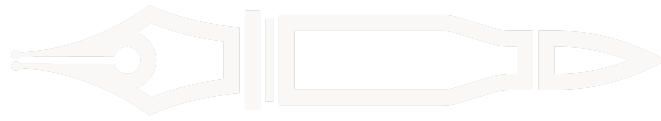




LIEBER INSTITUTE  
WEST POINT



# ARTICLES OF WAR



# RESPONSIBLE AI SYMPOSIUM - THE NEXUS BETWEEN RESPONSIBLE MILITARY AI AND INTERNATIONAL LAW

by Tobias Vestner | Nov 17, 2022



*Editor's note: The following post highlights a subject addressed at an expert workshop conducted by the Geneva Centre for Security Policy focusing on Responsible AI. For a general introduction to this symposium, see Tobias Vestner's and Professor Sean Watts's [introductory post](#).*

---

States and international organizations are increasingly adopting new principles on the responsible military and defense-related use of artificial intelligence (AI). While most take the form of policies, States and international organizations tend to call them principles to reflect that they are a set of guiding criteria. Often, they are referred to as principles on responsible AI (RAI), a term that originates from ethical guidance on the civilian development and use of AI.

This post explores the fundamental tenets of the nexus between international law and the principles on the responsible development, deployment, and use of AI for military and defense purposes. It first identifies the context, substance, and form of the RAI principles. The post then discusses how they relate to and potentially shape international law now and in the future. It commends such principles as



building blocks for further international legal developments and calls for their further refinement by States.

## RAI Principles

The **United States** adopted principles on RAI for military and defense purposes in 2020. Other States, including **France** in 2019, **Australia** in 2020, and the **United Kingdom** and **Switzerland** in 2022 issued principles, ethical guidelines, or elements thereof. NATO adopted its **Principles of Responsible Use of Artificial Intelligence in Defence** as part of its AI Strategy and the **European Union** established relevant guidance in 2021. Similarly, in 2019, the parties to the Convention on Certain Conventional Weapons (CCW) endorsed the **Guiding Principles Affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems**.

At the outset, the various principles on military and defense-related RAI are relatively homogeneous. States' and international organizations' RAI principles typically contain elements comparable to what NATO calls "Responsibility and Accountability," "Explainability and Traceability," "Reliability," "Governability," and "Bias Mitigation." The principles and their elements may be called differently but tend to consist of similar substance. Some States have additional criteria. Switzerland, for instance, includes "Agility" as distinct principle.

Interestingly, NATO includes the unique principle of "Lawfulness" in addition to the typical principles. This principle states: "AI applications will be developed and used in accordance with national and international law, including international humanitarian law and human rights law, as applicable."

This goes to the heart of the question of the RAI principles' link with international law. Indeed, States' and international organizations' principles do not represent international law. They are first and foremost national and institutional policies. Yet the various RAI

principles are inevitably part of the broader normative framework of international law. As such, there are six legal touchpoints.

### **Compliance with International Law**

A first legal touchpoint, as NATO's principle on lawfulness indicates, is that RAI principles must be in accordance with international rules. The principles cannot contradict States' (or, to the extent possible, international organizations') rights and obligations because international law is **legally binding** on its subjects. An assessment of the current RAI principles reveals that they do not contain elements that would contradict international law.

The requirement to comply with international law further implies that States' and international organizations' measures to implement and operationalize their RAI principles must also be in compliance with international law. The principles of accountability and transparency, for instance, may require the gathering and sharing of information. If a State has adhered to certain human rights obligations related to **data protection**, for instance, it must respect these rules even when its RAI principles do not explicitly refer to them. This means that the complexity of compliance is less at the level of the principles themselves, but rather at the level of their implementation and operationalization.

### **Application of International Law**

A second touchpoint is that the RAI principles help to apply international law to AI applications. Because international legal rules are often general and abstract, i.e. cover a large spectrum of actors and situations, the RAI principles guide States on how to apply the existing rules specifically to military and defense-related applications of AI. This is the case both where the rules are relatively clear and well-established



and where the law is silent, unclear, or the meaning of the law is debated.

The principle of accountability, for example, specifies that States should integrate liability and enforcement into the development and use of AI to enable the functioning of **international criminal law** and **military justice**. Similarly, the principle of explainability and traceability defines that transparency must be addressed in the application of AI to ensure responsibility and justice differently than with respect to traditional hardware and software.

The principle of governability, on the other hand, affirms that appropriate levels of human judgement or meaningful human control is essential in the context of AI. The Law of Armed Conflict (LOAC) does not explicitly state such a requirement because it was codified before the emergence of AI technology. But, this principle highlights the **issue** as well as indicates that States take it seriously and are invested to ensure this.

### **Implementation of International Law**

A further touchpoint is that the RAI principles support the implementation of international law. This notably applies to the principles that refer to technical preconditions for the effective and controlled deployment of AI. The principles of safety and reliability are the most obvious examples. An AI system without these features, i.e. a system that does not operate as generally **intended** or that can easily be interfered with, risks violating LOAC and other international rules.

Yet, due to their generality, the RAI principles are not granular enough to serve as proper implementation measures by themselves. They do, however, set the overall requirements for compliance with international law and guide its implementation by more specific standards and procedures. In this regard, the U.S. Department of Defense recently adopted its **Responsible Artificial Intelligence Strategy and**



**Implementation Pathway.** NATO is also working on **standards and procedures** based on its own principles. It recently adopted an **Autonomy Implementation Plan** and will establish a **Data and AI Review Board**.

### **Complement to International Law**

A fourth touchpoint refers to the possibility that the RAI principles, just like any policy, can complement legal requirements by going beyond what is legally required. Indeed, many principles refer to the notion of “ethical” use of AI or have resulted from reflections on its ethical use. This indicates the **moral** component that can overlap with ethical considerations enshrined in certain international rules, but which can also be distinct.

Views on the legally required degree of human judgement and control over AI systems **vary**, for instance. Even if a State holds the legal view that LOAC does not require stringent human control of AI, it can fix higher standards in its RAI principles for ethical reasons. This may include the prohibition of certain actions unless operators are in the decision-making loop. It is noteworthy, however, that States can unilaterally change national policies. NATO’s principles are harder to change because the North Atlantic Council adopted them by **consensus**. Yet they are nonetheless easier to change than international rules.

### **Clarification of International Law**

A further touchpoint is that the RAI principles help to clarify international law, at least in part, by indicating State practice. States’ principles do not communicate *opinio juris* or legal positions *per se* because they are simply policies which do not reflect States’ beliefs that they have international legal force. Yet a State’s principles may indicate that it values these criteria, from which can be deduced the State’s general preferences in terms of what the international rules could or should be.



More importantly, by the adoption of their RAI principles, States commit themselves to act on them. This indicates and leads to emerging State practice. In the context of NATO, this applies even to States that have not yet adopted any national RAI principles. Moreover, NATO institutionally coordinates such State practice. This involves seeking and managing the **interoperability** of related frameworks and mechanisms for the use of AI, which tends to support the coherence of State practice.

### Shaping International Law

The latter point on State practice leads to the question of whether the emerging RAI principles directly shape international law. If States indicate that their emerging practice reflects a legal obligation (*opinio juris*), this can lead to the emergence of customary rules specifically related to the military and defense-related development, deployment, and use of AI. For now, both State practice and related *opinio juris* have not crystallized.

Another outlook might conclude that the RAI principles adopted by intergovernmental institutions represent soft law. In the case of the principles adopted by NATO and the Group of Governmental Experts on lethal autonomous weapons systems, they do result from some degree of specified, mutual commitment among States. Yet, since soft law is neither clearly defined nor accepted as a **source** of international law, it makes more sense to simply treat these principles as politically, rather than legally, binding commitments among States. Indeed, States that have agreed to adopt these principles did not do so with the aim of creating new legal obligations.

### Conclusion

In sum, there are several angles by which the principles on RAI link to or interrelate with international law. In general, they must comply with States' legal obligations and support the application and implementation of international law for military and defense-related AI.



They may also complement international legal obligations and help to clarify existing rules. In the medium to long term, the various RAI principles may indicate and lead to the emergence of both State practice, and if coupled with *opinio juris*, to customary rules.

To what extent these principles interrelate with international rules more specifically deserves further analysis. This certainly applies to LOAC, which arguably is the most contentious branch of international law in the context of the military use of AI, because it governs the conduct of hostilities in armed conflicts. This also applies to international human rights law, fundamental rights, and the rule of law, notably because States will use AI applications in peace time in the context of information gathering and other types of operations both abroad and domestically. New legal developments related to this field, such as by the [Council of Europe](#), the [European Union](#), or the [United States](#), may affect States' operationalization of the RAI principles. Yet the principles may also apply and lead to other legal issues in international security relating to *jus ad bellum* and to [State responsibility](#).

\*\*\*

*Tobias Vestner is Head of the Research and Policy Advice Department and Head of the Security and Law Programme at the Geneva Centre for Security Policy (GCSP)*

Photo credit: Pexels

## SUBSCRIBE

### RELATED POSTS

#### [Introduction](#)

by [Tobias Vestner](#), [Sean Watts](#)

November 17, 2022







## LIEBER INSTITUTE

MADN-LAW

United States Military Academy

646 Swift Road

West Point, NY 10996

+1-845-938-2572

### Articles of War Disclaimer

*The views expressed are those of the authors, and do not necessarily reflect the official position of the United States Military Academy, Department of the Army, or Department of Defense.*

*Articles of War is a forum for professionals to share opinions and cultivate ideas. Articles of War does not screen articles to fit a particular editorial agenda, nor endorse or advocate material that is published.*



## CONNECT



## EMAIL