for a long-term incremental process of research, reform, and review to future-proof U.S. law for war crimes accountability in the digital era.

**Hammond, D. N. (2014). Autonomous weapons and the problem of state accountability. *Chicago Journal of International Law* *15*, 652.** [**https://chicagounbound.uchicago.edu/cjil/vol15/iss2/8**](https://chicagounbound.uchicago.edu/cjil/vol15/iss2/8)**.**

Currently in development and expected to become functional in the near future, fully autonomous weapons will have the capacity to operate entirely on their own, selecting targets and completing missions without human involvement. The prospective development of these weapons has raised concerns among some scholars who fear that the weapons would be unable to meet international legal standards. One criticism consistently raised is that in the event one of these weapons commits a war crime or human rights violation, it is not clear who should be held accountable. In this context, critics have focused primarily on whether military officers, designers, or manufacturers could (or should) be held individually liable. Few, however, have explored whether state liability is a viable option. This Comment takes up this inquiry, arguing that state liability would be preferable to individual liability because the state is in the best position to minimize its weapons' potential violations of international law and seems to be the most culpable actor in a moral sense. Nevertheless, although state liability is possible in the abstract, legal and practical barriers make the international legal regime as it stands ill-equipped to ensure that states would actually be held responsible for their weapons' crimes. If state liability is to resolve the accountability problem, the international community will need to make adjustments to this regime.

**Haney, B. (2020). Applied Artificial Intelligence in Modern Warfare and National Security Policy. *Hastings Science & Technology Law Journal, 11*(61).** [**http://dx.doi.org/10.2139/ssrn.3454204**](https://dx.doi.org/10.2139/ssrn.3454204)

Artificial Intelligence (AI) applications in modern warfare have revolutionized national security power dynamics between the United States, China, Russia, and the private industry. The United States has fallen behind in military technologies and is now at the mercy of big technology companies to maintain peace. After committing $150 billion toward the goal of becoming the AI technology world leader, China claimed success in 2018. In 2019, Chinese researchers published open-source code for AI missile systems controlled by deep reinforcement learning algorithms. Further, Russia’s continued interference in United States’ elections has largely been driven by AI applications in cybersecurity. Yet, despite outspending Russia and China combined on defense, the United States is failing to keep pace with foreign adversaries in the AI arms race. Previous legal scholarship dismisses AI militarization as futuristic science-fiction, accepting without support the United States’ prominence as the world leader in military technology. This inter-disciplinary article provides three main contributions to legal scholarship. First, this is the first piece in legal scholarship to take an informatics-based approach toward analyzing the range of AI applications in modern warfare. Second, this is the first piece in legal scholarship to take an informatics-based approach in analyzing national security policy. Third, this is the first piece to explore the complex power and security dynamics between the United States, China, Russia, and private corporations in the AI arms race. Ultimately, a new era of advanced weaponry is developing, and the United States Government is sitting on the sidelines.

**Hassan, S. A. M., & Ali, S. A. M. (2023). Limitations On Warfare Methods: A Brief Examination Under International Humanitarian Law. *Journal of Advances in Humanities Research*, *2*(4), 1-19.** [**https://doi.org/10.56868/jadhur.v2i4.189**](https://doi.org/10.56868/jadhur.v2i4.189)

The intricate landscape of modern warfare poses multi-faceted challenges to International Humanitarian Law (IHL). This research offers a nuanced examination of the limitations of warfare methods under IHL, underpinned by a systematic literature review (SLR) and an in-depth analysis of select case studies. By leveraging the PRISMA framework, a comprehensive search through notable databases yielded 42 salient articles. The study navigated through critical events, including the Rwandan Genocide, the Syrian conflict, the Mali cultural property crisis, and the environmental aftermath of the Gulf War. Through qualitative analysis, each case study was dissected for breaches, international responses, and the IHL framework's efficacy in addressing the challenges posed. Findings underscored significant gaps in IHL's applicability and enforcement. While the Rwandan genocide elucidated the international community's inertia, the Syrian conflict spotlighted IHL's frailties amidst multi-actor wars. Mali brought the urgent need for robust cultural property protection to the fore, while the Gulf War highlighted IHL's inadequacies in preempting environmental catastrophes. This study accentuates the pressing need for revisiting, fortifying, and expanding the IHL parameters to resonate with contemporary warfare realities. It advocates for a more agile, anticipatory, and encompassing IHL that delineates rights and wrongs and guarantees redress and reparation. The synthesis of historical case insights with present-day implications offers a roadmap for a more resilient IHL, paving the way for a harmonized global response to future conflicts.

**Heller, K.J. (2023, January 30). The concept of ‘the human’ in the critique of autonomous weapons. *Harvard National Security Journal 14.*** [**https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=4342529**](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4342529)

The idea that using “killer robots” in armed conflict is unacceptable because they are not human is at the heart of nearly every critique of autonomous weapons. Some of those critiques are deontological, such as the claim that the decision to use lethal force requires a combatant to suffer psychologically and risk sacrifice, which is impossible for machines. Other critiques are consequentialist, such as the claim that autonomous weapons will never be able to comply with international humanitarian law (IHL) because machines lack human understanding and the ability to feel compassion.

This article challenges anthropocentric critiques of AWS. Such critiques, whether deontological or consequentialist, are uniformly based on a very specific concept of “the human” who goes to war: namely, the Enlightenment subject who perceives the world accurately, understands rationally, is impervious to negative emotions, and reliably translates thought into action. Decades of research in cognitive psychology indicate, however, that the Enlightenment subject does not exist. On the contrary, human decision-making is profoundly distorted by cognitive and social biases, negative emotions, and physiological limitations — particularly when humans find themselves in dangerous and uncertain situations like combat. Given those flaws, and in light of rapid improvement in sensor and AI technology, it is only a matter of time until autonomous weapons are able to comply with IHL better than human soldiers ever have or ever will.

The article is divided into five sections. Section I critiques deontological objections to autonomous weapons. It shows that those objections wrongly anthropomorphize AWS by assuming they “decide” on targets in a manner similar to humans, overstate the inclination of humans to think about the consequences of killing and to risk sacrificing themselves for others, and are predicated on a romanticized and anachronistic view of war in which most killing takes place face-to-face.

Sections II and III critique consequentialist objections to autonomous weapons that focus on the jus in bello. Section II addresses the common argument that IHL compliance requires human understanding – particularly the ability to discern the intentions of potential targets and to make fact-sensitive and context-dependent determinations. The section begins by demonstrating that such understanding is far less necessary to IHL than AWS critics assume. It then explains why, in those situations where understanding is necessary, well-documented limits on human decision-making undermine the idea that human soldiers are more likely to comply with IHL than autonomous weapons. Finally, the section ends by discussing why the concept of “meaningful human control” is an undesirable solution to the supposed problems of AWS and should give way to the superior concept of “meaningful human certification.”

Section III responds to the claim that autonomous weapons should be prohibited because machines cannot feel compassion, an emotion that is both normatively and legally required on the battlefield. It makes three arguments. The first is that compassion is irrelevant to IHL compliance. The second is that the potential benefits of compassion in combat are far outweighed by the costs of negative emotions such as stress and anger. And the third is that compassion can lead to negative outcomes in combat as well as positive ones.

Section IV focuses on international criminal law, addressing the argument that the non-human nature of autonomous weapons makes it difficult, if not impossible, to hold humans responsible for war crimes AWS may commit. The section shows not only that the problem of “accountability gaps” is significantly overstated, but also that there is no significant difference between human soldiers and autonomous weapons in terms of criminal responsibility.

Finally, Section V explores a consequentialist objection to autonomous weapons that focuses on the jus ad bellum: namely, that replacing human soldiers with non-human machines will reduce the number of casualties during an armed conflict, making it easier for democratic states to go to war. The section argues that this is the most persuasive objection to AWS – and one that is actually understated, because it ignores the potential for such weapons to minimise civilian casualties, another factor that affects a state’s willingness to use armed force. As the section notes, however, the jus ad bellum critique is less an objection to AWS than to modern warfare itself, because most of the weapons developed over the past century have had precisely the same effect.

**Horowitz, M. C. (2020). Do emerging military technologies matter for international politics?. *Annual Review of Political Science*, *23*(1), 385-400.** [**https://doi.org/10.1146/annurev-polisci-050718-032725**](https://doi.org/10.1146/annurev-polisci-050718-032725)**.**

We live in a digital world. This fact has significant consequences for warfare. Two technologies in particular, cyber and drones, feature in military and intelligence operations and in scholarship. In addition, a new vein of scholarship is examining how advances in artificial intelligence have the potential to shape the future of warfare. While scholars disagree about the consequences of these technologies for international politics, they tend to agree that their consequences are mediated by the ability of military organizations, whether state or nonstate actors, to use them effectively in relevant military scenarios and in the pursuit of political ends. Studying newer military technologies, with less empirical evidence than is available for technologies that have been around for decades, also generates methodological challenges for research.

**Hua, Shin-Shin. (2020). Machine learning weapons and international humanitarian law: Rethinking meaningful human control. *Georgetown Journal of International Law.*** [**https://www.law.georgetown.edu/international-law-journal/wp-content/uploads/sites/21/2020/03/GT-GJIL200015.pdf**](https://www.law.georgetown.edu/international-law-journal/wp-content/uploads/sites/21/2020/03/GT-GJIL200015.pdf)**.**

AI’s revolutionizing of warfare has been compared to the advent of the nuclear bomb. Machine learning technology, in particular, is paving the way for future automation of life-or-death decisions in armed conflict. But because these systems are constantly “learning,” it is difficult to predict what they will do or understand why they do it. Many therefore argue that they should be prohibited under international humanitarian law (IHL) because they cannot be subject to meaningful human control. But in a machine learning paradigm, human control may become unnecessary or even detrimental to IHL compliance. In order to leverage the potential of this technology to minimize casualties in conflict, an unthinking adherence to the principle of “the more control, the better” should be abandoned. Instead, this Article seeks to define prophylactic measures that ensure machine learning weapons can comply with IHL rules. Further, it explains

how the unique capabilities of machine learning weapons can facilitate a more robust application of the fundamental IHL principle of military necessity.

**Jaynes, T. L. (2020). Legal personhood for artificial intelligence: citizenship as the exception to the rule. *AI & Society*, *35*(2), 343-354.** [**https://doi.org/10.1007/s00146-019-00897-9**](https://doi.org/10.1007/s00146-019-00897-9)

The concept of artificial intelligence is not new nor is the notion that it should be granted legal protections given its influence on human activity. What is new, on a relative scale, is the notion that artificial intelligence can possess citizenship—a concept reserved only for humans, as it presupposes the idea of possessing civil duties and protections. Where there are several decades’ worth of writing on the concept of the legal status of computational artificial artefacts in the USA and elsewhere, it is surprising that lawmakers internationally have come to a standstill to protect our silicon brainchildren. In this essay, it will be assumed that future artificial entities, such as Sophia the Robot, will be granted citizenship on an international scale. With this assumption, an analysis of rights will be made with respect to the needs of a non-biological intelligence possessing legal and civic duties akin to those possessed by humanity today. This essay does not present a full set of rights for artificial intelligence—instead, it aims to provide international jurisprudence evidence aliunde ab extra de lege lata for any future measures made to protect non-biological intelligence.

**Jaynes, T.L. (2021). Citizenship as the exception to the rule. *AI & Society 36,* 911-930.** [**https://doi.org/10.1007/s00146-020-01105-9**](https://doi.org/10.1007/s00146-020-01105-9)

This addendum expands upon the arguments made in the author’s 2020 essay, “Legal Personhood for Artificial Intelligence:Citizenship as the Exception to the Rule”, in an effort to display the significance human augmentation technologies will have on (feasibly) inadvertently providing legal protections to artificial intelligence systems (AIS)—a topic only briefly addressed in that work. It will also further discuss the impacts popular media have on imprinting notions of computerised behaviour and its subsequent consequences on the attribution of legal protections to AIS and on speculative technological advancement that would aid the sophistication of AIS.

**Jasanoff, S. (2017). Virtual, Visible, and Actionable: Data Assemblages and the Sightlines of Justice. *Big Data & Society,* *4*(2).** [**https://www.hks.harvard.edu/publications/virtual-visible-and-actionable-data-assemblages-and-sightlines-justice**](https://www.hks.harvard.edu/publications/virtual-visible-and-actionable-data-assemblages-and-sightlines-justice)

This paper explores the politics of representing events in the world in the form of data points, data sets, or data associations. Data collection involves an act of seeing and recording something that was previously hidden and possibly unnamed. The incidences included in a data set are not random or unrelated but stand for coherent, classifiable phenomena in the world. Moreover, for data to have an impact on law and policy, such information must be seen as actionable, that is, the aggregated data must show people both something they can perceive and something that demands interrogation, explanation, or resolution. Actionable data problematize the taken-for-granted order of society by pointing to questions or imbalances that can be corrected or rectified, or simply better understood, through systematic compilations of occurrences, frequencies, distributions, or correlations. The paper describes and analyzes three different modes of authorized seeing that render data on global environmental phenomena such as climate change both visible and actionable. It argues that the political force of environmental data compilations derives from the divergent epistemological standpoints and expert practices associated with producing views from nowhere, everywhere, and somewhere.

**Jensen, B.M., Whyte, C., Cuomo, S. (2020). Algorithms at War: The Promise, Peril, and Limits of Artificial Intelligence, *International Studies Review* *22*(3), 526–550.** [**https://doi.org/10.1093/isr/viz025**](https://doi.org/10.1093/isr/viz025)**.**

How might rapid advances in artificial intelligence (AI) technologies affect the construction and application of military power? Despite the emerging importance of AI systems in defense modernization initiatives, there has been little empirical or theoretical study from the perspective of the international relations (IR) and security studies fields. This article addresses this shortcoming by describing AI developments and assessing the manner in which AI is likely to affect military organizations. We focus specifically on military power, as new methods and modes thereof will alter the constitution of security relationships around the world and affect the ability of states to bargain, signal, and influence in the twenty-first century. We argue that, though rapid adoption of AI technologies stands to transform states’ ways of war on a number of fronts, an AI revolution brings with it new forms of risk that must be reconciled with the widespread integration of algorithmic systems across military functions. Where new technology promises a transformation of the character of military power in some veins, it also complicates the cognitive aspects of decision-making and bureaucratic interactions in security institutions. The speed with which complex integrated AI systems enable entirely new modes of war also stands to detach human agency in a potentially destabilizing fashion from the conduct of warfare on several fronts. Preventing the negative externalities of these “ghosts in the machine” will involve significant efforts to educate decision makers, promote accountability, and restrain irresponsible employment of AI,

**Johansson, S., & Chan, F. (2024, February 2). What ethical concerns might arise when robots are in charge of the decision making in warfare?. *NCO Journal.*** [**https://www.armyupress.army.mil/Journals/NCO-Journal/Archives/2024/February/The-Ethics-of-Robots-in-War/**](https://www.armyupress.army.mil/Journals/NCO-Journal/Archives/2024/February/The-Ethics-of-Robots-in-War/)

The world is evolving and expanding exponentially, with corporations, small businesses, and even individuals of all lifestyles fully integrating technology, specifically machines, and robots into their lives to make daily activities easier.

Machines can perform a variety of simple to complex operations. They range from inexpensive household items like coffeepots that brew individual cups of coffee when and how you want them to multimillion-dollar robotic systems programmed to defend our nation.

In robotics, innovation faces opposition despite its advantages, especially when it comes to military applications. Robots are cheaper to make than training and caring for living, breathing human Soldiers. Robot programming allows for specific guiding rules that allow for split-second decisions, minimizing errors caused by deficient human skills and reducing the potential for collateral damage.

If the U.S. Army used robotic Soldiers, it would also significantly increase military capabilities, further establishing our place as a world superpower by keeping us ahead of potential competition with near-peer adversaries.

Despite societal concerns, both in the U.S. and worldwide, the military needs to evaluate, understand, and fully embrace robotic technology to ensure our nation’s welfare and humankind’s future.

**Johnson, J. (2019). Artificial intelligence & future warfare: implications for international security. *Defense & Security Analysis*, *35*(2), pp. 147–169.** [**https://doi.org/10.1080/14751798.2019.1600800**](https://doi.org/10.1080/14751798.2019.1600800)

Recent developments in artificial intelligence (AI) suggest that this emerging technology will have a deterministic and potentially transformative influence on military power, strategic competition, and world politics more broadly. After the initial surge of broad speculation in the literature related to AI this article provides some much needed specificity to the debate. It argues that left unchecked the uncertainties and vulnerabilities created by the rapid proliferation and diffusion of AI could become a major potential source of instability and great power strategic rivalry. The article identifies several AI-related innovations and technological developments that will likely have genuine consequences for military applications from a tactical battlefield perspective to the strategic level.

**Johnson, J. (2020). Artificial intelligence in nuclear warfare: a perfect storm of instability?. *The Washington Quarterly, 43*(2), 197-211. DOI:**[**10.1080/0163660X.2020.1770968**](http://dx.doi.org/10.1080/0163660X.2020.1770968)**.**

A significant gap exists between the expectations and fears of public opinion, policymakers, and global defense communities about artificial intelligence (AI) and its actual military capabilities, particularly in the nuclear sphere. The misconceptions that exist today are largely caused by the hyperbolic depictions of AI in popular culture and science fiction, most prominently the Skynet system in The Terminator. Misrepresentations of the potential opportunities and risks in the military sphere (or “military AI”) can obscure constructive and crucial debate on these topics—specifically, the challenge of balancing the potential operational, tactical, and strategic benefits of leveraging AI, while managing the risks posed to stability and nuclear security. This article demystifies the hype surrounding AI in the context of nuclear weapons and, more broadly, future warfare. Specifically, it highlights the potential, multifaceted intersections of this disruptive technology with nuclear stability. The inherently destabilizing effects of military AI may exacerbate tension between nuclear-armed great powers, especially China and the United States, but not for the reasons you may think.

**Johnson, J. (2022). Delegating strategic decision-making to machines: Dr. Strangelove Redux?. *Journal of Strategic Studies*, *45*(3), 439-477.** [**https://doi.org/10.1080/01402390.2020.1759038**](https://doi.org/10.1080/01402390.2020.1759038)**.**

Will the use of artificial intelligence (AI) in strategic decision-making be stabilizing or destabilizing? What are the risks and trade-offs of pre-delegating military force to machines? How might non-nuclear state and non-state actors leverage AI to put pressure on nuclear states? This article analyzes the impact of strategic stability of the use of AI in the strategic decision-making process, in particular, the risks and trade-offs of pre-delegating military force (or automating escalation) to machines. It argues that AI-enabled decision support tools - by substituting the role of human critical thinking, empathy, creativity, and intuition in the strategic decision-making process - will be fundamentally destabilizing if defense planners come to view AI’s ‘support’ function as a panacea for the cognitive fallibilities of human analysis and decision-making. The article also considers the nefarious use of AI-enhanced fake news, deepfakes, bots, and other forms of social media by non-state actors and state proxy actors, which might cause states to exaggerate a threat from ambiguous or manipulated information, increasing instability.

**Johnson, J. (2022). The AI commander problem: Ethical, political, and psychological dilemmas of human-machine interactions in AI-enabled warfare. *Journal of Military Ethics*, *21*(3-4), 246-271.** [**https://doi.org/10.1080/15027570.2023.2175887**](https://doi.org/10.1080/15027570.2023.2175887)**.**

Can AI solve the ethical, moral, and political dilemmas of warfare? How is artificial intelligence (AI)-enabled warfare changing the way we think about the ethical-political dilemmas and practice of war? This article explores the key elements of the ethical, moral, and political dilemmas of human-machine interactions in modern digitized warfare. It provides a counterpoint to the argument that AI “rational” efficiency can simultaneously offer a viable solution to human psychological and biological fallibility in combat while retaining “meaningful” human control over the war machine. This Panglossian assumption neglects the psychological features of human-machine interactions, the pace at which future AI-enabled conflict will be fought, and the complex and chaotic nature of modern war. The article expounds key psychological insights of human-machine interactions to elucidate how AI shapes our capacity to think about future warfare's political and ethical dilemmas. It argues that through the psychological process of human-machine integration, AI will not merely force-multiply existing advanced weaponry but will become *de facto* strategic actors in warfare – the “AI commander problem.”

**Kent, R. (2015). The future of warfare: Are we ready?. *International Review of the Red Cross*, *97*(900), 1341-1378. doi:10.1017/S1816383116000412**

To what extent do the ways in which we anticipate threats, analyze their possible consequences and determine ways to mitigate them explain the causes of warfare in the future? This article – though never attempting to predict – poses plausible causes of future wars that may stem from transformative change over the next two decades. In asking the question “Are we ready?” to deal with such wars, the answer is framed in terms of the interrelationship between the prospect of profound change, emerging tensions, unprecedented violence and organizational capacities to deal with complexity and uncertainty. To be prepared to deal with the prospect of future wars, relevant organizations have to be more anticipatory and adaptive, while at the same time looking for new ways to engage the wider international community. The article concludes with a set of recommendations intended to meet such organizational challenges – with the aspiration that the question “Are we ready?” can be answered more affirmatively in the future.

**Khalil, A., & Krishna Raj, S.A. (2024). Deployment of autonomous weapon systems in the warfare: Addressing accountability gaps and reformulating international criminal law. *Balkan Social Science Review*, *23*(23), 261–285.** [**https://doi.org/10.46763/BSSR242323261kr**](https://doi.org/10.46763/BSSR242323261kr)

The importance of autonomous weapon systems (AWS) is particularly apparent in contemporary conflicts, in which all parties are striving to expand the autonomy of these systems, particularly in offensive contexts. The traditional understanding is that the deployment of AWS poses significant threats to civilian life, as well as a range of delicate legal concerns. The narrow understanding is evident in the principal drawbacks of AWS in the absence of human supervision, which results in unpredictable and unreliable outputs, particularly during the targeting and engagement stages, leading to a clear breach of international humanitarian law (IHL). This creates a legal vacuum regarding the attribution of criminal responsibility. This research paper aims to characterise AWS and the implications of their potential deployment, with the primary goal of better identifying and understanding emerging legal concerns. Furthermore, the authors seek to offer a comprehensive understanding of the concept of autonomy, which is internationally accepted. Additionally, they explore the validity of the statement that AWS create an accountability gap relating to the International Criminal Court (ICC) Statute by analysing various types of AWS that are categorised in this study and deployed in modern wars. While the ICC Statute does not explicitly address all issues raised by using AWS, the study presents an interpretation that criminal responsibility could be assigned to all individuals involved in the use of AWS, especially in thorny situations. This study employs an analytical approach to examine and analyse traditional theories of criminology and the ICC Statute in relation to the ability to attribute individual criminal responsibility when AWS may be used. This study aims to provide a new perspective on criminal legal rules that are compatible with modern technology.

**Koenig, A., Freeman, L. (2022). Cutting-edge evidence: Strengths and weaknesses of new digital investigation methods in litigation, *73 Hastings Law Journal,* 1233.** [**https://humanrights.berkeley.edu/publications/cutting-edge-evidence-strengths-and-weaknesses-of-new-digital-investigation-methods-in-litigation/**](https://humanrights.berkeley.edu/publications/cutting-edge-evidence-strengths-and-weaknesses-of-new-digital-investigation-methods-in-litigation/)

The increased use of digital technologies in daily life has led to a steep rise in the introduction of highly technical evidence and expert witness testimony in criminal and civil litigation. The growing use of novel, quickly-developing investigation methods for digital evidence presents several challenges related to the difficulty lay persons have in judging complex forensic methodologies. The lack of judicial and legal training in the underlying methods and their potential vulnerabilities can result in fact-finders who over-rely on experts’ conclusions without properly interrogating the evidence themselves. While many of the scientific and analytical methods employed by digital investigators can be promising additions to investigative toolkits, enthusiasm for these techniques should be tempered with healthy skepticism — and knowledge of the most helpful questions to ask about new investigative processes. In this Article, we identify the very real vulnerabilities in digital open source investigations and encourage careful analysis of each component in order to mitigate the risks. We recommend that investigators preserve digital material according to established forensic standards and carefully record the steps of their online investigation and analysis. Expert witnesses should be strictly prohibited from giving opinions on matters that stretch beyond the scope of their education, training, and well-established expertise. Lawyers and judges must be prepared to ascertain the reliability and validity of digital open source investigations and their findings through thorough interrogation of the underlying data. As a best practice, digital evidence should be triangulated with physical, testimonial, or other documentary evidence whenever possible. If conducted carefully and professionally, digital open source investigations can offer tremendous value for both civil and criminal proceedings.

**Kwik, J. (2023). The conceptual roots of the criminal responsibility gap in Autonomous Weapon Systems. *Melbourne Journal of International Law*, *24*(1), 1-26.** [**http://www8.austlii.edu.au/cgi-bin/viewdoc/au/journals/MelbJIL/2023/1.html**](http://www8.austlii.edu.au/cgi-bin/viewdoc/au/journals/MelbJIL/2023/1.html)

One major reason for the controversy around autonomous weapon systems (‘AWS’) is the concern that no criminal liability is possible for resulting war crimes. This article takes a comprehensive look at one factor, the cognitive element of mens rea, and how and when characteristics specific to artificial intelligence (‘AI’) can render it more difficult to assign criminal liability to the deploying commander. It takes a multidisciplinary approach, considering both technical characteristics of modern AI and realistic conditions under which AWS are used. The article finds that modern AI primarily induces reduced perceivability through imperfect tracking of human intuition, opacity and generic reliability metrics. It also finds that AWS make it easier to willingly avoid acquiring cognition simply through inaction. Subsequently, it attempts to locate the exact loci of the problem within criminal law’s spectrum of intent. This article finds that the epicentre of difficulty lies at the intermediate level of risk-taking, and particularly situations of generic risk: the condition where there is awareness only of a nondescript, indeterminate probability of ‘something going wrong’. In contrast, no-gap situations are identified higher up the ladder of intent where there is purpose or virtual certainty, and judicious gaps lower down where we want ‘impunity’ for justified risk-taking and genuine accidents. Additionally, this article also considers the dangers of manufactured ignorance, where the risk can theoretically be known but in practice was not, due to a prior, separate omission. It ends with recommendations to address these challenges, including reducing opacity, standardising iterative investigations and enforcing technical trainings.

**Lewis,** [**D.A.**](https://dash.harvard.edu/bitstream/handle/1/37366359/360-481-1-SM.pdf?isAllowed=y&sequence=1) **(2020).** [**International legal regulation of the employment of artificial-intelligence-related technologies in armed conflict**](https://dash.harvard.edu/bitstream/handle/1/37366359/360-481-1-SM.pdf?isAllowed=y&sequence=1)**. *Moscow Journal of International Law* *2*, 53–64.** [**https://doi.org/10.24833/0869-0049-2020-2-53-64**](https://doi.org/10.24833/0869-0049-2020-2-53-64)**.**

In recent years, increased attention has been dedicated at the international level to legal issues concerning the possible employment of artificial intelligence-related technologies in hostilities in armed conflict. Most prominently, discussions in the framework of the Convention on Certain Conventional Weapons (CCW) have addressed juridical aspects relative to emerging technologies in the area of lethal autonomous weapons systems.

I analyze contemporary intergovernmental debates in the context of the CCW, international legal frameworks pertaining to armed conflict, and developments in relevant technologies. I do so to trace current trajectories and generate an analytical framework to help apply legal responsibility.

A disagreement has arisen among certain States in the context of the CCW as to whether to develop a new primary legal norm or whether existing international humanitarian law is sufficient. Taking account of that current normative impasse, I propose an analytical framework aimed at ensuring the applicability of international legal responsibility in respect of the employment of AI-related technologies in armed conflict.

Given the range of relevant technologies, the employment of AI in armed conflict may occur across diverse thematic and functional areas: not only in the conduct of hostilities, including weapons, but also detention, humanitarian services,maritime systems, and many other areas. Identification of the general concepts and specific attributes necessary to apply international legal responsibility across the array of implicated areas may help provide a framework through which to respect the law, guide behavior, pursue accountability, and generate areas of greater normative consensus.

**Maas, M. M. (2019). International law does not compute: Artificial intelligence and the development, displacement or destruction of the global legal order. *Melbourne Journal of International Law*, *20*(1), pp. 29-57.** [**https://classic.austlii.edu.au/au/journals/MelbJIL/2019/3.html**](https://classic.austlii.edu.au/au/journals/MelbJIL/2019/3.html)

Within the coming decade, the deployment of artificial intelligence (‘AI’) appears likely to have a disruptive impact on global affairs. What will such ‘globally disruptive’ AI imply for the form, function and viability of international law? I briefly sketch the long history of technological innovation driving, shaping and destroying international law. Drawing on scholarship on the relation between new technologies and international law, I argue that new technology changes legal situations both directly, by creating new entities or enabling new behaviour, and indirectly, by shifting incentives or values. I argue that development of increasingly more disruptive AI may produce three types of global legal impacts. The first is ‘legal development’ (patching); the second is ‘legal displacement’ (substitution); the third is ‘legal destruction’ (erosion). I discuss the potential impact of AI in all three modalities, and the implications for international relations. I argue that many of the challenges raised by AI could in principle be accommodated in the international law system through legal development, and that while AI may aid in compliance enforcement, the prospects for legal displacement — a shift towards an ‘automated international law’ — look slim. However, I also conclude that technical and political features of the technology will in practice render AI destructive to key areas of international law: the legal gaps it creates will be hard to patch, and the strategic capabilities it offers chip away at the rationales for powerful states to engage fully in, or comply with, international law regimes. This suggests some risk of obsolescence of distinct international law regimes.

**Maas, M.M. (2019). Innovation-Proof Global Governance for Military Artificial Intelligence? *Journal of International Humanitarian Legal Studies* 10, no. 1, pp. 129-157.** [**https://matthijsmaas.com/uploads/Maas%20-%202019%20-%20Innovation-Proof%20Governance%20for%20Military%20AI%20how%20I.pdf**](https://matthijsmaas.com/uploads/Maas%20-%202019%20-%20Innovation-Proof%20Governance%20for%20Military%20AI%20how%20I.pdf)

Amidst fears over artificial intelligence ‘arms races’, much of the international debate on governing military uses of AI is still focused on preventing the use of lethal autonomous weapons systems (laws). Yet ‘killer robots’ hardly exhaust the potentially problematic capabilities that innovation in military AI (MAI) is set to unlock. Governance initiatives narrowly focused on preserving ‘meaningful human control’ over laws therefore risk being bypassed by the technological state-of-the-art. This paper departs from the question: how can we formulate ‘innovation-proof governance’ approaches that are resilient or adaptive to future developments in military AI? I develop a typology for the ways in which MAI innovation can disrupt existing international legal frameworks. This includes ‘direct’ disruption – as new types of MAI capabilities elude categorization under existing regimes – as well as ‘indirect’ disruption, where new capabilities shift the risk landscape of military AI, or change the incentives or values of the states developing them. After discussing two potential objections to ‘innovation-proof governance’, I explore the advantages and shortcomings of three possible approaches to innovation-proof governance for military AI. While no definitive blueprint is offered, I suggest key considerations for governance strategies that seek to ensure that military AI remains lawful, ethical, stabilizing, and safe.

**Matthias, A. (2003, November 30). The Responsibility Gap. Ascribing Responsibility for the Actions of Learning Automata. *Kluwer Academic Publishers.*** [**https://www.academia.edu/243900/The\_Responsibility\_Gap\_Ascribing\_Responsibility\_for\_the\_Actions\_of\_Learning\_Automata**](https://www.academia.edu/243900/The_Responsibility_Gap_Ascribing_Responsibility_for_the_Actions_of_Learning_Automata)

The paper discusses the concept of responsibility in relation to the actions of autonomous learning machines, observing the challenges that arise as these systems operate with reduced human control. It argues that traditional frameworks for ascribing responsibility fail to address scenarios where machines exceed operator's capabilities or malfunction. The implications of this shift on societal norms and individual accountability are explored, highlighting the need for new paradigms to manage responsibility in technologically advanced contexts.

**Maqbool, A., & Anwar, A. (2023). Warfare and Machines: An In-depth Study of Autonomous Weapons in the Context of International Humanitarian Law. *Society, Law and Policy Review*, *2*(1), 01-14. DOI:** [**https://doi.org/10.62585/slpr.v2i1.25**](https://doi.org/10.62585/slpr.v2i1.25)

This research article looks at the ramifications of autonomous weapon systems within the context of international humanitarian law. The author discusses the most recent breakthroughs in autonomous weapons, their features, and the benefits to the parties that use them. The study investigates how autonomous weapons comply with treaty law, focusing on the Geneva Conventions, the Convention on Certain Conventional Weapons, and Article 36 of Additional Protocol I. The study also examines how customary international law concepts be applied to autonomous weapons. To provide policymakers and other relevant authorities the clarity on the subject, the author has also offered some recommendations in the concluding section. In addition to analyzing the legal framework surrounding autonomous weapon systems, this research article delves into the ethical considerations associated with their deployment. The author explores the potential consequences of autonomous weapons on human rights, accountability, and the moral implications of delegating lethal decision-making to machines. By scrutinizing the intersection of international humanitarian law and ethical considerations, the article aims to contribute to a comprehensive understanding of the multifaceted challenges posed by autonomous weapon systems. The concluding section not only offers recommendations for legal compliance but also emphasizes the importance of a holistic approach that integrates ethical considerations into the regulatory landscape governing these advanced technologies. This holistic perspective aims to guide policymakers and relevant authorities in formulating balanced and responsible policies concerning the development and use of autonomous weapon systems.

**McDougall, C. (2019). Autonomous weapon systems and accountability: Putting the cart before the horse. *Melbourne Journal of International Law*, *20*(1), pp. 58-87.** [**https://classic.austlii.edu.au/au/journals/MelbJIL/2019/4.html**](https://classic.austlii.edu.au/au/journals/MelbJIL/2019/4.html)

Arguments that in many scenarios there will exist an ‘accountability gap’ where civilians are unlawfully killed through the use of an autonomous weapon system (‘AWS’) have been advanced to justify either the prohibition or restriction of AWS. This article examines the accountability problem through a critical review of the literature on accountability and AWS in order to identify why some experts say there will be no accountability gap while others argue there will, why some do not see this as a problem and others do, and why some consider this is a problem that has a solution while others see it as irresolvable. It is demonstrated that in large part these differing conclusions are the result of varying assumptions and preconditions. Without questioning the inherent value of accountability in the broad, it is argued that solutions to the debate over AWS will not be found in international criminal law, which should not be used as a backdoor to address perceived shortcomings in international humanitarian law. It is further argued that no analysis of the accountability problem will provide meaningful guidance as to whether the international community needs to prohibit or restrict AWS given that, one way or another, international criminal law can be amended to plug the accountability gap, if this is the desired policy outcome.

**Neslage, K. (2015). Does "Meaningful Human Control" Have Potential for the Regulation of Autonomous Weapon Systems?. *University of Miami National Security & Armed Conflict Law Review 6*, 151.** [**https://repository.law.miami.edu/umnsac/vol6/iss1/7**](https://repository.law.miami.edu/umnsac/vol6/iss1/7)**.**

The Convention on Certain Conventional Weapons ("CCW") is a treaty based upon the principles of international law that weapons used for engagement in armed conflict should not cause unnecessary injury and suffering. Eighty-two states are party to this treaty and continue to negotiate and ratify additional protocols that ban the use of specific weapons. In May 2014, the CCW held its first meeting on the topic of Lethal Autonomous Weapon Systems ("AWS"). The specific mandate of the meeting was to discuss emerging technologies in the area of AWS, and was generally regarded as a preliminary discussion, in an area of rapid technological change. Another meeting was held in April 2015 During the 2014 meeting, sessions were held on the ethical, sociological, military, and legal aspects of AWS, which included technical aspects on how to define autonomy and how autonomy can be measured. Amongst this discussion was the idea of adopting the treaty language of "meaningful human control" ("MHC") as a way to measure autonomy and/or regulate AWS. The purpose of this paper is to question the concept of MHC and how it may or may not be a useful legal tool for regulating AWS. Part II of this paper will provide a working definition of autonomy for the sake of creating a common understanding within this paper. It will then explain the current status of AWS technology and what possibilities the future holds for autonomy. Part III will discuss state reactions to MHC and the current legal system that already regulates AWS. Part IV will then critique what it may mean for states to comply with MHC and what challenges may be presented by implementation through taking a closer look at the "human control" in MHC. Part V of this paper will analyze the "meaningful" aspect of MHC, and to what degree humans can, will, or should delegate tasks to autonomous systems. Theories of automation bias, automation complacency, and their role in the field of AWS as well as other fields will be discussed. Part VI will offer concluding remarks on why MHC appears useful because of its simplicity, but fails to be useful in application because of its vagueness.

**Pandey, S. K., & Narayan, A. (2023). Means and Methods of Warfare and International Humanitarian Law in the Age of Artificial Intelligence and Machine Learning. *International Journal of Legal Science and Social Innovation, 5(3),* 160-181.** [**https://doij.org/10.10000/IJLSI.111607**](https://doij.org/10.10000/IJLSI.111607)

The development of AI and ML technologies has significantly altered how war is fought, which puts the existing legal system of international humanitarian law in jeopardy. The implications of AI and ML in combat are examined in this abstract, which emphasises the necessity for a thorough knowledge of their potential effects on IHL principles. Artificial intelligence (AI) and machine learning (ML) are being incorporated into weapon systems, targeting procedures, and decision-making, which has ramifications for distinction, proportionality, and precautions in assault. In the creation, implementation, and application of AI and ML technologies, the abstract emphasises the significance of ensuring accountability, human control, and compliance with IHL. Additionally, it emphasises the necessity of increased communication between nations, international organisations, and specialists to address the moral and legal issues raised. The goal is to increase awareness of the critical concerns involving AI, ML, and IHL and to promote additional study and conversations to make sure that these developments in combat adhere to the IHL's guiding principles of humanity, distinction, and proportionality.

**Peake, J. (2024). Challenges of Using Digital Evidence for War Crimes Prosecutions: Availability, Reliability, Admissibility. *AJIL Unbound*, *118*, 57–61.** [**https://doi.org/10.1017/aju.2024.5**](https://doi.org/10.1017/aju.2024.5)

Digital evidence has the potential to transform the accountability landscape. However, several obstacles must be overcome to use it effectively for war crimes prosecutions. The availability of digital evidence can be impacted by a range of factors, from Internet connectivity to removal of content by platforms to the difficulty of identifying relevant content through the fast pace of social media. The reliability of digital information is increasingly being called into question because of deliberate disinformation campaigns by parties to conflicts, as well as the fear of media manipulated by Artificial Intelligence (AI). And questions around chain of custody and admissibility of digital evidence have not fully been resolved by international courts. This essay unpacks some of these challenges. It suggests some ways in which governments and big tech should seek to ensure access to digital spaces and put in place measures to increase the integrity of online content, and how investigators and lawyers can gather, authenticate, archive, and establish chain of custody to ensure it can be used in accountability processes.

**Putro, Y. M., Tarigan, M. I., & Asyari, H. A. (2023). Artificial Intelligence in Indo-Pacific: Quo Vadis International Humanitarian Law and Regional Peace and Security in Southeast Asia. *Lentera Hukum*, *10*, 391.** [**https://doi.org/10.19184/ejlh.v10i3.43449**](https://doi.org/10.19184/ejlh.v10i3.43449)**.**

The use of Artificial Intelligence in the military is like two sides of a coin. It can provide convenience and aid in military operations but has the potential to hinder military operations. Dangerous and potentially catastrophic for humanity will be inevitable as no restrictions on its use. The United States, China, Australia, Japan, and India are examples of nations whose militaries have developed artificial intelligence technology. Geographically, Southeast Asia, which is located in the middle of these nations, will experience a significant impact due to its tight maritime borders if there is no international consensus on the military application of artificial intelligence technology. An autonomous or autonomous system to operate this technology will reduce the amount of human control and allow it to operate without any human intervention. It will be a threat to the application of the fundamental principles of international humanitarian law, such as the distinction principle, and proportionality principle. Where these principles are tightly intertwined with human command and control in making decisions regarding the execution of attacks. The article employs normative legal methodology. Furthermore, this paper endeavours to assess the pertinence of principles in international humanitarian law during the era of the artificial intelligence arms race. It also delves into the contribution of ASEAN in upholding stability, peace, and security in the Southeast Asia region, thereby reinforcing the importance of this research. This research emphasises the importance of aligning the progress of artificial intelligence in military contexts with core principles of international humanitarian law. It underscores the need for ASEAN to safeguard regional peace and security by establishing a novel regulatory framework that outlines restrictions on the development and deployment of artificial intelligence for military objectives.

**Renblad, N. A. (2023). International Humanitarian Law, Technology & Warfare: The Role of International Humanitarian Law in Regulating Emerging Military Strategies and Weapons Technologies. *Yonsei Journal of International Studies,* 15(2), 67-89.** [**https://doi.org/10.5281/zenodo.13124027**](https://doi.org/10.5281/zenodo.13124027)**.**

The international community has, in light of the new use of technology within the field of international humanitarian law, adopted several declarations and treaties in reaction to new technologies within weapons and warfare. Discussions in the international community revolve around the application of existing international humanitarian law to technology and war, including cyber military operations, military artificial intelligence, and the use of drones. This paper will examine the role of international humanitarian law in dealing with emerging military strategies and weapons technologies. Furthermore, the effectiveness of the law in regulating these technological advancements will be examined to assess its ability to adequately regulate the mentioned emerging issues. This paper concludes that international law has generally shown reluctance and remained silent in response to the emergence of new technologies. Existing norms of international humanitarian law can, to an extent, be applied to emerging military strategies and weapons technologies. However, the unique features of advanced technologies, the unpredictability and risks posed, can be seen to limit the application of existing norms.

**Roff, H.M. (2014). The Strategic Robot Problem: Lethal Autonomous Weapons in War. *Journal of Military Ethics*, *13*(3), 211–227.** [**https://doi.org/10.1080/15027570.2014.975010**](https://doi.org/10.1080/15027570.2014.975010)**.**

The present debate over the creation and potential deployment of lethal autonomous weapons, or ‘killer robots’, is garnering more and more attention. Much of the argument revolves around whether such machines would be able to uphold the principle of noncombatant immunity. However, much of the present debate fails to take into consideration the practical realities of contemporary armed conflict, particularly generating military objectives and the adherence to a targeting process. This paper argues that we must look to the targeting process if we are to gain a fuller picture of the consequences of creating or fielding lethal autonomous robots. This paper argues that once we look to how militaries actually create military objectives, and thus identify potential targets, we face an additional problem: the Strategic Robot Problem. The ability to create targeting lists using military doctrine and targeting processes is inherently strategic, and handing this capability over to a machine undermines existing command and control structures and renders the use for humans redundant. The Strategic Robot Problem provides prudential and moral reasons for caution in the race for increased autonomy in war.

**Roumate, F. (2021). Artificial Intelligence, Ethics and International Human Rights Law. *The International Review of Information Ethics, Vol 29.*** [**https://informationethics.ca/index.php/irie/article/view/422/406**](https://informationethics.ca/index.php/irie/article/view/422/406)

The ethics of artificial intelligence is the response to a new dilemma that demands international society to provide a legal response to the many ethical challenges artificial intelligence creates. COVID-19 accelerates the use of AI in all countries and all fields. The pandemic is accelerating the transition to a society that is increasingly based on the use of, and reliance on, AI, and this also enhances the threats and creates new risks related to human rights. Artificial Intelligence (AI) influences human rights and international humanitarian law. This paper addresses international mechanisms and ethics as new rules which can ensure the protection of human rights in the age of AI. Two arguments are discussed in this study. Considering the ubiquitous and global reach of AI, the challenges it imposes requires an international legal oversight, a requirement that highlights the importance of ethical frameworks. In conclusion, the paper emphasizes how optimal action is needed to protect human rights in the age of AI. Rethinking international law and human rights and enhancing the ethical frameworks have thus become obligatory rather than a choice.

**Schuppli, S. (2014). Deadly algorithms: Can legal codes hold software accountable for code that kills? *Radical Philosophy* *187*.** [**https://www.radicalphilosophy.com/commentary/deadly-algorithms**](https://www.radicalphilosophy.com/commentary/deadly-algorithms)**.**

Algorithms have long adjudicated over vital processes that help to ensure our well being and survival, from pacemakers that maintain the natural rhythms of the heart, and genetic algorithms that optimise emergency response times by cross-referencing ambulance locations with demographic data, to early warning systems that track approaching storms, detect seismic activity, and even seek to prevent genocide by monitoring ethnic conflict with orbiting satellites. However, algorithms are also increasingly being tasked with instructions to kill: executing coding sequences that quite literally execute. Guided by the Obama presidency’s conviction that the War on Terror can be won by ‘out-computing’ its enemies and pre-empting terrorists’ threats using predictive software, a new generation of deadly algorithms is being designed that will both control and manage the ‘kill-list,’ and along with it decisions to strike. Indeed, the recently terminated practice of ‘signature strikes’, in which data analytics was used to determine emblematic ‘terrorist’ behaviour and match these patterns to potential targets on the ground, already points to a future in which intelligence-gathering, assessment and military action, including the calculation of who can legally be killed, will largely be performed by machines based upon an ever-expanding database of aggregated information. As such, this transition to execution by algorithm is not simply a continuation of killing at ever greater distances inaugurated by the invention of the bow and arrow that separated warrior and foe, as many have suggested. It is also a consequence of the ongoing automation of warfare, which can be traced back to the cybernetic coupling of Claude Shannon’s mathematical theory of information with Norbert Wiener’s wartime research into feedback loops and communication control systems. As this new era of intelligent weapons systems progresses, operational control and decision-making are increasingly being outsourced to machines.

**Schuller, A. L. (2017). At the crossroads of control: The intersection of artificial intelligence in autonomous weapon systems with international humanitarian law. *Harvard National Security Journal* *8*, 379. SSRN:** [**https://ssrn.com/abstract=2978141**](https://ssrn.com/abstract=2978141)**.**

This Article explores the interaction of artificial intelligence (AI) and machine learning with international humanitarian law (IHL) in autonomous weapon systems (AWS). Lawyers and scientists repeatedly express a need for practical and objective substantive guidance on the lawful development of autonomy in weapon systems. This Article proposes five foundational principles to enable development of responsible AWS policy. The findings emerged from a research project conducted by a team of military and civilian professors at the Stockton Center for the Study of International Law at the U.S. Naval War College. The study is informed by experts in computer sciences, government and military, non-governmental organizations, think tanks, and academia.

Advances in AI will likely produce AWS that are different in kind from existing weapon systems and thus require a fresh approach to evaluating IHL compliance. First, this Article describes the technological details pertinent to understanding the distinction between current and future systems. It argues that the technological evaluation of the spectrum of autonomy should focus on the combination of authorities granted to the computer that controls an AWS, while also taking into account the physical capabilities of the system. Second, it argues that a key issue bearing on IHL compliance is whether an AWS has been granted some combination of authorities and capabilities that functionally delegate the decision to kill from human to machine. Third, it posits that predictability must be at the core of an evaluation into whether a particular AWS breaches this delegation threshold and examines how AI handles uncertainty, a critical component of the predictability analysis. Finally, the Article proposes five foundational principles to guide the development of AWS policy.

**Sharkey, N. E. (2012). The evitability of autonomous robot warfare. *International Review of the Red Cross*, *94*(886), 787-799.** [**https://international-review.icrc.org/sites/default/files/irrc-886-sharkey.pdf**](https://international-review.icrc.org/sites/default/files/irrc-886-sharkey.pdf)

This is a call for the prohibition of autonomous lethal targeting by free-ranging robots. This article will first point out the three main international humanitarian law (IHL)/ ethical issues with armed autonomous robots and then move on to discuss a major stumbling block to their evitability: misunderstandings about the limitations of robotic systems and artificial intelligence. This is partly due to a mythical narrative from science fiction and the media, but the real danger is in the language being used by military researchers and others to describe robots and what they can do. The article will look at some anthropomorphic ways that robots have been discussed by the military and then go on to provide a robotics case study in which the language used obfuscates the IHL issues. Finally, the article will look at problems with some of the current legal instruments and suggest a way forward to prohibition.

**Sharkey, N. (2011). Automating warfare: Lessons learned from the drones. *Journal of Law, Information & Science*, *21*, 140.** [**https://www5.austlii.edu.au/au/journals/JlLawInfoSci/2012/8.html**](https://www5.austlii.edu.au/au/journals/JlLawInfoSci/2012/8.html)

War fighting is currently undergoing a revolution. The use of robotics platforms for carrying weapons is coming on track at an increasing rate. US plans from all of the armed forces indicate a massive build up of military robots and at least 50 other countries have either bought them or have military robotics programmes. Currently all armed robots in the theatre of war are remotely controlled by humans; so called man-in-the-loop systems. Humans are responsible for both target selection and decisions about lethal force. But this is set to change. The role of the person in the loop will shrink and eventually vanish. But are we ready for this step? Do we understand the limits of the technology and how massive increases in the pace of battle will leave human responsibility in the dark? Before moving to autonomous operation we need to consider the lessons learned from the application of the current remotely piloted armed robots. Four areas are considered here: (i) moral disengagement; (ii) targeted killings in covert operations; (iii) expansion of the battle space; (iv) the illusion of accuracy.

**Shevchuk, V. (2023). The Role of Digital Technologies in the Investigation of War Crimes in Ukraine: Criminalistic Problems. *Grail of Science*, (25), 97-110. DOI:**[**10.36074/grail-of-science.17.03.2023.014**](http://dx.doi.org/10.36074/grail-of-science.17.03.2023.014)

The article is devoted to the study of current problems of the use of digital technologies in the investigation of war crimes in Ukraine. The emergence and formation of new areas of criminalistics due to scientific and technical progress, the introduction of the latest digital technologies and tools of digital criminalistics are substantiated. It is noted that it is important to take into account the modern trends in the development of forensics related to the formation of its separate branches (directions): medical, genotoscopic, aerospace, nuclear, digital and military criminalistics. In the conditions of the war in Ukraine, the problems of researching digital technologies in the investigation of war crimes and improving criminalistic methods and means of combating crime in Ukraine in modern conditions are gaining special relevance. It is substantiated that the process of digitization of criminalistics is a natural stage of development and formation of modern criminalistic knowledge, which involves the introduction of digital technologies in various fields of criminalistic science, forensic expertise and legal practice. At the same time, special attention should be paid to increasing the role of criminalistic didactics, in particular, forensic training of investigators, prosecutors, courts, detectives, criminalistic investigators, criminalistic experts in the field of digital technologies. The most urgent and promising problems of the study of the problems of digital technologies in the investigation of war crimes are highlighted.

**Sparrow, R. Killer Robots. *Journal of Applied Philosophy 24(1),* 62-77.** [**https://doi.org/10.1111/j.1468-5930.2007.00346.x**](https://doi.org/10.1111/j.1468-5930.2007.00346.x)

The United States Army's Future Combat Systems Project, which aims to manufacture a ‘robot army’ to be ready for deployment by 2012, is only the latest and most dramatic example of military interest in the use of artificially intelligent systems in modern warfare. This paper considers the ethics of the decision to send artificially intelligent robots into war, by asking who we should hold responsible when an autonomous weapon system is involved in an atrocity of the sort that would normally be described as a war crime. A number of possible loci of responsibility for robot war crimes are canvassed: the persons who designed or programmed the system, the commanding officer who ordered its use, the machine itself. I argue that in fact none of these are ultimately satisfactory. Yet it is a necessary condition for fighting a just war, under the principle of jus in bellum, that someone can be justly held responsible for deaths that occur in the course of the war. As this condition cannot be met in relation to deaths caused by an autonomous weapon system it would therefore be unethical to deploy such systems in warfare.

**Suchman, L. (2020). Algorithmic warfare and the reinvention of accuracy. *Critical Studies on Security, 8*(2), 175-187.** [**https://doi.org/10.1080/21624887.2020.1760587**](https://doi.org/10.1080/21624887.2020.1760587)**.**

This article aims to integrate two interrelated strands in critical security studies. The first is mounting evidence for the fallacy of claims for precision and accuracy in the United States ‘counterterrorism’ programme, particularly as it involves expanding aerial surveillance in support of operations of extrajudicial assasination. The second line of critical analysis concerns growing investment in the further automation of these operations, more specifically in the form of the US Department of Defense Algorithmic Warfare Cross-Functional Team, or Project Maven. Building upon generative intersections of critical security studies and science and technology studies (STS), I argue that the promotion of automated data analysis under the sign of artificial intelligence can only serve to exacerbate military operations that are at once discriminatory and indiscriminate in their targeting, while remaining politically and legally unaccountable.

**Troath, S. (2024). Trusting technology to wage war: the politics of trust and ethics in the development of robotics, autonomous systems, and artificial intelligence. *Critical Military Studies*, 1-19.** [**https://doi.org/10.1080/23337486.2024.2362074**](https://doi.org/10.1080/23337486.2024.2362074)**.**

Robotics, autonomous systems, and artificial intelligence (RAS-AI) are at the technological edge of militaries trying to achieve ‘ethical’ war. RAS-AI have been cast as essential technologies for defence forces to develop in order to sustain military advantage. Central to the success of this endeavour is trust: the technologies must be trusted for defence personnel to be willing to use them, for defence to be trusted by the public, and for allies and partners to have confidence in each other’s developments. Yet, trust is not merely a technocratic term. Its dominant role in the successful adoption of RAS-AI gives it power. This paper argues that the language of trust is being used to facilitate the development and adoption of military RAS-AI, often in concert with the language of ethics. Building on Maja Zehfuss’ concept of the politics of ethics, this paper contends that when it comes to RAS-AI there is also a politics of trust. Analysing British, American, and Australian military documents demonstrates that this politics manifests in side-lining political questions about how RAS-AI will be used – against whom and for what purposes – through focusing instead on the need to develop ethical and trustworthy RAS-AI to wage virtuous war.

**Ullah, M., Qureshi, G. M., & Shahid, A. (2023). Robotic warfare and international law: Implications for modern military. *International Journal of Contemporary Issues in Social Sciences,* 2:3, 337-347**

[**https://www.researchgate.net/publication/376521600\_ROBOTIC\_WARFARE\_AND\_INTERNATIONAL\_LAW\_IMPLICATIONS\_FOR\_MODERN\_MILITARY**](https://www.researchgate.net/publication/376521600_ROBOTIC_WARFARE_AND_INTERNATIONAL_LAW_IMPLICATIONS_FOR_MODERN_MILITARY)

This research examines the profound transformation occurring in modern warfare due to robotics and autonomous systems, with a focus on their impact on military organizations and international law. It explores the evolving nature of conflict, highlighting the rising role of robots and autonomous systems and their implications for military personnel, data management, and decision-making. The article investigates the potential applications of robotics within military structures, including enhancing performance, reducing risks to soldiers and marines, redefining force structures, improving institutional support, and fostering new operational concepts. Additionally, it addresses recruitment and retention challenges and the need for defense systems against robotic and AI adversaries. A core aspect of this research is the analysis of key military organizations leading the integration of robots and remote weapon systems. These organizations significantly influence the course of robotic warfare and its implications for international law.

**Ulrich, C. (2021). Forensic Architecture: Digital Citizen Intelligence in the Age of Urban Warfare. *PolisReflects*.** [**https://polisreflects.polis180.org/pub/ss9uabih/release/1**](https://polisreflects.polis180.org/pub/ss9uabih/release/1)

Forensic Architecture is an emergent investigatory practice and the name of a multidisciplinary research collective founded in 2010. They produce architectural evidence around human right violations and environmental destruction perpetrated by states and public actors. By integrating human rights research within architectural frameworks and technological innovation, they have developed a unique set of methods. Extending the discipline of open source intelligence, they empower victims to become agents of investigation. How effective is this approach for investigating contemporary political and environmental conflicts, and addressing state violence? Based on all investigations published by Forensic Architecture until March 2019, this article aims to evaluate the judicial and communicative impact of their work. It demonstrates that their effectiveness in courts is comparably low, and argues that their main quality is instead symbolic: with their investigations, they expose disinformation strategies, shape narratives around current conflicts and contribute significantly to a critical public discourse.

**Verdiesen, I., Santoni de Sio, F. & Dignum, V. (2021). Accountability and Control Over Autonomous Weapon Systems: A Framework for Comprehensive Human Oversight. *Minds & Machines* 31, 137–163.** [**https://doi.org/10.1007/s11023-020-09532-9**](https://doi.org/10.1007/s11023-020-09532-9)

Accountability and responsibility are key concepts in the academic and societal debate on Autonomous Weapon Systems, but these notions are often used as high-level overarching constructs and are not operationalised to be useful in practice. “Meaningful Human Control” is often mentioned as a requirement for the deployment of Autonomous Weapon Systems, but a common definition of what this notion means in practice, and a clear understanding of its relation with responsibility and accountability is also lacking. In this paper, we present a definition of these concepts and describe the relations between accountability, responsibility, control and oversight in order to show how these notions are distinct but also connected. We focus on accountability as a particular form of responsibility—the obligation to explain one’s action to a forum—and we present three ways in which the introduction of Autonomous Weapon Systems may create “accountability gaps”. We propose a Framework for Comprehensive Human Oversight based on an engineering, socio-technical and governance perspective on control. Our main claim is that combining the control mechanisms at technical, socio-technical and governance levels will lead to comprehensive human oversight over Autonomous Weapon Systems which may ensure solid controllability and accountability for the behaviour of Autonomous Weapon Systems. Finally, we give an overview of the military control instruments that are currently used in the Netherlands and show the applicability of the comprehensive human oversight Framework to Autonomous Weapon Systems. Our analysis reveals two main gaps in the current control mechanisms as applied to Autonomous Weapon Systems. We have identified three first options as future work for the design of a control mechanism, one in the technological layer, one in the socio-technical layer and one the governance layer, in order to achieve comprehensive human oversight and ensure accountability over Autonomous Weapon Systems.

**Warren, A., & Hillas, A. (2017). Lethal Autonomous Weapons Systems: Adapting to the Future Unmanned Warfare and Unaccountable Robots. *Yale Journal of International Affairs* *12*, 71.** [**https://researchrepository.rmit.edu.au/esploro/outputs/journalArticle/Lethal-autonomous-weapons-systems-Adapting-to/9921863142601341**](https://researchrepository.rmit.edu.au/esploro/outputs/journalArticle/Lethal-autonomous-weapons-systems-Adapting-to/9921863142601341)

In response to a push from civil society to confront the legal and ethical dimensions of lethal robotics, the UN Convention on Certain Conventional Weapons convened a four-day Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) in 2014, which was followed by five-day meetings in 2015 and 2016. This was the first occasion diplomats had openly discussed or even considered the prospect of lethal autonomy. Many issues remain unresolved. In response, this article seeks to address the question, how do lawmakers and policymakers in the United States envisage responding to the advent of LAWS? As a new addition to literature on lethal autonomy, the article considers whether a robot with either strong Artificial Intelligence (AI) or Artificial Consciousness (AC) could obtain moral agency and stand trial in the U.S. military justice system. The necessary reforms within the Uniform Code of Military Justice (UCMJ) are ultimately deemed too difficult to achieve, meaning that LAWS will not obtain personhood unless robots are conferred moral agency first under civilian criminal law. The status of Military Working Dogs (MWDs), which are alive and conscious, is then utilized as a case study to illustrate how unattainably high the bar for moral agency is for animals and robots alike, suggesting that the training and development of what we call the "Machine's Human Operator/Overseer (MHO)" -humans who will either share in the responsibility or be held solely accountable for the actions of LAWS during human-machine teaming missions-could utilize lessons gained from MWD handlers in previous conflicts.

**Wasilow, S., & Thorpe, J. B. (2019). Artificial intelligence, robotics, ethics, and the military: A Canadian perspective. *AI Magazine*, *40*(1), 37-48.** [**https://doi.org/10.1609/aimag.v40i1.2848**](https://doi.org/10.1609/aimag.v40i1.2848)

Defense and security organizations depend upon science and technology to meet operational needs, predict and counter threats, and meet increasingly complex demands of modern warfare. Artificial intelligence and robotics could provide solutions to a wide range of military gaps and deficiencies. At the same time, the unique and rapidly evolving nature of AI and robotics challenges existing policies, regulations, and values, and introduces complex ethical issues that might impede their development, evaluation, and use by the Canadian Armed Forces (CAF). Early consideration of potential ethical issues raised by military use of emerging AI and robotics technologies in development is critical to their effective implementation. This article presents an ethics assessment framework for emerging AI and robotics technologies. It is designed to help technology developers, policymakers, decision makers, and other stakeholders identify and broadly consider potential ethical issues that might arise with the military use and integration of emerging AI and robotics technologies of interest. We also provide a contextual environment for our framework, as well as an example of how our framework can be applied to a specific technology. Finally, we briefly identify and address several pervasive issues that arose during our research.

**Winter, E. (2022). The compatibility of autonomous weapons with the principles of international humanitarian law. *Journal of conflict and security law*, *27*(1), 1-20.** [**https://doi.org/10.1093/jcsl/krac001**](https://doi.org/10.1093/jcsl/krac001)**.**

The emergence of autonomous weapons remains a hot topic in international humanitarian law. Much has been said by States, international organisations, non-governmental organisations and academics on the matter in recent years. However, no agreement has been reached on how best to regulate this nascent technology. In the absence of any such agreement, the best approach is to analyse autonomous weapons through the lens of the principles of international humanitarian law. After humanity and military necessity are debunked as false principles, this article tests the compatibility of autonomous weapons with distinction, proportionality and precaution. It argues that autonomous weapons are not currently able to comply with these principles, owing primarily to a lack of sufficiently advanced artificial intelligence and contextual awareness. However, it also argues that the path to international humanitarian law-compliant autonomous weapons is shorter, and more clearly defined, than many realise. Indeed, the article highlights various ways in which future technological advances could enable autonomous weapons to comply with the principles of international humanitarian law more rigorously than human beings. The article does not seek to address the compatibility of autonomous weapons with other aspects of international law or the morality of such weapons.

**Worcester, M. (2015). Autonomous Warfare-A Revolution in Military Affairs’. *ISPSW Strategy Series: Focus on Defence and International Security*, *340*, 44-53.** [**https://www.files.ethz.ch/isn/190160/340\_Worcester.pdf**](https://www.files.ethz.ch/isn/190160/340_Worcester.pdf)

The world is on the cusp of an epochal shift from an industrial to an information based society and this is having a fundamental impact on the way war is conducted and what technologies are becoming available to the Military. The fundamental nature of war remains immutable. As Carl von Clausewitz characterized it “war is essentially an interactive clash or two-sided duel between independent, hostile, sentient wills dominated by friction, uncertainty, disorder and highly nonlinear interactions”. Nothing alters the fact that war is a human endeavor, with decidedly deadly consequences for all involved. New technology does not make war more clinical, it makes it more deadly. What technology does do is to make the battlefield more complex. Public debate is heating up over the future development of lethal autonomous weapon systems. Some advocate a complete ban on any further development, others a more gradual development and evolution of codes of conduct based on traditional legal and ethical principles governing weapons and warfare. On the other hand, there will always be those who will develop and deploy such future systems with scant regard of ethics and legality.

## 

## 

## **Non-peer Reviewed Articles and Journalism**

**Abdulrahim, R. (2021, February 13). AI emerges as crucial tool for groups seeking justice for Syria war crimes. *Wall Street Journal.***

[**https://www.wsj.com/articles/ai-emerges-as-crucial-tool-for-groups-seeking-justice-for-syria-war-crimes-11613228401**](https://www.wsj.com/articles/ai-emerges-as-crucial-tool-for-groups-seeking-justice-for-syria-war-crimes-11613228401)**.**

## Rights groups use machine learning to sort through a mammoth trove of video, photo and other evidence.

**Abraham, Y. (2023, November 30). ‘A mass assassination factory’: Inside Israel’s calculated bombing of Gaza*.* *+972 Magazine.*** [**https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/**](https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/)**.**

The Israeli army’s expanded authorization for bombing non-military targets, the loosening of constraints regarding expected civilian casualties, and the use of an artificial intelligence system to generate more potential targets than ever before, appear to have contributed to the destructive nature of the initial stages of Israel’s current war on the Gaza Strip, an investigation by +972 Magazine and Local Call reveals. These factors, as described by current and former Israeli intelligence members, have likely played a role in producing what has been one of the deadliest military campaigns against Palestinians since the Nakba of 1948.

The investigation by +972 and Local Call is based on conversations with seven current and former members of Israel’s intelligence community — including military intelligence and air force personnel who were involved in Israeli operations in the besieged Strip — in addition to Palestinian testimonies, data, and documentation from the Gaza Strip, and official statements by the IDF Spokesperson and other Israeli state institutions.

**Abu Sneineh, M. (2021, November 9). *Meet Blue Wolf, the app Israel uses to spy on Palestinians in the occupied West Bank.* Middle East Eye.** [**https://www.middleeasteye.net/news/israel-whats-blue-wolf-app-soldiers-use-photograph-palestinians**](https://www.middleeasteye.net/news/israel-whats-blue-wolf-app-soldiers-use-photograph-palestinians)

Cameras have long been an essential tool for Israel to monitor and watch Palestinians. But these days, technology has advanced, with sophisticated facial recognition technology now used to screen Palestinians at military checkpoints. But it's not just at checkpoints, according to [a report in the Washington Post](https://www.washingtonpost.com/world/middle_east/israel-palestinians-surveillance-facial-recognition/2021/11/05/3787bf42-26b2-11ec-8739-5cb6aba30a30_story.html) on Monday, which described a whole new level of surveillance in place across the West Bank called Blue Wolf. Middle East Eye takes a look at this technology and what makes it different - more dystopian and more terrifying - when compared to what the Israeli military was already known to have in place.

***AI helps scour video archives for evidence of human-rights abuses.* (2021, June 5). The Economist.** [**https://www.economist.com/international/2021/06/05/ai-helps-scour-video-archives-for-evidence-of-human-rights-abuses**](https://www.economist.com/international/2021/06/05/ai-helps-scour-video-archives-for-evidence-of-human-rights-abuses)**.**

Thanks especially to ubiquitous camera-phones, today’s wars have been filmed more than any in history. Consider the growing archives of Mnemonic, a Berlin charity that preserves video that purports to document war crimes and other violations of human rights. If played nonstop, Mnemonic’s collection of video from Syria’s decade-long war would run until 2061. Mnemonic also holds seemingly bottomless archives of video from conflicts in Sudan and Yemen. Even greater amounts of potentially relevant additional footage await review online.

**Amnesty International. (2023, May 2). *Israel and occupied Palestinian territories: Automated apartheid: How facial recognition fragments, segregates and controls Palestinians in the OPT.*** [**https://www.amnesty.org/en/documents/mde15/6701/2023/en/**](https://www.amnesty.org/en/documents/mde15/6701/2023/en/)**.**

In this report, Amnesty International explores how facial recognition technology is used extensively by the Israeli authorities to support their continued domination and oppression of Palestinians in the OPT. With a record of discriminatory and inhuman acts that maintain a system of apartheid, the Israeli authorities are able to use facial recognition software – in particular at checkpoints – to consolidate existing practices of discriminatory policing, segregation, and curbing freedom of movement, violating Palestinians’ basic rights.

**Biddle, S. (2022, July 24). *Documents reveal advanced AI tools Google is selling to Israel.* The Intercept.**

[**https://theintercept.com/2022/07/24/google-israel-artificial-intelligence-project-nimbus/**](https://theintercept.com/2022/07/24/google-israel-artificial-intelligence-project-nimbus/)

Google employees, who have been kept in the dark about the “Nimbus” AI project, have concerns about Israeli human rights abuses.

**Biddle, S. (2024, April 5). *Google won’t say anything about Israel using its photo software to create Gaza “hit list”.* The Intercept.** [**https://theintercept.com/2024/04/05/google-photos-israel-gaza-facial-recognition/**](https://theintercept.com/2024/04/05/google-photos-israel-gaza-facial-recognition/)**.**

Google prohibits using its tech for “immediate harm,” but Israel is harnessing its facial recognition to set up a dragnet of Palestinians.

**Biddle, S. (2024, May 1). *Israeli weapons firms required to buy cloud services from Google to Amazon.* The Intercept.** [**https://theintercept.com/2024/05/01/google-amazon-nimbus-israel-weapons-arms-gaza/**](https://theintercept.com/2024/05/01/google-amazon-nimbus-israel-weapons-arms-gaza/)**.**

Google downplays its military work with Israel, but “Project Nimbus” documents tie the American tech giants to Israel’s deadly military capabilities.

**Biesecker, M., Mednick, M. & G. Burke. (2025, February 18).** **Israel uses US-made AI models in war, concerns arise about tech’s role in who lives and who dies. *Associated Press.***

[**https://apnews.com/article/israel-palestinians-ai-technology-737bc17af7b03e98c29cec4e15d0f108**](https://apnews.com/article/israel-palestinians-ai-technology-737bc17af7b03e98c29cec4e15d0f108)

U.S. tech giants have quietly empowered Israel to track and kill many more alleged militants more quickly in Gaza and Lebanon through a sharp spike in artificial intelligence and computing services. But the number of civilians killed has also soared, fueling fears that these tools are contributing to the deaths of innocent people.

Militaries have for years hired private companies to build custom autonomous weapons. However, Israel’s recent wars mark a leading instance in which commercial AI models made in the United States have been used in active warfare, despite concerns that they were not originally developed to help decide who lives and who dies.

The Israeli military uses AI to sift through vast troves of intelligence, intercepted communications and surveillance to find suspicious speech or behavior and learn the movements of its enemies. After a deadly surprise attack by Hamas militants on Oct. 7, 2023, its use of Microsoft and OpenAI technology skyrocketed, an Associated Press investigation found. The investigation also revealed new details of how AI systems select targets and ways they can go wrong, including faulty data or flawed algorithms. It was based on internal documents, data and exclusive interviews with current and former Israeli officials and company employees.

“This is the first confirmation we have gotten that commercial AI models are directly being used in warfare,” said Heidy Khlaaf, chief AI scientist at the [AI Now Institute](https://ainowinstitute.org/) and former senior safety engineer at OpenAI. “The implications are enormous for the role of tech in enabling this type of unethical and unlawful warfare going forward.”

**Bo, Marta. (2020, December 18). *Meaningful Human Control over Autonomous Weapon Systems: An (International) Criminal Law Account.* Opinio Juris.** [**http://opiniojuris.org/2020/12/18/meaningful-human-control-over-autonomous-weapon-systems-an-international-criminal-law-account/**](http://opiniojuris.org/2020/12/18/meaningful-human-control-over-autonomous-weapon-systems-an-international-criminal-law-account/)**.**

Meaningful Human Control is at the core of [regulatory](https://documents.unoda.org/wp-content/uploads/2020/09/20200921-PoW-COR.pdf) and ethical debates on autonomous weapon systems. In international discussions and writings, the problem of meaningful human control has been addressed from different angles: from [philosophical](https://www.researchgate.net/publication/323459172_Meaningful_Human_Control_over_Autonomous_Systems_A_Philosophical_Account), ethical and legal ([here](https://link.springer.com/article/10.1007/s43154-020-00024-3#Sec3) and [here](https://blogs.icrc.org/law-and-policy/2018/04/03/autonomous-weapon-systems-ethical-basis-human-control/)), to [operational](https://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/), [cognitive](https://www.icrac.net/wp-content/uploads/2018/04/Sharkey_Guideline-for-the-human-control-of-weapons-systems_ICRAC-WP3_GGE-April-2018.pdf) and technical ([here](https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8924577) and [here](https://ieeexplore.ieee.org/abstract/document/7439432?casa_token=4q5umRGi6eMAAAAA:xpATcZz71mtSmmgFoyjij4BITmlInZ-P5715X2vwa20_UoXqeWsiWs_EraF_7uuY8RBas2FgK0s)), and recently from an [engineering, socio-technical and governance perspective](https://link.springer.com/article/10.1007/s11023-020-09532-9).

Despite the increasing use of parallel concepts such as ‘human element’ or ‘human oversight’, a recent, very comprehensive, [SIPRI-ICRC report](https://www.sipri.org/sites/default/files/2020-06/2006_limits_of_autonomy_0.pdf), which combines ethical, legal and operational considerations, adopts *human control* as a conceptual framework to shape limits to autonomy in weapons. In addition, a general obligation to maintain meaningful human control and positive obligations to ensure meaningful human control are the first and third pillars of the banning treaty advocated by [HRW and the Harvard Law School International Human Rights Clinic](https://www.hrw.org/report/2020/10/20/new-weapons-proven-precedent/elements-and-models-treaty-killer-robots).

This post provides an account of meaningful human control based on criminal law and on conditions to attribute principal criminal responsibility (rather than command responsibility). More specifically, it suggests that criminal law could provide elements not only to ground a theory of meaningful human control but also for the concrete operationalisation of this norm. A treaty obligation enshrining a duty to ensure meaningful human control should be adopted. This obligation should hinge on whether the individual had the capacity to fulfil the duty to take precautions in attacks and thus should provide for the attribution of criminal responsibility by omission (thus partly addressing responsibility gaps) limited to situations when it is fair to do so.

**Bo, Marta. (2024, June 6). Netanyahu and Gallant ICC Arrest Warrants: Tackling Modern Warfare and Criminal Responsibility for AI-enabled War Crimes. *Opinio Juris.*** [**https://opiniojuris.org/2024/12/06/netanyahu-and-gallant-icc-arrest-warrants-tackling-modern-warfare-and-criminal-responsibility-for-ai-enabled-war-crimes/**](https://opiniojuris.org/2024/12/06/netanyahu-and-gallant-icc-arrest-warrants-tackling-modern-warfare-and-criminal-responsibility-for-ai-enabled-war-crimes/)

In recent years, accountability for uses of artificial intelligence (AI) in warfare, especially under international criminal law, has progressively emerged as a critical issue in governance initiatives ([SIPRI](https://www.sipri.org/sites/default/files/2022-10/2210_aws_human_responsibility.pdf), [GGE’s Guiding Principles](https://ccdcoe.org/uploads/2020/02/UN-191213_CCW-MSP-Final-report-Annex-III_Guiding-Principles-affirmed-by-GGE.pdf), [REAIM Blue Print for Action](https://www.reaim2024.kr/home/reaimeng/board/bbsDetail.do?encMenuId=4e57325766362f626e5179454e6d6e4d4a4d33507a773d3d&encBbsMngNo=366e794c7a644d756342425668444f393053755142673d3d&encBbsNo=6f784e4542386f7735767465766a6531556f4b6149413d3d&ctlPageNow=1&schKind=bbsTtlCn&schWord=%23this#this)) and scholarly and civil society debates ([Matthias](https://www.academia.edu/243900/The_Responsibility_Gap_Ascribing_Responsibility_for_the_Actions_of_Learning_Automata), [Sparrow](https://onlinelibrary.wiley.com/doi/10.1111/j.1468-5930.2007.00346.x), [HRW](https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots), and more recently [Heller](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4342529), [Bo](https://academic.oup.com/jicj/article-abstract/19/2/275/6181757) [here](https://academic.oup.com/jicj/article-abstract/19/2/275/6181757) and [here](https://academic.oup.com/jicj/article-abstract/21/5/1057/7275803)). Initially the debate predominantly centered around [preventative measures](https://theglobal.blog/2019/05/28/lethal-autonomous-weapons-war-crimes-and-the-convention-on-conventional-weapons/), IHL compliance, and (more recently) on [responsible AI](https://hcss.nl/gcreaim/) development and use, framing the issue as a future concern. However, with AI technologies—particularly AI decision-support systems (AI-DSS)—reportedly being widely deployed in current conflicts (e.g., see Palantir’s [Artificial Intelligence Platform (AIP) Defense](https://www.palantir.com/platforms/aip/defense/), [Lavender](https://www.972mag.com/lavender-ai-israeli-army-gaza/), and here [Gospel](https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/)), it was only a matter of time before their use in targeting and the associated criminal responsibility came under the scrutiny of an international criminal court.

**Bridgeman, T. (2019, April 8). *The viability of data-reliant predictive systems in armed conflict detention.* ICRC Humanitarian Law and Policy Blog.** [**https://blogs.icrc.org/law-and-policy/2019/04/08/viability-data-reliant-predictive-systems-armed- conflict-detention/**](https://blogs.icrc.org/law-and-policy/2019/04/08/viability-data-reliant-predictive-systems-armed-)**.**

Are predictive algorithms helpful in determining who to detain and who continues to present a security threat in an armed conflict context? Or will use of these emerging technologies be irreconcilable with the requirements of international humanitarian law (IHL) in what are likely to be data poor, rapidly changing conflict settings? These are questions modern militaries will need to address as machine learning tools become more advanced and are used more widely by governments to make decisions about who to detain, and for how long, in modern conflicts.

**Bruun, L. (2023, January 5). *Investigating (mis)conduct in war is already difficult: Will the use of military AI make it harder?.* Just Security.**

[**https://www.justsecurity.org/84655/investigating-misconduct-in-war-is-already-difficult-will-the-use-of-military-ai-make-it-harder/**](https://www.justsecurity.org/84655/investigating-misconduct-in-war-is-already-difficult-will-the-use-of-military-ai-make-it-harder/)**.**

In war, civilians and civilian objects are rarely spared from harm. Sometimes the harm is a deliberate, malicious act, sometimes it is an accident, and most often, in the fog of war, it is hard to find out. Yet, understanding what (or who) caused the harm is critical to ensure compliance with the laws of war (international humanitarian law, or IHL) and to hold wrongdoers to account.

Now, as military technologies increase in complexity, the question is whether the ability to investigate (mis)conduct in war will be as well. For instance, if an algorithm has been used to inform militaries about who, when, and where to strike – who is to be held accountable if a civilian is wrongly attacked, and how would investigators even determine responsibility? While this question is yet to be addressed in public-facing material by modern militaries, a better understanding of the implications of AI on war crime investigations is critical to ensure that humans, not machines, are held to account for unlawful incidents in future wars.

To shed light on this question, this article zooms in on the implications of military use of artificial intelligence (AI) on the ability to *collect* and *assess* evidence in case of a harmful incident. It finds that while AI may have the potential to strengthen some aspects of investigations (for example in collecting and preserving evidence), the increased reliance on AI also presents significant challenges, especially around the ability to assess collected evidence.

**Buxbaum, J. (2022, October 5th). *Israel experiments on Palestinians with AI-powered gun at checkpoints.* Mint Press News.** [**https://www.mintpressnews.com/israel-smart-shooter-palestinians-lab-rats-ai-powered-gun/282129/**](https://www.mintpressnews.com/israel-smart-shooter-palestinians-lab-rats-ai-powered-gun/282129/)

The Israeli military installed an automatic weapon at a heavily-trafficked checkpoint in the occupied West Bank city of al-Khalil in September. While it was [initially reported](https://www.haaretz.com/israel-news/2022-09-24/ty-article/.premium/israeli-army-installs-remote-control-crowd-dispersal-system-at-hebron-flashpoint/00000183-70c4-d4b1-a197-ffcfb24f0000?utm_source=App_Share&utm_medium=iOS_Native) that the weapon will fire a wide range of projectiles, the army now states the device is only capable of firing sponge-tipped bullets. Reiterating the remote-controlled gun will not use live fire, the IDF hopes this system will be used to test approved crowd dispersal methods. But critics assert the device is yet another example of Israel using Palestinians as guinea pigs so they can market their military technology as field-tested to governments around the world.

**Davison, N. (2022, July 26). *What you need to know about autonomous weapons.* ICRC.** [**https://www.icrc.org/en/document/what-you-need-know-about-autonomous-weapons**](https://www.icrc.org/en/document/what-you-need-know-about-autonomous-weapons)

Autonomous weapons are not a work of science fiction from a distant dystopian future. They are an immediate cause of humanitarian concern and demand an urgent, international political response. Senior scientific and policy adviser at the ICRC, Neil Davison, explains.

**Deeks, A. (2019, March 25). *Detaining by algorithm*. ICRC Humanitarian Law and Policy Blog.** [**https://blogs.icrc.org/law-and-policy/2019/03/25/detaining-by-algorithm/**](https://blogs.icrc.org/law-and-policy/2019/03/25/detaining-by-algorithm/)**.**

Militaries may soon begin to develop and deploy predictive algorithms for use during armed conflicts to help them assess which actors are dangerous for purposes of detention and where future attacks are likely to occur for purposes of patrolling and targeting. The U.S. criminal justice system has already turned to predictive algorithms to help it make more objective judgments about who to keep in custody and more efficient decisions about where to deploy police resources. In a recent article called [*Predicting Enemies*](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3152385), I wrote about this possibility and discussed the parallels between goals such as these on the military side and those of the U.S. criminal justice system. Here, I build upon that article, highlighting important additional considerations that militaries should weigh as they evaluate how predictive algorithms can help them perform their missions.

**Dwoskin, E. (2021, November 8). Israel escalates surveillance of Palestinians with facial recognition program in West Bank. *Washington Post.*** [**https://www.washingtonpost.com/world/middle\_east/israel-palestinians-surveillance-facial-recognition/2021/11/05/3787bf42-26b2-11ec-8739-5cb6aba30a30\_story.html**](https://www.washingtonpost.com/world/middle_east/israel-palestinians-surveillance-facial-recognition/2021/11/05/3787bf42-26b2-11ec-8739-5cb6aba30a30_story.html)

The Israeli military has been conducting a broad surveillance effort in the occupied West Bank to monitor Palestinians by integrating facial recognition with a growing network of cameras and smartphones, according to descriptions of the program by recent Israeli soldiers.

The surveillance initiative, rolled out over the past two years, involves in part a smartphone technology called Blue Wolf that captures photos of Palestinians’ faces and matches them to a database of images so extensive that one former soldier described it as the army’s secret “Facebook for Palestinians.” The phone app flashes in different colors to alert soldiers if a person is to be detained, arrested or left alone.

**Gill, R. (2023, February 23). What is open-source intelligence? *SANS*.** [**https://www.sans.org/blog/what-is-open-source-intelligence/**](https://www.sans.org/blog/what-is-open-source-intelligence/)

Open-Source Intelligence (OSINT) is defined as intelligence produced by collecting, evaluating and analyzing publicly available information with the purpose of answering a specific intelligence question.

**Goodman, J., Korenyuk, M. (2023, May 31). AI: War crimes evidence erased by social media platforms. *BBC.*** [**https://www.bbc.com/news/technology-65755517**](https://www.bbc.com/news/technology-65755517)**.**

From May 2023, this article written by members of BBC’s Global Disinformation Team covers the ways AI is used to monitor and remove images of war from social media platforms, often to the detriment of archival evidence. The platforms are not responsible for storing these materials. It is imperative that the materials are saved using best practices, and this article also cites Mnemonic and their work in preserving evidence of human rights violations. The piece ends quoting an Ambassador for Global Criminal Justice arguing that social media platforms ought to work with accountability mechanisms to ensure that this evidence is preserved.

**Hao, K. (2020, June 25). *Human rights activists want to use AI to prove war crimes in court.* Technology Review.** [**https://www.technologyreview.com/2020/06/25/1004466/ai-could-help-human-rights-activists-prove-war-crimes/**](https://www.technologyreview.com/2020/06/25/1004466/ai-could-help-human-rights-activists-prove-war-crimes/)**.**

It would take years for humans to scour the tens of thousands of hours of footage that document violations in Yemen. With machine learning, it takes just days.

**Higgins, E. (2023, March 28). How open source evidence was upheld in a human rights court. *Bellingcat.*** [**https://www.bellingcat.com/resources/2023/03/28/how-open-source-evidence-was-upheld-in-a-human-rights-court/**](https://www.bellingcat.com/resources/2023/03/28/how-open-source-evidence-was-upheld-in-a-human-rights-court/)

Many questions remain as to how open-source data can be used or presented in court as evidence.

The field of open source investigations is still relatively new in the grand scheme of things, after all. But with the ongoing war in Ukraine presenting a mountain of open source data for prosecutors to work with, understanding those questions and how they can be answered is of vital importance.

***Israel uses data, algorithms to thwart terror attacks, but Palestinians say privacy, civil liberties breached.*  (2018, June 12).CBS News.** [**https://www.cbsnews.com/news/israel-data-algorithms-predict-terrorism-palestinians-privacy-civil-liberties/**](https://www.cbsnews.com/news/israel-data-algorithms-predict-terrorism-palestinians-privacy-civil-liberties/)**.**

Israeli authorities have foiled over 200 Palestinian attacks by monitoring social media and sifting through vast amounts of data to identify prospective assailants ahead of time, according to Israel's public security minister. These pre-emptive actions put Israel at the forefront of an increasingly popular -- and controversial -- trend used by intelligence and law enforcement agencies around the world that use big data technology to track would-be criminals.

While the technology appears to be effective, its tactics drew angry Palestinian condemnation and have raised questions about civil liberties.

**Kania, E.B. (2017, June 8). *Quest for an AI revolution in warfare.* The Strategy Bridge.**

[**https://thestrategybridge.org/the-bridge/2017/6/8/-chinas-quest-for-an-ai-revolution-in-warfare**](https://thestrategybridge.org/the-bridge/2017/6/8/-chinas-quest-for-an-ai-revolution-in-warfare)**.**

China aspires to surpass the U.S. in artificial intelligence, seeking to take advantage of the unique opportunities that this critical emerging technology could confer to its economic competitiveness and military capabilities. To date, the scale of Chinese research in artificial intelligence, as reflected by the number of papers published and cited, has already exceeded that of the U.S., and China also ranks second in AI patent applications. From speech recognition to computer vision, Chinese efforts in artificial intelligence are cutting-edge and evidently constitute a priority for China’s leadership at the highest levels. Within the past year alone, China has released a series of national science and technology plans that seek to advance artificial intelligence and established a national deep learning lab. In particular, China’s new national mega-project for artificial intelligence, known as “Artificial Intelligence 2.0,” will advance and direct an ambitious agenda for research and development, including economic and national security applications.

**Lewis, D.A. (2020, September 28). *An Enduring Impasse on Autonomous Weapons*. Just Security.** [**https://www.justsecurity.org/72610/an-enduring-impasse-on-autonomous-weapons/**](https://www.justsecurity.org/72610/an-enduring-impasse-on-autonomous-weapons/)

Regular readers of *Just Security* will know that the United States and Russia do [not](https://www.justsecurity.org/68892/norm-skepticism-in-cyberspace-counter-factual-and-counterproductive/) see eye to eye on many matters touching on war and peace, not least around cyber, information [security](https://doi.org/10.1007/978-3-030-29053-5_18), and the conflict in [Syria](https://www.justsecurity.org/33456/5-counting-russia-vetoes-no-fly-zone-security-council/). But you do not have to squint to glimpse how the two are, in several important respects, similarly positioned on one side of an enduring impasse on autonomous weapons. While there is no definition in international law of autonomous weapons, one shorthand is weapons that, once initiated, can nominate, select, and apply force to targets without further human intervention. The debate is not purely academic: a handful of systems falling into this relatively narrow definition are already in use, such as so-called [loitering munitions](https://dronecenter.bard.edu/loitering-munitions-in-focus/); once launched, those systems can linger in the air over several hours while scanning for targets and then strike without in-the-moment clearance by a human operator.

**Lewis, D.A. (2019, March 21). *Legal reviews of weapons, means and methods of warfare involving artificial intelligence: 16 elements to consider*, Humanitarian Law and Policy Blog.** [**https://blogs.icrc.org/law-and-policy/2019/03/21/legal-reviews-weapons-means-methods-warfare-artificial-intelligence-16-elements-consider/**](https://blogs.icrc.org/law-and-policy/2019/03/21/legal-reviews-weapons-means-methods-warfare-artificial-intelligence-16-elements-consider/)

What are some of the chief concerns in contemporary debates around legal reviews of weapons, means or methods of warfare involving techniques or tools related to artificial intelligence (AI)? One session of the December 2018 [*workshop*](https://pilac.law.harvard.edu/events//artificial-intelligence-at-the-frontiers-of-international-law-concerning-armed-conflict) on AI at the frontiers of international law concerning armed conflict focused on this topic. In this post, I outline a few key threshold considerations and briefly enumerate 16 elements that States might consider as part of their legal reviews involving AI-related techniques or tools.

**Lewis, D.A. (2020). *Three Pathways to Secure Greater Respect for International Law concerning War Algorithms*. Legal Commentary, HLS PILAC.** [**https://pilac.law.harvard.edu/three-pathways-to-secure-greater-respect-for-international-law-concerning-war-algorithms**](https://pilac.law.harvard.edu/three-pathways-to-secure-greater-respect-for-international-law-concerning-war-algorithms)**.**

Some themes and commitments cut across these three pathways. Arguably, respect for the law turns in no small part on whether natural persons can and will foresee, understand, administer, and trace the components, behaviors, and effects of relevant systems. It may be advisable, moreover, to institute ongoing cross-disciplinary education and training as well as the provision of sufficient technical facilities for all relevant actors, from commanders to legal advisers to prosecutors to judges. Further, it may be prudent to establish ongoing monitoring of others’ technical capabilities. Finally, it may be warranted for relevant international actors to pledge to engage, and to call upon others to engage, only in armed-conflict-related conduct that is sufficiently attributable, discernable, and scrutable.

**Lewis, D.A., Modirzadeh, N.K., and Blum, G. (2017, June 26). *The Pentagon’s New Algorithmic-Warfare Team*. Lawfare.** [**https://www.lawfareblog.com/pentagons-new-algorithmic-warfare-team**](https://www.lawfareblog.com/pentagons-new-algorithmic-warfare-team)

In April 2017, the Pentagon [created](https://www.govexec.com/media/gbc/docs/pdfs_edit/establishment_of_the_awcft_project_maven.pdf) an “Algorithmic Warfare Cross-Functional Team,” pending a transfer of $70 million from Congress. The premise of this initiative is that maintaining a qualitative edge in war will increasingly require harnessing algorithmic systems that underpin artificial intelligence (AI) and machine learning (ML).

**McKernan, B. & Sabbagh, D. (2023, December 1). *‘The Gospel’: how Israel uses AI to select bombing targets in Gaza.* The Guardian.** [**https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets**](https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets)**.**

Continuing upon the research done for the +972 and Local Call article, the Guardian provides a more detailed picture of the Habsora AI target-production tool. A brief article, it mostly summarizes the information shared in the +972 and Local Call investigation, being padded by reference to experts, mostly Richard Moyers of Article 36, a group that aims to minimize harm from weapons.

**Murgia, M. (2021, December 9). *Researchers train AI on ‘synthetic data’ to uncover Syrian war crimes.* Financial Times.** [**https://www.ft.com/content/8399873e-0dda-4c87-ba59-0e2678166fba**](https://www.ft.com/content/8399873e-0dda-4c87-ba59-0e2678166fba)**.**

Technique used to teach Tesla cars and Amazon Alexa has gathered evidence of weapons used against civilians.

**Lewis, D.A. (2020, April 28). *AI and Machine Learning Symposium: Why Detention, Humanitarian Services, Maritime Systems, and Legal Advice Merit Greater Attention*. Opinio Juris.** [**http://opiniojuris.org/2020/04/28/ai-and-machine-learning-symposium-ai-in-armed-conflict-why-detention-humanitarian-services-maritime-systems-and-legal-advice-merit-greater-attention/**](http://opiniojuris.org/2020/04/28/ai-and-machine-learning-symposium-ai-in-armed-conflict-why-detention-humanitarian-services-maritime-systems-and-legal-advice-merit-greater-attention/)**.**

The preservation of international legal responsibility and agency concerning the employment of artificial-intelligence techniques and methods in relation to situations of armed conflict presents an [array](https://blogs.icrc.org/law-and-policy/2019/03/19/expert-views-frontiers-artificial-intelligence-conflict/) of pressing challenges and opportunities. In this post, I will seek to use one of the many useful framings in the ICRC’s 2019 “Challenges” [report](https://www.icrc.org/en/publication/4427-international-humanitarian-law-and-challenges-contemporary-armed-conflicts)’s section on AI to widen the aperture further in order to identify or amplify four areas of concern: detention, humanitarian services, uninhabited military maritime systems, and legal advice.

**Pratt, S.F. (2024, May 2). *When AI decides who lives and dies.* Foreign Policy.**

[**https://foreignpolicy.com/2024/05/02/israel-military-artificial-intelligence-targeting-hamas-gaza-deaths-lavender/**](https://foreignpolicy.com/2024/05/02/israel-military-artificial-intelligence-targeting-hamas-gaza-deaths-lavender/)**.**

The Israeli military’s algorithmic targeting has created dangerous new precedents.

**Vilisov, V. (2022, May 27). *Logging a war. How digitalization has changed the perception of modern warfare and the documentation of war crimes.* The Insider.**

[**https://theins.ru/en/society/251631**](https://theins.ru/en/society/251631)**.**

From satellite imagery, geolocation services and mobile coverage to facial recognition, artificial intelligence and drones: over the past 10 years digital technology has dramatically changed the concept of modern warfare and has become an essential part of the process of identifying and documenting war crimes. Techniques honed during the Syrian conflict are already being actively used today to investigate the Russian attack on Ukraine aided by satellite imagery, drone video, big data analysis, AI-based facial recognition, and more. However, due to the inertia of the legal system, digital evidence is still difficult to use in an international court.

## **Reports and Protocols**

**Afina, Y. (2024, October 24). *Draft Guidelines for the Development of a National Strategy on AI in Security and Defence.* UNIDIR.** [**https://unidir.org/publication/draft-guidelines-for-the-development-of-a-national-strategy-on-ai-in-security-and-defence/**](https://unidir.org/publication/draft-guidelines-for-the-development-of-a-national-strategy-on-ai-in-security-and-defence/)

As innovation in artificial intelligence (AI) proceeds at breakneck speed, states’ appetite for devising frameworks for the governance of the research, development and deployment of these technologies is at its greatest. With calls for governance solutions increasing at both the national and international levels, the number of national strategy documents that frame the development, deployment and use of these technologies has started to grow across regions.

Yet, most of these policies exclude or barely touch upon security and defence applications. Only a handful of national strategy documents have a section dedicated to this realm; and even fewer are specifically dedicated to it. This scarcity is at odds with the United Nations Secretary-General’s recommendation for Member States to “urgently develop national strategies on responsible design, development and use of artificial intelligence”, as outlined in his [New Agenda for Peace](https://dppa.un.org/en/a-new-agenda-for-peace).

Against this backdrop, UNIDIR has launched a programme of work to establish guidelines for the development, adoption, implementation and review of national strategies on AI in security and defence. The purpose of the guidelines, both of procedural and substantive nature, is to capture, anticipate and dissect the key issues, considerations and needs that each state must address as it develops or seeks to develop, adopt, implement and review its national strategy on AI in security and defence. In recognition of the host of incentives stemming from the establishment of such strategies, it is hoped that these guidelines will serve as a useful tool for states and non-state stakeholders alike as they seek to address issues related to the responsible development, deployment and use of AI in security and defence.  
  
The present draft guidelines have been released to provide states and all relevant stakeholders involved in the development, adoption, implementation and review of national strategies on AI in security and defence with an opportunity to review and provide feedback to UNIDIR. The Institute aims to adopt a holistic and inclusive method to the establishment of the guidelines; it thus seeks to capture all the varying perspectives, viewpoints and approaches to this issue. We welcome feedback from stakeholders across sectors and domains.

**Afina, Y. (2024, September 5). *The Global Kaleidoscope of Military AI Governance.* UNIDIR.** [**https://unidir.org/publication/the-global-kaleidoscope-of-military-ai-governance/**](https://unidir.org/publication/the-global-kaleidoscope-of-military-ai-governance/)

In the run-up to the second iteration of the Responsible AI in the Military Domain (REAIM) Summit, to be held in Seoul, Republic of Korea, on 9-10 September 2024, the Governments of the Republic of Korea and the Netherlands organized, in partnership with Chile, Costa Rica, Kenya, Singapore and Türkiye, a series of five regional consultations on responsible artificial intelligence in the military domain.

This report captures UNIDIR’s main reflections on the key takeaways stemming from the five regional consultations. These consultations did indeed enable the dissection of local contexts, realities and approaches with regard to the responsible development, deployment and use of AI in the military and wider security domains – including the identification of areas of nuanced convergence and divergence at the regional level.

Specifically, this report first discusses the reflections shared by States on the unique characteristics of AI technologies and the opportunities that they provide in the military domain. In addition, States also discussed and exchanged views on the risks, challenges and implications stemming from the development, deployment and use of AI in the military and wider security domains. The report then covers six points of convergence that have emerged from the consultations, along with five main points of divergence observed across and within regions.

**Afina, Y. & Paoli, G.P. (2024, September 5). *Governance of Artificial Intelligence in the Military Domain: A Multi-Stakeholder Perspective on Priority Areas.* UNIDIR.** [**https://unidir.org/publication/governance-of-artificial-intelligence-in-the-military-domain-a-multi-stakeholder-perspective-on-priority-areas/**](https://unidir.org/publication/governance-of-artificial-intelligence-in-the-military-domain-a-multi-stakeholder-perspective-on-priority-areas/)

In March 2024, [the Roundtable for AI, Security and Ethics (RAISE)](https://unidir.org/raise/) was launched in Bellagio, Italy. A multi-year collaborative initiative led by UNIDIR and in partnership with [Microsoft](https://www.microsoft.com/en-us/), RAISE is intended to establish itself as the neutral, trusted and independent platform for inclusive, cross-regional and multisectoral engagement on artificial intelligence (AI) in security and defence.

The inaugural edition of RAISE convened participants primarily hailing from the industry and the research community, with select government representatives in light of the upcoming second edition of the Responsible AI in the Military Domain (REAIM) Summit (Seoul, 9-10 September 2024). Its objectives were two-fold:

1. Review the current state of applications of AI in security and defence contexts, across sectors and geographies but with a particular focus on the military domain; and
2. Identify key priority areas in which to develop specific guidance and policy recommendations on identified issues.

This inaugural edition of RAISE focused specifically on the military domain due to the momentum in this particular area. As a result of the convening and UNIDIR’s facilitation of the discussions, participants had identified, by its end, six key priority areas for RAISE to advance in the governance of AI in the military domain: building a knowledge base; trust building; the human element in AI uses; data practices; life cycle management; and destabilization.

This Policy Brief aims to lay the foundation for future work that will bring to life recommendations revolving around the six priority themes that the meeting identified, and which it agreed would serve as a basis for cooperation and collective action that transcends geopolitical rivalry, cross-sectoral divides and competition.

**Anand, A. & Deng, H. (2023, February 13).** ***Towards Responsible AI in Defence: A Mapping and Comparative Analysis of AI Principles Adopted by States.* UNIDIR.** [**https://unidir.org/publication/towards-responsible-ai-in-defence-a-mapping-and-comparative-analysis-of-ai-principles-adopted-by-states/**](https://unidir.org/publication/towards-responsible-ai-in-defence-a-mapping-and-comparative-analysis-of-ai-principles-adopted-by-states/)

Continuous advances in the field of artificial intelligence (AI) and efforts to integrate AI systems in critical sectors are gradually transforming all aspects of society, including in the defence sector. Although advancements in AI present unprecedented opportunities to augment human capabilities and improve decision-making in various ways, they also present significant legal, safety, security and ethical concerns. Thus, to ensure that AI systems are developed and used lawfully, ethically, safely, securely and responsibly, governments and intergovernmental organisations are developing a range of normative instruments. This approach is broadly known as "Responsible AI", or ethical or trustworthy AI. Presently, the most notable approach to Responsible AI is the development and operationalisation of responsible or ethical AI principles.

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. The project has three main aims. First, it aims to encourage states to adopt and operationalise tools that can enable responsible behaviour in the development and use of AI systems. It also seeks to help increase transparency and foster trust among states and other key AI actors. Finally, the project aims to build a shared understanding of the key elements of Responsible AI and how they may be operationalised, which may inform the development of internationally accepted governance frameworks.

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.

**Anand, A. & Puscas, I. (2023). *Proposals Related to Emerging Technologies in the Area of Lethal Autonomous Weapon Systems: A Resource Paper.* UNIDIR.** [**https://www.unidir.org/sites/default/files/2023-05/UNIDIR-Proposals\_Emerging\_Technologies\_Lethal\_Autonomous\_Weapons\_Systems\_2023.pdf**](https://www.unidir.org/sites/default/files/2023-05/UNIDIR-Proposals_Emerging_Technologies_Lethal_Autonomous_Weapons_Systems_2023.pdf)

The Elements paper provides an overview of the debate on LAWS and argues for the need for a holistic and multidimensional understanding of the effects of incorporating autonomy in weapons systems. Drawing on these considerations, the paper proposes a draft text containing prohibitions and regulations that may be part of a legally binding instrument, and which derive, according to The Elements, from the substantive discussions within the GGE on LAWS.

**Article 36. (2023, March). *Completely outside human control?*** [***https://article36.org/wp-content/uploads/2023/03/Completely-outside-human-control.pdf***](https://article36.org/wp-content/uploads/2023/03/Completely-outside-human-control.pdf)

The working paper contains valuable content relating to the control of systems in practice – notably around the need for “sufficient understanding of the weapons systems’ way of operating” and the need for “spatial and temporal limits” in the

deployment of a system.

**Article 36. (2013). *Killer robots: UK government policy on fully autonomous weapons.*** [**https://article36.org/wp-content/uploads/2013/04/Policy\_Paper1.pdf**](https://article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf)

Article 36 is a UK NGO concerned with policy and legal controls to prevent harm from weapons. In March 2012 the organization called for a ban on fully autonomous armed robots. This paper situates such a ban within a broader argument, for a positive obligation in international law for individual attacks to be under meaningful human control. Article 36 staff are part of the International Committee for Robot Arms Control (ICRAC) and the organization is a founder and Steering Committee member of the Campaign to Stop Killer Robots.

**Austero, M., et al. (2020, October). *AI, emerging technology and LAWS: Security, moral, and ethical perspectives in Asia*.** [**https://www.stopkillerrobots.org/wp-content/uploads/2021/11/NISEA-AI-Emerging-Tech-and-LAWS-Perspectives-in-Asia.pdf**](https://www.stopkillerrobots.org/wp-content/uploads/2021/11/NISEA-AI-Emerging-Tech-and-LAWS-Perspectives-in-Asia.pdf)

Artificial Intelligence (AI) development has been steadily expanding in the last decade, especially in the areas of economic development, rapid industrialization, increased productivity, and now, weaponry. Lethal Autonomous Weapons Systems (LAWS)

are gaining attention due to prominent advances in weapons development. LAWS are loosely defined as “weapons that can select and engage targets without human intervention.”1 Global concern over the use of LAWS on human beings is growing, especially in countries that suffer from various security issues. Internal insecurity and armed confrontations over territorial disputes have all increased circumspection about the weaponization of AI and its integration to LAWS, contributing to its moniker, “killer robots”. However, the threat does not merely lie in lethality, as an autonomous weapon system does not need to be “lethal” to inflict damage, and the weaponization of AI and the range of autonomous weapon systems that can inflict harm still pose a significant threat to human security. Beyond physical harm, the threat of force is enough to control the populace by discouraging certain actions. For now, the element of lethality remains unclear as most of these weapons systems are used to intercept and eliminate incoming projectiles.

**Bo, M., Boulanin, V. & Verbruggen, M. (2022). *Retaining Human Responsibility in the Development and Use of Autonomous Weapons Systems.* Stockholm International Peace Research Institute.** [**https://www.sipri.org/sites/default/files/2022-10/2210\_aws\_human\_responsibility.pdf**](https://www.sipri.org/sites/default/files/2022-10/2210_aws_human_responsibility.pdf)

It is undisputed that humans must retain responsibility for the development and use of autonomous weapon systems (AWS) because machines cannot be held accountable for violations of international humanitarian law (IHL). However, the critical question of how, in practice, humans would be held responsible for IHL violations involving AWS has not featured strongly in the policy debate on AWS. This report aims to support a deeper and more focused expert discussion on that very question. There are multiple legal frameworks through which human responsibility for IHL violations may be ensured. This report focuses on two central frameworks: the rules governing state responsibility for internationally wrongful acts and individual criminal responsibility for war crimes.

**Boulanin, V. & Verbruggen, M. (2017). *Mapping the Development of Autonomy in Weapons Systems*. Stockholm International Peace Research Institute.** [**https://www.sipri.org/sites/default/files/2017-11/siprireport\_mapping\_the\_development\_of\_autonomy\_in\_weapon\_systems\_1117\_1.pdf**](https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf) **[**[**https://perma.cc/XUT7-27AT**](https://perma.cc/XUT7-27AT)**].**

This report presents the conclusions of a one-year mapping study on the development of autonomy in weapon systems. It is intended to provide diplomats and members of civil society interested in the issue of lethal autonomous weapon systems (LAWS) with a better understanding of (a) the technological foundations of autonomy; (b) the state of autonomy in existing weapon systems; (c) the drivers of, and obstacles to, further increasing autonomy in weapon systems; and (d) the innovation ecosystems behind the advance of autonomy in weapon systems.

**Center for Humanitarian Dialogue. (2021). *Code of conduct on artificial intelligence in military systems.*** [***https://www.hdcentre.org/wp-content/uploads/2021/08/AI-Code-of-Conduct.pdf***](https://www.hdcentre.org/wp-content/uploads/2021/08/AI-Code-of-Conduct.pdf)

This draft Code of Conduct for AI-enabled military systems is the product of a two-year consultation process among Chinese, American, and international experts convened in person and online by the Centre for Humanitarian Dialogue (HD). The goal of the consultation process was to determine whether certain principles and limitations might be agreed regarding weapons and related military systems with significant AI components, especially among those international actors whose technology and deployment in this area is most advanced. Participants in the dialogue included current academics and former officials with military, diplomatic, intelligence, weapons design and legal backgrounds from the United States, China and an international delegation from Europe and Latin America. While some experts participated in previous UN conferences related to limitations on advanced weaponry – such as the UN CCW’s Lethal Autonomous Weapons Systems (LAWS) Group of Governmental Experts – the purpose of the consultation was to break away from public positions and to see if a discreet process could find common ground on limitations before AI-enabled military systems become so commonly used as to make future limitation impractical.

**Chandler, K. (2021, December 7). *Does military AI have gender? Understanding bias and promoting ethical approaches in military applications of AI.* UNIDIR.** [**https://doi.org/10.37559/GEN/2021/04**](https://doi.org/10.37559/GEN/2021/04)**.**

"Does Military AI Have Gender?" uncovers the significance of gender norms in the development and deployment of artificial intelligence (AI) for military purposes. The report addresses gender bias in data collection, algorithms and computer processing.

Drawing on research in ethical AI, the report outlines avenues for countering bias and mitigating harm, including a gender-based review of military applications of AI. In doing so, it seeks to chart a path for technology development that promotes – rather than hinders – gender equity and contributes to gender mainstreaming in the military.

**Cummings, M.L. (2017). *Artificial intelligence and the future of warfare.* Chatham House for the Royal Institute of International Affairs.** [**https://www.chathamhouse.org/sites/default/files/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings-final.pdf**](https://www.chathamhouse.org/sites/default/files/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings-final.pdf)**.**

Both military and commercial robots will in the future incorporate ‘artificial intelligence’ (AI) that could make them capable of undertaking tasks and missions on their own. In the military context, this gives rise to a debate as to whether such robots should be allowed to execute such missions, especially if there is a possibility that any human life could be at stake.

To better understand the issues at stake, this paper presents a framework explaining the current state of the art for AI, the strengths and weaknesses of the technology, and what the future likely holds. The framework demonstrates that while computers and AI can be superior to humans in some skill- and rule-based tasks, under situations that require judgment and knowledge, in the presence of significant uncertainty, humans are superior to computers.

**Dafoe, A. (2018). AI governance: a research agenda. *Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK*, *1442*, 1443.** [**https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf**](https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf)

Artificial intelligence (AI) is a potent general purpose technology. Future progress could be rapid, and experts expect that superhuman capabilities in strategic domains will be achieved in the coming decades. The opportunities are tremendous, including advances in medicine and health, transportation, energy, education, science, economic growth, and environmental

sustainability. The risks, however, are also substantial and plausibly pose extreme governance challenges. These include labor displacement, inequality, an oligopolistic global market structure, reinforced totalitarianism, shifts and volatility in national power, strategic instability, and an AI race that sacrifices safety and other values. The consequences are plausibly of a magnitude and on a timescale to dwarf other global concerns. Leaders of governments and firms are asking for policy guidance, and yet scholarly attention to the AI revolution remains negligible. Research is thus urgently needed on the AI governance problem: the problem of devising global norms, policies, and institutions to best ensure the beneficial development and use of advanced AI.

This report outlines an agenda for this research, dividing the field into three research clusters.

The first cluster, the technical landscape, seeks to understand the technical inputs, possibilities, and constraints for AI. The second cluster, AI politics, focuses on the political dynamics between firms, governments, publics, researchers, and other actors. The final research cluster of AI ideal governance envisions what structures and dynamics we would ideally create to govern the transition to advanced artificial intelligence.

**Davison, Neil. (2016, April 11). *A legal perspective: Autonomous weapons systems under international humanitarian law.* International Committee of the Red Cross.** [**https://www.icrc.org/en/download/file/65762/autonomous\_weapon\_systems\_under\_international\_humanitarian\_law.pdf**](https://www.icrc.org/en/download/file/65762/autonomous_weapon_systems_under_international_humanitarian_law.pdf)**.**

Written by the Scientific and Policy Adviser to the Legal Division of the Arms Unit of the International Red Cross (ICRC). Based on prior ICRC reports on autonomous weapons, Davison establishes the parameters by which the use of autonomous weapons must be judged under international law. Covering the rules of distinction, proportionality, and precaution, he admits that these rules cannot apply to machines and systems. He then covers the “principles of humanity” and the “dictates of the public conscience,” defined in the Martens Clause of the Geneva Convention as “customary international law,” or the ethical considerations that are posed against international humanitarian law. Davison tackles the issues of human control at the various stages of development, use, and accountability, outlining the importance of predictability as a major factor of judgment, recognizing the ambiguity of autonomous weapons in the determination of accountability.

**Defense Innovation Board. (2019). *AI principles: recommendations on the ethical use of artificial intelligence by the department of defense: Supporting document.* United States Department of Defense.** [**https://media.defense.gov/2019/oct/31/2002204458/-1/-1/0/dib\_ai\_principles\_primary\_document.pdf**](https://media.defense.gov/2019/oct/31/2002204458/-1/-1/0/dib_ai_principles_primary_document.pdf)**.**

The leadership of the Department of Defense (DoD) tasked the Defense Innovation Board

(DIB) with proposing Artificial Intelligence (AI) Ethics Principles for DoD for the design, development, and deployment of AI for both combat and non-combat purposes. Building upon the foundation of DoD’s existing ethical, legal, and policy frameworks and responsive to the complexities of the rapidly evolving field of AI, the Board sought to develop principles consistent with the Department’s mission to deter war and ensure the country’s security. This document summarizes the DIB’s project and includes a brief background; an outline of enduring DoD ethics principles that transcend AI; a set of proposed AI Ethics Principles; and a set of recommendations to facilitate the Department’s adoption of these principles and advance the wider aim of promoting AI safety, security, and robustness. The DIB’s complete report includes detailed explanations and addresses the wider historical, policy, and theoretical context for these recommendations. It is available at [innovation.defense.gov/ai](http://innovation.defense.gov/ai).

The DIB is an independent federal advisory committee that provides advice and

recommendations to DoD senior leaders; it does not speak for DoD. The purpose of this

report is an earnest attempt to provide an opening for a thought-provoking dialogue

internally to Department and externally in our wider society. The Department has the sole

responsibility to determine how best to proceed with the recommendations made in this report.

**Department of Defense. (2023, June 27). *Data, analytics and artificial intelligence adoption strategy: Accelerating decision advantage.*** [**https://media.defense.gov/2023/Nov/02/2003333300/-1/-1/1/DOD\_DATA\_ANALYTICS\_AI\_ADOPTION\_STRATEGY.PDF**](https://media.defense.gov/2023/Nov/02/2003333300/-1/-1/1/DOD_DATA_ANALYTICS_AI_ADOPTION_STRATEGY.PDF)**.**

As the 2022 National Defense Strategy (NDS) makes clear, the United States possesses strengths that our competitors cannot match, among them our diverse and open society, our culture of ingenuity, our innovation base, and our globe-spanning network of Allies and partners. The Department leverages these strengths by distributing authority, empowering leaders in our All-Volunteer Force to innovate at the edge and apply their own judgment to combine old and new capabilities into superior operational concepts. The latest advancements in data, analytics, and artificial intelligence (AI) technologies enable leaders to make better decisions faster, from the boardroom to the battlefield. Therefore, accelerating the adoption of these technologies presents an unprecedented opportunity to equip leaders at all levels of the Department with the data they need, and harness the full potential of the decision-making power of our people.

**Deputy Secretary of Defense. 2021. *Implementing Responsible Artificial Intelligence in the Department of Defense.* Department of Defense.** [**https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF**](https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF)

As the DoD embraces artificial intelligence (AI), it is imperative that we adopt responsible behavior, processes, and outcomes in a manner that reflects the Department' s commitment to its ethical principles, including the protection of privacy and civil liberties. A trusted ecosystem not only enhances our military capabilities, but also builds confidence with end-users, warfighters, and the American public. By leading in military ethics and AI safety, we reflect our Nation's values, encourage Responsible AI (RAI) development globally, and strengthen partnerships around the world. To that end, I reaffirm the DoD Al Ethical Principles adopted by the Department on February 21, 2020, for the design, development, deployment, and use of Al capabilities.

**Docherty, B. (2015, April 9). *Mind the gap: The lack of accountability for killer robots.* Human Rights Watch.** [**https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots**](https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots)**.**

Fully autonomous weapons, also known as “killer robots,” raise serious moral and legal concerns because they would possess the ability to select and engage their targets without meaningful human control. Many people question whether the decision to kill a human being should be left to a machine. There are also grave doubts that fully autonomous weapons would ever be able to replicate human judgment and comply with the legal requirement to distinguish civilian from military targets. Other potential threats include the prospect of an arms race and proliferation to armed forces with little regard for the law.

These concerns are compounded by the obstacles to accountability that would exist for unlawful harm caused by fully autonomous weapons. This report analyzes in depth the hurdles to holding anyone responsible for the actions of this type of weapon. It also shows that even if a case succeeded in assigning liability, the nature of the accountability that resulted might not realize the aims of deterring future harm and providing retributive justice to victims.

**Ekelhof, M. & Paoli, G.P. (2020). *Swarm Robotics: Technical and Operational Overview of the Next Generation of Autonomous Systems.* UNIDIR.** [**https://unidir.org/files/2020-04/UNIDIR%20Swarm%20Robotics%20-%202020.pdf**](https://unidir.org/files/2020-04/UNIDIR%20Swarm%20Robotics%20-%202020.pdf)

There is significant growing interest in many research laboratories and government agencies in developing swarms of robotic systems that have the ability to coordinate their actions to work collectively towards the execution of a shared goal. Working as a group, the swarm can perform both simple and complex tasks in a way that a single robot would be uncapable of. Each robotic unit within the swarm can be considered an autonomous member that reacts according to internal rules and the state of the environment. Nevertheless, it is precisely this ability of robots to autonomously make decisions (individually or as a group) that raises concerns among the international community.

**European Parliament. (2024, March 13). *Texts adopted - Artificial intelligence act.*** [***https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138\_EN.html***](https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html)

**European Parliament. (2018, September 12). *Texts adopted - Autonomous weapons systems.*** [**https://www.europarl.europa.eu/doceo/document/TA-8-2018-0341\_EN.html**](https://www.europarl.europa.eu/doceo/document/TA-8-2018-0341_EN.html)**.**

**Evaluating open-source imagery: A guide for judges and fact finders. *TRUE Project.*** [**https://www.trueproject.co.uk/osguide**](https://www.trueproject.co.uk/osguide)

Digital open source information – that is, information that is publicly accessible on the internet – is increasingly used as evidence before domestic and international courts, human rights bodies, and fact-finding bodies, where it has proven valuable in a variety of contexts.

​

This guide is organized around a number of key issues that a court or fact-finding body may need to address in their evaluation of open source information, including determining the authenticity of the digital image, and analyzing relevant metadata, source, location, and time information.

**Ferreira, A. (2023). Implications of Explainable Artificial Intelligence to the Ethics of Automated Warfare. Policy Brief, University of Waterloo.** [**https://uwaterloo.ca/defence-security-foresight-group/sites/default/files/uploads/documents/ferreira\_implications-of-explainable.pdf**](https://uwaterloo.ca/defence-security-foresight-group/sites/default/files/uploads/documents/ferreira_implications-of-explainable.pdf)

Explainable Artificial Intelligence (XAI) is a rapidly evolving field that aims to make the decision-making process of AI models more understandable, transparent, and accountable. As AI systems are increasingly adopted in critical defence and security applications, it is essential to ensure that these systems are explainable and justifiable to maintain trust, enable informed decision-making, and uphold ethical standards. However, as this brief will explore, although there are many promising recent advancements in the field of XAI that could be used in military domains, there are numerous downsides and ethical concerns that must be considered before deployment.

**Gender and Disarmament & Security and Technology Programmes.****(2024).*****Factsheet: Gender and Lethal Autonomous Weapons Systems.* UNIDIR.** [**https://unidir.org/publication/gender-and-lethal-autonomous-weapons-systems/**](https://unidir.org/publication/gender-and-lethal-autonomous-weapons-systems/)

This factsheet provides an overview of the issue of biases, especially on the basis of gender, that manifest in military applications of artificial intelligence (AI) such as lethal autonomous weapons systems (LAWS).

It also addresses how biases in LAWS have been discussed at relevant disarmament forums like the Group of Governmental Experts (GGE) meetings under the Convention on Certain Convention Weapons (CCW).

The factsheet further recommends areas of action for a host of stakeholders to ensure that gender biases are mitigated in military applications of AI.

**Grand-Clément, S. (2023). *Artificial intelligence beyond weapons: Application and impact of AI in the military domain.* UNIDIR.** [**https://unidir.org/publication/artificial-intelligence-beyond-weapons-application-and-impact-of-ai-in-the-military-domain/**](https://unidir.org/publication/artificial-intelligence-beyond-weapons-application-and-impact-of-ai-in-the-military-domain/)

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.

**Goussac, N. & Pacholska, M. (2025, March 6). *The Interpretation and Application of International Humanitarian Law in Relation to Lethal Autonomous Weapon Systems.* UNIDIR.** [**https://unidir.org/publication/the-interpretation-and-application-of-international-humanitarian-law-in-relation-to-lethal-autonomous-weapon-systems/**](https://unidir.org/publication/the-interpretation-and-application-of-international-humanitarian-law-in-relation-to-lethal-autonomous-weapon-systems/)

Much of the multilateral deliberations on lethal autonomous weapon systems (LAWS) over the last decade has been grounded in consideration of how international humanitarian law (IHL) is to be interpreted and applied to the development and use of these systems. The complexity of technologies in the area of LAWS challenges traditional understandings of IHL. Many contributions have grappled with what limits IHL places on the development and use of LAWS and what kinds of practical measures or limits might be or are being used to ensure that LAWS are used in compliance with these rules. Core topics among the views of States, scholars and other experts are the circumstances under which LAWS are permitted to be used in attacks and the measures that are required to be taken to minimize civilian harm due to the use of LAWS in attacks. In addition, the discourse has addressed the measures that must be taken before and after any attack involving the use of LAWS to prevent violations of IHL and ensure accountability in the case of any such violations.

To support these ongoing discussions, UNIDIR implemented a series of activities as part of the project "Towards a Common Understanding of the Application of IHL to Emerging Technologies in the Area of LAWS". This background paper summarizes publicly available views expressed by States, scholars and other experts participating in multilateral discussions on the applicability and interpretation of IHL with respect to the development and use of LAWS.

The background paper finds that, while all contributions to the discussion stem from the common starting point that IHL applies to the development and use of LAWS, divergences of both form and content persist in publicly available views. Despite the breadth of the discussions, a coherent comparison of views remains difficult to achieve and some IHL rules that govern the development and use of LAWS remain under-examined. Publicized views on measures that States can, do or should take with respect to the development and use of LAWS to avoid or minimize the effects of LAWS on civilian populations, civilians and civilian objects often do not specify whether such measures derive from an IHL principle or rule. The background paper underscores the considerations that arise in ensuring that LAWS are developed and used only in accordance with IHL and the challenges in achieving a level of certainty about the interpretation and application of IHL to these technologies.

**Heyns, C. (2013). *Report of the Special Rapporteur on extrajudicial summary or arbitrary executions.* UNHRC.** [**https://digitallibrary.un.org/record/755741?ln=en**](https://digitallibrary.un.org/record/755741?ln=en)

Lethal autonomous robotics (LARs) are weapon systems that, once activated, can select and engage targets without further human intervention. They raise far-reaching concerns about the protection of life during war and peace. This includes the question of the extent to which they can be programmed to comply with the requirements of international humanitarian law and the standards protecting life under international human rights law. Beyond this, their deployment may be unacceptable because no adequate system of legal accountability can be devised, and because robots should not have the power of life and death over human beings. The Special Rapporteur recommends that States establish national moratoria on aspects of LARs, and calls for the establishment of a high level panel on LARs to articulate a policy for the international community on the issue.

## **Human Rights Center: UC Berkeley School of Law & UN Human Rights Office of the High Commissioner. (2022, January 3). Berkeley Protocol on Digital Open Source Investigations: A Practical Guide on the Effective Use of Digital Open Source Information in Investigating Violations of International Criminal, Human Rights and Humanitarian Law.** [**https://www.ohchr.org/en/publications/policy-and-methodological-publications/berkeley-protocol-digital-open-source**](https://www.ohchr.org/en/publications/policy-and-methodological-publications/berkeley-protocol-digital-open-source)**.**

The Berkeley Protocol on Digital Open Source Investigations identifies international standards for conducting online research of alleged violations of international criminal, human rights, and humanitarian law. The Protocol provides guidance on methodologies and procedures for gathering, analysing, and preserving digital information in a professional, legal, and ethical manner. The publication also sets out measures that online investigators can and should take to protect the digital, physical, and psychosocial safety of themselves and others, including witnesses, victims, and first responders (such as citizens, activists, and journalists) who risk their own wellbeing to document war crimes and human rights violations so that those who are responsible are brought to justice. The Protocol is available on the websites of OHCHR and the Human Rights Center at the University of California, Berkeley, School of Law. It will also be issued in all six official UN languages, in print and electronic mediums.

**Human Rights Watch. (2022, November 10). *An agenda for action: Alternative processes for negotiating a killer-robots treaty.*** [**https://www.hrw.org/report/2022/11/10/agenda-action/alternative-processes-negotiating-killer-robots-treaty**](https://www.hrw.org/report/2022/11/10/agenda-action/alternative-processes-negotiating-killer-robots-treaty)

As the ninth year of CCW meetings on the topic comes to a close, the discussions have failed to produce the much-needed prohibitions and regulations on autonomous weapons systems that a majority of states desire. With major military powers getting ever closer to developing these dangerous systems, alternative options need to be pursued. It is time for states to initiate a process elsewhere to negotiate a new treaty on autonomous weapons systems.

**Human Rights Watch. (2012, November 19). *Losing humanity: The case against killer robots.*** [**https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots**](https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots)

With the rapid development and proliferation of robotic weapons, machines are starting to take the place of humans on the battlefield. Some military and robotics experts have predicted that “killer robots”—fully autonomous weapons that could select and engage targets without human intervention—could be developed within 20 to 30 years. At present, military officials generally say that humans will retain some level of supervision over decisions to use lethal force, but their statements often leave open the possibility that robots could one day have the ability to make such choices on their own power. Human Rights Watch and Harvard Law School’s International Human Rights Clinic (IHRC) believe that such revolutionary weapons would not be consistent with international humanitarian law and would increase the risk of death or injury to civilians during armed conflict. A preemptive prohibition on their development and use is needed.

A relatively small community of specialists has hotly debated the benefits and dangers of fully autonomous weapons. Military personnel, scientists, ethicists, philosophers, and lawyers have contributed to the discussion. They have evaluated autonomous weapons from a range of perspectives, including military utility, cost, politics, and the ethics of delegating life-and-death decisions to a machine. According to Philip Alston, then UN special rapporteur on extrajudicial, summary or arbitrary executions, however, “the rapid growth of these technologies, especially those with lethal capacities and those with decreased levels of human control, raise serious concerns that have been almost entirely unexamined by human rights or humanitarian actors.”[[1]](https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots#_ftn1) It is time for the broader public to consider the potential advantages and threats of fully autonomous weapons.

The primary concern of Human Rights Watch and IHRC is the impact fully autonomous weapons would have on the protection of civilians during times of war. This report analyzes whether the technology would comply with international humanitarian law and preserve other checks on the killing of civilians. It finds that fully autonomous weapons would not only be unable to meet legal standards but would also undermine essential non-legal safeguards for civilians. Our research and analysis strongly conclude that fully autonomous weapons should be banned and that governments should urgently pursue that end.

**Human Rights Watch. (2022). *Negotiating a treaty on autonomous weapons systems – The way forward.*** [**https://www.stopkillerrobots.org/wp-content/uploads/2022/06/Stop-Killer-Robots-Negotiating-a-Treaty-on-Autonomous-Weapons-Systems-The-Way-Forward.pdf**](https://www.stopkillerrobots.org/wp-content/uploads/2022/06/Stop-Killer-Robots-Negotiating-a-Treaty-on-Autonomous-Weapons-Systems-The-Way-Forward.pdf)

An internationally agreed legal framework on autonomous weapons systems is inevitable. After 9 years of discussions at the United Nations (UN), this outcome has immense global support including from the UN Secretary-General, at least 70 states, the International Committee of the Red Cross (ICRC), experts in technology and artificial intelligence, faith leaders, military veterans, and civil society organizations around the world.

**Human Rights Watch. (2020, October 20). *New weapons, proven precedent: Elements of and models for a treaty on killer robots.*** [**https://www.hrw.org/report/2020/10/20/new-weapons-proven-precedent/elements-and-models-treaty-killer-robots**](https://www.hrw.org/report/2020/10/20/new-weapons-proven-precedent/elements-and-models-treaty-killer-robots)

This report provides precedent for each of the treaty elements and shows that constructing a legally binding instrument does not require an entirely new approach. Earlier law and principles, often driven by similar concerns and objectives, can inform the structure of a treaty on fully autonomous weapons, and when negotiations start, facilitate crafting of language. The existence of relevant models should make it legally, politically, and practically easier to develop a new treaty.

# **International Council of the Red Cross. (2020). Artificial intelligence and machine learning in armed conflict: A human-centered approach. *International Review of the Red Cross: Digital technologies and war,* 102 (913), 463–479.** [**https://international-review.icrc.org/sites/default/files/reviews-pdf/2021-03/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913.pdf**](https://international-review.icrc.org/sites/default/files/reviews-pdf/2021-03/ai-and-machine-learning-in-armed-conflict-a-human-centred-approach-913.pdf)

At a time of increasing conflict and rapid technological change, the International Committee of the Red Cross (ICRC) needs both to understand the impact of new technologies on people affected by armed conflict and to design humanitarian solutions that address the needs of the most vulnerable. The ICRC, like many organizations across different sectors and regions, is grappling with the implications of artificial intelligence (AI) and machine learning for its work. AI is the use of computer systems to carry out tasks – often associated with human intelligence – that require cognition, planning, reasoning or learning; and machine learning systems are AI systems that are “trained” on and “learn” from data, which ultimately define the way they function. Since these are software tools, or algorithms, that could be applied to many different tasks, the potential implications may be far-reaching and yet to be fully understood. There are two broad – and distinct – areas of application of AI and machine learning in which the ICRC has a particular interest: first, its use in the conduct of warfare or in other situations of violence;1 and second, its use in humanitarian action to assist and protect the victims of armed conflict.2 This paper sets out the ICRC’s perspective on the use of AI and machine learning in armed conflict, the potential humanitarian consequences, and the associated legal obligations and ethical considerations that should govern its development and use. It also makes reference to the use of AI tools for humanitarian action, including by the ICRC.

# **International Committee of the Red Cross. (2014). *Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons.***

[**https://www.icrc.org/en/publication/4283-autonomous-weapons-systems**](https://www.icrc.org/en/publication/4283-autonomous-weapons-systems)**.**

This report includes a summary report of an expert meeting, summaries of expert presentations, and the background paper prepared before the expert meeting. Aspects of the report related to the present study include “Autonomous weapon systems under international humanitarian law,” “Accountability for use of autonomous weapon systems,” and “Ethical issues raised by autonomous weapon systems.”

**International Committee of the Red Cross. (2016). *Autonomous Weapon Systems: Technical, Military, Legal, and Humanitarian Aspects.*** [**https://www.icrc.org/sites/default/files/document/file\_list/4221-002-autonomous-weapons-systems-full-report.pdf**](https://www.icrc.org/sites/default/files/document/file_list/4221-002-autonomous-weapons-systems-full-report.pdf)**.**

Developing upon the 2014 expert meeting, another report was published after a 2016 Meeting of Experts. These conversations continue the debate regarding where limits can be placed on autonomous weapon systems. The legal and ethical conversations may provide greater insight into where accountability and evidence are being considered.

**International Committee of the Red Cross. (2019). *Autonomy, artificial intelligence and robotics: Technical aspects of human control.*** [**https://www.icrc.org/en/download/file/102852/autonomy\_artificial\_intelligence\_and\_robotics.pdf**](https://www.icrc.org/en/download/file/102852/autonomy_artificial_intelligence_and_robotics.pdf)**.**

In June 2018, the ICRC convened a round-table meeting with independent experts in autonomy, artificial intelligence (AI) and robotics to gain a better understanding of the technical aspects of human control, drawing on experience with civilian autonomous systems. This report combines a summary of the discussions at that meeting with additional research, and highlights the ICRC’s main conclusions, which do not necessarily reflect the views of the participants. Experience in the civilian sector yields insights that can inform efforts to ensure meaningful, effective and appropriate human control over weapon systems and the use of force.

**International Council of the Red Cross. (2020). *International Humanitarian Law and the challenges of contemporary armed conflicts: Recommitting to protection in armed conflict on the 70th anniversary of the Geneva Conventions.*** [**https://shop.icrc.org/international-humanitarian-law-and-the-challenges-of-contemporary-armed-conflicts-recommitting-to-protection-in-armed-conflict-on-the-70th-anniversary-of-the-geneva-conventions.html?\_\_\_store=en**](https://shop.icrc.org/international-humanitarian-law-and-the-challenges-of-contemporary-armed-conflicts-recommitting-to-protection-in-armed-conflict-on-the-70th-anniversary-of-the-geneva-conventions.html?___store=en)

In this report on international humanitarian law (IHL) and the challenges of contemporary armed conflicts, the International Committee of the Red Cross (ICRC) sets out its views on a number of pressing humanitarian and legal issues. The ICRC has submitted a report like this one – the Challenges Report – to every International Conference of the Red Cross and Red Crescent since 2003. The International Conference, which takes place every four years, brings together all the States party to the Geneva Conventions, all the National Red Cross and Red Crescent Societies, the International Federation of Red Cross and Red Crescent Societies, and the ICRC. The Challenges Report provides an overview of some of the challenges for IHL posed by contemporary armed conflicts, and outlines current or prospective ICRC action, positions, and areas of interest. One of its main aims is to stimulate reflection on all these matters. The 2019 Challenges Report focuses on the following topics: - urbanization of warfare - new technologies of warfare - the needs of civilians in increasingly long conflicts - IHL and non-State armed groups - terrorism, counterterrorism measures, and IHL - climate, armed conflict, and the natural environment - enhancing respect for IHL.

# **International Council of the Red Cross. (2021). *Legal review of new weapons.*** [**https://www.icrc.org/en/document/new-weapons-factsheet**](https://www.icrc.org/en/document/new-weapons-factsheet)

International humanitarian law (IHL) is a set of rules which seeks to limit the effects of armed con­flict. It restricts the means and methods of warfare permitted to combatants and protects those who are not, or are no longer, actively taking part in fighting. In particular, the right of the parties to an armed conflict to choose means and methods of warfare is limited by basic rules of IHL relating to the conduct of hostilities, many of which are found in Protocol I additional to the 1949 Geneva Conven­tions on the protection of victims of international armed conflicts (Additional Protocol I). Treaty and customary IHL also set prohibitions or limitations on the employment of certain weapons, means and methods of warfare. States are required to review the legality of new weapons, means and methods of warfare before deploying them to the armed forces. This obligation is found in Article 36 of Additional Protocol I, to which the vast majority of States are party.

**International Council of the Red Cross (2021). *Position on autonomous weapons systems.*** [**https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems**](https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems)

The International Committee of the Red Cross (ICRC) has, since 2015, urged States to establish internationally agreed limits on autonomous weapon systems to ensure civilian protection, compliance with international humanitarian law, and ethical acceptability.

**Kostopoulos, L. (2019). *The Role of Data in Algorithmic Decision-Making*. UNIDIR.** [**https://unidir.org/sites/default/files/publication/pdfs/the-role-of-data-in-algorithmic-decision-making-en-815.pdf**](https://unidir.org/sites/default/files/publication/pdfs/the-role-of-data-in-algorithmic-decision-making-en-815.pdf)

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.

**Lewis, D.A. (ed.). (2020). *A Compilation of Materials Apparently Reflective of States’ Views on International Legal Issues pertaining to the Use of Algorithmic and Data-reliant Socio-technical Systems in Armed Conflict.* Harvard Law School Program on International Law and Armed Conflict.** [**https://pilac.law.harvard.edu/a-compilation-of-materials-apparently-reflective-of-states-views-on-international-legal-issues-pertaining-to-the-use-of-algorithmic-and-data-reliant-socio-technical-systems-in-armed-conflict**](https://pilac.law.harvard.edu/a-compilation-of-materials-apparently-reflective-of-states-views-on-international-legal-issues-pertaining-to-the-use-of-algorithmic-and-data-reliant-socio-technical-systems-in-armed-conflict)**.**

This document is a compilation of materials that at least appear to be reflective of one or more states’ views on international legal issues pertaining to the actual or possible use of algorithmic and data-reliant socio-technical systems in armed conflict.

**Michel, A.H. (2021). *Known Unknowns: Data Issues and Military Autonomous Systems.* UNIDIR.** [**https://doi.org/10.37559/SecTec/21/AI1**](https://doi.org/10.37559/SecTec/21/AI1)**.**

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.

**Michel, A.H. (2020). *The Black Box, Unlocked.* UNIDIR.** [**https://doi.org/10.37559/SecTec/20/AI1**](https://doi.org/10.37559/SecTec/20/AI1)**.**

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.

**Ministry of Defence. (2022). *Ambitious safe and responsible: Our approach to the delivery of AI-enabled capability in defence.* GOV.UK.** [**https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\_data/file/1082991/20220614-Ambitious\_Safe\_and\_Responsible.pdf**](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082991/20220614-Ambitious_Safe_and_Responsible.pdf)**.**

The Defence AI Strategy sets out our view of the strategic opportunities and challenges presented by the emergence of AI as a transformative and disruptive new technology. Realising the benefits of AI – and countering threats and challenges associated with the use of AI by others – is one of the most critical strategic challenges of our time.

**Ministry of Defence. (2022). *Defence artificial strategy intelligence strategy.* GOV.UK.**[**https://assets.publishing.service.gov.uk/media/62a7543ee90e070396c9f7d2/Defence\_Artificial\_Intelligence\_Strategy.pdf**](https://assets.publishing.service.gov.uk/media/62a7543ee90e070396c9f7d2/Defence_Artificial_Intelligence_Strategy.pdf)**.**

The Integrated Review 1 (2021) highlights national excellence in AI as central to securing the UK’s status as a ‘Science and Technology Superpower’ by 2030. The National AI Strategy 2 (2021) notes its huge potential to rewrite the rules of whole industries, drive substantial economic growth and transform all areas of life. The Integrated Operating Concept 3 (2020) describes how pervasive information and rapid technological change are transforming the character of warfare. Across the spectrum of military operations, conflict is becoming increasingly complex and dynamic. New technologies generate massive volumes of data, unlock new threats and vulnerabilities and expand the scale of potential attacks through advanced next-generation capabilities (such as swarming drones, high-speed weapons and advanced cyber-attacks).

These technologies – and the operational tempo they enable – are likely to compress decision times dramatically, tax the limits of human understanding and often require responses at machine speed.

As the Defence Command Paper 4 (2021) notes, “future conflicts may be won or lost on the speed and efficacy of the AI solutions employed”. Simultaneously, information operations are increasingly important to counter false narratives that distract attention, provide cover for malign activities and undermine public support. In short, a radical upheaval in defence is underway and AI-related strategic competition is intensifying. Our response must be rapid, ambitious, and comprehensive.

**Minor, E. (2023, February). *Laws for LAWS: Towards a treaty to regulate lethal autonomous weapons.* Friedrich-Ebert-Stiftung.** [**https://library.fes.de/pdf-files/international/20013.pdf**](https://library.fes.de/pdf-files/international/20013.pdf)

The automation of decision-making has impacts and implications in many areas of society, particularly for human rights and digital dehumanisation, of which increasing autonomy in weapons systems represents the most deadly iteration. Away from the CCW, the potential risks to human rights of emerging military technologies using algorithms and machine learning were also recognised by states at the Human Rights Council in 2022. With concern remaining high amongst civil society, international organisations and the UN Secretary-General, the tech and robotics sectors and a large number of states, moving forward towards legal regulation is an urgent task. With convergence developing amongst states on many principles and points of policy, those states that are willing to participate must now take the next step towards developing these elements into international law that could win broad support and exert widespread influence.

**Nasser, N.B. (2025, February 7). *The Impact of Artificial Intelligence on Regional Security, Threat Perceptions and the Middle East WMD-Free Zone.* UNIDIR.** [**https://unidir.org/publication/the-impact-of-artificial-intelligence-on-regional-security-threat-perceptions-and-the-middle-east-wmd-free-zone/**](https://unidir.org/publication/the-impact-of-artificial-intelligence-on-regional-security-threat-perceptions-and-the-middle-east-wmd-free-zone/)

With significant advancements in artificial intelligence (AI), many countries have been seeking to integrate these technologies into military and defence industries, including in the Middle East. In this publication, the author examines and analyzes the impact of AI on regional security, weapons of mass destruction (WMD), proliferation-related risks in the Middle East, and its potential influence on the initiative to establish a WMD-Free Zone in the region.

The author examines plausible scenarios, such as the emergence of an arms race in military applications of AI among regional states, which could either increase WMD proliferation risks in the region or, conversely, help reduce them. The paper also discusses key factors AI may have in the negotiations to establish a WMD-Free Zone, including urgency and the potential technical benefits of AI in arms control processes.

**Naval Research Advisory Committee. 2017. *Autonomous and Unmanned Systems in the Department of the Navy.*** [**https://www.senedia.org/wp-content/uploads/2018/01/NRAC-Report-Autonomous-and-Unmanned-Systems-in-the-Department-of-Navy.pdf**](https://www.senedia.org/wp-content/uploads/2018/01/NRAC-Report-Autonomous-and-Unmanned-Systems-in-the-Department-of-Navy.pdf)

The slow pace of innovation in the Department of the Navy relative to the pace of innovation in our country’s private sector, including the private sector in adversary countries, shows the magnitude of the leadership challenge we face in this area. Potential adversaries, such as China and Iran, are aggressively advancing their civilian autonomy sectors. China already dominates the international drone market and intends to dominate the artificial intelligence market by 2030. The Naval Research Advisory Committee was tasked by the Secretary of the Navy in July 2016 to examine unmanned system strategies in the Department of the Navy (DoN) with a goal of providing recommendations on the application of autonomous and robotic systems, levels of autonomy, learning machines and human-machine teaming.

**Paoli, G.P. & Afina, Y. (2025, March 10). *AI in the Military Domain: A briefing note for States.* UNIDIR.** [**https://unidir.org/publication/ai-military-domain-briefing-note-states/**](https://unidir.org/publication/ai-military-domain-briefing-note-states/)

On 24 December 2024, the United Nations (UN) General Assembly adopted Resolution A/RES/79/239 on [Artificial intelligence in the military domain and its implications for international peace and security](https://unidir.org/wp-content/uploads/2025/03/UN_General_Assembly_A_RES_79_239-EN.pdf). The UN Secretary-General recently invited Member States, observer States, international and regional organizations, the International Committee of the Red Cross, civil society, industry and the scientific community to submit their views “on the opportunities and challenges posed to international peace and security by the application of artificial intelligence in the military domain, with specific focus on areas other than lethal autonomous weapons systems”.

This briefing note will contribute to a report submitted to the 18th session of General Assembly and aims to support States in the formulation of their national views on this topic. It seeks to ensure that the resulting report is as comprehensive, diverse and geographically representative as possible. The brief includes some contextual information on the topic of AI in the military domain, a set of considerations for States to refer to, and a list of suggested readings that draws on UNIDIR’s own research and selected external publications.

**Paoli, G.P., et al. (2020, August 31). *Modernizing Arms Control: Exploring responses to the use of AI in military decision-making.* UNIDIR.** [**https://unidir.org/publication/modernizing-arms-control/**](https://unidir.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.

**Paoli, G.P., Spazian, A. & Anand, A. (2021, July 30). *Table-Top Exercises on the Human Element and Autonomous Weapons System.* UNIDIR.** [**https://unidir.org/publication/table-top-exercises-on-the-human-element-and-autonomous-weapons-system/**](https://unidir.org/publication/table-top-exercises-on-the-human-element-and-autonomous-weapons-system/)

This report summarises the findings of UNIDIR's series of regional tabletop exercises conducted between September 2020 and June 2021. The project brought together 198 experts from 75 different 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).

**Pauwels, E. (2020). Artificial Intelligence and data capture technologies in violence and conflict prevention. *Global Centre on Cooperative Security. Accessed on*, *10*(07), 2022.** [**https://globalcenter.org/resource/artificial-intelligence-and-data-capture-technologies-in-violence-and-conflict-prevention/**](https://globalcenter.org/resource/artificial-intelligence-and-data-capture-technologies-in-violence-and-conflict-prevention/)

Threats to human rights and security triggered by artificial intelligence (AI) and data capture technologies will require peacebuilding and violence prevention actors to bridge the gap between early warning and response and anticipate new challenges. Conflict prevention is rapidly becoming an element of UN peacekeeper mandates in which technological and data governance will have powerful and unprecedented implications. UN actors need to integrate emerging technologies to digitize, share, and secure the information they collect from open sources, human informants, and data capture technologies. They also need to monitor how armed nonstate actors evolve and blend into civilian environments, collude with transnational criminal networks, and adapt their attack strategies to new domains, including cyberspace. Experts must scrutinize how online hate speech and incitements to violence contaminate the lifeblood of social media and private messaging applications in countries where ethnic and socioeconomic tensions prevail.

In this policy brief, Eleonore Pauwels examines how AI and data capture technologies can be positively harnessed and potentially misused, as the new paradigm of predictive behavioral analysis and population data capture is increasingly being presented as a solution to challenges in humanitarian action, conflict prevention, and counterterrorism. The brief advances recommendations for ensuring a do-no-harm approach to deploying these technologies in the field.

***Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy.* (2023, February). Responsible AI in the Military Domain Summit (REAIM 2023) in the Hague.**

[**https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/**](https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/)**.**

The Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy provides a normative framework addressing the use of these capabilities in the military domain. Launched in February 2023 at the Responsible AI in the Military Domain Summit (REAIM 2023) in the Hague, the Declaration aims to build international consensus around responsible behavior and guide states’ development, deployment, and use of military AI. The Declaration provides a basis for exchanging best practices and building states’ capacities, which will allow endorsing States to share experience and ideas.

The Department of State, in collaboration with the Department of Defense, convened the inaugural plenary meeting of States endorsing the Political Declaration on March 19-20, 2024. This plenary meeting was the first step to working with endorsing States to operationalize and implement this groundbreaking initiative. The United States will convene regular dialogue among endorsing States to further promote international support for and implementation of these responsible practices.

The United States encourages all states to support the Declaration and join with other endorsing States to promote responsible military use of AI and autonomy.

**Puscas, I. *Confidence-Building Measures for Artificial Intelligence: A framing paper.* UNIDIR.** [**https://www.unidir.org/sites/default/files/2022-12/Confidence-Building\_Final.pdf**](https://www.unidir.org/sites/default/files/2022-12/Confidence-Building_Final.pdf)

Artificial intelligence (AI) will shape the future of conflict and warfare in ways that are difficult to predict due to the high uncertainty that characterizes the development and integration of this transformative technology in military capabilities. What is certain is that the increased adoption of AI will introduce new risks to international security that traditional instruments of risk/incident prevention and management may not be adequate to address. Against this backdrop, UNIDIR is launching a project on confidence-building measures (CBMs) for AI, which seeks to explore options that states can consider to mitigate risks and build more confidence and transparency in the development and use of AI in military systems. The project will comprise of two main phases: 1. Risk mapping, which will aim to develop a comprehensive overview of the main categories of risks of AI systems, as well as their implications for international security. This evaluation of risks covers a broad taxonomy of risks, such as cybersecurity risks of AI systems, intrinsic risks of the technology (e.g. algorithmic brittleness), or risks related to human– machine interaction. 2. Exploring possible pathways for the development of CBMs, which will build on the research findings from the previous phase and will convene multi-stakeholder dialogues with a view to assess realistic options for the development of CBMs. While more work exists in relation to AI safety and risk management in the context of civilian applications of AI, UNIDIR’s project aims at filling a key gap by focusing specifically on military applications and the possible confidence-building framework that could be designed for this unique technology.

**Puscas, I. (2022). *Human-Machine Interfaces in Autonomous Weapons Systems: Considerations for human control.* UNIDIR.** [**https://unidir.org/files/2022-07/UNIDIR\_Human-Machine%20Interfaces.pdf**](https://unidir.org/files/2022-07/UNIDIR_Human-Machine%20Interfaces.pdf)

Human control over autonomous weapon systems (AWS) has been a core theme in the discussions of the Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE on LAWS), which has met formally since 2017 in the framework of the Convention on Certain Conventional Weapons (CCW). The meaning and operationalization of control have been among the most contentious topics in the Group’s debates. Three main modalities of control have emerged in recent years and are now widely considered to impose practical limits on AWS: control on the weapon parameters, control on the environment of use, and control through human– machine interaction during use.

**Puscas, I. & Anand, A. (2023, May 10). *Proposals Related to Emerging Technologies in the Area of Lethal Autonomous Weapons Systems: A Resource Paper (updated).* UNIDIR.** [**https://unidir.org/publication/proposals-related-to-emerging-technologies-in-the-area-of-lethal-autonomous-weapons-systems-a-resource-paper-updated/**](https://unidir.org/publication/proposals-related-to-emerging-technologies-in-the-area-of-lethal-autonomous-weapons-systems-a-resource-paper-updated/)

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. These include:

1. Application of International Humanitarian Law (IHL)
2. Weapons prohibitions and other regulations/restrictions
3. Application of International Human Rights Law (IHRL) and International Criminal Law (ICL)
4. Characterisation
5. General requirements regarding human-machine interaction and human control
6. Responsibility and accountability
7. Legal reviews
8. Risk mitigation
9. Ethical considerations
10. Peaceful uses of Artificial Intelligence (AI)
11. Potential benefits of autonomy in weapon systems

Also available: [Annex A](https://unidir.org/sites/default/files/2023-05/UNIDIR_Proposals_Emerging_Technologies_Lethal_Autonomous_Weapons_Systems_Annex_A_2023.pdf), which includes relevant excerpts from proposals related to emerging technologies in the area of lethal autonomous weapons systems.

*\* This Resource Paper is an updated version of the previous document UNIDIR released in July 2022, and includes the following additional submissions to the GGE on LAWS in 2022 that were not included in the previous version: Elements for a Legally Binding Instrument to Address the Challenges Posed by Autonomy in Weapon Systems; Protocol VI; Working Paper submitted by Finland, France, Germany, the Netherlands, Norway, Spain and Sweden; Working Paper of the People’s Republic of China on LAWS, and Working Paper of the Russian Federation “Application of International Law to Lethal Autonomous Weapons Systems (LAWS)”.*

**Sisson, M., et al. (2020, June 3). *The Militarization of Artificial Intelligence.* UNODA.** [**https://disarmament.unoda.org/the-militarization-of-artificial-intelligence/**](https://disarmament.unoda.org/the-militarization-of-artificial-intelligence/)

Artificial Intelligence (AI) has the potential to improve the health and well-being of individuals, communities, and states, and help meet the UN’s Sustainable Development Goals. However, certain uses of AI could also undermine international peace and security by raising concerns about safety and security of the technology, accelerating the pace of armed conflicts, or loosening human control over the means of war.

In 2019, the United Nations Office for Disarmament Affairs, the Stanley Center and the Stimson Center partnered in a workshop and series of papers to facilitate a multistakeholder discussion among experts from Member States, industry, academia, and research institutions, with the aim of building understanding about the peace and security implications of AI. This publication captures that conversation and shares assessments of the topic from US, Chinese, and Russian perspectives. It is intended to provide a starting point for more robust dialogues among diverse communities of stakeholders as they endeavor to maximize the benefits of AI while mitigating the misapplication of this important technology.

**Stanley Center for Peace and Security, UNODA & Stimson Center. (2019, August). *The militarization of artificial intelligence.*** [**https://front.un-arm.org/wp-content/uploads/2020/06/Stanley-Stimson-UNODA-2020-TheMilitarization-ArtificialIntelligence.pdf**](https://front.un-arm.org/wp-content/uploads/2020/06/Stanley-Stimson-UNODA-2020-TheMilitarization-ArtificialIntelligence.pdf)

To facilitate a conversation between disparate stakeholders on this topic, the UN Office for Disarmament Affairs, the Stimson Center, and the Stanley Center for Peace and Security convened an initial dialogue on the intersection of AI and national military capabilities. Over two days at UN headquarters in New York, experts from member states, industry, academia, and research institutions participated in a workshop on The Militarization of Artificial Intelligence.

**Taddeo, M. & Blanchard, A. (2021). *A comparative analysis of the definitions of autonomous weapons.* UNODA.** [**https://documents.unoda.org/wp-content/uploads/2021/10/20210721-Autonomous-Weapon-Systems-Definitions-TO-SHARE.pdf**](https://documents.unoda.org/wp-content/uploads/2021/10/20210721-Autonomous-Weapon-Systems-Definitions-TO-SHARE.pdf)

In this report we focus on the definition of autonomous weapons systems (AWS). We provide a comparative analysis of existing official definitions of AWS as provided by States and international organisations, like ICRC and NATO. The analysis highlights that the definitions draw focus on different aspects of AWS and hence lead to different approaches to address the ethical and legal problems of these weapons systems. This approach is detrimental both in terms of fostering an understanding of AWS and in facilitating agreement around conditions of deployment and regulations of their use and, indeed, whether AWS are to be used at all. We draw from the comparative analysis to identify essential aspects of AWS and then offer a definition that provides a value-neutral ground to address the relevant ethical and legal problems. In particular, we identify four key aspects – autonomy; adapting capabilities of AWS; human control; and purpose of use – as the essential factors to define AWS and which are key when considering the related ethical and legal implications.

**UC Berkeley School of Law Human Rights Center. (2021). *Digital lockers: Archiving social media evidence of atrocity crimes.*** [**https://humanrights.berkeley.edu/publications/digital-lockers-archiving-social-media-evidence-atrocity-crimes/**](https://humanrights.berkeley.edu/publications/digital-lockers-archiving-social-media-evidence-atrocity-crimes/)**.**

Given the use of social media by people living in areas of armed conflict or severe repression, social media platforms have become accidental and unstable archives for human rights content. The last two decades have witnessed a fundamental shift in how people around the world communicate. During this period, the proliferation of smartphones and the rise of social media platforms have enabled increased identification, collection, and sharing of digital information related to international crimes and human rights violations. Whereas human rights researchers once struggled to find online content relevant to their investigations, today researchers may find themselves drowned in a tsunami of content with potential evidentiary value, as well as utility for the documentation of atrocities more generally — including for advocacy, research, and development of an historical record of world events. With 6,000 tweets generated every second and 500 hours of video content uploaded to YouTube every minute, the challenge is figuring out how to find the “signal” by siphoning out the online “noise,” as well as how to find reliable information buried in a digital environment replete with misinformation and disinformation. “Digital Lockers” looks at models for archiving digital information, including the structures, funding, and management of archives, lessons from previous experience, and the legal, political, technical, financial, and operational challenges likely to arise in the creation of a digital evidence locker or new legal framework. The report offers recommendations for an effective and efficient way forward.

**UN General Assembly. (2024, December 31). *Artificial intelligence in the military domain and its implications for international peace and security.* UN Doc A/RES/79/239.** [**https://unidir.org/wp-content/uploads/2025/03/UN\_General\_Assembly\_A\_RES\_79\_239-EN.pdf**](https://unidir.org/wp-content/uploads/2025/03/UN_General_Assembly_A_RES_79_239-EN.pdf)

**UNHRC. *Human rights implications of new and emerging technologies in the military domain.* A/HRC/51/L.25.** [**https://daccess-ods.un.org/tmp/8317592.14401245.html**](https://daccess-ods.un.org/tmp/8317592.14401245.html)

**UNIDIR. (2018). *Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies.*** [**https://unidir.org/sites/default/files/publication/pdfs/algorithmic-bias-and-the-weaponization-of-increasingly-autonomous-technologies-en-720.pdf**](https://unidir.org/sites/default/files/publication/pdfs/algorithmic-bias-and-the-weaponization-of-increasingly-autonomous-technologies-en-720.pdf)

This primer is divided into three main sections. It begins with a general discussion of algorithmic biases – their nature, types and sources. The second section considers the impacts of algorithmic bias, with both real-world examples and examples of how bias could arise in future weapon systems. The third section considers potential mitigation strategies to address bias determined to be harmful or undesirable.

**UNIDIR. *The Human Element in Decisions About the Use of Force.*** [**https://unidir.org/files/2020-03/UNIDIR\_Iceberg\_SinglePages\_web.pdf**](https://unidir.org/files/2020-03/UNIDIR_Iceberg_SinglePages_web.pdf)

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.

**Work, R.O. (2021, April 28). *Principles for the combat employment of weapon systems with autonomous functionalities*. Center for a New American Security.** [**https://www.cnas.org/publications/reports/proposed-dod-principles-for-the-combat-employment-of-weapon-systems-with-autonomous-functionalities**](https://www.cnas.org/publications/reports/proposed-dod-principles-for-the-combat-employment-of-weapon-systems-with-autonomous-functionalities)

This paper is organized into four sections. The first details the history of U.S. weapon systems with autonomous functionalities. It is intended to give the reader a notion of how these weapons have historically been used, why autonomous functionalities are so useful, and why the DoD retains the right to use them. The second section explains why weapons with autonomous functionalities are now being improved through the addition of AI, an important development that aims to make the weapons more discriminate in the application of force. The third section explains why the DoD should consider publishing a new set of principles for the combat employment of weapon systems with autonomous functionalities. The final section outlines seven proposed principles for consideration.

## **Conference Papers, Proceedings, and Workshops**

**Burton, J., & Soare, S. R. (2019, May). Understanding the strategic implications of the weaponization of artificial intelligence. In *2019 11th international conference on Cyber Conflict (CyCon)* (Vol. 900, pp. 1-17). IEEE.**

Artificial Intelligence (AI) is expected to have a revolutionary impact across societies and to create economic displacement and disruption in security and defense. Yet the impact of AI on national security and military affairs has received relatively scant attention. The existing policy-focused literature has concentrated mainly on the technological, ethical or legal limitations of deploying AI and on the risks associated with it. This paper seeks to contribute to the debate by outlining the strategic implications of the weaponization of AI for international security. It explores how and in what ways AI is currently being utilized in the defense sector to enhance offensive and defensive military technologies and operations and assesses the ways in which the incorporation of AI into military platforms will affect war fighting and strategic decision-making. The paper is in four sections. Section one develops a typology of military AI that forms a foundation for the rest of the paper. The second section examines the uses of AI in cyberspace and the relationships between ‘cyber weapons’ and AI capabilities. The third section examines how the embeddedness of AI-based capabilities across the land, air, naval and space domains may affect combined arms operations. The final section distills the main strategic implications of weaponized AI, which include the speed of decision-making and action as well as enhanced domain situational awareness.

**Disruption Network Lab. *Investigating the kill cloud: Information warfare, autonomous weapons, and AI.*** [**https://www.disruptionlab.org/investigating-the-kill-cloud**](https://www.disruptionlab.org/investigating-the-kill-cloud)

# *Investigating The Kill Cloud* brings together the 2023-2024 research fellows of the [Disruption Network Institute](https://disruption.institute/) and international experts investigating the impact of artificial intelligence on new technologies of war, automated weapons and networked warfare. The Institute is a unique new hub that brings transparency to the complexity and opacity of war.

**Doshi-Velez, F. & Kortz, M. (2017). *Accountability of AI Under the Law: The Role of Explanation.* Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper.** [**https://dash.harvard.edu/bitstream/handle/1/34372584/2017-11\_aiexplainability-1.pdf?sequence=3&isAllowed=y**](https://dash.harvard.edu/bitstream/handle/1/34372584/2017-11_aiexplainability-1.pdf?sequence=3&isAllowed=y) **[**[**https://perma.cc/F44N-85RT**](https://perma.cc/F44N-85RT)**].**

The ubiquity of systems using artificial intelligence or “AI” has brought increasing attention to how those systems should be regulated. The choice of how to regulate AI systems will require care. AI systems have the potential to synthesize large amounts of data, allowing for greater levels of personalization and precision than ever before—applications range from clinical decision support to autonomous driving and predictive policing. That said, our AIs continue to lag in common sense reasoning [McCarthy, 1960], and thus there exist legitimate concerns about the intentional and unintentional negative consequences of AI systems [Bostrom, 2003, Amodei et al., 2016, Sculley et al., 2014]. How can we take advantage of what AI systems have to offer, while also holding them accountable?

In this work, we focus on one tool: explanation. Questions about a legal right to explanation from AI systems was recently debated in the EU General Data Protection Regulation [Goodman and Flaxman, 2016, Wachter et al., 2017a], and thus thinking carefully about when and how explanation from AI systems might improve accountability is timely. Good choices about when to demand explanation can help prevent negative consequences from AI systems, while poor choices may not only fail to hold AI systems accountable but also hamper the development of much-needed beneficial AI systems.

Below, we briefly review current societal, moral, and legal norms around explanation, and then focus on the different contexts under which explanation is currently required under the law. We find that there exists great variation around when explanation is demanded, but there also exist important consistencies: when demanding explanation from humans, what we typically want to know is whether and how certain input factors affected the final decision or outcome.

These consistencies allow us to list the technical considerations that must be considered if we desired AI systems that could provide kinds of explanations that are currently required of humans under the law. Contrary to popular wisdom of AI systems as indecipherable black boxes, we find that this level of explanation should generally be technically feasible but may sometimes be practically onerous—there are certain aspects of explanation that may be simple for humans to provide but challenging for AI systems, and vice versa. As an interdisciplinary team of legal scholars, computer scientists, and cognitive scientists, we recommend that for the present, AI systems can and should be held to a similar standard of explanation as humans currently are; in the future we may wish to hold an AI to a different standard.

**Du, X., et al. (2020, August). SoK: Exploring the state of the art and the future potential of artificial intelligence in digital forensic investigation. In *Proceedings of the 15th international conference on availability, reliability and security, Article 46*, 1–10.** [**https://doi.org/10.1145/3407023.3407068**](https://doi.org/10.1145/3407023.3407068)

Multi-year digital forensic backlogs have become commonplace in law enforcement agencies throughout the globe. Digital forensic investigators are overloaded with the volume of cases requiring their expertise compounded by the volume of data to be processed. Artificial intelligence is often seen as the solution to many big data problems. This paper summarises existing artificial intelligence based tools and approaches in digital forensics. Automated evidence processing leveraging artificial intelligence based techniques shows great promise in expediting the digital forensic analysis process while increasing case processing capacities. For each application of artificial intelligence highlighted, a number of current challenges and future potential impact is discussed.

**Freeman, L. (2023). Seeing through the fog: The impact of information operations on war crimes investigations in Ukraine. 2023 15th International Conference on Cyber Conflict: Meeting Reality, Jančárková, T., Giovannelli, D., Podiņš, K., Winther, I. (eds). NATO CCDCOE Publications, Tallinn.** [**https://humanrights.berkeley.edu/publications/seeing-through-the-fog-the-impact-of-information-operations-on-war-crimes-investigations-in-ukraine/**](https://humanrights.berkeley.edu/publications/seeing-through-the-fog-the-impact-of-information-operations-on-war-crimes-investigations-in-ukraine/)

As Russian forces closed in on Kyiv, a MiG-29 Fulcrum swooped in and took down six Russian jets. The next day, the same MiG shot down ten more. Stories of the hero fighter pilot spread like wildfire throughout Ukraine and across the internet, turning the “Ghost of Kyiv” into a living legend. But he was not living, or even real. The pilot and his exploits were a total fiction created as part of an influence campaign spread via social media to strike terror into Russian forces, fortify the resolve of Ukrainian citizens, and amaze the world with Ukraine’s unexpected strength and courage.The strategic use of the online information environment is only one facet of intangible warfare between Russia and Ukraine that makes this contemporary conflict particularly unique and complex. Propaganda, disinformation, and psychological operations are as old as warfare itself, but advanced digital technologies now reshape conflicts in often unanticipated, unforeseen, and surprising ways. These changing dynamics inevitably have an impact on those tasked with investigating war crimes and establishing the truth of what occurred on the battlefield. This paper examines the strategic use of digital information and communications technologies in the Russia–Ukraine conflict to better understand how they are changing the dynamics of war, war narratives, and war crimes investigations. The first section of the paper briefly explains how war crimes investigators and prosecutors are increasingly relying on digital material as evidence in their cases. The second section considers how digital information operations are being deployed and how these operations impact the investigation of war crimes. Finally, the third section highlights some of the tools that can help war crimes investigators fight back against a complex and chaotic information environment.

**International Institute of Humanitarian Law. (2019). Whither the human in armed conflict? IHL Implications of New Technology in Warfare. In G. Venturini. (ed.) & G.L. Beruto (assoc ed.), *42nd Round Table on Current Issues of International Humanitarian Law*, Sanremo. FrancoAngeli.**

The tight bond linking scientific progress, technological development and their military exploitation constitutes a recurring leitmotif in the current international security scenario.

In this context, the rule of law may appear blurred and international humanitarian law (IHL) incapable of keeping pace with technological progress. Armaments research and innovation programmes are among the issues at the top of the agenda of advanced national powers, bolstering an already unbalanced relationship between growing technological military capacity and the legal frameworks which limit their usage according to IHL fundamental principles.

The 42nd Round Table on current issues of international humanitarian law, jointly organized by the Sanremo Institute and the International Committee of the Red Cross, gathered together academics, legal experts, military commanders and government officials to discuss the crucial question of technology developments and their application in armed conflicts, including the challenges imposed by the military use of cyber arms and the widespread of autonomous weapons in warfare, focusing on the risks related to their use in urban contexts.

The Round Table provided the opportunity for fruitful and constructive debates on crucial topics, such as the potential human costs of cyber warfare and applicable IHL provisions; the potential support that cyber technology could provide to humanitarian operations; and how IHL represents an effective legal framework for warfare in outer space. The way forward in addressing the challenges of using new technologies and weapons within military operations were also discussed, highlighting the primary objective of IHL to protect and safeguard civilians and vulnerable groups from the violence of armed conflicts.

The proceedings of this Round Table aim to confirm, once again, the “humanitarian dialogue in the spirit of Sanremo” and to strongly reassert the importance of promoting the application of IHL, particularly when it comes to specific areas where regulatory gaps occur.

**Kermode, L., Freyberg, J., et al. (2019). Objects of violence: synthetic data for practical ML in human rights investigations. *AI for Social Good workshop at NeurIPS*, Canada.** [**https://aiforsocialgood.github.io/neurips2019/accepted/track1/pdfs/68\_aisg\_neurips2019.pdf**](https://aiforsocialgood.github.io/neurips2019/accepted/track1/pdfs/68_aisg_neurips2019.pdf)

We introduce a machine learning workflow to search for, identify, and meaningfully triage videos and images of munitions, weapons, and military equipment, even when limited training data exists for the object of interest. This workflow is designed to expedite the work of OSINT ("open source intelligence") researchers in human rights investigations. It consists of three components: automatic ren- dering and annotating of synthetic datasets that make up for a lack of training data; training image classifiers from combined sets of photographic and synthetic data; and mtriage [1], an open source software that orchestrates these classifiers’ deployment to triage public domain media, and visualise predictions in a web interface. We show that synthetic data helps to train classifiers more effectively, and that certain approaches yield better results for different architectures. We then demonstrate our workflow in two real-world human rights investigations: the use of the Triple-Chaser tear gas grenade against civilians, and the verification of allegations of military presence in Ukraine in 2014.

**Meetings of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System. (2017-2023). UNODA.** [**https://disarmament.unoda.org/meetings-of-the-group-of-governmental-experts/**](https://disarmament.unoda.org/meetings-of-the-group-of-governmental-experts/)

Membership of the Group is comprised of all High Contracting Parties to the CCW. The meeting is also open to any non-High Contracting Party, international organizations, civil society and academia.

The Secretariat of the Group is the Implementation Unit of the CCW, housed within UNODA Geneva.

**New Technologies on the Battlefield: Friend or Foe?*.* (2021). *Collegium 51: Proceedings of the 21st Bruges Colloquium.* College of Europe, CICR.** [**https://www.coleurope.eu/sites/default/files/uploads/page/collegium\_51\_web.pdf#page=80**](https://www.coleurope.eu/sites/default/files/uploads/page/collegium_51_web.pdf#page=80)

The Colloquium will be focused on new technologies on the battlefield. I believe that the different aspects that will be discussed, from the challenges that it puts on International Law to the new instruments it puts at the disposal of humanitarian actors, are extremely relevant, not only to experts and academia but also for policy makers. There are issues that need to be addressed that can shape the future of policy making for decades to come. I am sure that the quality of our discussions during the next five days will not only provide food for thought but also elements for decisions and actions to be taken on the ground for all those who are involved in this important field.

## **Dissertations and Theses:**

**Crosby, C. (2019). *Toward a Framework for Artificially Intelligent Warfare: Measures of Effectiveness for Unsupervised Deep Learning in DoD Autonomous Systems*. Dissertation, Capitol Technology University.**

Artificially intelligent (AI) warfare improves decision speed, simplifies dynamic battle spaces, and enables previously impossible missions. As an offset, AI will increasingly determine who wins, and loses, future conflicts. Yet, even as the U.S. advances AI in support of its own warfighters and to address adversarial overmatch, the machine learning techniques and structures that enable AI remain underdeveloped by the Department of Defense (DoD). Specifically, there is no standardized decision framework around the highly inventive, operationally appealing technology called for by government. In response, this study presents an architecture for measuring the effectiveness (MOE) of DoD autonomous battlefield systems. The architecture is founded on three theorems and described using a qualitative diagram. Each theorem proves a specific relationship within the MOE doctrine while the diagram assesses the existence of an end state that enables artificially intelligent warfare through minimally viable, unsupervised, deep learning-enabled, autonomous, and battlefield ready DoD systems. In doing so, the research addresses, for the first time, how artificially intelligent technology on the battlefield is assessed. As a novel assessment framework, this study amalgamates DoD doctrine with academic literature, accelerates DoD AI initiatives, fills a void in the current military paradigm, and presents a path away from conceptual focus vectors and toward tangible courses of action.

**Ekelhof, M.A.C. (2019, December). *The Distributed Conduct of War: Reframing Debates on Autonomous Weapons, Human Control and Legal Compliance in Targeting*. [Doctoral Dissertation, Vrije Universiteit]. VU Research Portal.** [**https://research.vu.nl/ws/portalfiles/portal/90547665/complete+dissertation.pdf**](https://research.vu.nl/ws/portalfiles/portal/90547665/complete+dissertation.pdf).

In this dissertation I argue that, to gain a better understanding of what such a human control standard means in the context of military operations, we need to move beyond abstract theories about ‘autonomous weapons’ and ‘meaningful human control’ and focus on military conduct and decision-making from the perspective of actual practice. Although global discourses on policy and governance are typically infused with ambiguity, abstract concepts are of little use if they ignore the operational context that confronts the military in their application. Taking this into account, the objective of this dissertation is to examine the current state of human control in military practice, specifically the targeting process, to inform ongoing discussions about autonomous weapons, in particular in relation to meaningful human control and legal compliance.

**Hägg, J. (2021). *The contemporary challenges of drone warfare: A descriptive and critical analysis on the contemporary challenge of integrating just war theory with artificial intelligence in warfare.* [Bachelor’s Thesis, Mälardalen University]. Mälardalen University DiVA. URN:** [**urn:nbn:se:mdh:diva-54400**](https://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Amdh%3Adiva-54400)**.**

Due to the rapid technological advancements of the 21st century the fundamental nature of warfare has changed. Drones along with autonomous weapon systems has presented new challenges to the traditional concept and internal interpretations of just war theory. The purpose of this thesis is to present an analytical summary of the academic debate surrounding the emergence of AI technology, and how it has challenged the core principles embodied within jus in bello and jus ad bellum. Furthermore, the thesis explores the ethical issues external to just war theory principles, with a focus on how AI technology has established unique challenges for drone operators as a consequence of this contemporary phenomenon of war. This is done through a descriptive idea analysis and a critical analysis based on existing empirical material on the current academic debate on this issue.

While the advantages of drones and LAWS are evidently presented throughout this thesis, the repercussions are equally as important to contemplate. Thus, the findings in this thesis concludes that it is difficult to argue in favour or against the emergence of AI technology in war, as relevant arguments exist on both sides of the spectrum. However, the challenges for future just war theorists will be to adjust and reinterpret the moral foundations embodied within the principles of jus in bello and jus ad bellum to adhere to this contemporary phenomenon of war.

**Ijebor, C. (2020). *Artificially intelligent warfare and the revolution in military affairs.* [Master’s Thesis, University of Manitoba]. University of Manitoba MSpace.** [**https://mspace.lib.umanitoba.ca/server/api/core/bitstreams/0519fd15-e698-4c09-b08c-503e2999f1a3/content**](https://mspace.lib.umanitoba.ca/server/api/core/bitstreams/0519fd15-e698-4c09-b08c-503e2999f1a3/content)**.**

This study examines the use of artificial intelligence (AI) by the United States military as a case study. It explores the role of AI, first by looking into the concept of “Revolutions in Military Affairs” (RMA), studying AI’s defining characteristics and the previous RMA that have occurred in history, thus clarifying whether or not the integration of AI into the military operations can be classified as an RMA. The study then continues with an analysis of the two previous Offset Strategies adopted by the United States military, both of which resulted in an RMA and investigates whether the present Third Offset Strategy largely defined by AI can be seen as an RMA. In conclusion, the findings lead to the suggestion that, while AI has the potential to cause a revolution in military affairs and has caused many changes in the global approach to warfare, it is still in its infancy stage, and thus cannot be labelled as an RMA at this time.

**Kvasňovský, T. (2020). Autonomous Weapon Systems as the next revolution in warfare and implications of technology deployment for global security. Master’s Thesis, Charles University.** [**https://dspace.cuni.cz/bitstream/handle/20.500.11956/116314/120341140.pdf?sequence=1**](https://dspace.cuni.cz/bitstream/handle/20.500.11956/116314/120341140.pdf?sequence=1)

This thesis addresses developments in Artificial Intelligence and the increasing trend of robotization and autonomization of military forces in the context of Revolution in Military Affairs. It examines and categorizes different approaches to concepts of AI, autonomy and RMA in the public debate and academic and military literature. It further explores potential impacts and challenges of AI and its weaponized subset – Autonomous Weapon Systems on civil-military relations, legal and ethical norms, arms control regime and general security domain. Building upon findings from previous chapters, AI and AWSs are analyzed in a context of RMA and broader socio-economic context. Specifically, AI-enabled autonomy is compared with aspects of existing remotely controlled systems. The thesis comes to a conclusion that AWSs are harbingers of the next RMA and AI has the potential to match the importance of Neolithic, Industrial and Information revolution.

**Palmcrantz, C. (2019). When we see something that is well beyond our understanding: The duty of States to investigate war crimes and how it applies to autonomous weapons systems. Thesis. Bachelor thesis, Swedish Defence University.** [**http://fhs.diva-portal.org/smash/record.jsf?aq2=%5B%5B%5D%5D&c=29&af=%5B%5D&searchType=LIST\_LATEST&sortOrder2=title\_sort\_asc&query=&language=sv&pid=diva2%3A1277021&aq=%5B%5B%5D%5D&sf=all&aqe=%5B%5D&sortOrder=author\_sort\_asc&onlyFullText=false&noOfRows=50&dswid=3866**](http://fhs.diva-portal.org/smash/record.jsf?aq2=%5B%5B%5D%5D&c=29&af=%5B%5D&searchType=LIST_LATEST&sortOrder2=title_sort_asc&query=&language=sv&pid=diva2%3A1277021&aq=%5B%5B%5D%5D&sf=all&aqe=%5B%5D&sortOrder=author_sort_asc&onlyFullText=false&noOfRows=50&dswid=3866)**.**

This thesis analyses States’ duty to investigate grave breaches of humanitarian law and how it applies to deep reinforcement learning autonomous weapons. It identifies basic technologic intricacies related to deep reinforcement learning and discusses what issues may arise if such software is used in weapons systems. The thesis applies a legal doctrinal method to study how the technology could frustrate the grave breaches regime and hamper States’ ability to investigate suspected incidents. Furthermore, investigative standards under humanitarian law and human rights law are examined in the context of autonomous weapons systems. The main argument is that deep reinforcement learning algorithms create a black box that is virtually impossible to investigate and consequently causes accountability issues.

**Pascual García, R. (2024). The impact of artificial intelligence in armed conflicts and International Humanitarian Law. [Master’s Thesis, Universidad Pública de Navarra]. Universidad Pública de Navarra D-Space.**

[**https://academica-e.unavarra.es/server/api/core/bitstreams/cda0feb9-2ebc-4310-a083-823fc9e6627f/content**](https://academica-e.unavarra.es/server/api/core/bitstreams/cda0feb9-2ebc-4310-a083-823fc9e6627f/content)**.**

Cybersecurity has become an increasingly threat for the world we know during the last decades. Individuals, states, and companies can be victims of a cyber-attack, but also parts to an armed conflict, challenging the rules of the game: International Humanitarian Law. This work aims to define how those rules of the game function and how cyberoperations are being employed by hackers. The applicability of International Humanitarian Law in case of cyberattacks will be examined in terms of territorial demarcation, agents involved in and type of violence employed. Traceability and decentralization of the attackers has proven to be one of the hardest challenges for IHL to persecute cyberwarfare, and must be addressed by a multidisciplinary approach, including cybersecurity experts on the traceability of the attack and law professionals on the problem of attribution.

**Tunysová, A. (2022). *Predicting and Preventing Terrorism with Artificial Intelligence and Machine Learning: Implications for Security in Israel*. [Diploma thesis, Prague: Charles University]. Charles University Digital Repository.** [**https://dspace.cuni.cz/handle/20.500.11956/174465**](https://dspace.cuni.cz/handle/20.500.11956/174465)**.**

The thesis examines the use of Artificial Intelligence and Machine Learning for predicting and preventing terrorism and the resulting security risks. At the conceptual level, the thesis examines the approach of predicting threats with a focus on predictive policing and presents risks associated with the use of predictive machine learning systems, which are then discussed within the context of counterterrorism. The paper aims to answer the question to which extent we can rely on machine learning systems used to predict and prevent terrorism and what are the implications of their use for security in Israel. The thesis points out that although the predictive tools seem to be faster and more precise than human analysts, they cannot be trusted to a full extent. If the results of these systems are used to employ strict measures such as the restriction of a suspect’s liberty, it may lead to the violation of human rights. Therefore, in the case of counterterrorism in Israel, which is sometimes presented as the only democracy in the Middle East, it is necessary to bear in mind the risks associated with the limits of predictive machine learning systems together with the up-to-date practice of Israeli security agencies and Israeli historical-social context, indicating that it would be very difficult, if not impossible, objectively and precisely predict terrorist activities and identify terrorists before they commit an attack.

**Wilson, N. A. (2020). *Understanding the Battle for AI in Warfare through the Practices of Assemblage: A Case Study of Project Maven* Master's thesis, Utrecht University.** [**https://studenttheses.uu.nl/handle/20.500.12932/37392**](https://studenttheses.uu.nl/handle/20.500.12932/37392)

Artificial Intelligence (AI) has been widely-heralded as a revolutionary technology across a range of domains, including defence. In this sphere, its potential for automating low-cost and low-risk forms of warfighting known as “remote warfare” has given rise to fears of Lethal Autonomous Weapons Systems (LAWS). While the use of LAWS has been subject to intense speculation, far less attention has been paid to their development. Notably, cutting-edge AI innovation is not to be found in government-funded research labs or traditional defence contractors, but consumer technology companies. Because of this, unique controversies have arisen over its development. In the case of Project Maven, a US Department of Defense initiative seeking to leverage AI for automating drone footage analysis, employee protests eventually forced one of its contractors, Google, to end their involvement. Using Project Maven as a case study, this research therefore seeks to understand the development of an emerging Military-Technological Complex developing these technologies and an oppositional Civil Society Coalition seeking to regulate them. Based on documentary analysis of key texts and five semi-structured expert interviews, this thesis uses an assemblage approach to examine the discourses surrounding Project Maven, the interplay of power between its elements, and their resulting configurations. It finds a tension between the compulsory powers of the Military-Technological Complex, exercised through structural, material relations, and the productive powers of Civil Society, exercised through the production and sanctioning of knowledge. It concludes that the development of consumer technologies for warfare and the associated emergence of a Military-Technological Complex reflects a broader unravelling of conventional ties between war, space and time.

## **Projects and Campaigns**

**Berkeley Human Rights Center: Technology, Law & Policy.**

***Accountability for Cyber-Enabled International Crimes.*** [**https://humanrights.berkeley.edu/projects/accountability-for-cyber-enabled-international-crimes/**](https://humanrights.berkeley.edu/projects/accountability-for-cyber-enabled-international-crimes/)**.**

Since before Russia’s full scale invasion of Ukraine in 2022, the TLP team has been documenting and investigating Russian cyber-attacks targeting Ukraine’s critical civilian infrastructure. After thorough legal research and an examination of the evidence and the facts, the team determined that several of these cyber-attacks could be charged as war crimes pursuant to Article 8 of the Rome Statute. In March 2022 and March 2023, we submitted Article 15 communications to the International Criminal Court (ICC) Office of the Prosecutor (OTP), identifying five key cyber-attacks that we believe could be charged as war crimes in the current investigation in Ukraine. Since then, we have received positive feedback from the ICC Prosecutor, who announced his intention to investigate cyber-enabled crimes that could fall within the jurisdiction of the Court in all situations. To this end, the OTP is drafting its first cyber policy.

We have also shared our confidential submissions with the Office of the Prosecutor General of Ukraine and provided advice to Ukrainian prosecutors on how such acts could be charged as war crimes under the Ukrainian Criminal Code. Given the unprecedented international cooperation for accountability in Ukraine, we have also shared our proposed case with the national war crimes units of other states to generate interest and support.

Cyber-enabled international crimes present novel legal issues that have yet to be adjudicated in court. As such, we are also actively engaged in legal analysis and advocacy on how international humanitarian law and international criminal law should be applied to cyberspace for the maximal protection of civilians and civilian objects.

***Preserving Social Media Evidence of Mass Atrocities.*** [**https://humanrights.berkeley.edu/projects/preserving-social-media-evidence-of-mass-atrocities/**](https://humanrights.berkeley.edu/projects/preserving-social-media-evidence-of-mass-atrocities/)**.**

Social media platforms run by private companies have become the de facto record-holders of atrocity documentation. They are the main means by which witnesses, journalists, citizen reporters and first responders share their observations, videos and photographs with the international community. They are also the channel through which perpetrators occasionally disclose information about their illegal activities (for recruitment, intimidation, or other purposes), and they are increasingly used as a tool to spread disinformation, distribute propaganda and incite violence. These platforms have become our historical records and evidence archives, even though they do not have the incentive or infrastructure to properly and indefinitely preserve this content, nor the power to proactively share it with the appropriate authorities.

In fact, many of these companies are incentivised to do the opposite. The rate at which potentially relevant and probative content is removed is rapidly increasing as political and public pressure to remove harmful content mounts against them. Private technology companies should not, nor do they want to be, the custodians of evidence of international crimes. At the same time, most international justice and accountability mechanisms lack the investigative authority to compel these companies to preserve social media content—even when its evidentiary value is obvious—and disclose it. Without full and fast-responding cooperation from state authorities in the country in which these companies reside, valuable digital evidence and historical records are permanently destroyed. International justice and accountability mechanisms should be empowered to preserve and compel evidence of war crimes, crimes against humanity, and genocide.

Building on the Human Rights Center’s work on digital open source investigations, and the [Oxford IPS’s work on improving investigations and documentation of atrocity crimes](https://www.bsg.ox.ac.uk/research/mass-atrocities-digital-age-preserving-social-media-evidence), this project:

* Addresses the relevance and probative value of user-generated content to international legal proceedings;
* Describes the significance of preserving the integrity of user-generated content in a forensic manner so that it can be used in legal proceedings, and the consequences of failing to do so; and
* Clarifies the requirements for and purposes of sharing relevant user-generated content with international and regional legal authorities.
* Looks at how rules around information-sharing and disclosure differ for technology companies based on whether the request is made by a domestic or international legal authority.
* Assesses whether an international investigative mandate could be vested with the power to compel content directly from technology companies through an innovative legal mechanism.

***Developing the Berkeley Protocol.*** [**https://humanrights.berkeley.edu/projects/developing-the-berkeley-protocol-on-digital-open-source-investigations/**](https://humanrights.berkeley.edu/projects/developing-the-berkeley-protocol-on-digital-open-source-investigations/)**.**

The Human Rights Center spearheaded a collaborative effort in partnership with the United Nations Office of the High Commissioner for Human Rights (OHCHR) to develop an international protocol outlining the minimum professional standards for the identification, collection, preservation, verification, and analysis of digital open source information, with an aim toward improving its effective use in international criminal, humanitarian and human rights investigations. Strengthening legal and ethical norms for open source investigations professionalizes the field and, in doing so, increases the likelihood that such information will be useful for justice and accountability purposes. [The Berkeley Protocol has been translated into all the languages of the United Nations for global impact.](https://humanrights.berkeley.edu/publications/berkeley-protocol-on-digital-open-source-investigations/)

The Berkeley Protocol is actively in use by Ukranian prosecutors documenting Russian war crimes, and was translated informally into Ukrainian within weeks of the invasion. [We have also developed practical guide to implementing the Protocol for lawyers and judges](https://humanrights.berkeley.edu/publications/evaluating-digital-open-source-imagery-a-guide-for-judges-and-fact-finders/) in partnership with Essex University, WITNESS, Human Rights Watch, and Mnemonic to further assist legal practitioners in the pursuit of justice. The Protocol is also the foundation of our [Professional Training Program.](https://humanrights.berkeley.edu/professional-trainings/)

**Forensic Architecture.** [**https://forensic-architecture.org/**](https://forensic-architecture.org/)**.**

Forensic Architecture (FA) is a research agency based at Goldsmiths, University of London. Our mandate is to develop, employ, and disseminate new techniques, methods, and concepts for investigating state and corporate violence. Our team includes architects, software developers, filmmakers, investigative journalists, scientists, and lawyers.

We are an interdisciplinary agency operating across human rights, journalism, architecture, art and aesthetics, academia and the law. In 2022, the Peabody Awards programme wrote that we had co-created ‘an entire new academic field and emergent media practice’; in 2024, the European Research Council assessed Forensic Architecture as ‘a scientific breakthrough (defined as a revolutionary work that led to deep change in existing paradigms or new methods opening a new stream of research)’.

Since 2020, FA has supported the growth of agencies worldwide that practice and apply our methods. [The Investigative Commons](https://investigative-commons.org/) is both a global network of practitioners, and a physical space in Berlin, within the offices of our sister agency [Forensis](https://counter-investigations.org/).

# **Global Commission on Responsible Artificial Intelligence in the Military Domain (GC REAIM).** [**https://hcss.nl/gcreaim/**](https://hcss.nl/gcreaim/)

GC REAIM has been established, for an initial period of two years, to help promote mutual awareness and understanding among the many communities working on issues related to the global governance of AI in the military domain. By linking dialogues between these communities, the Global Commission will contribute to an essential global task: supporting fundamental norm development and policy coherence in this field.

**Group on Earth Observations. *About GEOSS*.** [**https://old.earthobservations.org/geoss.php**](https://old.earthobservations.org/geoss.php)

A central part of GEO’s Mission is to build the Global Earth Observation System of Systems (GEOSS). GEOSS is a set of coordinated, independent Earth observation, information and processing systems that interact and provide access to diverse information for a broad range of users in both public and private sectors. GEOSS links these systems to strengthen the monitoring of the state of the Earth. It facilitates the sharing of environmental data and information collected from the large array of observing systems contributed by countries and organizations within GEO. Further, GEOSS ensures that these data are accessible, of identified quality and provenance, and interoperable to support the development of tools and the delivery of information services. Thus, GEOSS increases our understanding of Earth processes and enhances predictive capabilities that underpin sound decision-making: it provides access to data, information and knowledge to a wide variety of users.

**Harvard Law School Program on International Law and Conflict. *War-Algorithm Accountability.*** [**https://pilac.law.harvard.edu/waa**](https://pilac.law.harvard.edu/waa)

In *War-Algorithm Accountability* (August 2016), we introduce a new concept—war algorithms—that elevates algorithmically-derived “choices” and “decisions” to a, and perhaps *the,* central concern regarding technical autonomy in war. We thereby aim to shed light on and recast the discussion regarding “autonomous weapon systems” (AWS).

We define “war algorithm” as any algorithm that is expressed in computer code, that is effectuated through a constructed system, and that is capable of operating in relation to armed conflict. In introducing this concept, our foundational technological concern is the capability of a constructed system, without further human intervention, to help make and effectuate a “decision” or “choice” of a war algorithm. Distilled, the two core ingredients are an algorithm expressed in computer code and a suitably capable constructed system.

Through that lens, we link international law and related accountability architectures to relevant technologies. We sketch a three-part (non-exhaustive) approach that highlights traditional and unconventional accountability avenues. We focus largely on international law because it is the only normative regime that purports—in key respects but with important caveats—to be both universal and uniform. In this way, international law is different from the myriad domestic legal systems, administrative rules, or industry codes that govern the development and use of technology in all other spheres. By not limiting our inquiry only to weapon systems, we take an expansive view, showing how the broad concept of war algorithms might be susceptible to regulation—and how those algorithms might already fit within the existing regulatory system established by international law.

**No Tech for Apartheid.** [**https://www.notechforapartheid.com/**](https://www.notechforapartheid.com/)

No Tech For Apartheid is a worker-led campaign of Google and Amazon workers organizing against the companies' $1B Project Nimbus cloud computing contract with the Israeli government and military.

**Responsible AI in the Military Domain Summit.** [**https://www.reaim2024.kr/reaimeng/index.do**](https://www.reaim2024.kr/reaimeng/index.do)

REAIM is a platform for global discussion with all stakeholders on ways to ensure responsible military application of AI. The inaugural REAIM Summit was co-hosted by The Netherlands and The Republic of Korea(ROK), in The Hague, in February 2023. The REAIM Summit 2024 will be co-hosted by The Republic of Korea(ROK), The Kingdom of the Netherlands, The Republic of Singapore, The Republic of Kenya, The United Kingdom of Great Britain and Northern Ireland in Seoul, in September 2024.

**Stop Killer Robots.** [**https://www.stopkillerrobots.org/**](https://www.stopkillerrobots.org/)

With growing digital dehumanisation, the Stop Killer Robots coalition works to ensure human control in the use of force. Our campaign calls for new international law on autonomy in weapons systems.

Formed in October 2012 and publicly launched in 2013, we operate globally with 250+ member organisations. A united voice with national, regional and international affect. We are a vibrant, inclusive team driving change to protect our shared humanity for the future.

**UNIDIR. *Artificial Intelligence.*** [***https://unidir.org/focus-area/artificial-intelligence/***](https://unidir.org/focus-area/artificial-intelligence/)

Artificial intelligence will revolutionize how militaries operate and how future wars will be fought. Competitive pressures in today’s fraught global security environment will only accelerate this trend. With AI technology advancing at scale, the time to act is now.

AI will help to increase the speed and accuracy of military decision-making, planning and operations. In combination with robotic platforms and next-generation sensor technology, it will open up a vast spectrum of military applications, from AI-enabled logistical support, early warning systems, and intelligence gathering to AI-supported command structures, cyber operations and autonomous weapons systems.

An enabling technology as powerful and transformative as AI comes with a myriad of benefits but also with a long list of significant risks, not least inadvertent escalation, misperception and malfunction, and concerns about lack of transparency, discrimination and bias. Crucially, however, military AI also raises profound ethical and legal questions about human agency and human control, especially in matters of life and death. UN Secretary-General António Guterres has unequivocally stated that “human agency must be preserved at all cost”.

UNIDIR leverages its unique position at the nexus of security, technology and multilateral diplomacy to increase the understanding of the risks and implications of AI for international peace and security and works to support the development of tailored governance responses. Our [AI Policy Portal](https://aipolicyportal.org/) supports capacity and confidence-building by enabling transparency and information-sharing.

**UNODA. *Resources.*** [**https://disarmament.unoda.org/resources-2/**](https://disarmament.unoda.org/resources-2/)

This section offers an information-sharing platform where various types of resources are posted. Requests for consideration to add links to this webpage should be made to [ccw@un.org](mailto:ccw@un.org).

**WeRobotics.** [**https://werobotics.org**](https://werobotics.org)

WeRobotics supports and connects local experts in drone, data and AI technologies. By doing so, we co-create a proven and replicable model to successfully localize and shift power at a grassroots level.