NATO OTAN NORTH ATLANTIC TREATY ORGANIZATION

Summary of the NATO Artificial Intelligence Strategy

22 Oct. 2021 - | Last updated: 22 Oct. 2021 13:43

1. Artificial Intelligence (AI) is changing the global defence and security environment. It offers an unprecedented opportunity to strengthen our technological edge but will also escalate the speed of the threats we face. This foundational technology will likely affect the full spectrum of activities undertaken by the Alliance in support of its three core tasks; collective defence, crisis management, and cooperative security.
2. In order to maintain NATO's technological edge, we commit to collaboration and cooperation among Allies on any matters relating to AI for transatlantic defence and security. NATO and Allies can help accelerate these efforts by building on the existing adoption efforts of several NATO and Allied bodies.
3. The aim of this Strategy is fourfold:

- To provide a foundation for NATO and Allies to lead by example and encourage the development and use of AI in a responsible manner for Allied defence and security purposes;
- To accelerate and mainstream AI adoption in capability development and delivery, enhancing interoperability within the Alliance, including through proposals for AI Use Cases, new structures, and new programmes;
- To protect and monitor our AI technologies and ability to innovate, addressing security policy considerations such as the operationalisation of our Principles of Responsible Use; and
- To identify and safeguard against the threats from malicious use of AI by state and non-state actors.

4. In the future, the NATO Alliance aims to integrate AI in an interoperable way to support its three core tasks. Such use will be conducted in a recognised, responsible fashion across the enterprise, mission support and operational levels in accordance with international law. Recognising the leading role of the civil private sector and academia in AI development, this Strategy will be underpinned by: significant cooperation between NATO, the private sector[1] and academia; a capable workforce of NATO technical and policy-based AI talent; a robust, relevant, secure data infrastructure; and appropriate cyber defences.
5. Under the forthcoming Defence Innovation Accelerator for the North Atlantic (DIANA), national AI test centres could support NATO's AI ambition. NATO and Allies will also conduct regular high-level dialogues, engaging technology companies at a strategic political level to be informed and help shape the development of AI-fielded technologies, creating a common understanding of the opportunities and risks arising from AI.
6. Furthermore, NATO will remain the transatlantic forum for AI in defence and security, leveraging the potential of AI while safeguarding against its (AI) malicious use by state and non-state actors.

PRINCIPLES OF RESPONSIBLE USE

7. At the forefront of this Strategy lie the NATO Principles of Responsible Use for AI in Defence, which will help steer our transatlantic efforts in accordance with our values, norms, and international law. The NATO Principles of Responsible Use (the Principles) are based on existing and widely accepted ethical, legal, and policy commitments under which NATO has historically operated and will continue to operate under. These Principles do not affect or supersede existing obligations and commitments, both national and international.
8. The Principles below were developed based on Allied approaches and relevant work in applicable international fora. These Principles apply across all types of AI applications. They are aimed at providing coherence for both NATO and

Allies to enable interoperability. The Principles will be foundational to the discussion and adoption of more detailed AI best practices and should be considered a baseline for Allies as they use AI in the context of defence and security, noting that some Allies already have national principles of responsible use.

NATO Principles of Responsible Use of Artificial Intelligence in Defence

9. Allies and NATO commit to ensuring that the AI applications they develop and consider for deployment will be – at the various stages of their lifecycles – in accordance with the following six principles:

A. **Lawfulness:** AI applications will be developed and used in accordance with national and international law, including international humanitarian law and human rights law, as applicable.
B. **Responsibility and Accountability:** AI applications will be developed and used with appropriate levels of judgment and care; clear human responsibility shall apply in order to ensure accountability.
C. **Explainability and Traceability:** AI applications will be appropriately understandable and transparent, including through the use of review methodologies, sources, and procedures. This includes verification, assessment and validation mechanisms at either a NATO and/or national level.
D. **Reliability:** AI applications will have explicit, well-defined use cases. The safety, security, and robustness of such capabilities will be subject to testing and assurance within those use cases across their entire life cycle, including through established NATO and/or national certification procedures.
E. **Governability:** AI applications will be developed and used according to their intended functions and will allow for: appropriate human-machine interaction; the ability to detect and avoid unintended consequences; and the ability to take steps, such as disengagement or deactivation of systems, when such systems demonstrate unintended behaviour.
F. **Bias Mitigation:** Proactive steps will be taken to minimise any unintended bias in the development and use of AI applications and in data sets.

Ensuring the Safe and Responsible Use of Allied AI

10. To ensure the safe and responsible use of Allied AI, NATO will operationalise its Principles of Responsible Use. These Principles will apply across the lifecycle of an AI capability. Allies and NATO will therefore operationalise these principles across all lines of development.
11. To further inform this work, the NATO AI Test Centre(s) will develop best practices for Allies, which will include assisting overall interoperability and information security efforts.
12. Underpinning the safe and responsible use of AI, NATO and Allies will consciously put Bias Mitigation efforts into practice. This will seek to minimise those biases against individual traits, such as gender, ethnicity or personal attributes.
13. NATO will conduct appropriate risk and/or impact assessments prior to deploying AI capabilities.

Minimising Interference in Allied AI

14. Some state and non-state actors will likely seek to exploit defects or limitations within our AI technologies. Allies and NATO must strive to protect the use of AI from such interference, manipulation, or sabotage**,** in line with the *Reliability* Principle of Responsible Use, also leveraging AI-enabled Cyber Defence applications.
15. Allies and NATO should develop adequate security certification requirements for AI, such as specific threat analysis frameworks and tailored security audits for purposes of 'stress-testing'.
16. AI can impact critical infrastructure, capabilities and civil preparedness—including those covered by NATO's seven resilience Baseline Requirements—creating potential vulnerabilities, such as cyberspace, that could be exploited by certain state and non-state actors.
17. Some state and non-state actors may also leverage disinformation opportunities within Allied societies by creating public distrust of the military use of AI. Allies will seek to prevent and counter any such efforts within the context of the Principles of Responsible Use and utilise strategic communications, where appropriate. NATO will support Allies as required.

Standards

18. NATO will further work with relevant international AI standards setting bodies to help foster military-civil standards coherence with regards to AI standards.

[1] For purposes of the Strategy ‘private sector’ includes Big Tech, start-ups, entrepreneurs and SMEs as well as risk capital (such as venture and private equity funds).