Trustworthiness for AI in Defence (TAID)

-

Use Cases for AI in Defence Annex

Submission date: 09/05/2025

Document version: 1.0

Authors: TAID Working Group (TAID WG)

Table of Contents

List of Tables

1. Introduction

This Annex includes the list of Use Cases integrating AI for Defence, as shown in **Error! Reference source not found.**, and a detailed description of UC-01 to UC-09.

Table 1- List of AI Use Cases for Defence

ID	Title	Actors	Systems of interest	Level
UC01	Decision-making in multi domain operations	C4I operators, AI subject matter experts	C4I system in multi domain operations	Tactical
UC02	Failure of a decision support system	Commander interacting people, Commander training team, authorities	Sensor Mesh (passive/active radar, EO sensors, audio sensors) to Optimize Situational Awareness	Tactical
UC03	Collision avoidance/swarming of drones/emergence (tactics to neutralise targets)	Remote pilots, pilots in the surroundings	Detect & Avoid system	Tactical
UC04	Mission training	Military commanders and other military personnel	Combat Training System	Operational
UC05	Aerial refuelling	aerial refuelling operator	aerial refuelling system	Tactical
UC06	Data-centric security	Both human and non-human annotators as well as cross-domain solutions (information processors/guards)	Security domains and information processors	Strategical
UC07	Military Approval/Certification	Manufacturers, Data providers, Air traffic controllers, Approval/Certification Authorities	Equipment being used for both military and civil applications	Strategical
UC08	Meaningful Human Control	Drone operator, System designer	targeting and decision making on Firing (Autonomous Weapon System)	Tactical
UC09	Active Autonomous Cyber Defence	System developers, Cyberoperator, system owner	A cyber security system	Operational

2. UC01 – Decision-Making in Multi-Domain Operations

Overall Description

Command, control, communications, computers, and intelligence (C4I) systems enable effective military awareness, decision-making and operation. By integrating also surveillance and reconnaissance (C4ISR) capabilities, intelligence analysis benefits of additional information regarding adversary assets and capabilities, in peacetime and conflict.
AI can enhance the decision-making process both at strategic, operational, and tactical operation level.

This Use Case proposes AI to support decision-making in a tactical operation scenario where multiple threats participate with different impacts and speeds, which require a quick response using effective countermeasures.

AI tools find a promising application to:

- Data collection, correlation, and fusion from multiple platforms/sensors/probes in single or multi-domain environment (land, maritime, air, space, cyber);
- Data extraction providing the information of interest (e.g. adversary assets), in particular when the amount of data collected is huge, for tactical operation applications and for intelligence analysis purposes;
- Threat evaluation and weapon assignment, assisting the decision-making process to select the best defence resources against the threat in that timeframe.

Identified Impacts

- **Impact on the system:** These applications can have a positive impact on system performance (TAID-02:O, TAID-21:O, PER-03:O) and trustworthiness of AI (ASDC-01:O). AI technology has the potential to improve future C4I/C4ISR providing faster decisions, reducing the system response time (PER-01:O). However, the resources required for designing and developing such system and also for updating could be onerous (DCLC-02:R, DCLC-05:R). In addition, the reliability of the system needs to be ensured to prevent failure during mission (TAID-22:R).

- **Impact on the mission:** Military operations/specifics are impacted by improvement of goals achievement, mission situation assessment, reaction time and mission efficiency (OODA loop) (MOP-02:O, MOP-04:O, MOP-07:O). A better army management and allies' coordination can lead to a gain in supremacy vs. opponents at field (MSP-04:O, MSP-05:O). The missions can be improved by better assessment of own and opponents' intents, nonetheless they can be also compromised in case of system breach (MSP-06:N MSP-07:N).

- **Impact on the operator:** The fact that AI can handle a huge amount of data coming from a variety of sensors, platforms, systems operating in different domains, decreases human operator effort (MOP-06:O). However, challenges arise in the management and control of AI capabilities (TAID-11:R, HF-06:R) when ensuring that they are effective and up to date according to the changing needs typical of a distributed, evolving and contested environment (HUV-02:R, DCLC-02:R, DCLC-05:R). Relevant aspects also impacting the operator's activities are the need for having a clear and human-interpretable view of the system status provided by the AI models, so the decisions to be taken are well understood by the operator and the entailed consequences bearable (HF-08:R, TAID-12:R).

Involved Stakeholders:

According to the approach presented in Section 2.4.1 of TAID White Paper, this Use Case is on a **tactical level**. Therefore, the expected scenario takes place on a **short** time scale. Accordingly, the following main stakeholders are involved in this Use Case:

- AI customers
  - Legal responsibility
  - Technical responsibility
  - AI operators
    - Operator
    - Operator team members
    - Operators interacting with AI system
    - C4I operators
    - Commander training team
- AI partners
  - AI trainers
  - AI subject matter experts
- AI providers
  - Software platform providers
  - AI product providers
- AI producers
  - AI developers
- AI authorities
  - Executive power (Chain of command)

Applicable standardization/regulation documents for this scenario can be found in technical source like ISO and IEEE.

3. UC02 – Failure of a Decision Support System

Overall Description

Human operators (Remote Pilot, Military Air Controller, Commander in the field), use AI-based Collaborative Self-Organizing Cross-Platform Management of a Sensor Mesh (i.e. passive/active radar, electro-optical sensors (EO), audio sensors) to Optimize Situational Awareness. Future defence operations may operate multiple different sensor types integrated to manned and unmanned systems ("Sensor Mesh"). Each sensor type has its own characteristics and capabilities that are to create situational awareness. To allow tactical usage of the sensors on cross platform level, an AI-based sensor resource control will be mandatory.

The sensor mesh shall be online all the time and is used to cover the air-space of a nation ("the sensor dome"), detect adversarial aircraft, especially (unmanned) drones and drone swarms; classify air targets within a pre-defined probability range; track air targets; and instruct weapon stations to neutralise adversarial drones. To realize such a sensor resource control, AI-based techniques ("AI policies") which optimize decision sequences in a VUCA (volatile, uncertain, complex, ambiguous) environment to reach a predefined objective shall be used.
In case of system failure or connection loss, one or more sensors might be unavailable, leaving the operator unable to access the full range of information they normally use to make decisions. Therefore, they must communicate and collaborate to maintain best possible situational awareness.

Identified Impact:

- **Impact on the operator:** It is unlikely that all sensors would fail at once, but full operational picture would be impacted – especially situational awareness for those in the field would be compromised. Operators would be used to relying on the enhanced system feedback for situational awareness. Cognitive workload would increase for operators and additional time would be required to gain information from other sources (where available).

  Training at both individual and team level is warranted for operators to be able to make decisions both with and without the sensor mesh. Different combinations of failure conditions are required as well as sufficient frequency of training to ensure that the human operators are adequately trained when they are used as the main source of redundancy.

  The commander must consider the rules of engagement and do so within the confines of their operational, environmental, and cognitive capabilities at that time. Other operators should be aware of the rules of engagement and trained to both understand and anticipate what likely course of action would be.

- **Impact on the mission:** Additional time will be required for operators to communicate, collaborate and co-ordinate decisions and course of action. This additional time may increase likelihood of human casualties/fatalities and mission failure.
  An appropriate AI-based decision-making system does not only need to learn the environment's general dynamics ("context-aware") but is also requires anticipating how the environment is expected to react to its actions ("consequence-aware"). However, whether a control action is considered "good" generally depends on subsequent actions taken in the future ("credit assignment problem").

AI-based decisions regarding the appropriate use of various sensors must be both context-aware and consequence-aware and shall consider how the environment evolves under actions taken ("feedback loop"). This is especially important when using active sensors, which can be detected by enemy forces.

These applications have an impact on system performance (TAID-03:R , TAID-22:R, TAID-42:R PER-04:R), trustworthiness of AI (ASDC-01:R,TAID-22:R , TAID-28:R, TAID-37:R, TAID-38:R), Development Lifecycle, (DCLC-01:R), Advanced System Design Characteristics (ASDC-01:R), military operations/specifics (MOP-02:R, MOP-03:R, MOP-06:R, MOP-08:R, MSP-04:R, MSP-06:R, MSP-08:R), human-world values (HUV-01:R, HUV-02:R), and human factors (HF-11:R. HF-12:R, HF-13:R, HF-14:R, HF-15:R, HF16:R, HF-17:R, HF-18:R).

Involved Stakeholders:

According to the approach presented in Chapter 2 of TAID White Paper, this Use Case is on a **tactical level**. Therefore, the expected place takes on a **short** time scale. Accordingly, the following main stakeholders are involved in this Use Case:

- AI customers
  - Legal responsibility
  - Technical responsibility
  - AI operators
    - Operator
    - Operator team members
    - Operators interacting with AI system
    - Commander
    - Commander training team
- AI partners
  - AI trainers
- AI subjects
  - Allied
  - Neutral
  - Opponent

Applicable standardization/regulation documents for this scenario can be found in technical source like ISO and IEEE.

4. UC03 – Collision Avoidance/swarming of drones/emergence (tactics to neutralize targets)

Overall Description

Enhancing collision detection and avoidance is crucial for enabling UAV autonomy, especially in dynamic scenarios where uncertainties and non-cooperative intruders pose significant difficulties. In such environments, reinforcement learning (RL) stands out as a promising solution to effectively manage UAV manoeuvres by allowing the system to adapt to changing conditions in real-time.

Identified Impacts

- **System Development** (TAID-12 :R[1], DCLC-02:R): Training and verifying the RL model relies on simulation environments, providing controlled settings for generating various scenarios, including edge and corner cases. However, ensuring the accuracy and representativeness of the simulated data can be challenging. Moreover, deploying RL-based collision avoidance systems raises concerns about their reliability and robustness in unpredictable environments.
- **Impact on the Operator** (ASDC-01:R): Depending on the implementation, when the RL-based system takes control for collision avoidance, it may be challenging to communicate the entire trajectory to the operator in advance. Consequently, the operator must monitor performance metrics to assess whether the manoeuvre will avoid collision and maintain the aircraft within the safe flight envelope.
- **Impact on the Mission** (PER-01:O, PER-03:O, TAID-37:O, MOP-08:O): Integrating RL-based collision avoidance systems can enhance mission success rates and expanding operational capabilities by enabling assets to navigate complex and challenging scenarios and fulfil their objectives. By setting mission time reduction as a goal in RL-based system development, missions can be completed more efficiently.

Involved Stakeholders:

According to the approach presented in Chapter 2 of TAID White Paper, this Use Case is on a **tactical level**. Therefore, the expected scenario takes place on a **short** time scale. Accordingly, the following main stakeholders are involved in this Use Case:

- AI customers
  - Legal responsibility
  - Technical responsibility
  - AI Operator
    - Remote pilots
    - Pilots in the surroundings
- AI partners
  - AI trainers
  - AI subject matter experts
- AI producers
- AI developers
- AI authorities
  - Executive power
    - Aviation safety
    - Chain of command
- AI subjects

- Allied
- Neutral
- Opponent

Applicable standardization/regulation documents for this scenario can be found in technical source like ISO and IEEE.

5. UC04 – Mission Training

Overall Description

AI has an immeasurable potential to transform military mission training by offering capabilities that complement human skills and experience. Demonstrated by its mastery in strategic games like Chess [55], Go [56], Starcraft II [57], and Stratego [58], AI showcases its ability of analysing vast datasets, identifying patterns, and devising innovative tactics. It can serve as a multifaceted support system for military personnel, providing real-time analysis of training sessions to deliver personalized feedback and recommendations, thereby refining skills and enhancing decision-making abilities.

In addition, AI's capacity to generate strategic plans and scenarios tailored to specific training objectives can foster creativity and strategic thinking among trainees. Furthermore, acting as a dynamic adversary, AI can challenge trainees to navigate through unpredictable scenarios, exposing vulnerabilities in decision-making processes. Realizing its full potential necessitates the creation of realistic and immersive environments that mirror the complexities of actual missions, achievable through the integration of AI and the concept of Digital Twins, which involves crafting virtual replicas of physical assets or environments.

Identified Impacts

The inclusion of AI in military training would impact on various stakeholders and the mission itself:

- **Impact on military commanders and other military personnel** (MSP-04:O, MSP-07:O, TAID-11:R): AI presents opportunities to enhance decision-making and operational readiness. However, challenges might arise due to human-AI interaction, such as comprehension and accountability issues, potentially impacting trust in the AI system. In addition, increased reliance on AI-based systems may lead to over-reliance and potential blind spots in decision-making.
- **Impact on the mission** (MSP-04:O, MOP-06:O, MOP-07:O, MOP-08:O, HUV-02:R): AI-based systems can support effective decision-making in dynamic and complex operational environments, including the preparation to act against AI-powered adversaries in real-world conflicts. At the same time, it raises ethical concerns related to the potential for autonomous decision-making and the ethical use of force.

Involved Stakeholders:

According to the approach presented in Chapter 2 of TAID White Paper, this Use Case is on an **operational level**. Therefore, the expected scenario takes place on a **medium** time scale. Accordingly, the following main stakeholders are involved in this Use Case:

- AI customers
  - Legal responsibility
  - Technical responsibility
  - AI operators
    - Mission operators
    - Remote pilots
    - Pilots in the surroundings
- AI providers
- AI partners
  - AI trainers
  - AI subject matter experts

- AI producers
    - AI developers

Applicable standardization/regulation documents for this scenario can be found in technical source like ISO and IEEE as well as in regulatory frame works like the European AI Act and the High Level Expert Group on AI.

6. UC05 – Aerial Refuelling

Overall Description

Aerial refuelling is a process in which a tanker aircraft transfers aviation fuel to a receiver aircraft using either a boom or hose-and-drogue system. Successful aerial refuelling operations demand precise manoeuvring, real-time communication, and a well-coordinated approach. To alleviate operator workload while enhancing safety and mission efficiency, the operations involving the boom system have been upgraded to an automated version. This upgrade entails the implementation of a sophisticated system based on classical computer vision algorithms which can detect and track the receiver, as well as the calculation of optimal approach trajectory of the boom. Further improvements in terms of versatility, which benefit both the aerial refuelling with boom and hose-and-drogue, can be achieved by applying machine learning alongside the classical algorithms. This enhancement would enable the system to discern receivers across diverse scenarios characterized by variations in light conditions, background elements, and receiver types.

Identified Impacts

- **Impact on the mission** (TAID-02 :O, PER-01:O, TAID-21 :O, TAID-06:O, TAID-37:O, TAID-27:O): The implementation of automatic aerial refuelling ensures both quicker and more precise approaches during operations. Consequently, the increased availability of this system can significantly enhance mission success rates. Machine learning algorithms can achieve it due to its adaptability to changing conditions, such as variations in light and weather. Additionally, its scalability is a notable advantage, enabling it to easily accommodate growing demands such as larger and heterogeneous fleet of aircraft or operating in increasingly complex environments.
- **Impact on the operator** (ASDC-01:O, TAID-11:O, PER-04:O): It can reduce the cognitive workload on operators with the increased footprint of the automatic system, allowing them to focus on other critical aspects of the refuelling process. This can lead to improved operator performance and reduced fatigue, ultimately enhancing safety.

Involved Stakeholders:

According to the approach presented in Chapter 2 of TAID White Paper, this Use Case is on a **tactical level**. Therefore, the expected scenario takes place on a **short** time scale. Accordingly, the following main stakeholders are involved in this Use Case:

- AI customers
  - Legal responsibility
  - Technical responsibility
  - AI operators
    - Aerial refuelling operators
- AI producers
  - AI developers

Applicable standardization/regulation documents for this scenario can be found in technical source like ISO and IEEE.

7. UC06 – Data-Centric Security

Overall Description

Anticipating the increasing needs of multi-domain operations, scalable sharing of sensitive information and collaboration among allies is pivotal. Consequently, information protection is moving to the centre of attention too. As such, data-centric security (DCS) paves a promising direction ahead. DCS principles aim to ensure confidentiality, availability, and integrity of data over the entire lifecycle whereas access to encrypted data is controlled through fine-grained metadata that are strongly bound to one another. Therefore, DCS establishes a novel in-depth layer of defence close to the data while relaxing demands of perimeter protection at security domain level as frequently practised in military IT landscapes today. With these and further advantages at hand, DCS is strategically driven by NATO for the next decade and beyond through its implementation plan [59] [60]. Thus, it constitutes a paradigm shift with practical impact to existing military infrastructures, processes and staff between both the Alliance Federation and associated NATO enterprises where technology based on trustworthy AI plays a crucial role for reliable and secure information processing as well as support for human actors. In fact, trustworthy AI is the enabler for the transformation towards DCS.

Identified Impacts

Based on the overall description subsequent impact attributes apply in general: PER-02, PER-03, ASDC-02, MOP-02. Additionally, the following are significant:

- **Federation of security domains** (TAID-03, TAID-04, MOP-08): Within a DCS architecture, heterogeneous military IT landscapes converge into common data spaces with fewer security domains. Yet, next-generation cross-domain solutions must arise supporting the transformation process for various complex information sharing scenarios in both short and long term. Such technology implies the capability to learn and adapt particularly to prevent data leakages and to cope with cyberthreats that are unknown at design or deployment time.
- **Seeking and annotating legacy data** (ASDC-01, TAID-11): Huge quantities of existing information reside in legacy security domains today that must be inventoried for DCS. Seeking and annotating this information is costly and error-prone that cannot be driven by human analysts alone. This requires transparent AI-supported pre-labelling processes, where borderline cases are presented to human annotators ensuring consistent metadata quality eventually.
- **Granular confidentiality marking** (ASDC-01, MOP-08, TAID-11): Recent military practices in annotating information are based on coarse-grained labelling. Particularly in the context of classified information, it is common to set the degree of sensitivity globally for documents rather than on a finer level. As a result, sophisticated AI-based capabilities must emerge to discern sensitive regions including both classified and privacy information from less critical content within unstructured documents to explain and assist human annotators creating reliable and fine-grained metadata.

Involved Stakeholders:

According to the approach presented in Chapter 2 of TAID White Paper, this Use Case is on a **strategical level**. Therefore, the expected scenario takes place on a **large** time scale. Accordingly, the following main stakeholders are involved in this Use Case:

- AI customers
  - Legal responsibility

    - Technical responsibility
        - Individuals/Teams responsible for cybersecurity/strategic risk analysis
    - AI operators
        - Remote pilots
        - Pilots in the surroundings
- AI providers
    - Companies providing cross domain solutions
- AI partners
    - Human and non-human annotators
- AI producers
    - AI developers
- AI authorities
    - EDA
    - NATO

Applicable standardization/regulation documents for this scenario can be found in technical source like ISO, IEEE and STANAG.

8. UC07 – Military Approval/Certification

Overall Description

Dual civil/military certification of a military aircraft refers to the process by which an aircraft is certified to fulfil with both civil aviation regulations for civil use and dedicated military regulations for military operation. Such kind of certification provides great flexibility, especially for military transport aircraft, as it allows them to perform both civilian operations, such as humanitarian aid transport or evacuation of civilians from areas affected by natural disasters, and tactical operations such as transporting military equipment to conflict zones or tactically dropping loads in flight.

This certification process is more complex than a purely civilian or military process, as it involves both approval from the corresponding Civil Aviation Agency (EASA in Europe or FAA in USA) and the applicable Military Authority (National Military Airworthiness Authority of each country). Additionally, the process typically occurs sequentially: first, the Civil Type Certificate is obtained, which must be recognized by the NMAA, and then the military authority provides the Military Type Certificate (which includes the civil one). In Europe, the rules for the certification of military aircrafts are described in the European Military Airworthiness Certification Criteria (EMACC) Guidebook, developed by the European Defence Agency (EDA).

Identified Impacts

As innovation continually introduces new technologies and design features, existing airworthiness regulations must evolve accordingly as they may not fully cover some of the aspects for assessing the airworthiness of a new product. For this reason, industry and regulatory agencies must agree new requirements and standards coping with the particularities of new technologies. This is precisely the case with AI technology, which involves a range of specific risks that must also be addressed with specific regulations, means of compliance, and standards:

- **Civil Certification of AI-based systems:** EASA has proposed the anticipated rulemaking concept for AI applications in aviation as part of the EASA AI Roadmap 2.0, which consist in the new Part-AI (at the same level of existing Part 21), which would contain the requirements identified in the already published EASA concept papers, organized in three major provisions: Part-AI.AR: requirements for authorities Part-AI.OR: requirements for organizations, and Part-AI.TR: requirements on AI trustworthiness. Additionally, the future international standard EUROCAE ED-324 / SAE ARP-6983 [12] is expected to be recognized by the agencies as Acceptable Means to enable the certification of AI-based systems to be used in aviation.
- **Military Certification of AI-based systems:** Military applications pose specific characteristics such as relevance of operational context (mission orientation), interoperability, adaptability, cybersecurity, etc. which would require specific rules for certification. In particular for AI technologies, new rules must be agreed with the relevant Defence stakeholders (NMAAs, NATO, EDA, etc.), on top of the AI civil certification framework. R&D projects, such as EICACS, will produce dedicated guidelines to address the particularities of certifying and qualifying military capabilities using AI technology. These guidelines may be used in the future to support the establishment of AI military certification frameworks.

AI technology will have a significant impact on both Production and Design Organisation as well as Civil and Military Type Certificates. It will require the use of novel demonstration means

of compliance and standards to be agreed with the respective authorities. The use of AI technology for military applications will require, in addition to demonstrating airworthiness, the demonstration of mission performance, which involves an additional process known as Qualification. Additionally, AI may impact Continuing Airworthiness Management Organisations on planning the maintenance tasks.

Involved Stakeholders:

According to the approach presented in Chapter 2 of TAID White Paper, this Use Case is on a **strategical level**. Therefore, the expected scenario takes place on a **large** time scale. Accordingly, the following main stakeholders are involved in this Use Case:

- AI partners
  - Manufacturers
  - Data providers
  - Air traffic controllers
- AI producers
  - AI developers
- AI authorities
  - Certification authorities
  - International regulatory bodies

Applicable standardization/regulation documents for this scenario can be found in technical source like ISO and IEEE as well as in regulatory frame works like the European AI Act and the High Level Expert Group on AI (refer to Chapter 4 of TAID White Paper).

9. UC08 – Meaningful Human Control

Overall Description

A drone operator deploys an autonomous weapon system (via drone) to disable targets and or infrastructure in the field. This has the potential for greatly reduced risk to human life for both own defence forces but also for human lives in the surrounding target area. The use of lethal force requires clear ethical values and codes of conduct as well as ethics checks to be deployed throughout the design lifecycle. This use-case highlights the role and importance of International Humanitarian Law and Article 2 of the EU Treaty as well as Meaningful Human Control. It is prudent 7 key requirements for Trustworthy AI, are rigorously adhered to in future system design.

Identified Impacts

- **On the human operator:** There are no humans required to be in the field as the autonomous weapons system deploys weapons at the target site. Furthermore, fewer drone operators may be required in the case of multiple drones being deployed (i.e. swarming). Drones may be able to access areas or terrain previously considered “unreachable” without having troops on the ground thus potentially increasing the likelihood of disabling targets. The drone operator can have increased situational awareness from multiple system feeds (i.e. radar, video, etc.).

  Enabling of fully autonomous mode can afford additional safeguards to ensure the preservation of human life, by instructing the autonomous weapons system to never engage (i.e. fire) on humans within a specific target area, thus enabling disabling of infrastructure without targeting human life.

  There are critical ethical requirements for the use of lethal force and the support of the human operators. It is important that drone operators adhere to the ethical code of conduct as outlined by their own defence force and those of NATO, EDA, Geneva convention etc. Equally, it is pertinent that operators are supported by their organizations in the cases of both mission success and failure. Operators must be able to trust that the wider system (organizational, national levels) is supportive of their decision-making and actions (when carried out in accordance with ethical codes of conduct).

- **On the mission:** Being able to reach previously “unreachable” targets without troops on the ground affords more complex operations and may enable further strategic/ tactical opportunities. Having more strategic opportunities affords more timely decision-making at both the drone operator and command levels.

  Transparency should be such that the human is aware of when they are in control and when they have ceded control. The HMI design would benefit from making all stakeholders (operational and command level) aware of when adverse outcomes are experienced such as loss of control, Connection loss etc. and feedback to the human re likely intent of the autonomous weapon system e.g. continue to target, select new target, hover until connection resumes, abort mission etc. according to their own rules of engagement and Doctrine at national and EU levels.

These applications have an impact on system performance (PER-04:O) trustworthiness of AI (ASDC-01,TAID-01:R, TAID-12:R, TAID-01:R, TAID-15:R, TAID-20:R, TAID-28:R, TAID-35:R, TAID-36:R, TAID-38:R,) Advanced System Design Characteristics (ASDC-01:O/R, TAID-43:R) military operations/specifics (MOP-01 O/R, MOP-03:O, MOP-06:O MOP-08:O,

MSP-03:O), human-world values (TAID-11:R, HUV-01:O/R,HUV-2:R, HUV-03:R), and human factors (HF-01:R. HF-02:R, HF-06:R, HF-09:R HF-12:R, HF-15:R, HF-16:R, HF-17:R, HF-18:R) .

Involved Stakeholders:

According to the approach presented in Chapter 2 of TAID White Paper, this Use Case is on a **tactical level**. Therefore, the expected scenario takes place on a **short** time scale. Accordingly, the following main stakeholders are involved in this Use Case:

- AI customers
  - Legal responsibility
    - Procurement agencies
  - Technical responsibility
  - AI operators
    - Drone operators
    - Pilots in the surroundings
- AI producers
  - AI developers
  - AI system designers
- AI authorities
- AI Subject: Neutral

Applicable standardization/regulation documents for this scenario can be found in technical source like ISO and IEEE. as well as in regulatory frame works like the European AI Act and the High-Level Expert Group on AI.

10. UC09 – Active Autonomous Cyber Defence

Overall Description

Active cyber defence is a direct defensive action taken to destroy, nullify, or reduce the effectiveness of cyber threats against friendly forces and assets [8].

A cyber security system with AI is or will be state-of-the-art of military practices. It can be designed in a way that it responds to the events autonomously, for example AI detects an attacker in the system and initiates the response. The active response can be login blocking, blocking user behaviour patterns that are different than usual, antimalware response, intrusion prevention, etc. In some cases, this response can be of offensive kind - meaning that, for example, in a scenario where the attacker performs the Denial-of-Service attack, a similar action could be performed on the source server of the attacker. Active defence features include the scope of effects (internal/external networks), degree of cooperation (consent from network owner)

Defence actions are non-cooperative when they are performed without the network owners' consent.

Identified Impacts

System Development: Training AI models and testing cyber security systems is challenging due to constantly changing attack methods and patterns. Cyber security systems usually cover only a part of all the possible attacks. The scope of effects (available response actions) is determined at this stage. It's up to the software/service provide to ensure the trustworthiness and reliability of the cyber security system.

- **On the operator**: Operator must ensure the consent from the network owner to be able to perform actions of active defence. There is always a possibility that offensive defence actions will be interpreted as an attack from the network of the defending party. The degree of autonomy is also variable, usually the operator can provide more context for the decisions being made.
- **On the mission**: Human attacker element introduces a degree of uncertainty for the system's ability to detect the attacks and react. Some attacks might be undetected, also there is a chance of interpreting benign behaviour as a malicious action. Non-cooperative action raises ethical and legal issues too. Actions launched against the attacker might affect other parties that were not involved in the attack. It could result in noncombatant parties (civilian property) being harmed. Critical life-support systems, critical infrastructure (powerplants, etc.) can be damaged, taken over and stopped.

Involved Stakeholders:

According to the approach presented in Chapter 2 of TAID White Paper, this Use Case is on an **operational level**. Therefore, the expected scenario takes place on a **medium** time scale. Accordingly, the following main stakeholders are involved in this Use Case:

- AI customers
  - Legal responsibility
  - Technical responsibility
    - System owner
  - AI operators
    - Cyber operator

        - Pilots in the surroundings
- AI providers
    - AI system developers

All abbreviations exposed can be found at TAID White Paper.

11. Bibliography

[1] D. Silver, T. Hubert and J. Schrittwieser, "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play," *Sience,* vol. 362, no. 6419, pp. 1140-1144, 2018.

[2] D. Silver, A. Huang and C. Maddison, "Mastering the game of Go with deep neural networks and tree search," *Science,* vol. 529, pp. 484-489, 2016.

[3] O. Vinyals, I. Babuschkin and W. Czarnecki, "Grandmaster level in StarCraft II using multi-agent reinforcement learning," *Nature,* vol. 575, pp. 350-354, 2019.

[4] J. Perolat, B. De Vylder and D. Hennes, "Mastering the game of stratego with model-free multiagent reinforcement learning," *Science,* vol. 378, no. 6623, pp. 990-996, 2022.

[5] NATO, "Data Centric Security Implementation Plan 2.0," NATO, 2022.

[6] K. Wrona, "Towards Data-Centric Security for NATO Operations," *Springer - Communications in Computer and Information Science,* vol. 1790, no. Digital Transformation, Cyber Security and Resilience, pp. 75-92, 2023.

[7] EUROCAE/SAE, "ED-324/ARP-6983 - Process Standard for Development and Certification/Approval of Aeronautical Safety-Related Products Implementing AI," SAE International, (under development).

[8] D. E. Denning and B. J. Strawser, "Naval Postgraduate School Publications," 2015. [Online]. Available: https://faculty.nps.edu/dedennin/publications/Active%20Cyber%20Defense%20-%20Cyber%20Analogies.pdf. [Accessed 10 10 2024].