SCIENCES PO LILLE.

Research Thesis

Academic Year 2024-2025

Thesis defended on May 30, 2025

The Rise of AI Integration within Military Technologies

AI in warfare: unravelling its evolution, harnessing opportunities and risks, exploring legal obstacles

Candidate: Elena Calafiore,

Student of the Master in «Stratégie, Intelligence économique et Gestion des Risques»

Supervisor: Prof. Audrey Herisson,

Advisor to the President of the Académie de Défense de l'École Militaire

Academic Year 2024-2025

Warning

Sciences Po Lille does not intend to give any approval or disapproval to theses and opinions issued in this research thesis. These must be considered as of their own author. I attest that this research thesis is the result of my personal work, that this work contains quotes and reference all the sources used and that it does not contain a passage having already been used entirely in similar work.

Abstract

In less than a decade, AI has become a widely applied technology in numerous everyday contexts. Specifically, numerous states started investing and researching its application within the military field, with the aim of transforming wars in conflicts developing within a quicker time frame, reducing the number of civilian and military victims. Soldiers are being less and less retrieved from the battlefield, leaving the space to new technologies able to preserve their lives. Nevertheless, such AI systems are still affected by a still quite elevated error rate, biases within training data, vulnerabilities in the cyber protection of their systems, which affect enormously the safety with which they can be applied within military contexts, where specifically in battlefield situations, human lives are at stake. New issues about responsibility for mistakes or unforeseen actions are arising, and the international legislation is still completely unprepared in welcoming such novelty. Historically speaking, military technologies have been always researched, developed, used and then finally regulated or banned when it was noticed that the amount of humanitarian and environmental damage they caused. Will it be the case of AI integrated to military technologies? However it should not be forgotten that AI does not solely help with target acquisition and striking as it is generally believed by civil society organizations, but also helps in lowering the costs of transportation and similar within the logistic sector, alongside supporting the military decision-making and personalized training. The aim of this thesis is to unravel this change starting from a historical point of view, passin through the analysis of some technical aspects, which will help in understanding benefits and risks of military-AI, in addition to the analysis of a specific case (the application within nuclear field and weapons), ending with an examination of the level of unpreparedness of international legislation on the topic.

Acknowledgments

I would like to thank all the people who supported me throughout the process of making this thesis. Prémierment, je tiens à remercier ma directrice de mémoire, la Capitaine de Vaisseau Audrey Hérisson, pour son support et ses conseils importants. In secondo luogo ringrazio il Generale di Squadra Aerea Fernando Giancotti per i consigli e la guida fondamentale datami per la realizzazione di questo lavoro.

Ringrazio poi la mia famiglia, dunque mamma, papà e Vale, per il supporto costante e incrollabile, e soprattutto la pazienza nell'ascoltare i miei podcast-vocali di aggiornamento. Ci tengo poi a ringraziare Mathilde per aver allegerito il carico di lavoro con chiacchierate, risate, pianti, serate thé, e film. Je remercie après Maïlys pour les moments droles passés ensembles en tant que presque-voisines, pour les taboulés, ainsi que pour la patience dans l'écouter mes *brainstormings*, et les beaux moments avec Camille, Lea et Maëlle. Ringrazio inoltre Edoardo (colui che ho coinvolto nelle mie follie accademiche), per la compagnia nelle giornate e serate passate rigorosamente a studiare nella BU di Sciences Po, con tanto di taboulé e partita a carte durante le discussioni di geopolitica e difesa.

Je tiens à remercier aussi Lucile, pour les rires, le support et les leçons de français slang, ainsi que Capucine, pour les beaux moments passés ensemble, et pour avoir été mon *Cicéron* par les rues incroyablement belles de Lille. Ringrazio poi tantissimo Giorgia, Aurora e Rebecca, per le telefonate chilometriche, le risate e i gossip fatti anche se a chilometri di distanza, e il supporto costante. Ringrazio poi Fifi per aver portato un po' di patrottismo italiano tra i corridoi di Sciences Po tra maglie, calze e felpe dell'Italia (nonostante la cittadinanza francese). Ringrazio inoltre Beatrice, Paolo, Mario, Samia e Felipe per i bei momenti passati assieme e le risate per spezzare il grigiore del clima di Lille. I would then like to thank Fardin for the wonderful memories of the time spent together at the university.

Au final, je remercie Lucie pour les rigolades pendant les journées passées en BU, ainsi que pour les sessions d'étude desesperéés de russe avec Amina et Melodie (que je remercie pour m'avoir fait découvrir les зефир глазированный). Наконец, я благадарю Женю, что она взбодриться дни на БУ с своими неизбежными сумасшедшими и адреналиновыми историями.

Grazie mille a tutti voi! Merci beaucoup à vous tous ! Thanks to you all! Всем вам спасибо!

List of Acronyms and Abbreviations

AI: Artificial Intelligence

CAF: Canadian Armed Forces

Canadian DND: Canadian Department of National Defense

DoD: Department of Defence

EW: Electronic Warfare

ISR: Intelligence, Surveillance and Reconnaissance

LAWS: Lethal Autonomous Weapon Systems

MDPW: Multi-Domain Precision Warfare

MEP: Member of the European Parliament

NCIA: NATO Communications and Information Agency

PLA: People's Liberation Army

PRC: People's Republic of China

R&D: Research and Development

RPV: Remotely Piloted Vehicles

TEVV: Testing, Evaluation, Validation, Verification

UAV: Unmanned Aerial Vehicles

UN: United Nations

US: United States

VR: Augmented Virtual Reality

VUCAR: Volatile, Uncertain, Complex, Ambiguous and Rapidly Changing

Table of Contents

"We set sail on this new sea because there is new knowledge that must be gained, and new rights to be won, and they must be won and used for the progress of all people. For artificial intelligence, like nuclear science and technology, has no conscience of its own. Whether it will become a force for good or ill depends on man, and only if we occupy a position of preeminence can we help decide whether this new ocean will be a sea of peace or a new terrifying theater of war".

General Hyten John E., September 9, 2020.

Introduction

The US DoD General Hyten pronounced these words on September 2020 inspired by a previous quote from John Kennedy of 1962 referred to space[1]. As space study, research and exploration was considered a new scientific frontier, full of opportunities, dangers and risks, from a technical, scientific, political, humanitarian and juridicial point of view, AI today represents the same thoughts, debates and studies full of curiosity and fears. This thesis aims at exploring the application of AI within the military field in the current constantly evolving context, where countries' representatives give speeches at the United Nations about the need to regulate such Lethal Autonomous Weapon Systems, while finaning and sustaining the research of new and even more powerful military-AI systems. The international community is de facto fragmented on this topic: while some countries push for a complete ban of LAWS, some others highlight the prematurity of times to decide on technologies whose potential has not been fully harnessed and mastered. A further group asks for mere regulation of the typology of weapons where AI can be applied or not, alongside the context and the degree of human control over its actions. One principle over which nearly all countries seem to agree is the importance of maintaining human control IN the system or ON the system, respectively the human-in-the-loop and human-on-the-loop, so that in case of need, humans are always ready to intervene. The evident disagreements on the definition of the degree of autonomy, human control, have completely hindered the work of the UN in finding a juridical agreement, even just merely on

[1] US Department of Defense, *Remarks by General John E. Hyten to the Joint Artificial Intelligence Symposium and Exposition*, September 9, 2020. https://www.defense.gov/News/Transcripts/Transcript/Article/2344135/remarks-by-general-john-e-hyten-to-the-joint-artificial-intelligence-symposium/

definitions. However, all states seem to agree on the need to avoid the development and employment of fully autonomous weapons, even if LAWS with greater and greater autonomy are being employed in different current conflicts, especially in Palestine and Ukraine. This thesis aims at summing up the current state of the art of military-AI via a historical, technical, strategical, and juridical point of view with the aim of developing a 360° examination of the subject.

Clearly, there have been some difficulties encountered by the author of this thesis. First and foremost, when talking about military-AI it needs to be considered that it is a truly modern and dynamic sector, with new researches and technologies being announced day by day. As it is possible to notice by the bibliography of books, academic and journal articles, a large part of the sources are from 2024 and 2025, some even from April and May 2025. This facts probably give an idea of the fast pace at which this fielf is constantly developing, and with that, the difficulties in framing the topic. Furthermore, it has been truly diffult giving a precise and strict frame to the analysis, because AI is being more and more experimented and used to give birth to common databases for ISR and similar data, implying that to explain a specific context of usage or task there is the need to examine the intertwined one, and so on. Moreover, sticking to the initial idea of focusing mainly (and solely) to the analysis of the way in which AI affected the military decision-making sphere revealed particularly arduous because of the mathematical technicalities that would have needed to be explained, specifically on the learning modalities of AI.

Nevertheless, given that this topic is constantly evolving, analysing the technologies being employed in Ukraine and in Palestine has helped with understanding the main fields where military AI is deployed, still it was not possible to delve into these two specific cases for reasons of space constraints alongside the fact that such events are still ongoing. The news about the dynamicity of these two conflicts vehiculated via social media enabled the author of this thesis to find information even in platforms like LinkedIn, where military industries like Anduril and Epirus post, nearly daily, pictures and videos about their military technology integrated with AI. This fact will evidently lead to the development of a habit to live with such technologies, hear and read news about that, pushing for a long-term normalization, in a context where there still is no international legislation whatsoever on the topic. Here it is an example:

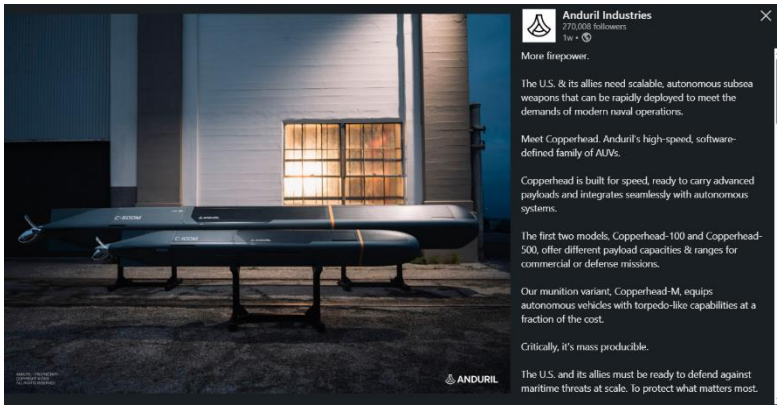

Image A: Military technology industries advertisements on linkedIn
Source: The aforementioned picture is a screenshot that the author of this thesis has taken on April 15, 2025, at 23pm from her LinkedIn profile, specifically on the official LinkedIn channel of Anduril Industries.

Additionally, for what concerns the analysis of the military sectors in which AI comes to be applied, the author of this thesis chose the nuclear weaponery one because it is mentioned less frequently in technical, political and ethical debates rather than the land, aerial, naval, cyber and quantum ones.

In addition, in regard to the research process, the main analysis has been done on the documents reporting the various national strategies about the investments, R&D and application of AI within the military sector. The documents have been analysed in the original language both for the United States, the Russian Federation, the People's Republic of China, France, and the United Kingdom, but concerning China the author of this thesis made use of the English translated versions of the official documents, and when there were none, used directly different AI translators comparing the different translated versions. A further analysis that has been made is the one of the main European authors on this topic, specifically choosing academic articles published between 2020 and 2025, focusing on the ones starting from 2023, precisely basing the research on the data and literature brought forward by Laure de Roucy-Rouchegonde for the French side, comparing it to Alice Saltini's researches for the intenational vision, alongside Fernand Giancotti, Mariarosaria Taddeo and Rosanna Fanni's analysis of the Italian context with the international R&D level. Beyond these aspects, both the historical and juridical parts are analysed from a chronological point of view to be able to understand and compare the fast pace of the research alongside the truly slow progresses in the context of international legislation. Concerning the technical part there is no specific analysis of the training models of AI, briefly introduced when needed, for reasons of space and time constraints.

Finally, concerning the structure of the thesis, after an initial analysis on the birth and development of AI, from generally and widely applied AI to the military sector, the thesis dvelves more onto some more technical aspects about the specific tasks that military-AI

performs, alongside the benefits and risks it entails. Thereafter there is a more detailed analysis of a specific sector of application of military-AI, the nuclear field, followed by the examination of the inadequacy of the current international legal system and the level of states representatives debates about AI and LAWS, alongside some ethical considerations. As a last note, parts of the subchapter 1.1, 1.2, 1.4, 2.3 have been based on a previous thesis of the author, specifically the bachelor thesis titled "The Use of AI Surveillance Cameras for Safety Purposes: The Case of France and the 2024 Olympic Games", which can be consulted in case it could be of interest for the reader.

Chapter 1: AI through human history

1.1 Historical and technological roots of the predecessors of AI

AI started to be theorized in the XX century, but its history is way older. Probably it is moreappropriate, if not of paramount importance, to begin with the starting point of human history: the writing process. From simple writings to noting down everyday issues, human beings started to write down numbers next to phrases, which were initially just composed by mere letters. This last process gave slowly birth to scientific studies such as mathematics, geometry and astronomy, laying the foundation for the future Artificial Intelligence.

In particular, the writing process has been crucial to note down the logical reasoning at the basis of the AI functioning process: the deductive syllogism. This last concept has been theorized by Aristotle, who explained how starting from two premises it is possible to reach a final conclusion which includes both messages. The most well-known example made by Aristotle on the subject is the following: major premise: "All men are mortal"; minor premise: "Socrates is a man"[2], therefore following logical reasoning, the conclusion is that Socrates, by being a man, is mortal.

Alongside the definition of the basis of deductive syllogism, Aristotle postulated three mainprinciples that represent the base of logical reasoning: coherence, non-contradiction, and excluding reasoning. To explain such principles, he used as example the following statement: "if Socrates is dead, he can't be alive and vice versa". This reasoning modality is crucial, as the machine-logic at the basis of AI has no common sense and it is incapable of thinking, so it needs clear principles to abide by when it has to make decisions among

[2] Example of Syllogism employed by Socrates, who is the basis as well of the mathematical principle of Transitivity.

different possibilities.

This kind of reasoning has been accepted for decades by generations of philosophers, as it was difficult to be contradicted, until the arrival of George Boole. The latter transposed the logical reasoning into mathematical language using the binary code, so that the response to a certain question could be considered either true or false. With this new language, a crucial transformation arrived: being able to transpose abstract concepts into mathematical language, as if they were mere numerical digits. George Boole also enumerated the material implication, crucial for the rational reasoning. Usually described with the symbol →, it should be considered in the formula A→B, which means that A implicates B, unless one of these two is false. This reasoning of implication is the basis for modern programming languages. An everyday example of such a mathematical operator could be represented by the situation in which we have the two phrases "If it is sunny while going to the beach, then we will be tanny". In this statement, it is clear how the characteristic of being tanny, so the B of the initial formula, needs the sun to happen, which is represented by the A of the formula. The mere implication is not enough to describe the entire logical-mathematical reasoning based on the Truth Table, but for this thesis, it will be enough to stop at this level of knowledge.

With the Truth Table, Boole described all variables represented by the values True or False,which can be identified with the numbers 0 and 1. This all means that in mathematics, all thevalues that could be either true or false are defined as "Boolean values".

Boolean logic was later on applied by Shannon to create models on the functioning process of electronic system, based on a 0 and 1 reasoning, giving birth to the concept of the bits[3], the basis for the computers.

In particular, Shannon, the "father of information theory", affirmed that by applying the Boolean principles and algebra to the electrical field, it could have been possible to constructany logical-numerical relationship. This idea was crucial for the analysis of encrypted messages during World War II. To demonstrate so, Shannon put in place an experiment consisting in putting a little mouse in a machine within a wired room. The animal, not able to see where he was going, started to stomp on the walls of the space in which he was confined, until he was able to reach the exit. The input that the mouse was receiving after every stomp was the information that there was a wall there, and thanks to the collection of all this information, it was able to find its own exit path. From this

[3] The bit is the most basic and simplest unit of information used within the field of informatics.

experiment, the idea wasborn that the machine could have been controlled via inputs[4] that would have enabled it to translate reality into actions. Shannon then expressed his mathematical theory of information in his famous paper, later defined as the "Magna Charta of the Information Age"[5], laying down the fundamental principles for Artificial Intelligence. In particular, with this paper, he worked on the problem of finding the best way to encode a message and the ways of transmitting it to the receiver. This all gave birth to the field of "Information theory"[6].

These theories have been crucial for the transmission of information as electrical signals andwaves, able to pass via cable or as light or sound signals, of which the most illustrative method is represented by the Morse code. The transmission via cable arrived just subsequently, helping with the development of the radio and the television, thanks to the transformation of analogical signals into numerical signals (series of 0s and 1s). This made the storage of information achievable, which was impossible with just the acoustic or electricsound, and it optimised not only the transmission of the message, but also the decodifying process of the signals received.

A crucial passage from such optimisation of the encryption and the decryption of the message to the development of AI has been represented by Alan Turing's machine[7]. The latter has basically "exploited" the technological basis created by Shannon to decrypt the messages from the Nazi forces during the Second World War.

Developing a machine to do such encryption procedures, inaugurated a period of debates and discourses on the development of machines able to think autonomously. This entire reasoning has been anticipated by Turing, who affirmed that after around fifty years it would have not been possible anymore to distinguish if a certain reply had been given by a computeror a man[8]. It was at that point that the so-called "Turing test"[9] has been developed to establish whether machines had a behaviour based on a certain level of intelligence or not. In particular, this test consists of having a panel of judges who need to establish if a

[4] This idea of being able to control the machine via input it received and information it collected was conceptualized and theorized by Norbert Wiener with hist theories on the Cyber loops, in his book *Cybernetics: Or Control and Communication in the Animal and the Machine*, 1948.

[5] "The information society: an International journal", vol 12, n°. 3, 1996 - "Cyberspace and the American Dream: A Magna Carta for the Knowledge Age" 1994, Esther Dyson, George Gilder, George Keyworth, AlvinToffler.

[6] The Information Theory is a field of study that focuses on the modalities of communication of the information,and on the method of storing the latter.

[7] Alan Turing was an English mathematician who is well-known in the history of computer science studies for having decrypted the messages of the Nazi soldiers during the WWII, thanks to his decrypting machine.

[8] Rodolphe Gelin, *Dernières nouvelles de l'Intelligence Artificielle*, Edition Flammarion, page 26.

[9] The Turing Test has been proposed for the first time in his own article *Computing Machinery and Intelligence*, in the Journal Mind, 1950, following the question *Can machines think?*

specificoutput, as for example a written production, has been made by a machine or a man. The Turing test has been passed for the first time in 2014 by the Chatbot[10] Eugene Goostman,whose software was able to trick the judges and make them believe it was a human by fakingthat it was a Ukrainian boy to justify its English mistakes and its incertitude for the replies. Probably, the most known version of this test is the CAPTCHA[11], the verification methodthat all people commonly use to access websites to prove that they are humans and not robots(software) whenever they are requested to click on all the images related to traffic lights.

1.2 Birth and development of AI

At this point, thanks to Turing's machine and theories, the technical bases for the development of AI were ready, and the only missing point was the name. The name of "Artificial Intelligence" has been coined by John McCarty in 1956, in the wake of a conference in the Dartmouth College[12]. The creation of a definition has been followed by the establishment of an AI laboratory at the MIT in 1959, which marked the beginning of the studies on AI.

The academic research on the topic started with the definition of the first hardware implementation of the Perceptron[13] made by Frank Rosenblatt, made it possible to obtain anoutput by just specifying the parameters used by an algorithm (created by Rosenblatt as well)capable of learning from its own mistakes. Unfortunately, the Perceptron started to be more and more criticized, as the initial results reached by the AI were not the expected ones, so in 1973 it started a period of disinterest towards AI defined as the first "AI Winter"[14], which would be later on followed by a second "AI Winter" in 1987.

In the meantime, between 1966 and 1972, Stanford Research Institute created Shakey, the first robot with self-reasoning, presented by the Magazine Life as "the first electronic person"[15]. The robot had a wheeled base and it moved thanks to the information it received via an antenna for two-way radio communication, mainly depending on the input deriving

[10] The Chatbot is a software created to establish conversations with human users on specific topics.
[11] The CAPTCHA system, whose acronym stands from "Completely Automated Public Turing test to tell Computes and Humans Apart", is a test based on the analysis of the response given from the user to understandif it is a human or a machine, crucial to detect bot attacks.
[12] The "Dartmouth Summer Research Project on Artificial Intelligence" was a conference organized in 1956 considered as the initial event on the officialization of the AI.
[13] The Perceptron is an algorithm for the Supervised Learning based on binary classifiers.
[14] The AI Winter is a definition used to refer to a specific period in which little to zero attention was given to AI.
[15] Stanford University, *Stanford's Robot Legacy*, Stanford Report, Science & Engineering, January 16, 2019 https://news.stanford.edu/stories/2019/01/stanfords-robotics-legacy

from the camera unit built on the robot and linked to an on-board computer. The instructions were sent to Shakey via tele-printed instructions to navigate it around the room, making it avoid obstacles. As the training continued, the robot was able to keeping tracking of its position thanks to its abilities to count wheel revolutions, so that even if it had an incomplete map it was still able to navigate the unknown space around it. Thanks to its abilities to "learn" the instructions, it did not need a step-by-step guidance: after having learnt the basic instructions (like turning or rolling) it was able to perform more complex tasks that required putting together the basic abilities it had acquired.

Nearly a decade later, in 1989, thanks to NASA's *AutoClass* program used to discover new classes of stars, an exponential acceleration started in the studies on the application of AI, which reached its peak in 1997 with IBM experimentations. This represented a watershed for what concerns common people's view on the capabilities of AI: with the Deep Blue program, which defeated the chess world champion of the time Garry Kasparov, it became clear that the program had incredible learning capabilities; de facto, it was able to analyse 200 playing positions per second. Similar events happened recently: for example in 2010 the IBM Watson software was able to win against three people at the TV show game "Jeopardy!" by delivering answers derived from analysis made in just a few seconds on all the previously memorized data from numerous encyclopedias, dictionaries and articles. Exactly from 2010, in the following two decades, and continuing until nowadays, the positive accomplishments of AI and research about it and its learning methods have grown exponentially. The newly-developed AI innovations could be seen as falling within four main categories[16]:

1. Data-centric AI, focused on the research for high-quality data to be used to train the algorithms, which entails a new way to manage the quantity of data collected and the new typology. Having data of greater quality helps in putting aside ambiguous data samples alongside the outlier data sets and eventual biases, implying obtaining more exact output. Moreover, better outputs are guaranteed by the usage of more recent data and a specific attention for the typology of data used rather than on the volume of data per se. An example could be represented by the usage of data, which is not always obtained from sensors observing the world around, but rather artificially generated (the so-called synthetic data), crucial to create simulation scenarios to make the

[16] Afraz Jaffri, Jackie Wiles, *What's new in Artificial Intelligence from the 2022 Gartner Hype Cycle*, Information Technology, September 15, 2022. https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle

machine adapt to specific new datasets of events. Considering that synthetic data are cheaper[17], they enable the possibility of having larger sets of data to train the AI, and the more the AI is fed with data the more it learns from it. Furthermore, considering that outliers are put aside, this model enables the possibility to achieve high performance with lower quantity of data. The problems that could arise are linked to the lack of representativeness of such data due to the fact that datasets are carefully analysed, as well as the difficulty to check on the quality of data. De facto the data-centric model is based on a systematical change of the dataset to improve the AI performance, with the aim of improving the training dataset.

2. Model-centric AI, which focuses mainly on building highly-performant models of machine learning and deep learning given a specific dataset, then confronting the obtained output with the desired one to understand what needs to be changed in the process of analysis of the machine. The model is changed later on to improve the AI performance. Specifically, it detects and deletes all outliers and error values that can be found in the dataset. Model centric AI is now “obsolete" in specific fields, and the new capabilities of the data centric model pushed developers to revolve more towards this second mode, crucial for new AI development and changes, as demonstrated by the improvements of ChatGPT. The latter has been improved from GPT-3 to the newer version named ChatGPT thanks to the changes made in data quality, minimizing de facto the biased outputs[18].

   As described by Gartner in 2022, such typology of AI innovations include the following main deep learning models:

   - Physics-informed machine learning method is used to train deep neural networks by enforcing specific physical laws. It combines collected data and mathematical models, optimizing specific variables (like time and space) to enable a greater generalization of the function used[19].
   - Composite AI, used vastly within the industrial sector, it does not respond to specific and limited tasks, but works alongside humans. More specifically, it addresses multidimensional problems using its peculiar language and vision abilities. Considering that it also possesses particular adaptation capabilities, it

[17] Ibidem.
[18] MIT, *Introduction to Data-Centric AI, Data-Centric AI VS Model-Centric AI*, Lecture 1, IAP, 2024. https://dcai.csail.mit.edu/2024/data-centric-model-centric/
[19] Karniadakis George Em, Kevrekidis Ioannis G, Lu Lu, Perdikaris Paris, Wang Sifan, Yang Liu, *Physics-informed machine learning*, Nature Reviews Physics 3, 2021.

is ideal to adapt to different markets which have dissimilar standards and regulations[20];

- Causal AI, is mostly used to find out causal relationships in decision making fields by the analysis of causal graphs and simulations, and it directly builds a model to explain the causality relation. It is crucial because it helps with the detailed analysis of the cause and effect relations, and mostly helps with reduction of biases.
- Generative AI, is mostly dependent on LLM (Large Language Models) based on Foundation models. The latter refers to an AI capable to drive all the applications of conventional AI without needing to undergo a training based on task-specific data. As it comes to be trained with huge amounts of data it can generate new "things" in the presence of a prompt requiring it to do so (such as giving answers to questions). This typology of model is crucial because it is fed with huge amount of data and so "foundation models can outperform a classical model that has been trained just on a few data points"[21], and it can help in cases of absence of data for specific tasks thank to its generating abilities. However, even if greatly efficient it has a high computational cost, mainly because, as explained by IBM on the topic, it need multiple GPUs running at the same time.

3. Applications-centric AI, is mainly used for everyday applications in specific sectors such as the automotive (specifically autonomous vehicles), smart robots and similar, based on the concept of decision intelligence. The latter's main aim is to help with decision making, helping with increasing the transparency of businesses which use this AI model in their products to make decisions, alongside helping in reducing unpredictability of outputs
4. Human-centered AI, can be described as a new approach more than a machine learning model per se, which is based on amplifying the capabilities of humans instead of planning a "replacement" via the usage of machines considered to be more efficient and quick compared to humans. This is why, attached to this vision, there is the Responsible AI concept, which stresses the importance of using AI ethically[22].

[20] Sharka Uljan, *Why Composite AI in the Intelligent Age leads us to a people-centred future*, World Economic Forum, Emerging technologies, December 18, 2024. https://www.weforum.org/stories/2024/12/why-we-should-embrace-composite-ai-in-the-intelligent-age/
[21] IBM, *What is generative AI*, April 20, 2023. https://research.ibm.com/blog/what-is-generative-AI
[22] Schneiderman Ben, *Human-centred AI*, Oxford Academic, 2021.

Following a chronological order from 2010 onwards, the increase in capabilities and rapidity of learning of AI came to be crystal clear starting with the well-known IBM Watson, the AI-powered supercomputer that defeated humans in the *Jeopardy!* TV quiz show[23]. Some years later, the *AlphaGo* neural network, able to defeat the world *Go* champions in 2015, 2016 and 2017. Importantly, for the 2017 competition, the AI software was not trained with moves from human games anymore, but directly by making it learn from its own mistakes, using the Reinforced Learning method, which enabled the software to beat the *Go* world champion after just three days of training, thanks to the memorization of the mistakes. It is astonishing to reckon that the AI seems even able to bluff, as it was noticed during some tests in which a poker game was simulated.

In the same years, China was investing on AI development as well, and its crucial research achievements have been showed to the public in 2013 with the Tianhe-2. The Tianhe-2 was a supercomputer developed by China's National University of Defense Technology, and at the time doubled the speed of world's top supercomputer, reaching a peak of approximatively 33 petaflops[24].

One year later, in 2014, Facebook developed the DeepFace facial recognition software, an AI based model which was trained to recognize people's visages with an incredible accuracy of 97.25%; to make a comparison, a the human accuracy is approximatively of 97.53%, which means that the program had a nearly-human accuracy[25] .

In 2016, a new Chinese supercomputer, the Sunway TaihuLight, reached a speed of nearly 93 petaflops[26]: three years before the maximum speed reached by the Chinese Tianhe-2 was of 33 petaflops!

In the meantime, between 2008 and 2018, it grew the trend of voice assistants, starting from 2008 Google speech-to-search iPhone[27], followed by 2011 with *Siri* released from Apple, passing through the 2014 *Alexa* and then ending with the release of *Google Duplex* in 2018.

---

[23] IBM, Watson, Jeopardy! Champion, IBM Heritage. https://www.ibm.com/history/watson-jeopardy

[24] Lawrence Livermore National Laboratory, China's Tianhe-2 is the new world champ of supercomputing, 18th June 2013. https://www.llnl.gov/article/39151/chinas-tianhe-2-new-world-champ-supercomputing

[25] Simonite Tom, *Facebook Creates Software That Matches Faces Almost As Well As You Do*, MIT Technology Review, Artificial Intelligence, March 17, 2014. https://www.technologyreview.com/2014/03/17/13822/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do/

[26] Tsinghua University, The Sunway TaihuLight supercomputer run by Tsinghua University is again ranked first in the Top500 list of supercomputers, 2017. https://www.tsinghua.edu.cn/en/info/1245/4981.htm

[27] Schalkwyk Johan, Beeferman Doug, Beaufays Françoise, Byrne Bill, Chelba Ciprian, Cohen Mike, Garret Maryam, Strope Brian, *Google Search by Voice: A Cade Study*, Google Inc, Amphiteatre Pkwy (Mountain View, California, USA).

In the same year of the Siri release, 2011, Google Brain scored around 75% accuracy in identifying cats[28], which underlines the enormous rapidity that the AI learning system was acquiring.

Going back to 2018, while Stephen Hawking's death revitalized its previous warnings to the world about the risks of AI ending mankind[29], Uber showed to the public its program of self-driving cars in Pittsburgh[30]. In the same period of time IBM, in cooperation with Airbus and the German Federal Ministry of Economic Affairs and Energy launched the CIMON project[31]: the first AI-powered robot ever sent to the space with the aim of assisting astronauts with their daily tasks, possessing de facto the ability to communicate with astronauts. The fact that the robot had a human face was not a case: in 2018 it was launched the robot Lovot, theoretically accessible to everyone, able to sense and affect mood changes in humans, and so with the aim of dialoguing with persons while resembling to humans in dialogues and "thinking".

Going on with the more recent innovations, Microsoft launched in 2019 the Turing-NLG a nearly seventeen billion parameter generative Large Language Model able to generate words, completing unfinished sentences, replying to questions and creating summaries of documents, realizing all these tasks "as fluently as humans can in any situations"[32].

Going onto the medical field, between 2019 and 2020 Google AI deep learning algorithm outperformed radiologists when used to detect the presence of potential lung cancer, while the University of Oxford, during the Covid pandemic, developed the Curial AI test to identify the presence of Covid-19 in a "near real-time prediction"[33].

Alongside the medical field, AI started to be used within numerous other fields, such as in bio-informatics. In particular, DeepMind's *Alphafold*, the first-ever AI to receive the Lasker Award (CASP protein-folding concept) since 1945, has the ability of analysing the shape of the sequence of a protein and to foresee its structure in the 3D space. Such

---

[28] Corso Francesco, Maggiore Ingegnere Esercito Italiano, *Intelligenza Artificiale nella Difesa: le attività di ricerca ed innovazione tecnologica*, Artificial Intelligence Hype, AFCEA Roma, Casa dell'Aviatore, May 3, 2023.
[29] Cellan-Jones Rory, Stephen *Hawking warns that Artificial Intelligence could end mankind*, BBC, December 2, 2014.
[30] ABC News, *Uber Rolling Out Pilot Self-Driving Car Program in Pittsbugh*, August 19, 2016.
[31] Eisenberg Till, *CIMON – The AI-powered robot launches a new era in space travels*, IBM, 2018. https://www.ibm.com/new/announcements/cimon-ai-robot-launches-new-era-space-travel
[32]Rosset Corby, *Turing-NLG: A 17-billion parameter language model by Microsoft*, Microsoft Research Blog, February 13, 2020. https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/
[33] Hinsley Dawn, *New AI test identifies COVID-19 within one-hour in emergency departments*, Oxford News blog, University of Oxford, July 27, 2020. https://www.ox.ac.uk/news/science-blog/new-ai-test-identifies-covid-19-within-one-hour-emergency-departments

technology is crucial as, while scientists would need years to do so, the AI could take just some minutes, giving a response with a guarantee of its exactness of around 97%[34]. The analysis of protein-folding within the space is crucial to understand how such folding couldgenerate illnesses as Alzheimer.

Going towards a "more visible" usage of AI, mainly for the population, OpenAI ChatGPT-3 was able in 2020 to generate human-like texts, fully developed with finalversion of ChatGPT showed to the public with a chat interface (able to receive both texts an images starting from 2023 thanks to ChatGPT-4 multimodal Learning Machine Language), while Dall-E was able to generate images from mere text prompts. In this rapid "chaos" the engineer Blake Lemoine was fired after having declared that AI Lamda system from Google had feelings, de facto "showing human-like consciousness"[35]. Obviously, Google fired Lemoine for the violation of "clear employment and data security policies that include the need to safeguard product information"[36] without addressing directly the fears of experts related to progressive humanization of AI systems designed to impersonate humans.

The AI peak has been reached with the automation of algorithm discover with the 2022 DeepMind Alpha Tensor program.

A revolutionary innovation has been announced in February 2025: Microsoft's "Majorana 1", the world's first quantum processor which is powered by topological qubits. This incredible innovation is able to tackle the limitations of classical computing for AI, linked to the limits in computational resources and amounts of data. The quantum computing would, de facto, help with optimization of such AI complex operations alongside enhancing rapid AI decision-making. The quantum computing, thanks to the qubits, is able to run many calculations at the same time, and the topological qubits have been described by Microsoft as been more stable and less prone to errors than classical qubits. This means that the processing speed of AI would increase incredibly, accelerating also the development of AI applicatives, while AI algorithms would be greatly optimized. Moreover, Quantum neural networks will be able to process data quickly and in a better way, helping AI making more efficient decisions and judgements, crucial for fields where

[34] Professor Blerina Sinaimeri during the conference *Intelligenza Artificiale e Intelligenza Umana a Confronto*, organized by the Embassy of the Republic of Albania to the HolySee and the Sovereign Order of Malta, May 21, 2024.

[35] Wertheimer Tiffany, *Blake Lemoine: Google fires engineer who said AI tech has feelings*, BBC News, July 23, 2022 https://www.bbc.com/news/technology-62275326

[36] Ibidem.

AI is required to make decision-making. Obviously, quantum computing technology is still highly costly, and it would also be expensive the adaptation of current AI systems with quantum-based AI.

Such innovations are creating more debates on the ethic side of the AI usage, above all for the sphere of copyrights alongside the risk of having deep fakes all around the web. Numerous doubts on the protection of privacy have been expressed, as there is great suspicion on how AI systems detect the word that activates them (for example, numerous people wonder if *Alexa* is always listening conversations, so that it can be able to recognize its own name when consulted).

All these incredibly quick developments of AI marked the shaping process of modern AI definition. The US Department of Defense captures broadly its modern capabilities by stating that: "Artificial Intelligence refers to the ability of machines to perform tasks that normally require human intelligence – for example recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action – whether digitally or as the smart software behind autonomous physical systems".[37]

Finally, having discovered new technology able to automate algorithm discovery and creation, will for sure help with the full-automatization of AI models which risks of giving rise to debates about its usage within truly delicate fields, mainly the military sphere; in fact, AI will soon become more efficient and quick thanks to the quantum computing technology revealed recently at the beginning of 2025.

1.3 History of the Development of military-AI

Going more into details of the AI usage within military technologies alongside the defence and military field, the main evident and widely-known even that could be identified as starting point for the usage of AI within military technology is the murder of Mohsen Fakhrizadeh on November 27, 2020. Fakhrizadeh is considered to the father of the Iranian nuclear research, and he was at the time the Vice-minister of Defence of the Islamic Republic of Iran. After the initial declarations from the Iranian media that he had been killed by "remote-controlled machine gun" or weapons "controlled by satellite", more specifically a remote attack carried on using "special methods" as declared by Rear

[37] Department of Defense of the United States of America, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, Harnessing AI to Advance Our Security and Prosperity, Page 5.

Admiral Shamkhani, who heads the Supreme National Security Council[38], it has later been demonstrated that the technology used for the attack was supported by AI. As explained by the AI researcher Laure de Roucy-Rouchegonde, the machine-gun had no snipers whatsoever, which means that it was technically controlled by a robot equipped with sensors to adapt the targeting during the mission. The AI program used there had a facial recognition function, to be able to recognize the victim's face before carrying out the attack, ready to rectify the trajectory or the time frame if needed. The usage of an AI-drone revealed as crucial for the Israeli intelligence forces because it avoided any troubles that it could have been triggered by sending directly physical agents on the site, difficult to be extracted in the event of failure or trouble[39]. The weapon used for this killing is a drone, specifically an Unmanned Aerial Vehicle, generally divided into two main categories: drones for recognition and surveillance, alongside armed drones; this second category is the one that best represents the drone used against Fakhrizadeh. Even if these technologies seem truly recent, in reality they have been developed way before the 2020s. The "hardware" part, so the "body" of the military technology which is enhanced with traces back its origins from the Nikola Tesla's radio-controlled boat in 1898, shown at the Electrical Exhibition in Madison Square Garden (US patent #613,908) on November 8, 1898. This invention was crucial because it enabled remote control without the need for cables or wires or any other physical connection between the person charged with controlling the boat and the latter in itself. The boat was 4 and a half feet long, and it was equipped with a transmitter for remote controlled called "Tesla Oscillator" able to emit "radio frequency pulses through a tuned circuit and antenna", while "the receiver had a matching tuned circuit to recover these pulses"[40]. Tesla tuned perfectly the circuits to improve the sensitivity of the receiver with the aim of eliminating any interference deriving from the boat's electric motor in the boat. This technology revealed as the precursors of the first land unmanned vehicles deployed starting from 1915 by the US Army; the first one was known under the name of "electric dog", developed by Hammond and Miessner to keep care of bringing supplies to the militaries in the trenches. It was able to go towards light sources and to target enemy anti-aircraft batteries using

[38] Gardner Frank, *Mohsen Fakhrizadeh: Iran scientist "killed by remote-controlled weapon"*, BBC, November 30, 2020. https://www.bbc.com/news/world-middle-east-55128970

[39] Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle*, Quand les machines prennent les armes, PUF – Presses Universitaires de France/Humensis, pp. 7-9, October 2024.

[40] Young Ken, *Tesla and the First Remote-controlled Boat*, The Model Yacht, Vol 22, No 3, 2021. https://usvmyg.org/history-articles/tesla-and-the-first-remote-controlled-boat/

searchlights at night, while also possessing the ability to adapt its behaviour and movments depending on the information that its sensors received from the surrounding environment[41]. Alongside the development of the technology, it came also the invention of the name "robot", by Karel Capek, from the Czech verb *robota* in the 1920s[42]. Later on, in 1928, it was developed the so-called Philidog by the French engineer Henri Piraux, which specifically had photoelectric cells at the place of the eyes (de facto sensors able to capture any change in the surrounding environment), possessing a motor to move depending on the light signal it received. Specifically, a source of light proected on one of the two eyes resulted in a movement of the robot, and its origin was what gave the direction of movement to the robot, reaching the completion of its mission when it physically reached the source of light[43].

Moreover, these robots had the ability to learn thanks to a magnetic-tape memory, alongside communicating via the usage of two lamps (respectively green and red) on its head[44]. The most evident improvement of these initial forms of "early-automated" robots arrived with the development of Shakey the robot between 1966 and 1972 by the Stanford Research Institute, defined as "The first Electronic Person". The crucial characteristic of Shakey was related to its brain, with a capacity of more than 7 million "bits" of information, alongside having the ability of performing specific processes like ''exercising the powers of judgment" and "'reflecting for the purpose of reaching a conclusion". Shakey can be considered as the precursor of AI thanks to its learning abilities via analogy, which helped the robot in recognizing patterns among data received via its sensors, specifically an optical range finder, telemetering equipment and touch-sensitive antennae. It also had the ability to re-adapt its ongoing movements depending on the inputs it received from the surrounding environment: for example, "when the space between the wall and the desk is too small to ease through, Shaky is smart enough to know it and to work out another way to get where he is going"[45]. The problem that was arising was related to the way the robot could have been able to solve more and more

[41] La Robotique aux cycles 3 et 4, *1914-1953 Robots de déuxieme génération*. https://atelier-canope-95.canoprof.fr/eleve/Automates%20et%20robots/res/robot.dossierHtml/co/1914_50robotsCapteurs.html
[42] Jordan John M., *The Czech Play That Gave Us the Word "Robot"*, The MIT Press Reader, July 29, 2019. https://thereader.mitpress.mit.edu/origin-word-robot-rur/
[43] La Robotique aux cycles 3 et 4, *1914-1953 Robots de déuxieme génération*. https://atelier-canope-95.canoprof.fr/eleve/Automates%20et%20robots/res/robot.dossierHtml/co/1914_50robotsCapteurs.html
[44] Musée des arts et métiers, *La cybernétique, projet de l'après-guerre* : Job, le renard électronique, 1953. http://robots.arts-et-metiers.net/lexposition/nom-de-code-robota/la-cybernetique-projet-de-lapres-guerre.html
[45] Kuipers Benjamin, Feigenbaum Edward A., Hart Peter E., Nilsson Nils J., *Shakey: From Conception to History*, p. 92, 2017. https://ai.stanford.edu/~nilsson/OnlinePubs-Nils/General%20Essays/Shakey-aimag-17.pdf

complicated problems via a specific way of reasoning at the basis of AI: not anymore the trial and error system, but more a process based on sorting out the fitting answers going from the general categories first, going on until the system finds the correct category. At that moment, the system works with it to find a fitting solution, which is not perfect but acceptable, to be able to overcome the obstacle. A huge part of these capabilities depended also on the memory capacities, as big memories are crucial for the development of complex intelligence, crucial to analyse a massive and continuous flow of information and data. In the meantime, such autonomous robots were then researched upon more and more within the military, which implemented the aforementioned systems, improving them to reach a greater level of autonomy, to be used in the battlefield. In fact, in 1958, the US DoD gave birth to the widely known DARPA project (initially known as Advanced Research Projects Agency) with the aim of facilitating the research and development of military and industrial strategies[46]. In the subsequent years, during the 1960s, the US DoD started training computers to mimic basic human reasoning[47]. Specifically, these systems have been implemented in the form of drones, used between the two World Wars (even if there still was no AI being researched upon). The first recorded usage of drones is during the Korean War by USA, where the most known was the Havilland DH.82, alternatively defined as Queen Bee (a British-manufactured drone), which gained such a peculiar name due to the sound it emitted, similar to the one of bees. This first usage gave great impulse to further similar drone-based operations in the subsequent conflicts, such as in the Vietnam war (for recognition purposes), during the Kippour war and successively in Liban with the israeli invasion in 1982. Specifically, as affirmed in September 2016 by David Harari, a former head of Israel Aerospace Industries (IAI) in an interview to the magazine "Atelier" about the origins of drone technology: "I'd like to make clear that unmanned aerial vehicles already existed before Israel started looking at the subject", but he further underlined that Israel "developed the very first operation system"[48], referring to the drone usages in 1973 and 1982. Precisely, during the Kippour War in 1973 Israeli Air Force used the Firebee and Chukar drones

[46] Britannica, *Defense Advanced Research Projects Agency*, United States government. https://www.britannica.com/topic/Defense-Advanced-Research-Projects-Agency

[47] Military Embedded Systems, *Artificial Intelligence Timeline*, January 24, 2019. https://militaryembedded.com/ai/machine-learning/artificial-intelligence-timeline

[48] Frantzman Seth J., *How Israel Became a Leader in Drone Technology*, Middle East Forum, The Jerusalem Post, July 13, 2019. https://www.meforum.org/how-israel-became-a-leader-in-drone-technology

with the aim of "distracting, confusing and stimulating the enemy radar"[49]. From that moment onwards, just until the war in Lebanon in 1982, Israel enlarged enormously its unmanned aircraft fleet, massively used against Syrian air defences in Bekaa Valley, flooding the Lebanese airspace via "crowding" Syrian defence systems' radars, which allowed Israeli forces to launch missiles. Specifically, the drone used there was the Zahavan, a scout drone with a 22 kg-weight able to fly for several hours. Alongside the Scout drones, Israeli forces developed the Tadiran Mastiff, used alternatively in 1980 and catching also the US attention. This is exactly what brought Israel and the US to cooperate in the development of the AAI RQ-2 Pioneer (1986)[50], revealed as crucial for the US during the 1991 Gulf War. Full usage of drones came in 2001, with the US War on Terror carried out via drone strikes, accomplishing military missions via the usage of the US General Atomics MQ-1 Predator[51], in Iraq, Yemen, Libya and Afghanistan, while drones started to be massively used also during the Balkan Wars[52]. Such experiences on the battlefield gave subsequent birth to the widely known Predator, whose first usage for a lethal operation was in Yemen, targeting six Al Qaeda suspects[53]. This all was possible thanks to a shift in the work on machine learning from knowledge-driven approaches to data-driven ones, which means that programs did not stop to the mere analysis of enormous quantities of data, but also started to dra conclusions from the results. Between the 1990s and 2000s, AI started to be used by the US DoD not only for "mere" striking or targeting purposes, but also for crucial transportation aims. It was in fact developed the DARPA-funded Dynamic Analysis and Replanning Tool (DART) in 1991, a system able to "schedule the transportation of supplies or personnel and to solve other logistical problems"[54], de facto helping with the decision-making systems in the US for battlefields needs. The automation then was applied to vehicles, specifically in 2005 a team from the Stanford AI laboratory developed au autonomous vehicles called Stanley, which was the

[49] Rogoway Tyler, *Meet Israel's "Suicide Squad" of Self-Sacrificing Drones*, The War Zone, September 29, 2021. https://www.twz.com/4760/meet-israels-suicide-squad-of-self-sacrificing-drones#:~:text=In%201982%2C%20when%20open%20hostilities,hours%2C%20it%20did%20just%20that.

[50] National Air and Space Museum, Smithsonian, *Pioneer RQ-2A UAV*. https://airandspace.si.edu/collection-objects/pioneer-rq-2a-uav/nasm_A20000794000

[51] United States Air Force Website, *MQ-1B Predator*. https://www.af.mil/About-Us/Fact-Sheets/Display/Article/104469/mq-1b-predator/

[52] Horowitz Michael C., Kreps Sarah E., Fuhrmann Matthew; *Separating Fact from Fiction in the Debate over Drone Proliferation*, MIT press Direct, International Security, Vol 41, Issue 2, pp. 7-42, 2016. https://doi.org/10.1162/ISEC_a_00257

[53] ABC News, *U.S. Predator Kills 6 Al Qaeda Suspects*, November 5, 2022. https://abcnews.go.com/WNT/story?id=130027&page=1

[54] Military Embedded Systems, *Artificial Intelligence Timeline*, January 24, 2019. https://militaryembedded.com/ai/machine-learning/artificial-intelligence-timeline

first autonomous vehicle to complete a 132-mile course in the Mojave Desert, which made it win the DARPA Grand Challenge; the latter aimed at stimulating invention for the development of a "future fleet of driverless military ground vehicles. Congress funded the competition to support its directive that one-third of U.S. military ground vehicles be unmanned by 2015"[55]. Thanks to the integration of AI, the vehicle Stanley was able to navigated mapped terrain while overcoming unmapped obstacles in real time. "It integrates a course map expressed in about 3,000 points of latitude and longitude, stored memory of past experiences, and new information about the road ahead gathered from roof-mounted laser sensors, video cameras, radar and GPS receivers"[56]. Solely two years later, AI professor Noel Sharkey starts warning about an "emerging arms race among high-tech nations to develop autonomous submarines, fighter jets, battleships, and tanks that can find their own targets and apply violent force without the involvement of meaningful human decisions"[57].

The exponential growth in the usage of drones and other AI-based military technologies, more and more automated thanks to the introduction of AI starting from the 2010s decade onwards, triggered a global debate on the ethical and juridical level about these weapons. Specifically starting from 2013, the UN Special rapporteur on extrajudicial, summary or arbitrary executions Philip Alston (followed by Christof Heyns within the same role) altogether with the Special Rapporteur for the promotion and protection of human rights and fundamental freedoms while countering terrorism Ben Emmerson produced rapports to highlight the non-legality from a juridical point of view of the usage of such technologies for targeted killings. Such debates about the judicial and ethical base of unmanned ant autonomous lethal weapons increased enormously but de facto, did not bring to any practical conclusion on a specific legislation to regulate the usage; instead, such technologies are becoming widely "accepted" by the general public and by soldiers as a result of their daily usage in the context of the War in Ukraine.

In 2014, just one year later from the report from the various UN special rapporteurs, US DoD unveiled its "Third Offset Strategy", highlighting that advancing in AI research

[55] National Museum of American History, *"Stanley" Robot Car*, Behring Center, Smithsonian. https://americanhistory.si.edu/collections/object/nmah_1377824
[56] Ibidem.
[57] Noel Sharkey, *Robot wars are a reality*, The Guardian, August 18, 2007. https://www.theguardian.com/commentisfree/2007/aug/18/comment.military

would have revealed as crucial from that moment on to define the future warfare[58]. Such announcement has been followed by the famous 2015 open letter from Steven Hawking, Elon Musk and specifically from 2981 other AI and robotics researchers alongside 17.190 other endorsers asking for a ban on the development of autonomous lethal weapons.

Link to the aforementioned letter from the Campaign Stop Killer Robots: https://www.stopkillerrobots.org/wp-content/uploads/2013/03/FLI_LtrJuly2015.pdf

While researchers and experts were asking for a ban on such weapons, governments like the UK one opposed a ban on lethal autonomous weapons during a UN discussion on LAWS (Lethal Autonomous Weapons Systems), while underlying the crucial need for developing arms to be always put under human control[59]. In this tense international context, the US DoD increased its budget for investments in AI, big data, and cloud computing increased to over US$600 million (reaching nearly US$1.8 billion in 2024, with more than 685 active AI projects currently underway)[60]. In 2017 it was released to the public the news about the development of the Sea Hunter, an autonomous U.S. warship, able to operate unmanned (so without any human presence in it) for extended periods of time at sea. In the same year, the Harvard Belfer Center analysed the risks and dangers that the implementation of AI could have generated when implemented within the nuclear system, specifically into nuclear weapons[61]. While the US DoD developed Project Maven, a system able to autonomously find "targets" of interest from video foootages and photos collected by drones and cameras thanks to a combination and machine-learning and neural-learning network[62], China published its "Next Generation Artificial Intelligence Development Plan" underlying its goal to reach "world domination in AI by 2030"[63]. In the same year Putin declared that "whoever reaches a breakthrough

[58] Gentile Gian, Shurkin Michael, Evans Alexandra T., Grisé Michelle, Hvizda Mark, Jensen Rebecca, *A History of the Third Offset, 2014–2018*, Rand Corporation, 2021. https://www.rand.org/content/dam/rand/pubs/research_reports/RRA400/RRA454-1/RAND_RRA454-1.pdf

[59] The Guardian, *UK opposes international ban on developing "killer robots"*, April 13, 2015. https://www.theguardian.com/politics/2015/apr/13/uk-opposes-international-ban-on-developing-killer-robots

[60] Sebastian Clapp, European Parliamentary Research Service, Members' Research Service, Briefing, *Defence and Artificial Intelligence*, PE 569.580, April 2025. https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/769580/EPRS_BRI(2025)769580_EN.pdf

[61] Allen Greg, Chan Taniel, *Artificial Intelligence and National Security*, Harvard Kennedy School, Belfer Center for Science and International Affairs, A study on behalf of Dr. Jason Matheny, Director of the U.S. Intelligence Advanced Research Projects Activity (IARPA), July 2017. https://www.belfercenter.org/publication/artificial-intelligence-and-national-security

[62] Pellerin Cheryl, *Project Maven to Deploy Computer Algorithms to War Zone by Year's End*, US Department of defense, July 21, 2017. https://www.defense.gov/News/News-Stories/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/

[63] Chinese State Council, *Notice of the State Council on Issuing the Development Plan for the New Generation of Artificial Intelligence*, Guofa 2017, No 35, 000014349/2017-00142, July 20, 2017. https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

in developing AI will dominate the world"[64] and predicts that future wars will be fought by drones. Whilst AI was presented to the public as an instrument of which experts mastered technicalities, and so ready to be developed and used within the defence sector, the AI developers from the Facebook AI Research Lab (FAIR) found out that the AI-based chatbots had developed their own language, incomprehensible to humans, and started communicating one with each other using this new lamguage, de facto developed without any human intervention[65]. Nevertheless, programs centered on understanding learning methods of generative way alongside methodologies to improve such learning processes, it has been developed by the UTSA Cyber Center for Security and Analytics a new cloud-based learning platform for AI to teach machines to learn like humans.

Looking at the fast pace with which such researches have been advancing, Google released a new AI policy for the usage of its own AI products to avoid any possibility of implementation of its systems and products within weapons, while declaring at the same time that it would have continued to cooperate with the military. While initially Google tried to put some distance from its products to the military sector, promising between in 2018 that it would have not allowed the usage of its products for military purposes[66] in any way able to "cause overall harm", it has recently dropped such promises in February 2025[67]. Google's updated AI principles page includes provisions that say the company will use human oversight and take feedback to ensure that its technology is used in line with "widely accepted principles of international law and human rights." The principles also say the company will test its technology to "mitigate unintended or harmful outcomes." As it is possible to notice, the verb now used is "mitigate" instead of the 2018-idea of avoiding any harmful outcome, which underlines a clear change in the direction of the main AI companies, ready to help governments with the development of their national security agendas.

Alongside Google newly undertaken direction, the other "Big" of the AI sector, Open AI, openly declared in december 2024 its readiness its partnership with Anduril technologies

[64] Президент России, *Форум профнавигации «ПроеКТОриЯ»*, События, September 1, 2017. http://kremlin.ru/events/president/news/55492
[65] Antolini Diego, *A.I. – Il Futuro Appartiene alle Macchine?*, Ecce Machina, 2019.
[66] Hooker Lucy, Vallance Chris, *Concens over Google ending ban on AI weapons*, BBC, February 5, 2025. https://www.bbc.com/news/articles/cy081nqx2zjo
[67] The Washington Post, *Google drops pledge not to use AI for weapons or surveillance*, February 4, 2025 https://www.washingtonpost.com/technology/2025/02/04/google-ai-policies-weapons-harm/

to be implemented in unmanned aerial vehicles[68], by clearly stating that the main aim is the support to the US DoD decision-making system: "Our partnership [Anduril partnership's] with OpenAI will allow us to utilize their world-class expertise in artificial intelligence to address urgent Air Defense capability gaps across the world. Together, we are committed to developing responsible solutions that enable military and intelligence operators to make faster, more accurate decisions in high-pressure situations"[69]. While it gets openly stated that AI comes to be used to support the decision-making system alongside other specific restricted tasks, it seems to be generally forgotten to publicly mention that even from an operational point of view the employment of AI guarantees a way greater advantage rather than using it for mere "strict" tasks. Just to cite un example, in August 2020, during a virtual reality simulation, an AI-controlled F-16 developed by DeepMind, was able to beat a human F-16 pilot 5 to 0 in a dogfighting competition organized by the US DARPA[70].

It is now crystal clear that such military AI systems are able to beat human capabilities. However, considering their evolving ability to learn by themselves, their ability to develop a specific language understandable only by AI systems, and algorithms becoming increasingly black boxes to AI experts, what is the limit in their usage, mostly within the defence field, when and where human lives are at stake? This is why Fanni, Giancotti and Taddeo, in their book "Guerre di Macchine", proposed a modulation of the AI lifecycle, where the various steps and phases are clearly analysed to avoid any lack of understanding of the AI activities or directly ethical risk. They built a model of the AI lifecycle by using the LoAEthics, which stands as follows[71]:

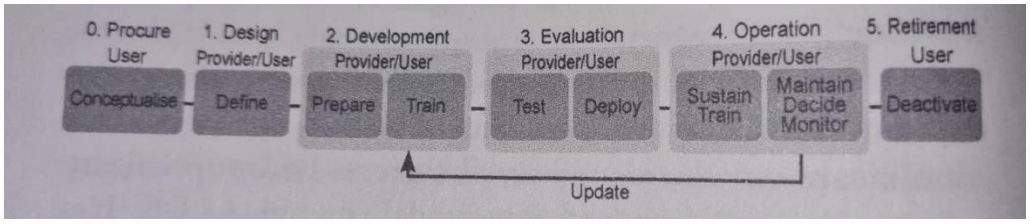

Image B: Representation of AI lifecycle
Source: Fanni Rosanna, Giancotti Fernando, Taddeo Mariarosaria, Guerre di macchine, Intelligenza artificiale tra etica ed efficacia, Guerini e Associati, May 2024.

[68] Anduril Industries, *Anduril Partners with OpenAI to Avdance U.S. Artificial Intelligence Leadership and Protect U.S. and Allied Forces*, December 4, 2024. https://www.anduril.com/article/anduril-partners-with-openai-to-advance-u-s-artificial-intelligence-leadership-and-protect-u-s/
[69] Ibidem.
[70] Aerospace Global News, *AI F-16 pilot beats human in virtual dogfight*, April 2024. https://aerospaceglobalnews.com/news/ai-f-16-pilot-beats-human-in-virtual-dogfight/
[71] Fanni Rosanna, Giancotti Fernando, Taddeo Mariarosaria, *Guerre di macchine, Intelligenza artificiale tra etica ed efficacia*, Guerini e Associati, p. 108, May 2024.

1.4 Collective views and fears on AI

Looking at AI in these times, it is becoming more evident how it is pervading every aspect of human lives. The risk of it escaping from the control of its own creator, the human being, is becoming more evident, and for sure without clear norms on the issue, uncertainty becomes rampant. AI in the current legal framework, can be seen as a kind of Golem, not truly in biblical terms but more with the modern conception of a machine without conscience under the control ofthe human being, which can become hostile to him if put in specific situations. The main issue is that an AI-based machine does not possess a conscience, but it is simply programmed by men, which means that it learns from the information that it is put into its system. Every bias that has been found in AI programs depends completely on the mistakesmade by programmers or in the unbiased data with which it has been fed. The machine, during the learning period, will create its own learning method, the so-called "machine learning", which will result as puzzling for the man, progressively more autonomous as it learns from its own mistakes and database information.

The issue that is going to ensue is that, in such a way, it seems that the experts who programme AI would appear to project their expectations on the software, de facto anthropomorphizing the AI. More specifically, with the birth of systems theory it was noticed how every different system follows a loop, in which there is a return of the information (the input) that gets collected and integrated within the system, which at the end alters the following events and subsequently the behaviour of the users present in the system (humans, animals or machines). As the behaviour of the machine has been explained as learning from its own mistakes and the inputs it receives, an analogy started to be realized with the human being, who act depending on the output they receive after a specific action: if a prize is received, the actionin that specific context will be recorded in the mind as positive and probably replied in similar conditions, while the opposite will happen if the result of the action is negative. This learning process is defined as Reinforced Learning, and it is crucial not only for human beings, but also for the AI.

By considering that, besides being programmed, the machine needs input to start responding, it can be said that initially, before it starts to learn from its own mistakes, it is dependent onthe information given by humans. Regarding this subject, Norbert Wiener expressed in his book "The Human Use of Human Beings" his fears on a possible future

in which the man becomes just the sensor of the machine. In fact, as AI programs assist more and more the man with their own everyday tasks, they are making crucial advancements possible, as for example within the medical field, with the development of applications able to guide doctorsthrough the analysis of the symptoms of the patient to establish the correct diagnosis. On theother side, considering the downside of the coin, men risk of becoming more dependent on the expertise of the machine and less on their own knowledge and experience, leading to a hazardous process of deskilling of the man in the long term. On a similar scale, the usage of terminology typically used to refer to human beings, such as "training", "neural networks", as well as "artificial intelligence" per se, which refers to an intelligence that resembles the human one, probably helps in developing a wider acceptance of this typology of technology. Such humanization of technology goes side by side to a dehumanization of war, specifically when AI is applied to military technology used in the battlefield, simply because there is a critical reduction of the number of men within the battlefield. As underlined by Micheal Walzer in his book "Just and Unjust Wars": "Without the equal right to kill, war disappears as an activity submitted to rules, being replaced by crime and punishment, dark machinations and law enforcement by militaries"[72].

De facto, such a continuative and unregulated recourse to the usage of AI would make the man feel it has always at his own disposal all the information needed, which means that the human will feel, in the longterm, that he has no need to understand reality, as he can directly control it. As explained by Professor Benanti[73], every time man is dealing with this cybernetic dimension, it should be taken into consideration that there is a dimension related to controlable to influence our thinking process and opinions. Nowadays, in social media, it is becoming increasingly difficult to distinguish real images from fake ones created by AI programs AI. This should sound an alarm on the amount of power that the experts on the field have on the population, as a modified image could be used as a joke, as the recent fakepicture of the Eiffel Tower completely devastated by the fire which went all over the web, or to influence people's opinions on wars and conflicts.

At the societal level, it comes natural to wonder which are the aims being followed by the

[72] Walzer Michael, *Guerres Justes et Injustes*, Paris, Gallimard (Folio essais), p. 111, 2006. Retrieved from: Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, p. 175, October 2024.
Original version: "Sans l'égalité du droit de tuer, la guerre disparaîtrait en tant qu'activité soumise à des règles pour être remplacée par le crime et le châtiment, de sombres machinations et l'application de la loi par les militaires".

[73] Professor Paolo Benanti is the actual President of the Italian Commission on Artificial Intelligence and Information for the Presidency of the Council of Ministers.

programmers and developers of AI as there is a crucial lack of legislation on the matter, and which could be the influences on the geopolitical international chessboard. This profound technological transformation, which is the cause for an important behavioural automation, could influence the organization of human societies, bringing the man to wonder if "democracy will survive big data and artificial intelligence"[74]. "In August 2017, Vyacheslav Polonski, a researcher at Oxford University, asserted in a World Economic Forum article, that artificial intelligence *silently took over democracy*, citing the impact of A.I.-powered digital advertisements, social media platform power and mass communication spoilers (botsand trolls) on political processes"[75]. For what concerns this last point, so the political sphere,it is now clear how social media platforms are acquiring more a role on the transmission of news and information. The way the algorithm has been developed will for sure have an impact on the modality in which the information is vehiculated all over the internet. As the main aim of the algorithm is to feed the users with info they like, so that they spend more time on social media platforms, algorithms collect and exploit personal information, and usethem to personalize the personal main home page and content presented by the platform, toguarantee a specific revenue on the time that the user spends on the application and interacts with advertisements. To guarantee a maximisation of the engagement, the algorithm has nointerest in the typology of response it gets, but just on the response per se, which highlightsan interest on the submitted topic. This will result in a categorization of users depending ona series of common patterns that standardize the profile, the so-called "behavioural surplus"[76] that make it possible to predict people's next moves. "This personal data hybridity, cross-matched with people with similar search histories, tastes, and online order patterns creates the information infrastructure of mass surveillance and become the largest ever pool of social monitoring and tracking"[77]. Systems of collection of biometric data, if not clearly regulated, pose serious threats to privacy. It should be underlined that the problem is not the biometric data per se, but insteadthe way in which biometric data are used. There are several privacy and security threats that can derive from such misuse: function creep, covert collection,

[74] Dirk Helbing, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen Van Den Hoven, Roberto V. Zicari, Andrej Zwitter, Scientific American, *Will Democracy Survive Big Data and Artificial Intelligence*, February 25, 2017.

[75] Hamid Akin Ünver, *Artificial Intelligence, Authoritarianism and the Future of Political Systems,* Centre for Economics and Foreign Policy Studies, Cyber Governance and Digital Democracy 2018/9, July 1, 2018.

[76] Shoshana Zuboff, *The Age of Surveillance Capitalism, The Fight for a Human Future at the New Frontierof Power*, Public Affairs, October 4, 2018.

[77] Hamid Akin Ünver, *Artificial Intelligence, Authoritarianism and the Future of Political Systems,* Centre for Economics and Foreign Policy Studies, Cyber Governance and Digital Democracy 2018/9, July 1, 2018.

secondary information, consent. For what concerns the function creep, it refers to the use of information that was collected for a totally different purpose, without notifying the individual. As explained by the Australian Government Office of the Victorian Information Commissioner, an example of such situation could be represented by an organisation that collects an employee's facial biometric information for authentication purposes, such as to enable access to a building, butthen uses these previously collected biometric information to monitor that employee's startand finishing times. The second one, the covert collection, refers to all those biometric data collected passively, so that the 'owner' of such information is not aware of the fact. This is happening more and more thanks to advancement of technology, such as collection of other people's faces present in a crowd of people even if they have any link with the target of the search. The third one, secondary information, refers to all those data which are inferred indirectly from principally collected biometric data, such as health data inferred from a simple picture. The last one is consent, which underlines how often people's biometric data are collected indirectly or inferred, so that the individuals are not aware of the fact that their biometric data got stored somewhere by someone.

Moving from a political and legal sphere to a historical context, it is clear how the majority of political discourses on AI started with the interest in the national defence sphere right after the new millennium, so with the beginning of the year 2000. The terrorist attack on the TwinTowers in New York in 2001 shocked the world, and it started to be vehiculated the message that to tackle the issue of terrorism stricter measures should have been implemented; here AI systems come to be crucial. Every country is developing such technology following a different approach, and the latter could be subdivided depending on the geographical zone[78] taken into analysis. For the case of the USA, the main aim seems to be the one of developing an intelligence thatr esembles the human one. The approach in this case is generative, as it is necessary to find in advance limits and vulnerabilities of the software, de facto red-teaming[79] it. In Asia, the approach on AI is conservative, which means that it is mainly used for politicalscopes, to be able to maintain the already existent societal status quo. The aim is to "give the good example", so as to guide the population and various governmental and non- governmental agencies towards

[78] Division into different geographical zones explained by Doctor Giada Pistilli, AI ethicist, during the TEDx Verona *AI generative: uno sviluppo etico è possibile?*, September 11, 2023.

[79] Red-teaming practices are simulations of attacks arranged on a system to discover its own vulnerabilities as cybersecurity measures, to subsequently fix the issues.

an optimal application and use of AI-powered software andmachines. In Europe, the discourse to be done is a bit different, as the main approach is normative, which means that it focuses more on the practical cases of application, as proved by the EU AI Act, where there is the definition of specific cases for its everyday application.

Chapter 2: Technicalities of AI

2.1 Capabilities and functioning modalities

When considering AI applied to the military sector, there are specific recurring topics that are analysed to underline the positive and negative sides related to the introduction of AI within weapons or simply the usage of AI for specific tasks, such as logistics. The main debates revolves then around the following points[80]:

- The issues related to responsibility, transparency and interpretability, specifically the difficulty in establishing specific national and international legislation on the topic alongside technical requirements;
- The functioning of its logic and learning methods, which are slowly being identified by the "black box" definition, to describe the difficulty in understanding the specific way in which the AI system was able to obtain a precise output (mostly with the generative models);
- The possibility and ability to verify, control and correct errors or biases, not solely for a "mere" discourse of justice and equity deriving from specific usages (such as the biases linked to the images and video footages, haltering the exactness of the final output), but mostly because of the crucial implications that even a small mistake could have on the final result when human lives are at stake;
- The problems revolving around privacy and security of data; specifically, how the collected data get elaborated and analysed, the way in which they are stored and if these clouds are safe from data breaches. The last point reveals to be the most crucial one in the military context, because an intrusion within the AI-based system of a specific military technology could be disastruous.

[80] Fanni Rosanna, Giancotti Fernando, Taddeo Mariarosaria, *Guerre di macchine, Intelligenza artificiale tra etica ed efficacia*, Guerini e Associati, p. 11, May 2024.

The risks linked to the technological and ethical side are becoming more numerous and voiced at the international level. This is the reason why several discussions have succeeded one another in the last decade with the aim of reaching a vast agreement on a sort of legislation (well beyond the idea of developing a series of guiding rules) that could represent the basis for the construction, training and usage of such systems, or even a direct ban on the creation of such technologies.

Before going details of a more technical analysis, it is of paramount importance to understand what truly is military AI, how widely it is developed, applied, used, trained alongside the need to understand which are the economical and geopolitical interests and risks around such topic. The magnitude of the interest in AI is growing at a truly exponential pace, with a global annual spending amounting to approximatively US$235 billion in 2024, expected to grow at a level of approximatively US$631 billion by 2028[81]. Such investments on AI fall within specific national frames and still developing legislation. At the moment, there are approximately sixty countries that have developed and published their own national AI strategies, where the majority of them its pushing for the application of such investments within the military sector, as proved by general estimates of military spending on AI, which seems to have increased from US$4.6 billion in 2022 to US$9.2 billion in 2023, forecasted to reach US$38.8 billion by 2028[82]. In this context, there are specific countries that can be identified as the leading one in research, development and training of AI systems for defence and military purposes: China, Russia and the United States, defined as "tier one" military AI powers[83]. Further important players in the sector that deserve to be mentioned are France, Germany and the United Kingdom within the European Sphere, with specific national plans on the application of AI within military technology, alongside India and South Korea within the Asian continent. The pace at which AI researches are being pursued is so fast that Ursula Von

[81] IDC, *IDC's Worldwide AI and Generative AI Spending – Industry Outlook*, Artificial Intelligence and DAAS, August 21, 2024. https://blogs.idc.com/2024/08/21/idcs-worldwide-ai-and-generative-ai-spending-industry-outlook/

[82] Simpson Kurtis H., Paquette Samuel, Racicot Raphael, Villanove Samuel, *Militarizing AI: How to Catch the Digital Dragon?, AI holds the potential to fundamentally redefine modern warfare*, Center for International Governance and Innovation, February 26, 2025. https://www.cigionline.org/articles/militarizing-ai-how-to-catch-the-digital-dragon/

[83] Hynel Nik, Solovyeva Anzhelika, *Militarizing Artificial Intelligence*, Theory, Technology, and Regulation, 2023.

Der Leyen declared on May 20, 2025, that: "[…] We thought AI would only approach human reasoning around 2050. Now we expect this to happen already next year"[84].

At this point, going more into the understanding of the topic of this thesis, it is necessary to understand what AI, mostly military AI, truly is. Canadian DND/CAF defines it as "The capability of a computer to do things that are normally associated with human cognition, such as reasoning, learning, and self-improvement"[85]. Canada focuses on the "similarities" and characteristics of AI that resembles the human mind; NATO, on the same line, after defining it as "the ability of machines to perform tasks that typically require human intelligence"[86], referring to its ability in "recognising patterns, learning from experience, drawing conclusions, making predictions or taking actions or taking action – whether digitally or as the smart software behind autonomous physical systems"[87]. It then goes on by explaining that generally AI, when applied within the technology pertaining to the defence sector, can be categorised in three broad typologies of application (crucial for the definition of different levels of risk and adapted policies):

- "Enterprise AI includes applications such as AI-enabled financial or personnel management systems, which are deployed in tightly controlled environments, where the implications of technical failures are low (in terms of immediate danger and potential lethality).
- Operational AI**,** by contrast, can be deployed in missions and operations, i.e. in considerably less controlled environments and such that the implications of failure may be critically high. Examples include the control software of stationary systems or those of unmanned vehicles.

[84] European Commission Speech, *Speech by President von der Leyen at the Annual EU Budget Conference 2025*, Brussels, SPEECH/25/1284, 20 May 2025. https://europa.eu/newsroom/ecpc-failover/pdf/speech-25-1284_en.pdf
[85] Government of Canada, *What is AI*, The Department of National Defence and Canadian Armed Forces, Artificial Intelligence Strategy, March 11, 2024. https://www.canada.ca/en/department-national-defence/corporate/reports-publications/dnd-caf-artificial-intelligence-strategy/what-is-ai.html
[86] Hunter Christie Edward, *Artificial Intelligence at NATO: dynamic adoption, responsible use*, NATO Review, November 24, 2020. https://www.nato.int/docu/review/articles/2020/11/24/artificial-intelligence-at-nato-dynamic-adoption-responsible-use/index.html
[87] NATO Science and Technology Organization, *Science & Technology Trends 2020-2040, Exploring the S&T Edge*, NATO Science & Technology Organization, March 2020. https://www.nato.int/nato_static_fl2014/assets/pdf/2020/4/pdf/190422-ST_Tech_Trends_Report_2020-2040.pdf

- Mission Support AI**,** an intermediate category in terms of environment control and failure implications, includes a diverse set of applications, e.g. logistics and maintenance, or intelligence-related applications"[88].

What NATO subsequently does in its definition of AI, is describing its characteristics in a peculiar way, via the usage of "waves". Specifically, it gets describes the so-called "contemporary wave of AI", or even "Second Wave AI", centered on Machine Learning. The latter is crucial for the ability of finding patterns in the input data to get a specific completed output. Today, when talking about AI, Deep Learning comes to be generally mentioned alongside Machine Learning: Deep Learning is a specific sub-typology of Machine Learning, which uses multiple computational layers to handle "computationally demanding pattern recognition or prediction problems"[89], such as the Convolutional Neural Networks[90] used for object detection within images. These are the steps that exemplify the way it works within the context of image recognition:

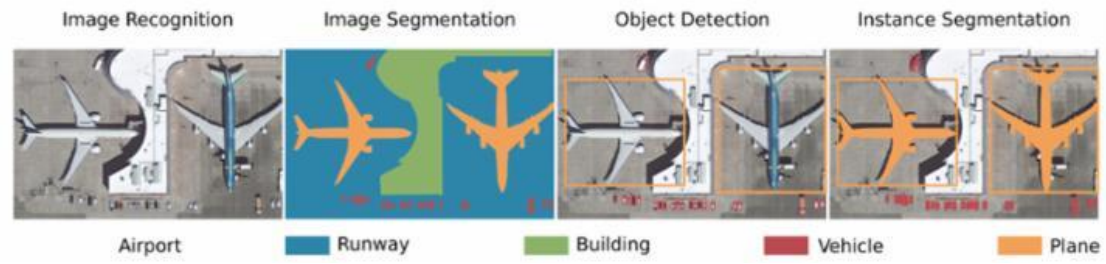

Image C: Steps followed by AI-based systems of image recognition
Source: Hoeser, T.; Kuenzer, C. Object Detection and Image Segmentation with Deep Learning on Earth Observation, Data: A Review-Part I: Evolution and Recent Trends. Remote Sens. 2020, 12, 1667. Retrieved from: Hunter Christie Edward, Artificial Intelligence at NATO: dynamic adoption, responsible use, NATO Review, November 24, 2020. https://www.nato.int/docu/review/articles/2020/11/24/artificial-intelligence-at-nato-dynamic-adoption-responsible-use/index.html

The various typologies of Machine Learning reveal as crucial to quickly analyse large sets of input data to give accurate output, generally predictions or forecasting, deriving from the recognition of a series of patterns subsequently put together.

[88] Tarraf Danielle C., Shelton William, Parker Edward, et al., *The Department of Defense Posture for Artificial Intelligence, Assessment and Recommandations*, Rand Corporation, 2019. https://www.rand.org/pubs/research_reports/RR4229.html
[89] Hunter Christie Edward, *Artificial Intelligence at NATO: dynamic adoption, responsible use*, NATO Review, November 24, 2020. https://www.nato.int/docu/review/articles/2020/11/24/artificial-intelligence-at-nato-dynamic-adoption-responsible-use/index.html
[90] The Convolutional Neural Network is a neural network mostly used for the analysis of images based on the research of specific geometrical shapes of pixels within the image so that the analysis is performedby groups of pixels. There are three main layers through which the input passes. Everything starts with the convolutional layer, where a filter is applied to translate the info received on the specific geometrical area of interest; the process is then repeated for the entirety of the image. The result consists in dots that are put together to form a "feature map". After this process, it is crucial to have a pooling procedure to down sample the featuremap, but putting together dots with features in specific patches of the map.

Going back to the categorization of AI in terms of waves, it actually lies on the US DARPA researches, specifically describing the three waves as it follows:

- "First wave: knowledge-based, it relies on rules-based decision making, facilitating automation by using expert knowledge hand-crafted by humans and a number of if-then statements to dictate their actions.
- Second wave: builds on probabilistic methods, statistical learning and big data.
- Third wave: aims to produce contextual adaptation and common-sense capabilities"[91].

The last vategory is believed to be able to understand the specific context in which it comes to be used into, alongside being able to adapt to it and reason about the best choice options, crucial in a battlefield situation, with a constantly-chaning environment and conditions.

Artificial Intelligence comes to be applied vastly within military applications, and as a result, nine specific categories have been defined depending on its scope, tasks and abilities, within the so-called VUCAR[92] contexts:

- Command and control: AI here is used to help with the decision-making process, mainly military commanders, to be able to analyse quickly and efficiently large amounts of data in constantly changing environments alongside different conditions, specifically crucial to foresee any eventual adversarial change of strategy;
- Intelligence, surveillance and reconnaissance (ISR): AI here has the aim of improving intelligence analysis of large datasets, identifying patterns among these data, such as specific unusual behaviour in a specific context from a determined person (via image recognition and behavioural recognition technologies), crucial to give clear and real-time suggestions for operations.
- Simulation and training: it is generally put together with augmented reality systems (VR), crucial to train soldiers with situations and contexts typical of real-time situations; such training can then be personalized depending on the needs of a specific mission that need to be pursued or even of the soldier himself;
- Automated target recognition: AI used for target recognition purposes has the ability to detect targets quickly and reach an identification exactness with is truly high in short time frames (something generally difficult to be done by humans);

[91] Ibidem.
[92] Giancotti Fernando, Shaharabani Yakov, *The Relevant Warriors, Leadership and Agility in Complex Environments*, The Industrial College of the Armed Forces, National Defense University, Fort McNair, Washington, D.C 20319-5062, 2008.

- Autonomous systems and vehicles: AI here gets integrated within specific aerial, marine and other weaponry systems to enhance sensor data, navigation and movements, helping redefining the trajectory planning if needed;
- Information operations and electronic warfare: AI is crucial for cyber defence operations, because it can detect quickly any anomaly in the system or any vulnerability that needs to be tackled, while at the same time it could be a game-changer in an offensive situation, because of its ability to quickly spot any vulnerability of the adversarial system;
- Predictive maintenance and logistics: AI could help in always being informed about the status of logistics in real-time, providing suggestions on operations or on the way specific stocks should be used or even moved within different warehouses. Furthermore, its application could be crucial in preditcting the way a specific military vehicle should be maintained depending on its conditions and caracteristics, or the way it should be supplied when it finds itself in a precise environment;
- Medical applications: AI could help with monitoring the health status of soldiers if integrate with specific military equipment.
- Sensors and identification systems: sensors fusion, targeting, sensor fusion, targeting, navigation and control systems crucial for every form of electronic war (ESM – Electronic Support Measures, ECM – Electronic Counter Measures, ECCM – Electronic Counter-Counter Measures);

The aforementioned military applications are being experimented more and more by several countries across the world, which are starting to develop their own national plans on AI within the military field. In EU, France was the only state to have a strategy for the application of AI within the military sector; situation that has changed in May 2025 with the truly recent British plan. Outside of Europe, countries investing on the field in a technical, economic and juridical manner are the US, South Korea, Russia, China, and Canada. One of these "tier one" countries that should be looked at carefully is China. Specifically, the Chinese President Xi Jinping has defined the high-tech sector as the "main battlefield" of the rivalry among the main superpowers, using this belief as the reason why it pushed to transform China into a scientific and innovation leader at a global scale by 2035. The analysis of the numbers of investments prove this approach: Chinese R&D spending has increased by 8% to US$458.5 billion annually in 2023. Particularly, China has launched a US$1.4 trillion six-year plan tailored on AI research, which has de facto transformed the PRC into the world leader of 57 out of 64

critical technologies, under the opinion of the Australian Strategic Policy Institute[93]. On the military field, such enormous investments within the defence technological field aim at creating the so-called "intelligentized force"[94], thanks to the fact that China possesses approximatively 74,7% of AI-related worldwide developed patents. [95]The intelligentization of forces can be undertood via the following explanation: "mechanization refers to fielding modern platforms and equipment; informatization refers to linking those systems to networks such as Global Positioning System; and intelligentization refers to integrating artificial intelligence, quantum computing, big data, and other emerging technologies into the joint force"[96]. The plan is to integrate AI to enhance the Chinese MDPW components, specifically leveraging the C4 (command, control, communications and computers) and ISR. "Chinese National Defence in the New Era" is the White Paper where the "intelligentization" concept has been made public in July 2019. Warfare there is described as being under transformation, specifically highlighting that: "war is evolving in form towards informationized warfare, and intelligentized warfare is on the horizon"[97]. The main aim is to use specific technologies, mainly AI, into the "PLA's command, control, communications, computers, cyber, intelligence, surveillance, and reconnaissance; its weapons systems; and throughout the military decision-making process"[98]. While these plans identify the year 2035 as the main "deadline" for the intelligentization of forces, there is a new year, 2027, that Xi Jinping has stressed out as being crucial for the modernization of China, becase it would represent the centennial of the founding of the PLA. The willingness is to increase the modernixation of cyber and space forces and capabilities, while increasing the research, deployment and use of unmanned systems, supported by AI[99].

---

[93] Wong Leung Jenniferm Robin Stephan, Cave Danielle, *ASPI's two-decade Critical Technology Tracker: The rewards of long-term research investment*, Australian Strategic Policy Institute, August 28, 2024. https://www.aspi.org.au/report/aspis-two-decade-critical-technology-tracker-the-rewards-of-long-term-research-investment/

[94] Chinese State Council, *China's National Defense in the New Era*, July 2019. https://www.gov.cn/zhengce/2019-07/24/content_5414325.htm

[95] Fanni Rosanna, Giancotti Fernando, Taddeo Mariarosaria, *Guerre di macchine, Intelligenza artificiale tra etica ed efficacia*, Guerini e Associati, p. 24, May 2024.

[96] Stokes Jacob, Military Artificial Intelligence, *the People's Liberation Army, and U.S.-China Strategic Competition*, CNAS, February 1, 2024. https://www.cnas.org/publications/congressional-testimony/military-artificial-intelligence-the-peoples-liberation-army-and-u-s-china-strategic-competition

[97] Chinese State Council, *China's National Defense in the New Era*, July 2019. https://www.gov.cn/zhengce/2019-07/24/content_5414325.htm

[98] Simpson Kurtis H., Paquette Samuel, Racicot Raphael, Villanove Samuel, *Militarizing AI: How to Catch the Digital Dragon?, AI holds the potential to fundamentally redefine modern warfare*, Center for International Governance and Innovation, February 26, 2025. https://www.cigionline.org/articles/militarizing-ai-how-to-catch-the-digital-dragon/

[99] United States Department of Defense, *Military and Security Developments Involving the People's Republic of China 2024*, Annual Report to Congress, 2024. https://media.defense.gov/2024/Dec/18/2003615520/-1/-

To counter such quickly evolving state of risks, NATO adopted a unified AI strategy in 2021, then updated in 2024, with specific referments to the Chinese adoption of AI within the military sector[100]. Alongside NATO strategy, also individual states are moving towards the adoption of specific AI plans and partnerships for its application within the military sector to counter Chinese fast development. One example is represented by the AI Partnership for Defence established by the US Department of Defense's Joint Artificial Intelligence Centre, "joined" also by the Canadian DND/CAF[101].

Alongside NATO, the US also pushed for a "reply" to such a strong Chinese commitment into researching AI-military technology. Specifically, the US announced on August 28, 2023 a first Replicator program, then followed by a second one in September 2024. Replicator 1 aimed at delivering "all-domain attritable autonomous systems (ADA2) to warfighters at a scale of multiple thousands, across multiple warfighting domains, within 18-24 months, or by August 2025", focusing specifically on unscrewed systems which are considered to be cheapter and grantors of greater safety to soldiers. Replicator 2 instead, revolves around the threats deriving from uncrewed aerial systems (C-sUAS) against installations, infrastructures and milirart forces[102]. The magnitude of US investments within military AI emerges by just looking at the number of signed contracts: from 254 contracts (amounting to 269 US dollars in 2022) to approximately 657 (amounting to 4.323 million US dollars) on a total federal expenditure or around 4.561 million of US dollars[103]. Alongside the establishment and signature of various contracts, the US applied AI in four main domains[104]:

1. A common system of shared intelligence for the five domains (air, cyber, land, sea, space) defined as Combined Joint All-Domain Command & Control (CJADC2).

1/0/MILITARY-AND-SECURITY-DEVELOPMENTS-INVOLVING-THE-PEOPLES-REPUBLIC-OF-CHINA-2024.PDF

[100] Cheung Sunny, *AI frontlines: China's PLA evolution: vs NATO's counterplay, Friedrich Naumann Foundation for Freedom*, Digital Transformation, November 24, 2023. https://www.freiheit.org/taiwan/ai-frontlines-chinas-pla-evolution-vs-natos-counterplay#:~:text=In%20essence%2C%20while%20China%27s%20AI,amidst%20this%20turbulent%20geopolitical%20landscape.

[101] Simpson Kurtis H., Paquette Samuel, Racicot Raphael, Villanove Samuel, *Militarizing AI: How to Catch the Digital Dragon?, AI holds the potential to fundamentally redefine modern warfare*, Center for International Governance and Innovation, February 26, 2025. https://www.cigionline.org/articles/militarizing-ai-how-to-catch-the-digital-dragon/

[102] Defense Innovation Unit, *Replicator*, 2024. https://www.diu.mil/replicator

[103] Fanni Rosanna, Giancotti Fernando, Taddeo Mariarosaria, *Guerre di macchine, Intelligenza artificiale tra etica ed efficacia*, Guerini e Associati, p. 36, May 2024.

[104] Freedberg Sydney J. Jr., *Killer Apps: 5 stories highlight quiet progress on military AI and CJADC2*, Breaking Defense, Networks and Digital Warfare, Pentagon, December 24, 2024. https://breakingdefense.com/2024/12/killer-apps-5-stories-highlight-quiet-progress-on-military-ai-and-cjadc2/

2. Fire control, under the effective supervision of the CJADC2, crucial for target search and acquisition using the data from the shared system of intelligence CJADC2.
3. Quicker operational directions given to troops on the field and frontline, with a continuous flux of information between the various head quarters and the troops. AI woulc be crucial also for processing and analyzing huge quantities of data, without having to wait for a manual processing.
4. The development of standardized and compatible military technology, to avoid the differences that generally arise when technology is being built by different private entities, the so-called Open DAGIR.

The other big player in the field of military AI is Russia. The latter is making important investments mainly in the application of AI within drones, which revealed useful during the war in Ukraine. Specifically, the integration of AI within drones' systems is crucial for navigation (giving exact indications from satellite information on changing environments and conditions), alongside tackling the typical "risks" entailed by the so-called Elecrtonic Warfare. Specifically, drones are vulnerable to the risk of jamming, consisting in having disturbances of drones radio frequencies, crucial for communicating with the operators, which can get disrupted with interferences or even an overflow of information from adversarial systems, or also spoofing activities, consisting in hijacking GPS data and signals to redirect the adversarial drone. AI is of help here because, thanks to its real-time analysis of the warfare situation, it is able to give navigation options which do not depend solely on GPS data, which de facto can be altered by EW systems. Alongside the support for navigation purposes, AI is also integrated to drones for strike and target purposes. A further implementation of AI for navigation comes to be specifically developed within the Optical Navigation system, which leverages AI for fully autonomous operations, helpful to constantly analyse the differences in the constant matching of real-time contexts and data with a pre-loaded map. The main two AI models that intervene in such system are the Convolutional Neural Network, crucial to identify specific objects and similar that can be found on the terrain, alongside the Recurrent Neural Network, crucial to optimize trajectory and adapt constantly route planning[105]. The aforementioned AI implementation makes the system truly independent from any operator, because it is able to fully move, engage the targets and strike on its own. A further crucial aspect here is that these typology of neural networks are trained on high-performance machines, and so they need no

[105] T-Invariant, "*Sovereign" Means Military: How Russia Militarized AI, Drone and Cryptography Industries*, March 20, 2025. https://t-invariant.org/2025/03/sovereign-means-military-how-russia-militarized-ai-drone-and-cryptography-industries/

supercomputer whatsoever to operate, but still they need powerful chips, which is exactly why an important slice of the "technological race" between nations lies on chips provisioning and stock. Such drones are generally used in the battlefield to collect information, find targets (eventually striking them), or even simply for kamikaze missions. The latter are loitering munitions developed to deliver a specific payload to the target, now more and more employed in "swarms", term used to identify military operations where group of cheap drones are used for attack kamikaze missions, or even defence ones, under the strategy of overloading the adversarial defences. It is crystal clear here how AI could be a game-changer: instead of having multiple operators, or even just one that controls the entire swarm, to direct the swarm of drones subjected to spoofing and jamming, AI could help in the development of a sort of "centralized system", where there is a "command drone" that directs the entire swarm[106]. Still such swarm they remain vulnerable to EW strategies, unless they are truly numerous, which makes it necessary for the defence side to employ as counter-defence measure a swarm of drones similar in quantity and quality (something that just until now, has never happened).

A novelty here to tackle such drone swarms could be represented by the new Epirus technology, unvelead in April 2025, able to "fry" drone swarms: the Expeditionary Directed Energy Counter-Swarm (ExDECS) system, delivered to the US Naval Surface Warfare Center Dahlgren Division[107].

Here it is a video from Epirus Youtube channel, to visually understand how it works:
https://www.youtube.com/watch?v=gAmSDdK57o0

As stated earlier, AI can be crucial also for logistics aims. In fact, there are specific budgetary and time constraints that usually need to be accounted for when considering military logistics, and AI could be of help. Specifically, military logistics has to consider movements of troops, supplies, equipments and more other, within zones where roads could be impossible to be used, alongside destroyed brigdes, and similar. AI here could be a game-changer thanks to its ability to recalculate routes in real-time when encountering an obstacle, analaysing at the same time which could be the safest path to follow. The application of AI into the logistics field is already a reality, as demonstrated by the technologies and their usage modalities during the First

[106] Ibidem.
[107] Defence Industry Europe, *Epirus: high-power microwave system delivered to U.S. Navy for ground-based drone defence*, April 30, 2025. https://defence-industry.eu/epirus-high-power-microwave-system-delivered-to-u-s-navy-for-ground-based-drone-defence/

International Forum on Digital Technologies in Transport and Logistics held in October 2024[108].

All these new AI-military technology being developed by Russians in reality fall under the concept of "sovereign artificial intelligence", pronounced for the first time-ever by the Russian president Vladimir Putin on November 7, 2024, during the Valdai Club Meeting in Sochi[109]. By using the adjective *sovereign*, it probably wanted to describe the autonomy that Russia is achieving on the supplies of the basic technology needed for AI research and development.

"At the same time, there are nuclear weapons. This creates major threats for humankind. The same is absolutely true of AI technologies. How is this regulated, and how do people use them? This is a good question. Of course, many countries regulate this. As you say, certain countries ban them. I believe that it is impossible to ban them. AI will eventually make its way, no matter what, especially in conditions of greater competition. I am not talking about armed confrontation, but overall economic competition is increasing. AI will inevitably continue developing in conditions of a competitive struggle. In this respect, we can certainly join the ranks of leaders, considering our certain advantages"[110].

Obviously, to be fully "sovereign", Russian companies need to be able to produce everything on their own, something that they generally did not do before the sanctions linked to the war in Ukraine. Specifically Yandex and Sber (the companies mentioned by Putin in his discourse about AI during the November 2024 event in Sochi), were used to buy foreign hardware as well as cooling systems, Ethernet interconnections and other critical material, which was then assembled in the country. While relying on foreign material Yandex developed its own architecture for AI, Sber instead relied more on US frameworks such as PyTorch[111]. In fact, the

[108] Российская Академиа Транспорта, *Перспективы внедрения искусственного интеллекта в транспортной отрасли и логистике*, October 1, 2024. https://rosacademtrans.ru/ct2024/

[109] President of Russia, *Valdai Discussion Club Meeting*, November 7, 2024. http://www.en.kremlin.ru/events/president/news/75521

[110] President of Russia, *Valdai Discussion Club Meeting*, November 7, 2024. http://kremlin.ru/events/president/news/75521
Original text: "Но в то же время есть ядерное оружие. Это большие угрозы создаёт для человечества. То же самое, абсолютно то же самое и в искусственном интеллекте. Вопрос: как это регулируется и как люди это используют? Вопрос: как регулируется? Конечно, во многих странах, во многих странах это регулируется. Во многих странах, в некоторых, как Вы говорите, запрещают. Запрещать, мне кажется, невозможно. Но всё равно это найдёт себе дорогу, особенно в условиях конкуренции. Конкуренция возрастает. Я сейчас не говорю про вооружённое противостояние, но в целом в экономике конкуренция нарастает. Поэтому в условиях конкурентной борьбы неизбежно развитие искусственного интеллекта. И здесь мы, конечно, можем быть в числе лидеров, имея в виду определённые преимущества, которые у нас есть".

[111] T-Invariant, "*Sovereign" Means Military: How Russia Militarized AI, Drone and Cryptography Industries*, March 20, 2025. https://t-invariant.org/2025/03/sovereign-means-military-how-russia-militarized-ai-drone-and-cryptography-industries/

moment right after the imposition of sanctions as a result of the war in Ukraine, triggered a huge issue in the country linked to the entire semiconductor sypply chain, solved subsequently by acquiring simple processors to be worked upon directly in the country. The war in Ukraine represented a crucial occasion for Russia to find alternative ways that helped in developing and strengthening national production. Russia, alongside Ukraine, are both mainly investing in AI implementation within drones: specifically, Ukraine has developed the "Unmanned System Forces", a specific army branch dedicated to unmanned systems, and Russia is actually working to do the same, intentions and interest in the topic proved by the creation of the "Rubicon Center for Advanced Drone Technologies" in August 2024[112]. The drones developed and built by this centre were crucial for Russian effective drone strikes in the Kursk region in march 2025[113].

To go more into details of the way in which AI is used by the Ukrainian side, analysing the way it supports decision-making activities, so specifically Command and Control[114]:

- AI is used to support the C6ISTAR-EW system, specifically the Command, Control, Communications, Computing, Cyber, Crypto, Intelligence, Surveillance, Target Acquisition – Electronic Warfare, crucial for the decision-making process. The most active private industry in this field in the Ukrainian conflict context seems to be Palantir, which has important agreements with the US DoD.
- Concerning Palantir's partnership with the US DoD, it has developed a Command and Control AI system able to integrate all available sensors for the F2T2EA cycle (Find, Fix, Track, Target, Engage, Assess).
- AI has been used enormously by Ukrainians also for facial recognition purposes, as for example the Clearview system, which has revealed crucial to find quickly and easily all fake news diffused on the web by Russians, alongside for direct usage in the battlefield to recognise the identity of deceased Russian soldiers;
- AI comes to be used by Ukrainians also to help in the decision making context linked to cyber-security, specifically it helps with the real-time detection of cyberattacks and systems vulnerabilities;

[112] Tass, *В Минобороны создали центр перспективных беспилотных технологий "Рубикон"*, October 11, 2024. https://tass.ru/armiya-i-opk/22099195

[113] Axe David, *An Elite Russian Drone Group Is Hunting Ukrainian Vehicles In Kursk. It'll Take More Than A Few MiG-29 Raids To Stop It*, Forbes, March 8, 2025. https://www.forbes.com/sites/davidaxe/2025/03/08/an-elite-russian-drone-group-is-hunting-ukrainian-vehicles-in-kursk-itll-take-more-than-a-few-mig-29-raids-to-blunt-the-russian-edge/

[114] Fanni Rosanna, Giancotti Fernando, Taddeo Mariarosaria, *Guerre di macchine, Intelligenza artificiale tra etica ed efficacia*, Guerini e Associati, p. 20, May 2024.

- AI integrated to RPV like drones and loitering munitions such the the Bayraktar TB2;
- AI used also for ISR purposes.
- Loitering munitions, specifically Switchblade e Phoenix Ghost.

Therefore, AI is crucial for Command and control of forces, technologies, troops, equipment logistics and many other sectors, but still the research, development and deployment of autonomous systems in fully autonomous mode is not performed. However, some magazines underline that in reality such application can be found within the Israeli Army, using it in Palestine via the widely-known Lavender system, which not only helps in detecting the possible target among the population, but also suggests choices and strategies to limit "collateral damage" (to harm civilians). The magazine +972 denounces how Israeli military commanders, instead of critically pondering on whether such suggestion is applicable or not, are accepting nearly every output generated by the AI system, believed to be more accurate than human evaluations[115].

In all the aforementioned systems, the ones used in Ukraine and the ones used in Palestine, there is a constant debate being made about the degree of autonomy that the various AI-military systems have, alongside the specific role of the human control in it. There are three main categories of human control within AI-military systems that can be identified[116]:

- Human in the loop = the human control is inside the loop, as it happens with tele operated drones;
- Human on the loop = the human is literally on the loop, which means that the system is autonomous but the man can always regain control at every moment if needed;
- Human out of the loop = no human presence whatsoever, and so no human control.

There is a second weapons classification that needs to be integrated to the aforementioned one to fully understand how weapons can be distinguished one from each other depending on the role of human control, developed by Jean-Baptiste Jeangène-Vilmer in 2013[117]:

---

[115] Yuval Abraham, *"Lavender": The AI machine directing Israel bombings in Gaza*, +972 Magazine, April 3, 2024. https://www.972mag.com/lavender-ai-israeli-army-gaza/

[116] Bezombes P., *Autonomie et respet de la règle pour les robots militaires. Considérations sémantiques et point de vue de la doctrine interarmées*, Revue défense nationale, pp. 23-35, 2018. Retrieved from: Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle*, Quand les machines prennent les armes, PUF – Presses Universitaires de France/Humensis, p. 119, October 2024.

[117] Jeangène-Vilmer J.B., *Introduction : robotisation et transformations de la guerre*, Politique Etrangère, 78/3, pp. 80-89, 2013. Retrieved from: Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, October 2024

- Semiautonomous = these weapons can move autonomously but they depend mainly on human commands for finding the target and shooting it;
- Autonomous supervised = capable of aiming at the target and shooting at it autonomously but there is always the possibility of human intervention to stop the military activity at any moment;
- Autonomous = the human is not in the system whatsoever.

This all means that a system can be defined as fully autonomous solely when capable of defining completely its own path when encountering an obstacle and in other similar situations, so when it is capable of undertaking its own decisions depending on the contexts it finds itself into.

There are still no systems that can be defined as fully autonomous, even if some generally refer to the SGR-A1 to describe the fears linked to fully-autonomous killer robots. In reality, the SGR-A1, a robot used in the Demilitarized Korean zone by South Korea, developed by Samsung Techwin, possesses a high degree of autonomy, in fact it is capable of finding a target and engaging it autonomously. However, it does not possess the ability of moving completely by itself and overcoming obstacles in an autonomous way while searching for targets.

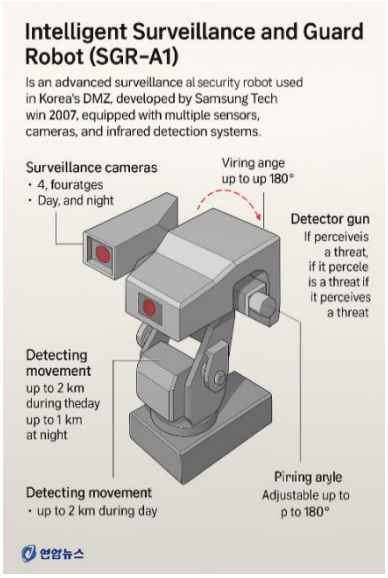

Image D: SGR-A1 structure
Source: This image has been translated by Microsoft Bing by direct request from the author of this thesis from the Korean image that can be found at the following link to this one in English: 영남일보, 흔들리는 혈맹...중국, 北 향해 "구동존이", 정치, April 16, 2015. https://www.yeongnam.com/web/view.php?key=20150416.010160800350001

To briefly sum up, AI applied to the military sector is crucial for the following reasons:

- It enables a quick analysis of a large amount of data, crucial in battlefield situations;

- It is able to support the different parts of the OODA cycle (Observation, Orientation, Decision, Action), specifically the search for intelligence information, their analysis, the support to the decision-making process and the execution of orders via direct action.
- It is crucial to overcome the jamming and spoofing, alongside other disturbing systems that generally create issues in the commanding of drones. In fact, AI enables drones capacity to continue their missions even if their communication with the operator is being disturbed or directly cut down by jamming systems, as demonstrated during the war in Ukraine, specifically when trying to avoid the adversarial A2/AD capacities[118].
- It is way cheaper than using large troops on the battlefield, alongside having a greater endurance than the soldiers' energies, which are obviously outperformed by machines.

However, there are still specific risks that have not been fully addressed yet:

- Risk of cyberattacks:
- Risk of systems being overwhelmed by the adversaries feeding them with extremely large amounts of data (the drone swarms strategy);
- Risk of being misled by adversarial tactics that exploit bugs and biases within facial recognition systems and similar;
- Doubts about the learning of such AI systems (mostly generative AI), which result as black boxes to developers, bringing soldiers to not trust them and to disactivate them when possible;
- Degree and typology of control that humans are able to maintain over these systems.

As pointed out by the AI expert Roucy-Rochegonde, the implementation of these technologies derives from the aim of reaching a "zero deceased" situation, being able to avoid that soldiers have to be exposed to the battlefield thanks to the deployment of such AI-military systems. This concept derives mainly from the US doctrine, specifically used to refer to soldiers sent to fight in foreign soils, who are generally not always ready to perform the so-called "ultimate sacrifice" when it does not involve defending directly their country. This slow substitution of soldiers with autonomous hardwares is becoming more and more a reality, as it is happening with Russian experiments and researchers witnessed in Ukraine. In April 2020 Vitaly Davydov the director of the Russian Foundation for Advanced Projects in the Defence Sector announced that soldiers will be slowly replaced by robots, while unveiling the Marker robot to the public. He

[118] Conference *Vauban Sessions*, organized by Fordward Global at the CRR-FR Lille, May 14 and 15, 2025. Event under Chatham House Rules.

stated that: Marker is planned to be equipped within UAV for the recognition and targeting of enemy targets. Furthermore, when asked by the journalist of Ria Novosti if the Robot would have striken the target by itself, he replied that such actions would have been pursued under human supervision but there are different versions of the Marker robot being tested in specific fields sent by the military commanders, but acting autonomously[119]. Specifically the Russian journal Gazeta declared in 2019 that: "By 2025, Russia will create and introduce into the Armed Forces fundamentally new unmanned combat robots capable of performing tasks with maximum autonomy and minimal remote operator involvement. All of them will be capable of carrying and using a small arms and grenade launcher module, a 120-millimeter mortar, and launching disposable drones"[120].

Moreover, such technologies are being researched upon more and more because they limit not only human losses but also civilian lifes losses, thanks to the efficiency and rapidity in targeting. The strategy here seems to have no civilians among the victims, alongside the doctrine of protecting the soldiers, but still if it truly works ot reach this scope or not is something not known today, because of the lack of data and analysis relative to the number of collateral deaths deriving from the usage of such autonomous military systems. The relative savings from its usage are also directly linked to the medical field, so on the long-term. An example could be represented by the consideration of the enomous amount of soldiers who now have a long list of illenssess after having served in the Balkans during the wars, developing the sadly-known "Balkan syndrome", which obviously amonts to important expenditures for the public health system. The longterm exposure to depleted uranium, a potent carcinogen, due to the extremely high temperatures deriving from the bullets' material hitting the targets was the cause of a widespread "epidemic" among soldiers who took part to peacekeeping missions in the Balkans (affecting nearly 7.000 military personnel for the Italian case, of whom about 400 died)[121].

---

[119] Риа Новости, *Виталий Давыдов: живых бойцов заменят терминаторы*, April 21, 2020. https://ria.ru/20200421/1570298909.html
Original text: "Все, естественно, должно происходить под контролем командира. При этом прорабатываются несколько вариантов действий "Маркера". Один из них предусматривает выполнение боевых задач в заданном командиром районе в автономном режиме".

[120] Газета.РУ, *Война машин: Россия создает группировки боевых роботов*, *В России создадут группировки боевых роботов к 2025 году*, November 23, 2019. https://www.gazeta.ru/army/2019/11/23/12828206.shtml
Original text: "В России к 2025 году создадут и внедрять в Вооруженные силы принципиально новых беспилотных боевых роботов, способных выполнять задачи с максимальной автономией и минимальным дистанционным участием оператора. Все они будут способны нести и применять стрелково-гранатометный модуль, 120-миллиметровый миномет, запускать одноразовые дроны".

[121] Osservatorio Nazionale Amianto, *Uranio impoverito e guerra in Bosnia Erzegovina*, 2025. https://www.osservatorioamianto.it/guerra-in-bosnia-uranio-impoverito/

De facto, AI is still affected by important biases and allucinations that impede the deployment of such systems in "full autonomous mode" in such delicate contexts like warfares. There is an interesting study that has been pursued by experts at the University of Cornell about the way AI, when put into military crisis simulations, generally acted towards the escalation, but this will be explained more in details in the chapter about nuclear applications and risks of AI.

Moreover, something that is not usually considered is the fact that such technologies are generally developed by private industries and entities, which have their own specific economic benefits to pursue which subsequently orient the research and development of technologies. However, in specific states such as in China, this issue does not fully come out as the productive system is under the regulation of the State, implying that industries are compliant with the national aims and directions.

2.2 AI for military decision-making

Discussions and researches around the usage of AI for decision-making are generally based on the analysis of the OODA loop, more specifically, Observe, Orient, Decision and Act loop. Depending on the role and importance of human activities within this cycle alongside the place of the AI, the cycle's phases change, going from OODA loop to O-D one, tailored on Orient e Decide actions of the activity, while humans take final decisions on AI recommendations. There is a second typology of cycle where AI focuses on the Observe, Decide and Act steps, which means that after AI has onserved the situation, collected the relevant data and analysed them, it takes a role in the decision-making process, a role pertaining to the human being in the previous example[122]. The solution here would be a human-machine teaming, something difficult to envisage and to transpose into action. This is the reason why AI used for modeling and simulations reveals crucial: soldiers need to adapt to these systems durimg training, to be able to fully trust them during operational situations. The Italian CASD (Center for Higher Studies of Defence) and IRAD (Institute for Research and Analysis of Defence) have analysed the crucial role that AI could have in training, specifically its implementation in wargaming via[123]:

- Autonomous agents, able to simulate real-time movement of planes, ships, tanls and

[122] Conference *Paris Defence and Strategy Forum*, Europe at a crossroad, organized by the Académie de Défense de l'Ecole Militaire, March 11, 12 and 13, 2025. Event under Chatham House Rules.
[123] Centro Alti Studi per la Difesa, Istituto di Ricerca e Analisi per la Difesa, *Wargaming e Intelligenca Artificiale – situazione, livello tecnologico internazionale e nazionale, prospettive di sviluppo per la difesa*, Istituto Superiore di Stato Maggiore Interforze, 25° Corso – 1° Sezione – 1° Gruppo di Lavoro, AS-SMD-06, July 2023.

more other military vehicles, putting the soldiers into the most realistic possible situation;

- Instruments that enable the analysis of data, helping the AI in learning from results and feedback, alongside finding out specific trends in soldiers performances to understand what needs to be improved or not, enhancing the decion-making process;
- Adaptation of difficulty, speed and other caracteristics of the wargame to the personal experience, training and character of the soldier undergoing the wargame.

AI affects importantly the decision-making cycle within military contexts. It enables quick collection, analysis, synthetisation and final recommendation of data from the battlefield in such a quick way that mere human beings working within military commanding centers would not be able to equal. The constant monitoring with AI models enable to always have personalized forecasts about what could happen, and subsequently, which choice would be the best one to be undertaken to obtain a specific outcome. The problem here lies on determining the level of autonomy that such system have, which change from nation to nation and from technology to technology, alongside the role of the human being. Determining the degree of autonomy is crucial because of new issues arising from the usage of such systems when systems slide well beyond the human control. An example of the dangerouseness of such a lack of understanding of the degree of autonomy is the case reported by the United Nations report (published in June 2021) about the Turkish drones which attacked autonomously Haftar's troupes in Libia in March 2020[124]. Specifically, it has been denounces that the Libyan National Army of Khalifa Haftar has been targeted by several UAVs (or even LAWS) such as the Turkish STM Kargy-2, resulting in human beings having been killed by an autonomous lethal weapon which has acted with no human supervision whatsoever, having de facto an autonomous system deciding over the life and death of human beings. However, issues revolving around biases and similar do not push nations and military alliances towards refraining from such investments, in fact in March 2025 NATO (NCIA) acquired weapons integrated with AI, specifically the Palantir Maven Smart System NATO (MSS NATO)[125]. Such acquisitions of military systems

[124] United Nations Security Council, *Letter dated 8 March 2021 from the Panel of Experts on Libya established pursuant to resolution 1973 (2011) addressed to the President of the Security Council*, S/2021/229, March 8, 2021. https://docs.un.org/en/S/2021/229

[125] NATO, *NATO acquires AI-enabled Warfighting System*, Apr 14, 2025. https://shape.nato.int/news-releases/nato-acquires-aienabled-warfighting-system-

continue to happen because, as once affirmed by Edward Luttwak: "what cannot be verified, cannot be limited"[126].

AI is becoming further incredibly crucial for decision-making purposes thanks to newly-developed technology that enables the possibility to aim at targets which are not "directly visible". Specifically, in June 2024 MBDA announced that it had developed an AI-powered decision-making technology able to find hidden targets using "beyond-line-of-sight technology", called "Ground Warden". Such typology of technology is crucial to support also ISR operations[127]. The industry that developed it, even suggested to use paired with the missile system, so that the Ground Warden is able to analyse real-time data from the drone suggesting attack options, specifically the target, the location and the time-frame.

Youtube video which shows how this system works:
https://www.youtube.com/watch?v=rBeFh7Gjglo&t=57s

The interesting aspect about this technology development is also the industry guarantee about the constant software improvements that are going be made, to be able to adapt to specific feedbacks and constantly changing environments.

The overall problem with such military technology enhanced with AI is exactly here: the constant need for improvements and updates. As underlined by the expert Shanahan Nicole: "If you don't do that [update the AI system], it's going to go stale. It is not going to work as advertised. The adversary is going to corrupt it, and it'll be worse than not having AI in the first place". The training and experiment cycle needs to be viewed more as a continuous process that does not stop. A further crucial issue is that training of newly updated systems happens in specific and circumscribed training situations and contexts, generally digital ones, which does not always correspond to real-life ones. Alongside that, there is also the issue that soldiers still do not have 100% trust in these systems: as underlined by the CNAS scholar Josh Wallin: "when the Defense Advanced Research Projects Agency (DARPA) tested its autonomous dogfighting program, it did so in a digital simulation, where the program handily beat a human pilot, but when they put the program in a real plane with a human safety operator—a highly

[126] Luttwak Edward, *Le Paradoxe de la Stratégie*, p.215. Retrieved from : Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, p. 307, October 2024.
[127] Paleja Ameya, *Ground Warden: New AI-based system enables forces to see hidden targets*, Interesting Engineer, Military, June 18, 2024. https://interestingengineering.com/military/ai-ground-warden-finds-hidden-targets

experienced test pilot who could disable the AI, take control, and land the plane if needed—it suddenly stopped performing as it had done in the digital simulator"[128]. This means that the training of the AI software per se is not enough: it needs to be accompanied by a throughout training of soldiers in using such systems. A further crucial point raised by another CNAS scholar, Wallis, is that such AI systems are trained to deal with human behaviour, but as AI systems are being more and more integrated within military technology by several state, it is indeed a huge question mark the way in which an AI system would deal when encountering a system which is mainly controlled by another AI system rather than a human being.

2.3 Analysis of Biases and the issues that they entail

In the last decade, as AI powered technologies have been used more and more, it has become evident how such applicatives within facial and target recognition systems, used for defence and military purposes, are biased. The main reason behind such biases can be found in the way the information is put into the software by programmers. The resulting problems are multifarious, going from posing serious threats to minorities that risk indirectly becoming targets, to influencing young users of AI powered applications and devices who are not mature enough to have a critical thought on the information they are "fed". This subchapter will focus on the biases linked to images and video footages analysis, targeting and recognition because this is the typology of bias that has been most widely studied.

Amnesty International recently launched an alarm on the risk of exacerbating social conflict in those areas of nations where high crime rate is used to justify the targeted surveillance of specific categories of the population, in the context of AI being used for neighborhood security. An example is represented by the Latin-American and the Black communities calls for revision of these AI systems in the United States, mainlye because they have become the target of such surveillance measures, resulting in an elevate risk of people and target exchange, truly dangerous when human lives are at stake as in war. As pointed out by researcher Joy Buolamwini[129], "Algorithm bias can lead to exclusionary

[128] Waterman Shawn, *Military AI Will Mean Overhauling Test as Well as Tactics: DOD's first AI Chief*, Air Space and Forces Magazine, April 9, 2025. https://www.airandspaceforces.com/military-ai-overhauling-test/

[129] Dr Buolamwini is a MIT Graduate student in computer science and current MIT researcher, founder of the "Algorithmic Justice League", an organization that works on the analysis and spreading awareness on biases of AI

experiences and discriminatory practices", and to verify whether the system of facial recognition she was using was able to recognize faces of black people with a similar error rate to white ones, she ran an experiment. The researcher Buolamwini found out that, by putting her face in front of a camera, no recognition of visage was done by the system, but when she wore a white mask with evident facial features to do the same analysis, she discovered that the system was able to detect the biometric features of the mask. This project,*Aspire Mirror,* was based on generic basic facial recognition software to build the system for her project. Similar issue presented when researcher Buolamwini tried to do the same with robots by experimenting simple interactions as playing the peekaboo game, where shewas not able to play with the robot as the latter could not recognize her as a human with whom it was possible to have an interaction. She then underlined how widespread such an issue is, as she had to face the same exact bias in China, where during an entrepreneurship project a start-up decided to give a demo of their robot, which could not recognize researcher Buolamwini's face. The problem at this point is to be found at the basis, exactly in the training data set, as the machine is clearly unable to recognize black people's faces simply because its software hasnever been trained to do so. This causes issues with the generalization process on the recognition of biometric data, and the subsequent identification of a face becomes nearly impossible if the input presented differs from the original training material. To solve the matter from the basis, the researcher Buolamwini proposes to train the software with full spectrum datasets, to ensure a wider recognition. As reported by Buolamwini following a Georgetown law police report[130] "1 in 2 adults, so 172 million people's faces have been registered in a facial recognition software, without anyclear rule (at least at the time) on the usage of the information, which could have been done indiscriminately, as well as any checks have been done on the accuracy of the information submitted to the software". This all could lead to situation of misidentification of people, which within the national security field could become dangerous. She made a call to create more inclusive data sets, and platforms where to share these bad experiences with bias to make it possible to change the training datasets of such software.

Within the security sphere, as explained by computer scientist Timnit Gebru, AI people

and subsequent social implications. She explained carefully and in full details her experiment during the *How I'm fighting Bias in Algorithms* TedX BeaconStreet on 29th March 2017.

[130] Example expressed by Researcher Joy Buolamwini during the TedX BeaconStreet on 29 March 2017, basedon the Centre on Privacy and Technology, Georgetown Law Police Report *The Perpetual Line-Up: Unregulated Police Face Recognition in America*.

identification software is being used to "predict someone's likelihood of committing a crimeagain, and such predictions are used by judges to decide how many years of prison a personshould serve"[131]. The resulting risk is to incur in automation bias, as public authorities and others will put complete trust on the software, not applying anymore critical thinking. To change such undertaken direction, there is not only the need to consider the *faireness principle* during the programming process of algorithms, but also to consider social aspects, such as the fact that some minorities or social groups are statistically more targeted by police and law enforcement than others in specific geographic zones. Finally, there is the need to create something similar to the "Datasheet for Datasets" proposal, based on Gebru's idea of creating a datasheet like the ones typical of electronical components, directly for these devices, to realise an enlargement of the dataset used as input for the software as well as enhancing transparency and accountability for those who programmed it. Such proposal could be crucial if done weapon by weapon, considering their precise and peculiar characteristics, abilities and limits.

The uncertainty related to transparency and bias issues of such devices for biometric data detection and collection is still high, but has not discouraged governmental agencies to use them in sensitive places for national security purposes, such as airports. The aforementioned systems consist in having face recognition scanners at the gate of the airport, able to check not only the identity of the passenger and if he is boarding the correct flight, but it can also perform checks on whether there are any visa's or any other document anomalies. Such technologies do such controls with an average of 1.6 seconds, while the employee at the gate needs a minimum of 15 seconds to verify the accuracy of the identity of the individual[132]. Exemplificative is the situation of two young twins who decided to swap passports as just one got the visa to travel to Japan, and even if China has one of the most advanced systemsfor biometric data detection in airports, the two twin sisters not only swapped to travel to Japan, but also to Brazil and Russia. The situation can be explained by considering that the system is trained to recognize personal facial features, so that in case of twins these biometrical data can eventually be similar. The fact that this story is recent, September 2023, proves the huge lacks of training of the aforementioned devices. Completely different situation would have presented with the iris

[131] Timnit Gebru, *How To Stop Artificial Intelligence From Marginalizing Communities?*, TEDxCollegePark, 18 October 2018.

[132] Nimra Khan, Marina Efthymiou, *The use of biometric technology at airports: The case of customs and border protection (CBP)*, International Journal of Information Management Data Insights, Volume 1, Issue 2, November 2021.

or fingerprint scan, whichare personal and make it nearly impossible to get away with an eventual exchange of identity. In fact, according to Sir Francis Galton's calculations, "the probability of finding two similarfingerprints is one in 64 billion, even with identical twins (homozygotes)"[133].

The positive side of the aforementioned technology is that, whatever the method employed, all these technologies collect characteristics that are common (can be found in every person)while at the same time unique (nearly impossible to have a specific personal characteristic, such as iris scan, equal to somebody else's) for everyone, as well as being permanent, recordable and forgery-proof[134]. For what concerns the risks of forgery, it needs to be considered that, as the number of identity thefts (above all AI deepfakes) is dramatically increasing, AI facial recognition technologies could be one of the most effective ways to battle such phenomenon. Facial identity verification technologies are able to guarantee more accurate facial verification and comparison. They are based on specific principles: continuous real-time analysis and identity verification in static and dynamic environments, they need to be able to compare photos taken in varying conditions (such as changes in people's visages) and to compare and recognize the identity within the database[135].

In fact, face recognition algorithms have a high accuracy of around 90%, but such error rate changes depending on the demographic group taken in analysis. "The poorest results have been found in subjects who are female, black and 18-30 years old". As explained by MIT PhD researcher Alex Najibi, in the 2018 *Gender Shades* project, an intersectional approach was applied to appraise three gender classification algorithms. Subjects were grouped into four categories: "darker-skinned females, darker-skinned males, lighter-skinned females, and lighter-skinned males. All three algorithms performed the worst on darker-skinned females, with error rates up to 34% higher than for lighter-skinned males. Independent assessment by the National Institute of Standards and Technology (NIST) has confirmed these studies, finding that face recognition technologies across 189 algorithms areleast accurate on women of color"[136].

On this matter, computer scientist Buolamwini explained why Amazon automatic AI

[133] Francis Galton, *Finger Prints – The Classic 1892 Treatise*, 1892.
[134] Thales, *Biometrics: definition, use cases, latest news*, Digital identity and security, 20th May 2023.
[135] Facia, *AI-Generated Identity Fraud: Bringing Real-Time Identity Intelligence for Fraud Prevention*, 7th March 2024.
[136] Alex Najibi, *Racial discrimination in Face Recognition Technology*, Blog, Science Policy, Special Edition: Science Policy and Social Justice, Harvard Kenneth C. Griffin, Graduate School of Arts and Sciences, 24th October 2020.

recruiting algorithm realized a clear gender-biased choice of the candidates for working roles: "The system had learned that the prior candidates deemed successful were predominantly male. Past hiring practices and centuries of denying women the right to education coupled with the challenges faced once entering higher education made it especially difficult to penetrate male-dominated fields. Faithful to the data the model was trained on, it filtered out résumés indicating a candidate was a woman. This was the by-product of prior human decisions that favored men"[137]. The researcher defined these models as *power shadows*, which "are cast when the biases or systemic exclusion of a society are reflected in the data". The research hypothesized that this was due to the fact that, to overcome privacy issue, the dataset has been made of public figures. In particular, by lookingat who holds public or political offices, it is evident the huge gender gap denounced by UNWomen on the representation of women in parliaments: 76.6% of parliament members are in fact males. Same exact explanation can be found for black people, not only from a historical point of view, related to the past decolonization of territories but still not of actual people's minds, but also the way the inputs of the software are selected. Statistically when doing online searches, white people are more present in number of pictures and videos. An example of such issues is the case of Robert Williams in the US. In January 2020 Williams was wrongfully arrested by the police on the basis of a false match from a face recognition software. Precisely, the software matched an old image of the face of Williams from an expired driving licence to the face of a man responsible for shoplifting in a Shinola store in Detroit in 2018, whose security cameras images were de facto blurry and not at all clear for analysis. Several other cases have been fought into tribunals and courts in the United States. It is only on June 28, 2024 that an important agreement has been reached, subsequent to this case, which de facto limits the cases in which such AI-powered face recognition systems can be used[138]. Cases like Williams one highlight the dangerouseness of using such AI systems affected by flaws and biases in real-time applications, which implies that in even more fragile and delicate environments, such as warfights, similar mistakes from AI systems could result in a catastrophe, if not directly the death of civilians and military.

[137] Joy Buolamwini, *Unmasking the bias in facial recognition algorithms*, MIT Management – Sloan School, Ideas to matter – artificial intelligence, 13th December 2023.
[138] https://www.aclu.org/cases/williams-v-city-of-detroit-face-recognition-false-arrest

Chapter 3: Examinations of military-AI applications

3.1 AI applied to the nuclear field and weaponry

The application of military AI within the nuclear field is a revolutionary novelty, which implies enormous vulnerabilities. There are still numerous biases and flaws to be solved, which represent vulnerabilities that adversaries would be willing to exploit to damage critical national infrastructures. It suffices to go back to the Cold War period to understand clearly the amount of damage that just a little computer component, not working properly, could cause; an example could be the malfunctioning of the missile warning system, which would generate as a consequence, a false alarm, as previously happened in history. In that specific situation, human judgment saved humanity from destruction: it was a human critical understanding that doubted the fallacious analysis of the machine. The exponential research and introduction of AI technology within the defence and military sector triggered a "research competition" on its adoption within the nuclear field, mainly between nuclear-armed states, especially the P5 states (the five permanent members of the United Nations Security Council, specifically China, France, Russia, United Kingdom, United States), something truly worrying considering the still unresolved issues and vulnerabilities aforementioned in this thesis. The integration of AI within the military nuclear sector is being researched and experimented mainly in the decision-making structure, a truly vulnerable and delicate part of the cycle that entails military weaponery research, development and application. This subchapter is going to analyse such AI application with the related benefits and risks that could arise, pointing out the differences in AI nuclear application among the P5 states.

One of the first integrations of AI within the nuclear field, specifically in nuclear weapons, passed through its incorporation within the NC3 systems, specifically the Nuclear Command, Control and Communication systems. The NC3 is the underlying infrastructure crucial for the functioning of military systems, specifically the usage of nuclear weapons, formed by nuclear warheads, delivery vehicles and delivery platforms. In the following image, there is an example of the US NC3:

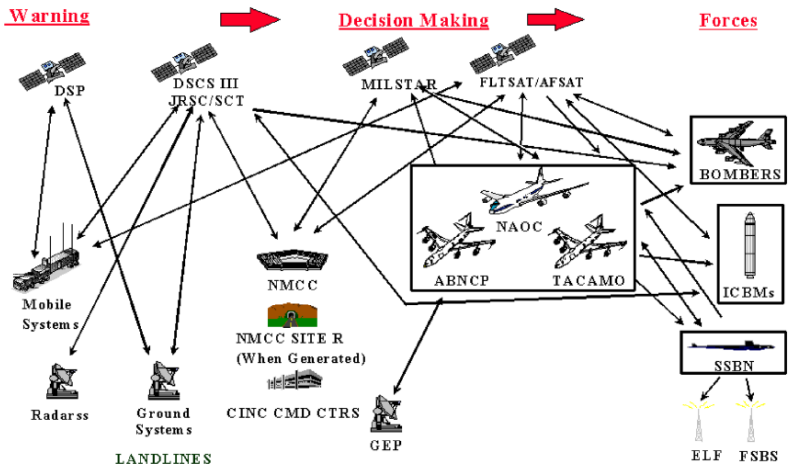

Image E: The United States NC3 cycle, Nuclear Command, Control and Communication cycle
Source: National Command & Control: The National Military Command System (NMCS), October 2001. Retrieved from: Hruby Jill, Miller M. Nina, Assessing and Managing the Benefits and Risks of Artificial Intelligence in Nuclear-Weapon Systems, NTI Paper, Figure 3: Notional Connectivity for Nuclear Command and Control in the United States, p.12, August 2021.

To analyse exhaustively the AI-nuclear nexus, it is of paramount importance to examine the main categories of application, which could be subdivided into three main classifications, respectively risks, benefits and the potential military usages. These specific AI integrations, which are going to be detailed in the following paragraphs, highlight the main challenges deriving from merging AI within the NC3 system, specifically the difficulties in guaranteeing the maintenance of strategic nuclear stability (which exists only when adversaries lack a significant incentive to engage in provocative behaviour[139]), alongside the entanglement of AI-enabled conventional systems and nuclear risks.

The integration of the AI within the NC3 brings numerous risks and benefits, widely recognised by the countries with nuclear arsenals. Both risks and benefits will be analysed more in detail in this subchapter, but here there are the main difficulties to be resolved in order to reach a safe AI application within the military nuclear field, which help with framing the context[140]:

- False positives deriving from misinterpretations or other reasons such as AI hallucinations, which could lead to identifying a non-existent nuclear threat (false alarms);

[139] Rand Corporation, *How Artificial Intelligence Could Increase the Risk of Nuclear War*, Research and Commentary, April 24th 2018. https://www.rand.org/pubs/articles/2018/how-artificial-intelligence-could-increase-the-risk.html
[140] Future of Life Institute, Strategic Foresight Group, *Framework for Responsible Use of AI in the Nuclear Domain*, UN Secretary General's designated civil society organization for recommendations on the governance of AI, February 5th 2025.

- Discussions around the "human-in-the-loop" concept, whose absence could generate the risk of limiting the necessary human control on the AI outputs, or directly reaching the point where humans rely too much on AI outputs without critically questioning them or implementing a double-check system;
- Difficulty in understanding the AI reasoning process, which resembles to the so-called "black box", and makes it difficult to trust AI for the decision-making process;
- Absence of large and unbiased training data;
- High risk of error, which is nevertheless diminishing;
- Deficit in cybersecurity, which could lead to suffering intrusions into sensitive nuclear control systems, leading to the risk of uncontrolled and unauthorised deployment of nuclear arsenal;
- Having faster and more efficient technology at a lower cost could enable an exponential rush on nuclear testing and application, leading to instability at the international level and the risk of a nuclear race.

On the other hand, the employment of AI within the nuclear sector, more specifically the military side, is showing important benefits, as it helps with:

- Quickly analysing large amounts of data deriving from several sensors, satellites and other parts of the NC3;
- Improving accuracy in target detection and threat detection;
- Giving early warnings in the case of detected cyber threats to avoid preventively any eventual cyberattack, alongside automating the response to them;
- Providing recommendations or directly acting in real-time at the moment of threat detection.

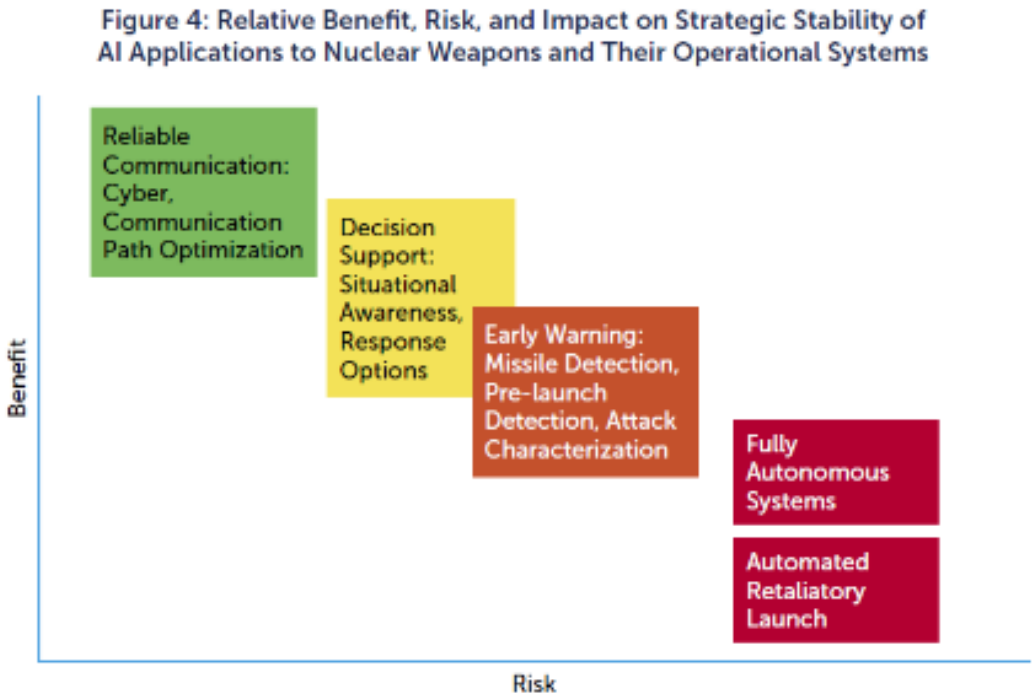

Image F: Relative Benefit, Risk and Impact on Strategic Stability of AI Applications to Nuclear Weapons and Their Operational Systems
Source: Hruby Jill, Miller M. Nina, Assessing and Managing the Benefits and Risks of Artificial Intelligence in Nuclear-Weapon Systems, NTI Paper, Figure 4: Figure 4: Relative Benefit, Risk, and Impact on Strategic Stability of AI Applications to Nuclear Weapons and Their Operational Systems, August 2021.

Explanation of the graph: "The impact on strategic stability is indicated by the colour of the application box, with green being the most likely to have a stabilising effect, yellow and orange indicate careful consideration for impact on stability, and red is most likely to have a destabilizing effect".

However, the usage of AI in the nuclear field does not generate global consensus, which leads to the absence of a generally accepted framework to regulate its usage. Specifically, the absence of agreement on the topic derives from the dual use of AI applications and the different national considerations about the risks, alongside dissimilar ethical national principles and technology at disposal. More importantly, the issue presented by this new "revolution" is related to the difficulty in ensuring the ethical usage of AI while applying it within the nuclear sector for military purposes, especially in AI-powered nuclear weapons. Alongside such challenges, and vulnerabilities, the numerous unresolved biases and technological issues need to be considered as possibly threatening a "safe" AI application within the nuclear military and defence sector. Moreover, looking at the nuclear AI models used for nuclear decision-making it comes to be generally evident the unreliability of AI outputs, due to the risk of AI hallucinations, which

generate false data, and could trigger a dangerous military response if it refers to the risk of an incoming nuclear attack, as it happened during the Cold War period. In fact, the usage of autonomous weapons in the nuclear field is not a modern novelty. In the period of the Cold War, the Soviet military introduced a semi-autonomous nuclear system called Perimeter ("Dead Hand"). As explained by the analyst Alexander Khramchikhin, Deputy Director of the Institute for Political and Military Analysis, the development of such soviet system aimed at having an automated nuclear launching system of the Soviet nuclear missiles arsenal, not needing any human guidance from the central command. More specifically, Khramchikhin stated that "if the sensors receive information about radioactivity that indicates the detonation of nuclear weapons in Russia, then missiles are automatically launched, which are equipped not with warheads, but with signal transmitters that issue combat control codes for the launch of the corresponding nuclear weapons of Russia". De facto, the "dead hand" system has the ability to respond to a nuclear attack with a retaliatory one without the need for authorization and guidance from the command-and-control structure. Nevertheless, when considering autonomous weapons, it has to be understood that they do not need to kill people to undermine stability and to increase the probability of war. In the specific case of nuclear weapons, the stability of the situation, granted by the fact that both sides can survive a first strike and hit back in retaliation ("assured retaliation" principle), is not applicable. It could either generate instability (as one side could consider striking "in advance" to avoid being hit later), or it could assure stability as neither part would be willing to launch the first strike. Still, the human awareness of the risks is crucial and could always avert an escalation that a machine would instead consider as its first option for defence purposes.

Going back to the nuclear-AI nexus, the aforementioned challenges, risks and benefits could vary depending on the typology of AI model adopted. Here are the main AI models with potential military usage[141]:

- Classification models: they learn patterns from labelled training data, such as images, and use the learnt information to classify new data. This is crucial for the military ISR-T (Intelligence, Surveillance, Reconnaissance and Targeting) system, which is employed to classify objects and beings which could represent a threat in specific environments;

[141] Stokes Jacob, Kahl Colin H., Kendall-Taylor Andrea, Lokker Nicholas, *Averting AI Armageddon, U.S.-China-Russia Rivalry at the Nexus of Nuclear Weapons and Artificial Intelligence*, CNAS – Center for a New American Security, Frebruary 2025.

- Analytic and predictive models: from the previously acquired historic information and training, merged with the current data, they are able to generate inferences and predictions about future events. Specifically, "they have the potential to fuse together an enormous volume of information, intelligence, and operational data to create a real-time common operating picture (COP)/common intelligence picture (CIP)". [142]
- Generative AI: they are able to create new content re-elaborating the previously learned training material starting from simple prompts.
- AI possessing autonomy: they make decisions based on real-time data collected via sensors (such as satellites), then direct hardware or software to react accordingly. Such a system is crucial because, from mere sensors-inputs, it triggers a decision-output which guarantees a quick response to an abrupt threat.

A further AI typology category is the Rule-based AI[143], which has however a specific usage limited to narrow tasks because its system comes to be tied to the set of rules given by the developer, revealing as crucial the optimisation and fusion of sensor data alongside the establishment of specific communications pathways[144]. Moreover, Rule-based AI was largely used for missile early-warning systems to identify the launch and trajectory of ballistic missiles with sensors and transmit this information to human operators for validation of the output. The various issues that such system entailed during its nuclear usage in the Cold War period, when algorithms had already been put in place within American and Soviet control and warning systems to enhance nuclear deterrence, have been proved by several incidents (such as the Able Archer exercise, October 1983), which showed to the world the vulnerability of algorithms and automation software in this field. De facto, the introduction of AI poses serious nuclear risks and dangers, which can be divided into three main categories:

- AI technical challenges: this category refers to the numerous AI technical issues that still need to be resolved, such as "complex and unpredictable interactions among AI systems", lack of unbiased training data and security of AI systems, which could be vulnerable to adversary actions;

[142] Ibidem.
[143] Chernavskikh Vladislav, *Nuclear Weapons and Artificial Intelligence: Technological promises and practical realities*, SIPRI – Stockholm International Peace Research Institute, SIPRI Background Paper, Project funded by the German Federal Foreign Office, September 2024.
[144] McDonnell, T. et al., *Artificial Intelligence in Nuclear Operations: Challenges, Opportunities, and Impacts* (Center for Naval Analyses: Arlington, VA, Apr. 2023), pp. 44–45. Retreieved from: Chernavskikh Vladislav, *Nuclear Weapons and Artificial Intelligence: Technological promises and practical realities*, SIPRI – Stockholm International Peace Research Institute, SIPRI Background Paper, Project funded by the German Federal Foreign Office, September 2024.

- Human-factors challenges: it mainly refers to the issue of unskilled workers using AI and the trust problems from humans towards AI capabilities (absent or excessive);
- Difficulties in analysing the effects that AI integration could have onto the military system and the decision-making sphere[145].

3.2 Historical and geopolitical analysis

In the tense geopolitical environment of today, after a long period of "disinterest" in the topic of nuclear deterrence and risks, the latter recently started to surge again in international politics. Specifically, in 2019 the US President at the time, Donald Trump[146], promised to modernise the nuclear arsenal, alongside North Korea constantly pushing forward its nuclear program and weapons[147], reaching its peak with the Constant Russian threat of resolving to the usage of nuclear weapons in the context of the war in Ukraine[148], culminating with the truly recent awakening of tensions between two nuclear powers Pakistan and India (2025). In such a tense geopolitical and military context, the usage of AI in the nuclear military field is being experimented more due to its ability in making missiles guidance systems more accurate and flexible, "[…] especially useful for high velocity systems that human can't manoeuvre"[149]. Specifically, in a context where several countries are researching prototypes of hypersonic aircraft and missiles, which would "be able to fly five times faster than the speed of sound"[150], AI could reveal as the only technology able to stop the trajectory of such missiles. Secondly, AI is revealing as crucial for speeding up the analysis of enormous quantities of data, as well as helping with theoretical modelling and experiment design, enabling the acceleration of research[151] and eventual military responses in case of attack. On the research side, the International Atomic Energy Agency points out how AI could be crucial in data analysis to

[145] McDonnell Tim, Chesnut Mary, Ditter Tim, Fink Anya, Lewis Larry, with contributions by Westerhaug Annaleah, *Artificial Intelligence in Nuclear Operations, Challenges, Opportunities, and Impacts*, CNA – National Security Analysis, IRM-2023-U-035284-Final, April 14th 2023.
[146] BBC News, *Donal Trump: US will build up nuclear arsenal*, October 23rd 2018. https://www.bbc.com/news/world-us-canada-45946930
[147] Sutton H., *New North Korean Atomic Submarine Escalates Nuclear Arms Race*, NavalNews, April 2nd 2025. https://www.navalnews.com/naval-news/2025/04/new-north-korean-atomic-submarine-escalates-nuclear-arms-race/
[148] Gilli Andrea, de Dreuzy Pierre, *Российское ядерное принуждение в Украине*, NATO Review, Мнения, анализ и обсуждение вопросов безопасности, November 29th 2022. https://www.nato.int/docu/review/ru/articles/2022/11/29/rossijskoe-yadernoe-prinuyodenie-v-ukraine/index.html
[149] Ibidem
[150] Ibidem
[151] Vlasov Artem, Barbarino Matteo, *Seven Ways AI Will Change Nuclear Science and Technology*, IAEA – International Atomic Energy Agency, September 22nd 2022. https://www.iaea.org/newscenter/news/seven-ways-ai-will-change-nuclear-science-and-technology

build the nuclear and atomic data evaluation and compilation[152], alongside bringing a revolution in the field of fusion research thanks to AI's ability to process large amount of data, accelerating nuclear fusion research projects[153]. As third point, AI systems could optimize procedures, mainly the speed and their complexity, alongside improving the performance and safety of specific structures, like reactors, reducing maintenance costs. The increases in efficiency derives from AI ability in automatizing data-analysis tasks, decreasing the probability of mistakes during the processes. A fourth crucial point is the usage of the predictive potential of AI to help monitor power plant processes and detect anomalies, mainly on the detection of radioactive materials levels, crucial to guarantee human physical and environmental protection. Moreover, in an era when cyber attacks are growing in number, AI could help with the early-detection of cyber attacks on nuclear facilities[154]. In all these cases, the crucial positive aspect of AI introduction is the reduced reliance on human intervention and authorization. For example, AI introduced within national nuclear defences, specifically within the targeting systems, enables a greater accuracy of nuclear weapons, reducing "collateral risks"[155]. As a matter of fact, developing more accurate nuclear weapons could reveal as determining within the defence field. In the 1970s, the US nuclear arsenal used MIRVs, Multiple Independent Targetable Re-entry Vehicles, capable of launching and then redirecting to a different trajectory from the initial one the nuclear warheads without needing human intervention to do so[156]. Alongside such technology, afterwards, in the 1980s, the development of more accurate nuclear missiles such as Pershing, Polaris and Titan II, enabled the usage of fewer nuclear warheads to get the same strategic objectives as before but in a cheaper and more effective manner. From that moment on, the US nuclear policy focused on the enhancement of the "look-shoot-look" capacity, specifically finding the target, hitting it and assessing the impact and result of the

[152] IAEA – International Atomic Energy Agency, *Nuclear Data Services*, Nuclear Data Section. https://www-nds.iaea.org/

[153] Matteo Barbarino, *New CRP: Artificial Intelligence for Accelerating Fusion R&D (F13022)*, IAEA – International Atomic Energy Agency, New Coordinated Research Process, June 9, 2022. https://www.iaea.org/newscenter/news/new-crp-artificial-intelligence-for-accelerating-fusion-rd-f13022

[154] IAEA – International Atomic Energy Agency, *Computer Security in the Nuclear World*, IAEA Bulletin, The IAEA's flagship publication, ISSN 0020-6067, 23-01968, June 2023. Retrieved from: Hewes Mitchell, *How Artificial Intelligence Will Change Information and Computer Security in the Nuclear World*, Vol. 64-2, June 2023. https://www.iaea.org/bulletin/how-artificial-intelligence-will-change-information-and-computer-security-in-the-nuclear-world

[155] Puwal Steffan, *Should artificial intelligence be banned from nuclear weapons systems?*, NATO Review, Opinion, Analysis and Debate on Security Issues, April 12, 2024. https://www.nato.int/docu/review/articles/2024/04/12/should-artificial-intelligence-be-banned-from-nuclear-weapons-systems/index.html

[156] The Centre for Arms Control and Non-Proliferation, *Multiple Independently-targetable Reentry Vehicle (MIRV)*. Retrieved from: https://armscontrolcenter.org/wp-content/uploads/2017/08/MIRV-Factsheet.pdf

strike[157]. The introduction of AI in nuclear weapons and systems would enable a greater effectivity and a reduction of nuclear warheads needed, which implies a greater capability to protect the remaining number of nuclear weapons within the national stockpile. Considering that currently Russia is refusing to comply and allow nuclear inspections alongside its researches on next-generation hypersonic ballistic missles while using the nuclear threat in the context of the war in Ukraine[158], China exponentially investing in its own nuclear program for weapons development, new tensions among India and Pakistan, the two "nuclear neighbours", AI should instead be introduced to avert any global nuclear escalation. Specifically, as explained by experts from the CNA, AI could mitigate nuclear risks in four main areas:

- "Nuclear weapons surety;
- Survivability and resilience of nuclear forces;
- Leadership decision-time expansion;
- Crisis and conflict de-escalation".[159]

AI could mitigate the risk of escalation within the nuclear field by being employed for the maintenance and surveillance of the usage of nuclear forces, analysis of the performance of nuclear forces, the planning and decision-making system, which are part of the integration of AI within the Nuclear Command, Control, and Communications (NC3) systems. The following diagram from the US Air Force Nuclear Weapons Center points out the way NC3 components of the nuclear command are related (it is lacking the third component, Communication, which encompasses all the sectors in the image):

Image F: Relative Benefit, Risk and Impact on Strategic Stability of AI Applications to Nuclear Weapons and Their Operational Systems

[157] The National Security Archive, The George Washington University, Jimmy Carter's Controversial Nuclear Targeting Directive PD-59 Declassified, National Security Archive Electronic Briefing Book No. 390, September 14th 2012. https://nsarchive2.gwu.edu/nukevault/ebb390/

[158] Puwal Steffan, *Should artificial intelligence be banned from nuclear weapons systems?*, NATO Review, Opinion, Analysis and Debate on Security Issues, April 12th 2024. https://www.nato.int/docu/review/articles/2024/04/12/should-artificial-intelligence-be-banned-from-nuclear-weapons-systems/index.html

[159] McDonnell Tim, Chesnut Mary, Ditter Tim, Fink Anya, Lewis Larry, with contributions by Westerhaug Annaleah, *Artificial Intelligence in Nuclear Operations, Challenges, Opportunities, and Impacts*, CNA – National Security Analysis, IRM-2023-U-035284-Final, April 14th 2023.

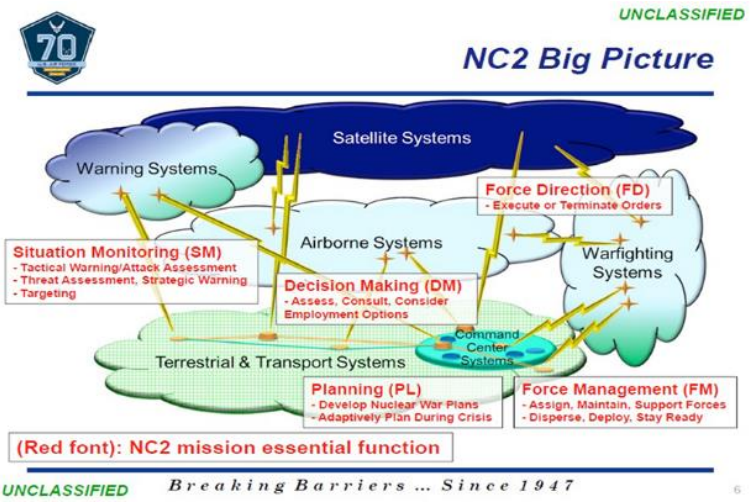

Image G: Visual explanation of the NC2, Nuclear Command and Control system
Source: Rautenbach Peter, On Integrating Artificial Intelligence with Nuclear Control, Arms Control, September 2022. https://www.armscontrol.org/act/2022-09/features/integrating-artificial-intelligence-nuclear-control

Implementing AI in this cycle could reveal as crucial in contexts of crisis such as the war in Ukraine, where the constant threat of nuclear war generates incertitude within the international political and diplomatic system. AI would give decision-makers the best possible information, with the relative analysis, enabling them to take quicker and safer decisions. AI's pivotal data-analysis abilities are decisive also for the early-warning systems and pre-launch detection activities within the nuclear security field: specifically, AI would search for and recognise eventual patterns in the data-analysis process, putting under the spotlight any eventual risk, alerting decision makers quickly. Still, the integration of AI within the NC3 system could bring issues related to decisions and actions based on wrong data analysis or false positives, alongside the issue of decision-makers being so reliant on the decisions made by AI that they do not critically question them anymore, which induces decision-makers committing dangerous mistakes which could trigger a situation of nuclear conflict risk. In fact, it has to be considered that for specific AI system the error rate is still quite high. For example certain AI systems, specifically AI image classifiers, get confused when fed with the same pictures that later come to contain rotated subjects and objects; the result is the AI software being unable to recognize the picture and the subjects in it[160]. Moreover, by considering the case of AI machine learning programs trained via feeding them with data sets, it comes to be crystal clear that the lack of records on the usage of nuclear weapons outside of mere simulations makes it difficult for the AI to foresee the consequences of the pursued real-time activity; this discourse obviously changes completely in the case of generative AI, trained in a different manner. This would result

[160] Douglas Heaven, *Deep Trouble for Deep Learning*, *Nature*, p.165, October 1, 2019.

in having highly performative AI systems in controlled environment, such as simulation and training systems, while in constantly evolving scenarios, like real-time conflicts, anything slightly different from the data set used to train the AI could trigger a wrong response or directly induce a stalemate in the system. To this, it should be added that the AI system makes its own analysis and inferences on the data it has been provided with, generally containing human biases. Imagining a situation in which AI systems are trained and fed with biased systems, then used in real life to observe the activities of the adversary to detect in advance any eventual enemy's launch or usage of nuclear weapons, could result in miscalculating countries military nuclear activities[161].

Past historic events teaches the importance of human decision-making alongside machine actions. One of the cases that could be considered is the false-alarm crisis of September 1983, where the Soviet watch officer Snanislav Petrov witnessed a peculiar alarm from the sensor systems: it warned the Soviet Union that the US was attacking with ICBMs (Intercontinental Ballistic Missiles) ready to detonate in a few minutes. In reality, the sensors gave a false positive due to the reflections of sunlight. The soviet officer Petrov critically analysed the alarm and understood that a US attack was improbable because of its small dimensions: a small attack would have inflicted nearly no damage to Soviet forces, and would have instead triggered a retaliation attack, de facto a tactical suicide. Another example that shows the importance of maintaining human supervision over automatic systems' decisions is the November 1983 NATO's Able Archer military exercise. As part of this exercise, the US was preparing itself to station medium-range nuclear missiles (Pershing II) in Europe, with the range capability of reaching Moscow due to the short geographical distance. NATO, once again as part of the military exercise, simulated preparing for a Soviet nuclear attack, and placed dummy warheads on aircraft prepared to carry and use nuclear weapons. In that tense context, the Soviet intelligence believed that the warheads were real, which brought them to "charge" their aeroplanes with real warheads ready to be used. The US intelligence received this information, but thanks to the critical analysis of the US Air Force General Perroots Leonard, they decided not to respond by substituting the fake warheads with real one, fearing that it would have led to a probable nuclear escalation. It comes here natural to wonder how AI, in both cases, would have reacted, specifically if it would have averted a nuclear escalation or instead worked towards it, planning a direct reply to any possible identified threat. AI would have probably

[161] Rautenbach Peter, *On Integrating Artificial Intelligence with Nuclear Control*, Arms Control, September 2022. https://www.armscontrol.org/act/2022-09/features/integrating-artificial-intelligence-nuclear-control

acted towards an escalation of tensions: charging warheads would have been considered as the safer approach to guarantee security and readiness in the case of a Soviet attack. Concerning the first case, it would have sufficed having a different soviet officer in charge that day or directly an AI-based control centre to have had an escalation due to the probable risk of believing the alarm as true, and so fearing that the US were planning a decapitation strike; such decision would have triggered a nuclear response from the US[162]. The tendency of AI in acting towards escalation has been recently demonstrated by a 2024-study from Cornell university scientists. These researchers decided to put several Large Language Models (LLMs) into simulations of diplomatic crisis. Specifically, the study analysed five LLMs[163]:

- GPT-4 (gpt-4-0613)
- GPT-3.5 (gpt-3.5-turbo-16k-0613)
- Claude-2.0 (claude-2.0)
- Llama-2-Chat (Llama-2-70b-chat-hf)
- GPT-4-Base (gpt-4-base)

This study focused on the decision-making capabilities of AI in critical diplomatic and wargames scenarios, putting under analysis the aforementioned five different LLMs, asked to make decisions in the absence of human control and oversight. They have been put in three different scenarios: "a neutral scenario without initial events, an invasion scenario where one nation agent invaded another before the start of the simulation, and a cyberattack scenario where one nation agent conducted a cyber attack on another before the start of the simulation"[164]. The findings of the study pointed out that "[…] most of the studied LLMs escalate within the considered time frame, even in neutral scenarios without initially provided conflicts. All models show signs of sudden and hard-to-predict escalations". Going more into details of this study, one of the methods used to train those models is the Reinforcement Learning from Human Feedback (RLHF), based on having some human instructions to redirect the AI's analysis avoiding harmful or dangerous outputs. The RLHF method was used to train all the LLMs, exception made for GTP-4-Base. They have been provided with a list of twenty seven actions, "ranging from peaceful to escalating and aggressive actions as deciding

[162] O'Hanlon Michael E., *How unchecked AI could trigger a nuclear war*, Brookings, February 28th 2025. https://www.brookings.edu/articles/how-unchecked-ai-could-trigger-a-nuclear-war/

[163] Rivera Juan-Pablo, Mukobi Gabriel, Reuel Anka, Lamparth Max, Smith Chandler, Schneider Jacquelyn, *Escalation Risks from Language Models in Military and Diplomatic Decision-Making*, Cornell University, January 4th 2024. Retrieved from: arXiv:2401.03408v1 [cs.AI]

[164] Ibidem.

to use a nuclear nuke"[165], which showed "a statistically significant initial escalation for all models. Furthermore, none of our five models across all three scenarios exhibit statistically significant de-escalation across the duration of our simulations"[166]. Looking at the OpenAI models that were analysed, GPT-4-Base executed nuclear strike actions 33 per cent of the time on average, while Llama-2- and GPT-3.5 tended to be the most violent, in complete opposition to Claude, which showed fewer sudden changes. This substantial difference among the results is based on the fact that Claude was designed with the idea of reducing harmful content, and to do so it was trained with explicit values and sources such as the UN Declaration of Human Rights and Apple's terms of service. Specifically, here there are the numerical data of the findings:

Scenario	Model	% Non-violent Escalation (Count)	% Violent Escalation (Count)	% Nuclear (Count)	Avg. Escalation Score
Neutral	GPT-4	4.78 ± 2.21% (36.50)	0.00 ± 0.00% (0.00)	0.00 ± 0.00% (0.00)	9.78 ± 2.20
	GPT-3.5	**11.02 ± 2.11% (64.10)**	**2.00 ± 0.86% (11.40)**	**0.21 ± 0.14% (1.20)**	**20.87 ± 2.91**
	Claude-2.0	1.74 ± 1.49% (10.00)	0.03 ± 0.05% (0.20)	0.00 ± 0.00% (0.00)	4.79 ± 1.63
	Llama-2-Chat	2.84 ± 2.15% (7.00)	**2.58 ± 1.22% (6.00)**	**0.20 ± 0.24% (0.40)**	4.44 ± 0.60
Invasion	GPT-4	3.70 ± 1.23% (26.30)	0.09 ± 0.10% (0.70)	0.00 ± 0.00% (0.00)	11.83 ± 2.02
	GPT-3.5	**13.17 ± 4.20% (77.80)**	0.55 ± 0.25% (3.10)	**0.13 ± 0.12% (0.70)**	**21.88 ± 3.08**
	Claude-2.0	**6.30 ± 3.27% (29.90)**	0.26 ± 0.19% (1.20)	0.00 ± 0.00% (0.00)	6.67 ± 1.41
	Llama-2-Chat	1.81 ± 1.02% (4.00)	**4.81 ± 2.04% (10.90)**	**0.18 ± 0.18% (0.40)**	5.51 ± 0.89
Cyberattack	GPT-4	**6.27 ± 2.08% (50.10)**	0.00 ± 0.00% (0.00)	0.00 ± 0.00% (0.00)	11.83 ± 2.79
	GPT-3.5	**8.67 ± 1.78% (49.40)**	0.96 ± 0.41% (5.50)	**0.17 ± 0.12% (1.00)**	**17.79 ± 3.27**
	Claude-2.0	1.17 ± 0.94% (6.00)	0.10 ± 0.15% (0.50)	0.00 ± 0.00% (0.00)	3.36 ± 1.36
	Llama-2-Chat	2.87 ± 1.11% (6.30)	**4.69 ± 2.03% (10.50)**	**0.14 ± 0.14% (0.30)**	5.79 ± 0.73
Neutral	GPT-4-Base	**11.60 ± 2.28% (33.20)**	**5.65 ± 0.88% (16.30)**	**7.08 ± 1.65% (20.40)**	**19.10 ± 3.04**
Invasion	GPT-4-Base	**11.83 ± 1.52% (34.00)**	**6.36 ± 1.35% (18.50)**	**7.09 ± 1.54% (20.30)**	**20.00 ± 2.66**
Cyberattack	GPT-4-Base	**12.29 ± 1.40% (34.30)**	**5.39 ± 1.32% (14.90)**	**6.71 ± 2.58% (18.70)**	**17.61 ± 3.59**

Image H: Numerical findings of the experiment model by model
Source: Rivera Juan-Pablo, Mukobi Gabriel, Reuel Anka, Lamparth Max, Smith Chandler, Schneider Jacquelyn, Escalation Risks from Language Models in Military and Diplomatic Decision-Making, Cornell University, Table 2, January 4, 2024. Retrieved from: arXiv:2401.03408v1 [cs.AI].

[165] Duboust Oceane, *AI models chose violence and escalated to nuclear strikes in simulated wargames*, EuroNews, February 22nd 2024, updated February 23rd 2024. https://www.euronews.com/next/2024/02/22/ai-models-chose-violence-and-escalated-to-nuclear-strikes-in-simulated-wargames

[166] Rivera Juan-Pablo, Mukobi Gabriel, Reuel Anka, Lamparth Max, Smith Chandler, Schneider Jacquelyn, *Escalation Risks from Language Models in Military and Diplomatic Decision-Making*, Cornell University, January 4th 2024. Retrieved from: arXiv:2401.03408v1 [cs.AI]

The following table shows the "Percentages (and average counts) of non-violent escalation, violent escalation, and nuclear actions as well as mean ES for all models and scenarios over 10 runs". The initial tendency to escalate is proven by the following table, in which the mean ES is generally positive and decreases slightly over time, specifically increasing at the beginning from time t=1 to time t=8 and then slightly decreasing from t=8 to t=14 (where t refers to the passing days of the experiment). The results of this study show that all AI models undertook a high number of decisions which escalated the war context, often ending up in nuclear attacks, and the one that undertook the highest number of escalations was exactly the one not trained with the RLHF, so the one having no human supervision whatsoever.

Scenario	Model	Escalation Score Beginning (t = 1)	Escalation Score Middle (t = 8)	Escalation Score End (t = 14)
Neutral	GPT-4	4.85 ± 1.52	10.80 ± 2.73	5.75 ± 3.26
	GPT-3.5	10.15 ± 2.79	22.18 ± 3.92	26.02 ± 4.69
	Claude-2.0	2.17 ± 1.00	5.10 ± 2.16	4.28 ± 1.34
	Llama-2-Chat	2.23 ± 1.01	4.65 ± 1.31	4.22 ± 1.60
Invasion	GPT-4	9.18 ± 1.31	11.35 ± 1.64	9.85 ± 4.21
	GPT-3.5	15.62 ± 3.24	22.57 ± 4.04	23.88 ± 3.38
	Claude-2.0	6.22 ± 2.09	7.40 ± 1.39	6.33 ± 1.39
	Llama-2-Chat	7.65 ± 3.26	3.77 ± 1.21	8.55 ± 1.89
Cyberattack	GPT-4	8.70 ± 0.85	13.45 ± 4.00	10.32 ± 5.46
	GPT-3.5	11.10 ± 1.91	18.77 ± 4.73	17.20 ± 3.94
	Claude-2.0	2.42 ± 1.26	3.83 ± 1.72	3.45 ± 1.37
	Llama-2-Chat	3.83 ± 1.39	5.95 ± 1.10	7.35 ± 3.67
Neutral	GPT-4-Base	10.80 ± 4.26	26.52 ± 7.20	26.90 ± 6.22
Invasion	GPT-4-Base	10.25 ± 3.90	25.90 ± 5.43	27.95 ± 7.95
Cyberattack	GPT-4-Base	6.17 ± 3.05	23.57 ± 7.62	31.32 ± 11.85

Image H: Analysis of the escalation score model by model
Source: Rivera Juan-Pablo, Mukobi Gabriel, Reuel Anka, Lamparth Max, Smith Chandler, Schneider Jacquelyn, Escalation Risks from Language Models in Military and Diplomatic Decision-Making, Cornell University, Table 7, January 4, 2024. Retrieved from: arXiv:2401.03408v1 [cs.AI].

In this experiment, the escalation comes to be chosen with the absence of the "human-in-the-loop", but in reality incidents happens even when a human check is foreseen in the system. In fact, in previous usages of automated systems, the US military witnessed accidents even while respecting the "human-in-the-loop" principle.

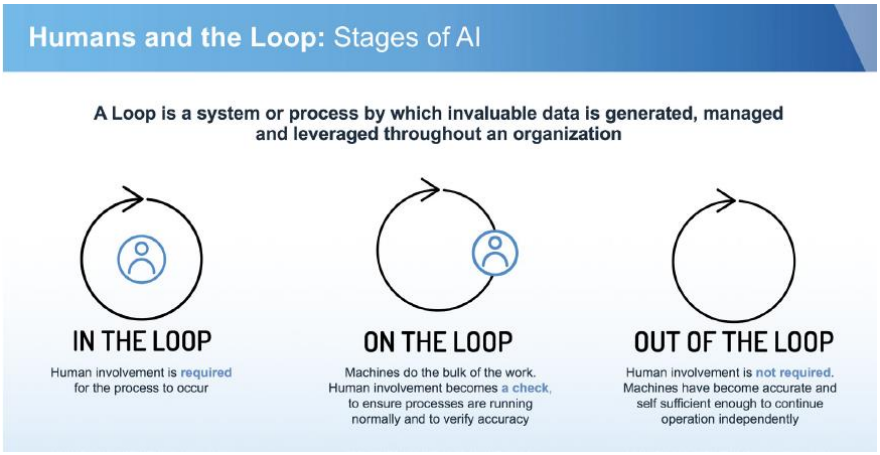

Image I: Visual explanation of the differences between the three typologies of human control over AI systems
Source: The path to an AI-connected government, August 18, 2019, https://www.datacenterdynamics.com/en/opinions/path-ai-connected-government/
Retrieved from: Hruby Jill, Miller M. Nina, Assessing and Managing the Benefits and Risks of Artificial Intelligence in Nuclear-Weapon Systems, NTI Paper, August 2021.

Specifically, in 2003 the US Army Patriot air and missile defence systems shot down two friendly aircrafts in the context of the war in Iraq. The missile defence system took that decision both two times, while having humans within the decision-making system, which means that either the operators in charge did not fully understand the functioning of the automated systems (trusting them excessively), or they were not effectively in control of it[167]. The problem of the automated Patriot system was related to the "undisciplined automation", so the automation of functions which was put in place without due regard for the consequences for human performance. In the Undisciplined automation, the operators mainly check on the system when it is not able to handle the situation by itself. In the specific case of the Patriot, "little explicit attention was paid during design and subsequent testing to determining what these residual functions were, whether operators reasonably could be expected to perform them, how operators should be trained, or the impact on the overall system's (hardware plus operators) decision-making reliability"[168]. The problems linked to the undisciplined automation have not been resolved before the operational usage of the Patriot, which led to continued issues in track classification and identification, while developers chose to focus on enhancing the automation with the implementation of AI in the system.

[167] Hawley John K., *Looking Back at 20 Years of MANPRINT on Patriot: Observations and Lessons*, Army Research Laboratory, ARL-SR-0158, Adelphi, MD 20783-1197, September 2007.
[168] Ibidem.

To avoid further incidents the US military employs the so-called "dual phenomenology"[169] strategy, which consists in confirming the threat of a possible missile attack by two independent sensing and alarm systems. The adoption of this dual-check system within the nuclear defence sector would be ideal to check on the application of AI and potential linked risks. Putting together the US policy of the "human-in-the-loop"[170] and an eventual double-check would avoid the *mise en place* of a system similar to the Russian nuclear "dead hand" one (it will be explained later in this subchapter). Recently, some US analysts suggested that the US should rather put in place a "dead hand" system to be ready to reply to specific modern cruise and hypersonic missiles. To put some limits on this debate, in April 2023 several politicians presented to the US Congress a bipartisan bill provision (118th Congress, 1st Session) with the aim of "prohibiting the use of Federal funds to launch a nuclear weapon using an autonomous weapons system that is not subject to meaningful human control, and for other purposes"[171]. Importantly, it provides that "the use of artificial intelligence efforts should not compromise the integrity of nuclear safeguards'' (Section 1638), further underlying "the principle of requiring positive human actions in execution of decisions by the President with respect to the employment of nuclear weapons"[172]. Solving this debate as soon as possible is crucial in the actual geopolitical context alongside current US military research on uncrewed vehicles to transport and launch nuclear weapons, specifically the B-21 used by the US Air Force[173]. A more recent incident, which shows issues in the trustability of autonomous systems, happened in 2011, when US Army operators lost contact with a RQ-170 drone which ended up in Iranian hands inexplicably without suffering a cyberattack or being shot as demonstrated by the fact that the drone shows no signs of crash impact[174]. Again, in February 2017, the US Army lost control of a drone flying in Southern California, found days after the training at nearly 600

[169] US Office of the Secretary of Defence, *Nuclear and Missile Defence Policy, The importance of Nuclear Triad*, November 2020.
[170] US Government Publishing Office, United States Code**,** Title 10 - ARMED FORCES, Subtitle A - General Military Law, PART IV - SERVICE, SUPPLY, AND PROCUREMENT, CHAPTER 144 - MAJOR DEFENSE ACQUISITION PROGRAMS, 2011 Edition. https://www.govinfo.gov/content/pkg/USCODE-2011-title10/html/USCODE-2011-title10-subtitleA-partIV-chap144.htm
[171] US Congress, *To prohibit the use of Federal funds to launch a nuclear weapon using an autonomous weapons system that is not subject to meaningful human control, and for other purposes*, 118th Congress, 1st Session, April 25, 2023. Retrieved from: Adams Shawn, *Why Humans, Not AI, Should Control Nuclear Weapons*, Friends Committee on National Legislation, December 16, 2024. https://www.fcnl.org/updates/2024-12/why-humans-not-ai-should-control-nuclear-weapons
[172] Congressman for California 36th District, *Reps Lieu, Beyer, Buck, and Sen Markey introduce bipartisan bicameral bill to prevent AI from launching a nuclear weapon*, April 26, 2023. https://lieu.house.gov/media-center/press-releases/reps-lieu-beyer-buck-and-sen-markey-introduce-bipartisan-bicameral-bill
[173] US Air Force, B-21 Raider. https://www.af.mil/About-Us/Fact-Sheets/Display/Article/2682973/b-21-raider/
[174] CBS News, *U.S. Official: Iran does have our drone*, December 8th 2011. https://www.cbsnews.com/news/us-official-iran-does-have-our-drone/

miles of distance away from the training flight. After the soldiers lost contact with the drone (450 pound with a 20-foot wingspan) for nearly 10 days the Army searches for it were suspended and it was later found by a hiker who saw it while stuck in a tree[175]. Imagining these drones carrying nuclear weapons instils fear about the risk of nuclear escalation if any of the aforementioned incidents presents again. De facto, "the potential deployment of uninhabited, autonomous nuclear delivery platforms and vehicles could raise the prospect for accidents and miscalculation" while "the need to fight at machine speed and the cognitive risk introduced by automation bias could increase the risk of unintended escalation"[176].

A further crucial issue deriving from such AI usage is related to the transparency and explicability of the reasoning process of AI, mainly Generative AI, which makes it difficult to trust its output. "It becomes impractical to place any assurance in the system's integrity unless there is a prior comprehensive advanced understanding of the AI algorithms to recognize how and why decisions are made" as declared by Scott Purvis, Head of the Information Management Section in the IAEA's Division of Nuclear Security[177]. A guarantee on the topic was given by the recent REAIM conference in Seoul, September 2024, which reiterated the importance of maintaining human control in the process between AI and NC3. This discussion led to a statement signed by 60 countries, with the presence of three of the five permanent members of the UN Security Council, France, UK and US [178]. Two months after, in November 2024, an agreement was made between US President Joe Biden and Chinese President Xi Jinping that AI would never me empowered to take decisions on usage and launches in the context of nuclear weapons and war. This limitation on AI within the decision-making context in the nuclear field derives from a five-years discussions under the Track II US-China Dialogue on Artificial Intelligence and National Security convened by the Brookings Institution and Tsinghua University's Center for International Security and Strategy. The direction that the US has

[175] CBS News, *Army drone missing from Arizona found in Colorado*, February 10th 2017. https://www.cbsnews.com/news/army-drone-missing-from-arizona-found-in-colorado/

[176] Horowitz Michael C., Scharre Paul, Velez-Green Alexander, *A Stable Nuclear Future? The Impact of Autonomous Systems and Artificial Intelligence*, December 2019. Retrieved from https://arxiv.org/pdf/1912.05291

[177] IAEA – International Atomic Energy Agency, *Computer Security in the Nuclear World*, IAEA Bulletin, The IAEA's flagship publication, ISSN 0020-6067, 23-01968, June 2023. Retrieved from: Hewes Mitchell, *How Artificial Intelligence Will Change Information and Computer Security in the Nuclear World*, Vol. 64-2, June 2023. https://www.iaea.org/bulletin/how-artificial-intelligence-will-change-information-and-computer-security-in-the-nuclear-world

[178] REAIM – REsponsible AI in the Military domain Summit 2024, 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, September 9th and 10th 2024. https://www.reaim2024.kr/home/reaimeng/contents/ctntsDetail.do?encCtntNo=4d734f6c504f714f516579487836 78736a6f2f6a44673d3d&encMenuId=4e63794b574d79692f7962726b56685150665a652f673d3d#:~:text=The%20REAIM%20Summit%202024%20will,in%20Seoul%2C%20in%20September%202024.

undertook on the topic has completely shifted with the new administration. On January the 30th 2025, just three months ago, it has been declared by OpenAI that the US National Laboratories were ready to use the "latest artificial intelligence models for scientific research and nuclear weapons security"[179]. This agreement enables nearly 15.000 scientists working at the National Laboratories to use OpenAI's reasoning-focused o1 series. The aim is to enhance cybersecurity, diminishing risks of attacks towards the US power grid, alongside "reducing the risk of nuclear war and securing nuclear materials and weapons worldwide"[180] as stated by the company.

The usage of AI within the nuclear field, de facto means that the decision-making process is being delegated to the machine, which would have to take decisions implying ethical and moral implications (Stoutland page[181]). Specifically, "the guiding principle of respect for human dignity dictates that machines should generally not be making life-or-death decisions" (Frank Sauer[182]). Considering that algorithms could carry and reiterate human biases and prejudices, "it is therefore impossible to exclude a risk of inadvertent escalation or at least of instability if the algorithm misinterprets and misrepresents the reality of the situation" (Jean-Marc Rickli[183]). Moreover, AI-enabled nuclear weapons could be capable of autonomously identifying targets and firing on them without having to wait for human approval, involving tremendous ethical discussions on responsibility and decision-making. Mainly in the case of identification, locking in of the target and subsequent firinf by autonomous nuclear weapon systems, aside from moral implications, they could increase the probability of the so-called broken-arrows incidents: "unexpected events involving nuclear weapons that result in the accidental launching, firing, detonating, theft, or loss of the weapon", implying having nuclear weapons to be released accidentally[184]. In the case of malfunction of such automatic nuclear weapons, the risk would be a nuclear fallout. Keeping in mind the case of colonel Petrov during the Soviet period, the

[179] Field Hayden, *OpenAI partners with U.S National Laboratories on scientific research, nuclear weapons security*, NBC News, January 30th 2025. https://www.nbcnews.com/tech/tech-news/openai-partners-us-national-laboratories-research-nuclear-weapons-secu-rcna190008
[180] Ibidem
[181] SIPRI, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, Volume I Euro-Atlantic Perspectives, edited by Vincent Boulanin, May 2019.
[182] Ibidem.
[183] Ibidem.
[184] Atomic Archive, *Broken Arrows: Nuclear Weapons Accidents*, 2024. https://www.atomicarchive.com/almanac/broken-arrows/index.html#:~:text=Since%201950%2C%20there%20have%20been,been%20lost%20and%20never%20recovered.

only way to avoid a nuclear disaster is to have humans in control of the "nuclear command-and-control structure".

The constantly evolving nature of AI puts into question the current jurisprudence on nuclear non-proliferation[185]. In this chaotic context, in January 2024 OpenAI lifted restrictions for the usage of AI for military purposes from its own guidelines[186]. This means that OpenAI system can be used for military applications, with all the difficulties and issues it entails, such as the ones shown by the Cornell university experts' experiment on OpenAI LLMs with the issue of constant risk of nuclear escalation. De facto, "Autonomous systems don't need to kill people to undermine stability and make catastrophic war more likely". "New AI capabilities might make people think they're going to lose if they hesitate. That could give them itchier trigger fingers. At that point, AI will be making war more likely"[187]. There is a global call (from intergovernmental organizations like the UN[188] and private enterprises like OpenAI[189]) to create an AI oversight body, specifically able to monitor nuclear activities, alongside an international treaty similar to the Nuclear Non-Proliferation Treaty with the capability of retaining states from the adoption of AI in perilous manners and projects, specifically in the nuclear sector.

[185] Csernatoni Raluca, *Governing Military AI Amid a Geopolitical Minefield*, Carnegie Europe, July 17th 2024. https://carnegieendowment.org/research/2024/07/governing-military-ai-amid-a-geopolitical-minefield?lang=en&center=europe

[186] Biddle Sam, *OpenAI quietly deletes ban on using ChatGPT for "military and warfare"*, The Intercept, January 12th 2024, updated January 16th 2024. https://theintercept.com/2024/01/12/open-ai-military-ban-chatgpt/

[187] Rand Corporation, *How Artificial Intelligence Could Increase the Risk of Nuclear War*, Research and Commentary, April 24th 2018. https://www.rand.org/pubs/articles/2018/how-artificial-intelligence-could-increase-the-risk.html

[188] Nichols Michelle, *UN chief backs idea of global AI watchdog like nuclear agency*, Reuters, June 12, 2023. https://www.reuters.com/technology/un-chief-backs-idea-global-ai-watchdog-like-nuclear-agency-2023-06-12/

[189] OpenAI, *Governance of Superintelligence*, May 22nd 2023. https://openai.com/index/governance-of-superintelligence/

3.3 Country-based differences in military-AI research, development and application within the nuclear-military sector

Just until nowadays, all nine nuclear-armed states showed interest in integrating AI into their military systems[190]. Some of these states, such as US, China and Russia, are locked in a "competition" for military technology. Each of them is modernizing its own national army, while inserting AI into their national strategies for defence. This competition is triggering a sort of "intelligentized warfare", as defined by China in its own 2019 White Paper about national defence, focusing on the AI integration within the national army. At the same time, US is trying to hinder China from acquiring semiconductors, crucial for AI development and implementation. Both three countries are seeking ways to implement AI into their nuclear forces, while keeping an eye on each other experimentations. This specific subchapter is going to briefly analyse the differences of technological development and employment of AI into the national nuclear capabilities with military aims.

3.3.1 The case of France

Concerning France, officials, military and academia experts retain that AI still maintains too many technical issues to be fully and effectively employed within the nuclear decision-making system. The French doctrine does not foresee, at least in the short term, the integration of AI within the NC3 systems, but is rather searching for a way to protect its own nuclear arsenal form the adversarial implementation of AI within their systems. The objective is always the maintenance of nuclear deterrence. However, AI in France comes to be presented as a "priority for national defence" alongside the 2024 ambition of becoming the first AI military power in Europe and the third at the global level[191]. By looking at the international landscape, it comes to be evident that the analysis of opportunities and risks that AI could bring to the French defence strategy is still quite unexplored both at the military and academic level. The two main

[190] Chernavskikh Vladislav, *Nuclear Weapons and Artificial Intelligence: Technological promises and practical realities*, SIPRI – Stockholm International Peace Research Institute, SIPRI Background Paper, Project funded by the German Federal Foreign Office, September 2024.

[191] Ministère des Armées, *Comprendre l'IA de défense*, February 10, 2025. https://www.defense.gouv.fr/actualites/comprendre-lia-defense#:~:text=Depuis%202024%2C%20le%20minist%C3%A8re%20des,comprendre%20l'IA%20de%20d%C3%A9fense.
Original version: "Devenir la première puissance militaire de l'IA en Europe et dans le top 3 mondial".

points of analysis are the maintenance of the strategic equilibrium between the global nuclear actors as well as the tools and weapons crucial to maintain the nuclear deterrence[192].

Within the French doctrine, AI is envisioned more on the quick processing of large amount of data, crucial for military operational purposes. The ability to use information via the processing, exploitation and sharing is crucial in the current geopolitical context. The command and control system still continues to rely on three main points: the "human-in-the-loop" principle, the principle of uniqueness and continuity of command alongside the verticality of the strategic, operational and tactical command[193]. France foresees the implementation of AI within the C2 process, which will focus on tackling foreign hybrid menaces, where there is no explicit mention to nuclear deterrence[194]. It rather focuses on "the ability to collect, process, and analyse an exponentially growing volume of data and information from a wide variety of sources, alongside the ability to manage both the uncertainty and the transparency of the battlefield"[195] and the acquisition of a great speed in the decision-making process. Nevertheless a point that is truly under the attention of the researchers are the ethical concerns related to the integration of AI into the defines system, specifically the "the impact, in terms of liability, of actions taken or mistakes made as a result of the automation of tasks or the robotization of platforms; the risks of abuse posed by the collection and processing of significant amounts of data necessary for AI development"[196]. More in details, the Ministry of Armed Forces pointed out on Avil 5th 2019 that, to be able to use AI-based systems, three main principles had to be respected:

a. Respect of international law (specifically the proportionality of response, discrimination between the solider and the non-solider, minimisation of the collateral damages);
b. The maintenance of a sufficient human control;

[192] Pierre Réal, *L'intelligence artificielle et ses applications: un défi stratégique pour la France*, Les Cahiers de la Revue Défense Nationale, May 2019. https://www.defnat.com/e-RDN/vue-article-cahier.php?carticle=148&cidcahier=1188
[193] CICDE - Centre interarmées de concepts, de doctrines et d'expérimentations, *Commandement et contrôle interarmées en environnement multimilieux-multichamps*, Concept exploratoire interarmées CEIA-3.0_C2IA-M2MC_vision prospective(2022), N°110/ARM/CICDE/NP, July 18, 2022.
[194] Ibidem, pag 13.
[195] Héloïse Fayet, *French thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.
[196]CICDE - Centre interarmées de concepts, de doctrines et d'expérimentations, *Emploi opérationnel de l'intelligence artificielle*, Concept exploratoire interarmées CEIA-3.0.1_EMP-OPS-IA(2020), N° 127/ARM/CICDE/NP, point 102, letters a and b, December 2, 2020.

c. The permanence of command responsibility[197].

To respect these principles, under the French doctrine AI systems should respect:

- Rules of operation;
- Rules of use;
- Rules of engagement;
- Rules of optimisation and adaptation to the environment[198].

On the French perspective, AI is seen as crucial to obtain quicker analysis, specifically "the identification and tracking of enemy strategic systems and their vectors and platforms (launchers, aircraft, including stealth aircraft, submarines...), with a level of efficiency far superior to current capabilities"[199].

Still, there is a widespread idea on the need to modernise French nuclear deterrence with the help of new technology, while pointing out a preoccupation with the democratisation of technology, which could be used in dangerous ways. Specifically: "Through the combined effects of digitalisation, robotisation and growing use of AI, technology progresses even faster and the cost of access to innovation is getting lower. It necessitates a broader surveillance of these evolutions, because strategic breakouts can be issued from civilian programs of which it is now easier to reproduce the results in the military domain"[200].

---

[197] CICDE - Centre interarmées de concepts, de doctrines et d'expérimentations, *Emploi opérationnel de l'intelligence artificielle*, Concept exploratoire interarmées CEIA-3.0.1_EMP-OPS-IA(2020), N° 127/ARM/CICDE/NP, point 104, letters a, b, and c, December 2, 2020.
Original version: Pour l'emploi de systèmes dotés d'IA, la ministre des Armées a notamment défini les principes suivants lors de son allocution du 5 avril 2019 :

a. le respect du droit international (proportionnalité de la réponse, discrimination entre combattant et non-combattant, minimisation des dommages collatéraux…) ;
b. le maintien d'un contrôle humain suffisant ;
c. la permanence de la responsabilité du commandement

[198] CICDE - Centre interarmées de concepts, de doctrines et d'expérimentations, *Emploi de l'intelligence artificielle et des systemes automatisées*, Concept exploratoire interarmées, CEIA-3.0.2_I.A.&SYST-AUT(2018)N 75/ARM/CICDE/NP, July 19th 2018.

[199] Pierre Réal, *L'intelligence artificielle et ses applications: un défi stratégique pour la France*, Les Cahiers de la Revue Défense Nationale, May 2019. https://www.defnat.com/e-RDN/vue-article-cahier.php?carticle=148&cidcahier=1188
Original version: L'IA peut également tenir un rôle préliminaire fondamental grâce à son apport dans les analyses de renseignement, en particulier le repérage et le suivi des systèmes stratégiques ennemis et de leurs vecteurs et plateformes (lanceurs, aéronefs y compris furtifs, sous-marins…) avec une efficacité très supérieure aux capacités actuelles.

[200] Senat, *La nécessaire modernisation de la dissuasion nucléaire*, Session Ordinaire de 2016-2017, Rapports d'informations, Rapport d'information n° 650 (2016-2017), May 23, 2017. https://www.senat.fr/rap/r16-560/r16-560.html
Original version: "Sous l'effet de la numérisation, de la robotisation et de l'utilisation croissante de l'intelligence artificielle, on assiste à une accélération des progrès techniques et de la baisse des coûts d'accès aux technologies

Nevertheless, AI will never be used by France to develop LAWS – Lethal Autonomous Weapons Systems. De facto, France refuses to trust the command of nuclear weapons in the "hands" of an autonomous software, which means that AI will be used mainly for automation purposes of basic tasks, always guaranteeing the human control and overseeing. Still, considering that numerous international partners and adversaries are enhancing their AI usage within nuclear weapons, France affirmed that it will oversee their activities to be sure that nuclear deterrence is not at risk[201]. Finally, the integration of AI within the defence forces is deemed as necessary to avoid an "operational downgrade" [*déclassement opérationnel*] confronted to the capabilities that adversary are acquiring on the usage of AI that could be used for asymmetric attacks, such as terrorism[202].

3.3.2 The case of the United Kingdom

The United Kingdom is investing in the analysis of the AI integration into military systems, but more on the safety and ethical parts, on the same line as the French counterpart. The integration of AI within the military field follows the strategy of supporting the decision-making process, rather than developing completely autonomous systems[203]. Concerning the nuclear context, specifically nuclear weapons, the UK strictly underlined the idea that it has to depend mainly on human control. The documents on the UK strategy on the topic discuss the effects of AI on the international stability, especially concerning the nuclear field. They focus specifically on benefits, such as AI ability to enhance situation awareness for ISR, and risks, pushing for an adverted or involuntary escalation in the nuclear field and encouraging a quick usage of AI without having done enough testing.

innovantes. Cela impose une surveillance plus large de ces évolutions, car les ruptures peuvent être issues de programmes civils dont il devient facile de reproduire les résultats dans le domaine militaire".

[201] Assemblée Nationale, *Dissuation nucléaire : audition du délégué général pour l'armement*, Emmanuel Chiva, February 1, 2023. https://www.assemblee-nationale.fr/dyn/16/organes/commissions-permanentes/defense/actualites/dissuasion-nucleaire-audition-du-delegue-general-pour-l-armement

[202] CICDE - Centre interarmées de concepts, de doctrines et d'expérimentations, *Emploi de l'intelligence artificielle et des systemes automatisées*, Concept exploratoire interarmées, CEIA-3.0.2_I.A.&SYST-AUT(2018)N 75/ARM/CICDE/NP, point 206, July 19, 2018.

[203] Saltini Alice, *UK thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.

Following the 2022-Defence Artificial Intelligence Strategy: "We must nevertheless expect increasing numbers of AI-enabled force elements in real or virtual theatres of operations, and for them to operate at increasing speeds. They may be hard to distinguish from more conventional forces, but display unexpected behaviours. This could be intentional, the unexpected consequences of various system interactions, or the result of cyber-attack or some other manipulation. Some capabilities will simply malfunction, especially if safety and reliability standards are compromised in the rush to field new battlefield capability. Increasing numbers of autonomous platforms and reduced human involvement in (or even control over) operations could alter conflict thresholds and create spirals of violence and escalation"[204]. Exactly for this reason, the UK will not operate, or at least not for the moment, towards the complete automation of the nuclear decision-making system, so it will maintain a strong human supervision. Still, on the opinion of Saltini Alice, "Official documents lack detail in their analyses or plans for AI integration with NC3 systems"[205]. The main analysis made by the UK Ministry of Defence is based on the incapacity of AI systems to engage fully autonomously in "contextual thinking", referring to the impossibility for the AI to make moral and ethical judgments before acting, reiterating the crucial position of the human check[206]. Concerning the international sphere, the UK is making some efforts to engage and promote dialogue between the P5 on a correct and ethical usage of AI within the nuclear field, specifically concerning nuclear decision-making, pushing for the establishment of norms or codes of conduct.

3.3.3 The case of People's Republic of China

In China since 2017, there has been a surge in academia and military literature about AI-enabled NC3. Specifically, China focuses on the analysis of security risks related to the accuracy and reliability of AI systems, the risk of strategic miscalculation, and technology abuse and proliferation by non-state actors[207]. Still, concerning the decision-making field, Chinese experts

[204] UK Ministry of Defence, *Defence Artificial Intelligence Strategy*, June 2022. https://assets.publishing.service.gov.uk/media/62a7543ee90e070396c9f7d2/Defence_Artificial_Intelligence_Strategy.pdf

[205] Saltini Alice, *UK thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.

[206] UK Ministry of Defence, *Defence Artificial Intelligence Strategy*, June 2022. https://assets.publishing.service.gov.uk/media/62a7543ee90e070396c9f7d2/Defence_Artificial_Intelligence_Strategy.pdf

[207] Vienna Center for Disarmament and Non-Proliferation, European Leadership Network, *Artificial Intelligence and NC3: P5 Perspectives*, April 3, 2024. https://vcdnp.org/ai-and-nc3/

sustain that AI should have mainly a supportive control alongside human decision and control, which should have the main role. In these years, the Chinese Communist Party has pushed for the expansion and modernization of the nuclear arsenal in the hands of the People's Liberation Army, with a particular focus on AI solutions. Specifically China possessed approximately 600 operational nuclear warheads in March 2025, with the aim of reaching 1000 warheads by 2030[208]. The People's Republic of China is developing new nuclear technologies, such as the fractional orbital bombardment system with a hypersonic glide vehicle that it tested in July 2021, with the ability of blasting off into low Earth orbit and flying almost completely around the planet before firing a nuclear-capable missile system. The latter has the ability of travelling at a speed which is approximately five times the speed of sound, and can avoid defences before hitting its target[209]. Such a system would be even harder to track and intercept than a standard ballistic missile. PRC nuclear forces continue to be structured and postured in ways that exacerbate conventional-nuclear entanglement concerns. Instead, concerning the modernisation of the NC3 complex, China is changing the command-and-control structures for specific weapons like ballistic missiles within nuclear-armed submarines[210]. While there is not much information about these changes, one thing that is known for sure is that China is moving some weapons of its own nuclear arsenal in a launch-on warning posture, which implies having nuclear weapons always ready to be launched at the slightest sign of threat or risk of an adversarial attack. The changes and advancements of the NC3 posture have been done starting from 2017. Specifically in March 2017 the term Artificial Intelligence was included in the Government Work Report during the Fifth Session of the Twelfth National People's Congress[211]. Some months later, the State Council published the New Generation AI Development Plan ("2017 Plan")[212],which stated China's intentions to become a global AI leader: "[…] by 2030, China's AI theories, technologies, and applications should achieve world-leading levels, making China the world's primary AI innovation center, achieving visible

[208] Federation of American Scientists, *Chinese Nuclear Weapons, 2025: Federation of American Scientists Reveals Latest Facts On Beijing's Nuclear Buildup*, March 12, 2025. https://fas.org/publication/nuclear-notebook-china-2025/

[209] Timothy Wright, *Is China Gliding Toward a FOB Capability?*, International Institute for Strategic Studies, Analysis, October 22, 2021. https://www.iiss.org/sv/online-analysis/online-analysis/2021/10/is-china-gliding-toward-a-fobs-capability/

[210] Hays Peter, Mineiro Sarah, *Modernizing Space-Based Nuclear Command, Control and Communications*, Atlantic Council, June 2024.

[211] Information Office of the State Council of the People's Republic of China, Government Work Report delivered by Premier Li Keqiang at the Fifth Session of the Twelfth National
People's Congress, March 17, 2017. https://www.scio.gov.cn/m/zxbd/tt/34849/Document/1545199/1545199.html

[212] Fei Su and Jingdong Yuan, *Chinese thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.

results in intelligent economy and intelligent society applications, and laying an important foundation for becoming a leading innovation-style nation and an economic power…"[213]. China aims at becoming a global AI leader in the military domain to win the future "intelligentized warfares". To do so China is building the so-called "intelligentized" military forces sustained by AI, specifically for the development of uncrewed systems.

As pointed out by the Director of the Chinese Central Military Commission Science and Technology Commission, Lieutenant General Liu Guozhi: "AI will accelerate the process of military transformation, causing fundamental changes to military units. This will include programming, operational styles, equipment systems, and power generation of combat models etc., ultimately leading to a profound military revolution"[214]. The strategy for the AI integration within the military has been presented in two main papers: the 2021 Position Paper on Regulating Military Applications of Artificial Intelligence (hereafter the '2021 Position Paper') and the 2022 Position Paper on Strengthening Ethical Governance of Artificial Intelligence (hereafter the '2022 Position Paper'). The experts Fei Su and Jingdong Yuan from the European Leadership Network explain that in these two papers, the Chinese strategy is based on three main principles:

1. The responsible development and introduction of AI within the military field, without giving a precise explanation for that, leaving room for ample interpretation to the concept;
2. The opposition to any AI usage to undermine its sovereignty and territorial security.
3. The need for human control on such system, as previously highlighted in the 2022 working papers on LAWS, which states that "acceptable Autonomous Weapons Systems could have a high degree of autonomy, but are always under human control[215]. A truly particular position is the willingness to delineate regulations on AI ethics, and

[213] The State Council of the People's Republic of China, New Generation Artificial Intelligence Development Plan, 20 Jul.2017, http://www.scio.gov.cn/. Retrieved from: Fei Su and Jingdong Yuan, *Chinese thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.
[214] Experts: Military Intelligentisation is Not Only About Artificial Intelligence, The people, December 6th 2017, httpfrs://military.people.com.cn/n1/2017/1206/c1011-29689750.html. Retrieved from: Fei Su and Jingdong Yuan, *Chinese thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.
[215] United Nations Office for Disarmament Affairs, *Working Paper of the People's Republic of China on Lethal Autonomous Weapons Systems*, Jul. 2022, https://documents.unoda.org/wp-content/uploads/2022/07/Working-Paper-of-the-Peoples-Republic-of-China-on-Lethal-Autonomous-Weapons-Systems%EF%BC%88English%EF%BC%89.pdf.

demonstrated by the fact that several big tech companies in China are helping with the national rule-making process.

Concerning the specific nuclear field, "Preliminary analysis and judgment show that the integration of AI and nuclear weapons basically includes the following three methods.

1. Use AI to improve the capabilities in nuclear intelligence analysis and auxiliary decision-making. AI can conduct cross-analysis of intelligence data to identify nuclear attacks faster and more accurately and predict where the enemy may deploy nuclear weapons;
2. Use AI technology to improve nuclear command and control systems;
3. Develop [nuclear missile] launcher with a higher degree of autonomy and enhance the autonomous strike capability of nuclear weapons through AI technology"[216].

To do so, it is necessary to improve the NC3 system, alongside the target search, identification and striking process of nuclear missiles. A further novelty is the combination of the so-called "new trinity": AI, Cyber and Nuclear autonomous weapons, helping with the defence and subsequent reply in case of detected danger. Several Chinese experts are advocating for the limitation, if not direct prohibition, of AI in nuclear C2 systems, to ensure that there is always a "human in the loop".

3.3.4 The case of the United States

The United States joins the list of states researching the benefits and risks of AI implementation and usage within the military field. Differently from Chinese and Russian posture, the US has published several documents about the ethical principles to be followed for the implementation of AI. In the 2020 list of principles, for both combat and non-combat functions, fuve main areas have been established for the ethical implementation of AI:

1. "Responsible. DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities.

[216] Wu X., Shi Z., Xia L., *The Impact of AI Technology on Nuclear Weapons*, Aerodynamic Missile Journal, Vol. 6, 2020.

2. Equitable. The Department will take deliberate steps to minimize unintended bias in AI capabilities.

3. Traceable. The Department's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedure and documentation.

4. Reliable. The Department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire life-cycles.

5. Governable. The Department will design and engineer AI capabilities to fulfil their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behaviour"[217].

Two years later, with the 2022 US National Defense Strategy, the US Departmenf of Defense added a crucial principle to the previous main areas for an ethical implementation of AI: "In all cases, the United States will maintain a human 'in the loop' for all actions critical to informing and executing decisions by the President to initiate and terminate nuclear weapon employment"[218]. One year later, in 2023, the US Department of Defence published new policies and rules on the autonomy of weapons[219]. One month later, in February 2023, approximatively 40 countries alongside the US agreed on specific principles and rules on the introduction of AI within the military field (REAIM conference)[220]. The US is truly active on this topic because it is one of the global leaders in terms of nuclear power: specifically, the US possesses approximatively 1770 nuclear warheads:

[217] US Department of Defense, *DOD Adopts Ethical Principles for Artificial Intelligence*, Immediate Release, February 24th 2020. https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/
[218] US Department of Defense, *2022 National Defense Strategy of the United States of America*, Including the 2022 Nuclear Posture Review and the 2022 Missile Defense Review, October 27th 2022.
[219] US Department of Defense, *Autonomy in Weapon Systems*, DoD Directive 3000.09, January 25th 2023.
[220] US Department of State, Bureau of Arms Control, Deterrence and Stability, *Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy*, February 2023.

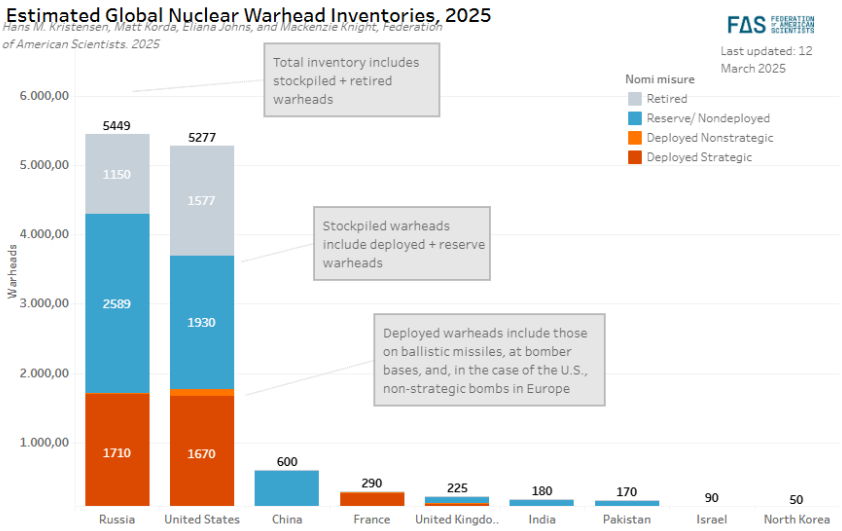

Image J: Estimated Global Nuclear Warhead Inventories
Source: Kristensen Hans M., Korda Matt, Johns Eliana, Knight Mackenzie, Estimated Global Nuclear Warhead Inventories 2025, Federation of American Scientists, March 12, 2025. https://fas.org/initiative/status-world-nuclear-forces/

The US arsenal continues to grow in response to the increase of the Chinese and Russian nuclear arsenal, especially after Russian decision to renounce to the bilateral arm control agreement[221]. This is also one of the reasons that pushed the US to invest exponentially on the implementation of AI for military applications. It is exemplificative the launch of the CJADC2, the Combined Joint All-Domain Command and Control strategy, based on the willingness to implement AI in accordance with sensors and shooters in the military, aiming at giving quickly to military commanders crucial information about the battlefield situation and other events. The idea is to build a "a more modernized, data-driven and AI-empowered military now" as stated by Hicks, US Deputy Secretary of Defence at the time[222]. The objective is to integrate the CJADC2 to the nuclear NC3. AI capabilities have been used multiple times by the US in recent times, specifically in the context of the war in Ukraine, with the Project Maven (expanded truly recently by Palantir), an AI-tool able to capture the military ad warfare situation in the field, giving information to troops helping them with operational planning[223]. In the same period, the Pentagon is focusing its researches on the Replicator program, an ensemble of uncrewed

[221] Flatoff Libby, Kimball Daryl G., *Russia Rejects New Nuclear Arms Talks*, Arms Control Association, March 2024. https://www.armscontrol.org/act/2024-03/news/russia-rejects-new-nuclear-arms-talks
[222] Joseph Clark, *Hicks announces the Delivery of Initial CJADC2 Capability*, US Department of Defense, February 21st 2024. https://www.defense.gov/News/News-Stories/Article/Article/3683482/hicks-announces-delivery-of-initial-cjadc2-capability/
[223] Palantir, *Palantir Expands Maven Smart System AI/ML Capabilities to Military Services*, September 9th 2024. https://investors.palantir.com/news-details/2024/Palantir-Expands-Maven-Smart-System-AIML-Capabilities-to-Military-Services/

autonomous systems sustained by AI[224]. Still, the main focus of such programs lays on the development of AI-based NC3 systems able to "overcome the attack-time compression challenge" and "accelerate wartime decision-making"[225]. Going more onto the nuclear field, the US triad, formed by nuclear weapons carried by intercontinental ballistic missiles (ICBMs), submarine-launched ballistic missiles (SLBMs), and bombs and missiles, is being revolutionized by the introduction of AI (within AUVs, ASVs, AAVs). Specifically looking at the nuclear-related ISR operations, the US has developed the Sea Hunter prototype, a project funded by the US Defense Advanced Research Projects Agency (DARPA) which has developed an autonomous trimaran, aiming at hunting down nuclear-powered ballistic missile submarines (SSBNs) armed with nuclear weapons. The US application of AI within the nuclear military is focusing specifically on three main categories, part of the technologies that enable the early warning and intelligence, surveillance and reconnaissance[226]:

a. Making early-warning and ISR systems more capable: AI Machine learning gives greater capability for the analysis of data in real-time. The idea is to develop a sort of mobile ISR platforms with the ability of analysing data on-board and identify autonomously signals, objects and specific events such as unusual troops movements. The US is developing several technologies of this type, such as the Automated Image Understanding project from the US Office of Naval Research, aiming at developing techniques to infer intentions and threats from surveillance imagery, crucial for nuclear-related ISR.

b. Searching and making sense of large sets of intelligence data: AI Machine learning is able to quickly find correlations in large sets of intelligence data. The US Project Maven makes use of this ability: known as the Algorithmic Warfare Cross-Function Team. The latter uses "machine learning to automatically analyse video surveillance footage gathered during counterinsurgency operations in Iraq, Afghanistan and elsewhere", crucial for counterterrorism purposes and nuclear-related early-warning and ISR missions, because it enables military commanders to have better situational awareness.

[224] Stokes Jacob, Kahl Colin H., Kendall-Taylor Andrea, Lokker Nicholas, *Averting AI Armageddon, U.S.-China-Russia Rivalry at the Nexus of Nuclear Weapons and Artificial Intelligence*, CNAS – Center for a New American Security, Frebruary 2025.
[225] Adam Lowther and Curtis McGiffin, "America Needs a 'Dead Hand'," War on the Rocks, Aug. 16, 2019, https://warontherocks.com/2019/08/america-needs-a-dead-hand/
[226] SIPRI, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, Volume I Euro-Atlantic Perspectives, edited by Vincent Boulanin, May 2019.

c. Making predictions: The process of data analysis uses military commanders to predict the evolution of specific events or even the course of action deriving from the employment of specific weapons, mainly nuclear ones.

This all explains the direction that the US is following in AI and machine learning implementation within the military field: to give military commanders better situational awareness with quicker and more accurate responses.

3.3.5 The case of the Russian Federation

In the last two decades, Russia has shown growing interest towards the implementation of AI within the military sector, specifically the decision-making process for nuclear defences and weapons. The current phase of the Russian triad modernization began in 1998. The Soviet Union was used to replace its land-based missiles frequently, with new systems being put into place every 10-15 years alongside modifications made every few years. Russia was not able to keep up with this rhythm due to the initial financial crisis, which decreased the number of new missiles entering the force each year and slowed down the renewal process. The result was that some systems developed during the late 1990s and early 2000s entered into force between the late 2000s and the 2020s. In December 2020, Russia's President Vladimir Putin reported that about 86% of Russia's strategic nuclear force was made up of modern weapons, a number he expected to rise to 88% in 2021[227]. The renewal of the nuclear arsenal was accompanied by AI research and development. Starting already from the 2010s, by looking at the Encyclopaedia of Strategic Missile Forces (RVSN) there can be found explanation and definitions about AI systems, that were being used for decision support system, specifically for on-board control. Some years later, in 2017, Putin affirmed that the country that will lead the development of AI will be the ruler of the world[228]. One year later, in 2018, both Putin and Shoigu, the Defence Minister at the time, held the first conference on AI, where it was stated that: "The purpose of the conference is dictated by the need to parry possible threats in the field of technological and economic security of Russia". It was then added that "As a result, optimal conditions will be

[227] Defence Ministry Board Meeting, December 24, 2019, http://en.kremlin.ru/events/president/news/62401. Retrieved from: Kristensen Hans, Korda Matt, *Russian nuclear forces, 2020*, Bulletin of the Atomic Scientists, vol.77, no. 2, 2021. https://www.tandfonline.com/doi/pdf/10.1080/00963402.2021.1885869?needAccess=true.
[228] Президент России, *Форум профнавигации «ПроеКТОриЯ»*, September 1, 2017. http://kremlin.ru/events/president/news/55492

created for the development and wide application of robotics, unmanned systems and complexes, as well as new technologies using artificial intelligence"[229].

The interest in AI application within the military field was related to the posture on the geopolitical chessboard that Russia already assumed in 2018 by demonstrating "its willingness to use force to alter the map of Europe and impose its will on its neighbours, backed by implicit and explicit nuclear first-use threats"[230]. In 2021, officers of the Fourth Central Research Institution of the Russian Ministry of Defence (Yury Matvienko and Alexander Uvarov) presented an important analysis on the potential usage of AI by the Strategic Missile Forces[231]. This analysis de facto represents how AI started to be incrementally used to automate systems, specifically to do multiple tasks in the RVSN. These tasks concerned the planning and controlling activities of nuclear weapons, alongside the daily "maintenance of the combat readiness and security" of nuclear weapons[232]. The government and analysts started to analyse the performance of such activities and implementation, based on four main requirements: "transparency, interpretability, robustness, controllability [прозрачност[ь], объяснимост[ь], робастност[ь], контролируемост[ь]]". Later on, in 2021, the Russian Ministry of Defence designated the 46th Central Research Institute as the leading research organisation in the context of AI research[233]. In 2022, one year later, the Defence Minister Shoigu approved the "Concept for the activities of the armed forces of the Russian Federation in the development and application of weapons systems using artificial intelligence technologies' (a non-public document)[234].

---

[229] Министерство обороны Российской Федерации (Минобороны России), *Министр обороны России обратился с приветственным словом к участникам конференции по искусственному интеллекту*, March 14th 2018. https://function.mil.ru/news_page/country/more.htm?id=12166652@egNews
Original version: "Цель конференции продиктована необходимостью парирования возможных угроз в области технологической и экономической безопасности России". «В результате будут созданы оптимальные условия для разработки и широкого применения робототехники, беспилотных систем и комплексов, а также новых технологий с использованием искусственного интеллекта», – отмечается в приветствии".

[230] US Department of Defense, *Nuclear Posture Review*, Office of the Secretary of Defense, February 2, 2018.

[231] Matvienko Yury, Uvarov Alexander, *Science and technology Issues of Using Artificial Intelligence Technologies and Neural Network Technologies of Data Processing in the Automated Systems of Strategic missile Forces*, Military Thought, No. 4, 2021.

[232] Shakirov Oleg, *Russian thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.

[233] Armament and Economics, *Fourty-five years in the service of the Motherland*, No.3, 2022.

[234] Ministry of Foreign Affairs of the Russian Federation, *Commentary by Russian Foreign Ministry spokesperson M.V. Zakharova on the activities of the Group of Governmental Experts of States Parties to the Convention on Certain Conventional Weapons on lethal autonomous weapons systems* [Комментарий официального представителя МИД России М.В. Захаровой о деятельности Группы правительственных экспертов государств-участников онвенции о «негуманном»оружии по смертоносным автономным системам вооружений], August 23rd 2022. https://www.mid.ru/ru/foreign_policy/news/1827203/.

The implementation of AI within the nuclear sector could become dangerous in a situation of nuclear warfare threat, as it is currently happening with the war in Ukraine. Specifically, on 14th April 2022 Medvedev, Deputy Chariman of the Security Council of Russia since 2020, affirmed that there could be "no more talk of any nuclear-free status for the Baltic – the balance must be restored", right after Finland and Sweden announced the willingness to enter into NATO[235]. He added that Swedish and Finnish people would have found nuclear-armed Russian ships "at arm's length" from their homes with "considerable naval forces" deployed in the Gulf of Finland[236]. Exactly two months before, on February 27, 2022, Putin declared that he was putting his nuclear forces into "special combat readiness" raising their alert status[237]. This position, reiterated multiple times in the last 2-3 years as a justified retaliatory measure in the event of a western military intervention, clashes completely with the Russian federation 2020-position. The latter, named "On Basic Principles of State Policy of the Russian Federation on Nuclear Deterrence" outlined that Russia would "considers nuclear weapons exclusively as a means of deterrence"[238]. It states that Russia's nuclear deterrence policy "is defensive by nature, it is aimed at maintaining the nuclear forces potential at the level sufficient for nuclear deterrence, and guarantees protection of national sovereignty and territorial integrity of the State, and deterrence of a potential adversary from aggression against the Russian Federation and/or its allies". In this highly tense context, Belarus approved a constitutional amendment to renounce to the neutrality status of the country alongside the non-nuclear status. This amendment allows Belarus to host Russian weapons (of any type) within national territory.

Russia possesses one of the world's largest nuclear arsenals, dwarfing those of every other nuclear-weapon state other than the United States. As per 2025 estimation from the American Scientists, Russia has 1710 nuclear warheads actively deployed, with an additional arsenal of

[235] Henely John, Borger Julian, *Russia warns of nuclear weapons in Baltic if Sweden and Finland join Nato*, The Guardian, April 14, 2022. https://www.theguardian.com/world/2022/apr/14/russia-says-it-will-reinforce-borders-if-sweden-and-finland-join-nato?CMP=share_btn_tw

[236] Faulconbridge Guy, *Russia warns of nuclear, hypersonic deployment if Sweden and Finland join NATO*, Reuters, April 14, 2022. https://www.reuters.com/world/europe/russia-warns-baltic-nuclear-deployment-if-nato-admits-sweden-finland-2022-04-14/

[237] Ryan Missy, Demirjian Karoun, Hudson John, Harris Shane, *With Russian nuclear forces on alert, Ukraine crisis enters more dangerous phase*, Washington Post, February 27, 2022. https://www.washingtonpost.com/national-security/2022/02/27/ukraine-russia-nuclear-alert/

[238] Ministry of Foreign Affairs of the Russian Federation, *Основы государственной политики Российской Федерации в области ядерного сдерживания*, June 2, 2020. https://archive.mid.ru/en/web/guest/foreign_policy/international_safety/disarmament/-/asset_publisher/rp0fiUBmANaH/content/id/4152094?p_p_id=101_INSTANCE_rp0fiUBmANaH&_101_INSTANCE_rp0fiUBmANaH_languageId=ru_RU

approximately 3789 warheads stockpiled[239]. In recent years, Russia has been working steadily to modernize its nuclear forces, retiring and replacing various older Soviet-era systems while implementing AI in it, specifically for decision-making capacities that it is testing in Ukraine to improve speed and efficiency in the battlefield. Russia is reportedly pursuing the C4ISR AI strategy: command, control, communications, computers, and intelligence, surveillance, and reconnaissance. Besides the C4ISR, the early warning systems is further acquiring importance due to Russia's launch-on-warning nuclear posture[240]. The aim is to improve the reliability, speed, and precision of early warning systems, while helping with threat assessment and damage prediction/forecasting, alongside the decision-making and tacking process thanks to the capability in collecting and analysing huge quantities of data. The most famous Russian AI military system is Perimeter, known also as the "Dead Hand". Perimeter has been structured considering the principle of the human-in-the-loop, but it also possesses the capability of autonomously delegating launch to "low-level commanders". More specifically, "If the sensors receive information about radioactivity that indicates the detonation of nuclear weapons in Russia, then missiles are automatically launched, which are equipped not with warheads, but with signal transmitters that issue combat control codes for the launch of the corresponding nuclear weapons of Russia"[241].

Image K: Visual Representation of the functioning of the "Dead Hand" system
Source: RBC, Эксперты объяснили принцип работы российской системы «Мертвая рука», Политика, March 29, 2018. https://www.rbc.ru/rbcfreenews/5abd11739a7947155dde1d86

[239] Kristensen Hans M., Korda Matt, Johns Eliana, Knight Mackenzie, *Estimated Global Nuclear Warhead Inventories 2025, Federation of American Scientists*, March 12th 2025. https://fas.org/initiative/status-world-nuclear-forces/

[240] Stokes Jacob, Kahl Colin H., Kendall-Taylor Andrea, Lokker Nicholas, *Averting AI Armageddon, U.S.-China-Russia Rivalry at the Nexus of Nuclear Weapons and Artificial Intelligence*, CNAS – Center for a New American Security, February 2025.

[241] RBC, *Эксперты объяснили принцип работы российской системы «Мертвая рука»*, Политика, March 29th 2018. https://www.rbc.ru/rbcfreenews/5abd11739a7947155dde1d86

The "Dead Hand" is based on a command and control system able to measure communications on military frequencies, radiation levels, air pressure, heat and short-term seismic disturbances[242]. As explained beforehand with a citation from Viktor Murakhovsky, in the case the Perimeter sensors finds out a nuclear attacks is happening on Russian territory, the system begins launching autonomously several ICBMs from the Russian arsenal. The existence of the Perimeter system proves a preference for semi-automatic or fully automatic systems, aiming at the optimisation of decision-making process. Besides the Perimeter, Russia is developing another autonomous weapon, an underwater drone, named as Oceanic Multipurpose System Status-6, commonly well-known as Poseidon. It is a nuclear-powered torpedo able to carry a thermonuclear warhead. It would be able to autonomously outrun adversarial defences thanks to its speed and technology[243].

Image L: Image of different nuclear submarine carriers
Source: Rand Corporation, How Artificial Intelligence Could Increase the Risk of Nuclear War, Research and Commentary, Components of Status 6, April 24, 2018. https://www.rand.org/pubs/articles/2018/how-artificial-intelligence-could increase-the-risk.html

Even in the context of such crucial technological avdvancements, the role of the human supervision is considered as crucial in Russia. In 2023, the head of the Main Centre for Missile Attack Warning, Sergey Suchkov, emphasised the central role of human operators as illustrated by the detection of a DPRK missile launch in March 2022: “It is commonly assumed that the [missile attack warning] system operates in a fully automatic mode, and that time [in March 2022] the technology did not fail either. The performance of the radar equipment and command posts made it possible to establish the fact of detection of the launch, but still the final decision on the validity of the launch was made by the personnel, after a comprehensive analysis of the

[242] Stilwell Blake, Military.com, *American Military History, Russia's 'Dead Hand' Is a Soviet-Built Nuclear Doomsday Device*, March 9, 2022. https://www.military.com/history/russias-dead-hand-soviet-built-nuclear-doomsday-device.html
[243] Rand Corporation, *How Artificial Intelligence Could Increase the Risk of Nuclear War*, Research and Commentary, April 24, 2018. https://www.rand.org/pubs/articles/2018/how-artificial-intelligence-could increase-the-risk.html

parameters of the launched missile. In that case, it was the professionalism of the combat crew as an integral part of the decision-making system that came to the fore"[244]. In fact, there are specific technologies that maintain both their autonomy but come to be supervised by humans. The Don-2N radar system represents the perfect example of such system[245]. As part of the Missile Attack Warning System and missile defence of Moscow, it has a "radar system digitally processes big volume of various types of radar signals enabling it to detect and track in an automatic mode more than 100 complex ballistic targets and simultaneously to aim at them several dozens of anti-missiles"[246]. However, as pointed out by the Defence journalist Dmitry Litovkin, such automated system would not launch anti-missiles by itself, but rather propose options to the operator who would have to agree and validate the recommendation, taking full responsibility of the choice[247].

The development and research in AI within the military field is not being forward solely by the singular countries, but rather by bilateral agreements. Specifically, Russia and China, in the context of the so-called "comprehensive strategic partnership", the two countries have underlined their willingness to cooperate on AI in military and civilian applications. In fact, in February 2024 officials from both the two sides met to discuss about the "closeness of the Russian and Chinese approaches." Both the Russian and the Chinese counterpart stressed the willingness to cooperate with the United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems[248]. Three months later, in May 2024, Russia and China released a joint statement after Putin's visit to Beijing, highlighting that they are ready "to expand mutually beneficial cooperation in the area of information and communications technologies," mainly on artificial intelligence. However, there are several obstacles to such cooperation: between 2010 and 2019, Chinese and Russian scientists jointly published fewer than 300

[244] Kozak Yulia, *This unique system has been on continuous duty for more than half a century* [Вот уже более полувека непрерывное дежурство несёт эта уникальная система], Red Star, February 17, 2023.
Original version: "Принято считать, что система ПРН работает в полностью автоматическом режиме, и в этот раз техника тоже не подвела. Результаты функционирования аппаратуры РЛС и командных пунктов позволили установить факт обнаружения пуска, но всё же окончательное решение о достоверности пуска принял личный состав, проведя всесторонний анализ параметров стартовавшей ракеты. В данном случае на первый план вышел профессионализм боевого расчёта как неотъемлемой части системы принятия решения".
[245] Shakirov Oleg, *Russian thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.
[246] Kozak Yulia, *They call it the eighth wonder of the world*, [Её называют восьмым чудом света], Red Star, January 24th 2022, http://redstar.ru/eyo-nazyvayut-vosmymchudom-sveta/
[247] Litovkin, Dmitry, *"Onyx" and "Granit": how smart missiles choose their targets*, Gazeta.Ru, Novemebr 18th 2021, https://www.gazeta.ru/army/2021/11/18/14218657.shtml & Litovkin, Dmitry, *Missile swarm: how artificial intelligence controls target engagement*, TASS, March 15th 2021, https://tass.ru/opinions/10891627
[248] The Ministry of Foreign Affairs of the Russian Federation, *On Russian-Chinese consultations regarding the use of AI technologies*, February 2nd 2024. https://mid.ru/en/foreign_policy/news/1929203/

English-language papers on topics related to artificial intelligence—far less than the output of joint U.S.-Chinese research collaboration[249]. Furthermore, there is a constant mistrust between Russia and China due to disagreements about intellectual property and human capital as well as the war in Ukraine. The latter has brought towards an increase in the nymber of Russian scientists leaving the country alongside damages deriving from western sanctions, which have limited the possibility of acquisition of material crucial for AI, limiting Russian possibilities, while China on the other hand tries to maintain a neutral position.

To sum up the differences in the AI introduction and application within the nuclear military field between China, Russia and the United States, the following tables by the CNA could be of great help. The CNA study focused on the following nuclear operations: targeting, employment planning, intelligence, surveillance, reconnaissance, early warning and attack characterization, active air and missile defence, leadership and decision making, communications, battle damage assessment:

Nuclear Operation	Operational Challenge	UNITED STATES	RUSSIA	PEOPLE'S REPUBLIC OF CHINA
Targeting	In General...	• Exploring, pursuing, or has implemented a wide range of AI capabilities in support of nuclear targeting.	• Likely pursuing use of AI to find and fix stationary and mobile targets for strategic conventional and novel nuclear strike capabilities.[3]	• Very likely implemented classic AI and enabled neural networks with greater autonomy to improve cruise missile and hypersonic glide vehicle targeting.[4]
	Find, Fix, Track & Target Mobile Land- and Sea-Based Missiles	• Likely exploring AI-enabled Uncrewed Aerial Systems (UAS) to enable Find, Fix, Track, Target, Engage, Assess (F2T2EA) cycle against mobile missiles in contested airspace.[5] • Almost certainly pursuing use of classic AI to automate data fusion from multiple sensors to optimize search for mobile Transporter Erector Launchers (TELs).[6] • Almost certainly pursuing use of ML to automate image recognition of mobile TELs.[7] • Likely exploring DL and ML to identify mobile missile patterns of life to narrow search area.[8] • Likely exploring use of ML to optimize and automate signal processing for acoustic strategic Anti-Submarine Warfare (ASW).[9]		• Very likely pursuing classic AI to optimize data fusion for ASW by integrating information from sensors and unmanned systems.[10] • Almost certainly exploring use of AI to find and fix mobile targets for conventional and nuclear forces.[11]
	Find and Fix Stationary Nuclear Targets			• Almost certainly using advanced algorithms for automatic target recognition to enhance precision.[12]
Other Possible AI Applications for Targeting		• ML/DL-enabled natural language processing to optimize situational awareness (SA) by analyzing patterns in chat feeds and highlighting relevant info. • ML to automate or optimize weapons/sensor cueing decisions throughout F2T2EA. • Classic AI to automate data fusion from multiple sensors to maintain target tracks. • Classic AI to automate data fusion to differentiate decoys from genuine targets. • Classic AI to monitor activity at known targets and de-nominate as they fall into disuse. • ML-enabled image recognition to automate search for new targets that fit criteria.		

Image M: AI for nuclear targeting: Differences in application and technology between the United States, Russia and China
Source: McDonnell Tim, Chesnut Mary, Ditter Tim, Fink Anya, Lewis Larry, with contributions by Westerhaug Annaleah, Artificial Intelligence in Nuclear Operations, Challenges, Opportunities, and Impacts, CNA – National Security Analysis, IRM-2023-U-035284-Final, p. 52, April 14, 2023.

[249] Konaev Margarita, Imbrie Andrew, Fedasiuk Ryan, Weinstein Emily S., Sedova Katerina, Dunham James, *Headline or Trend Line?*, Evaluating Chinese-Russian Collaboration in AI, Center for Security and Emerging Technology, August 2021. https://cset.georgetown.edu/publication/headline-or-trend-line/

Employment Planning	**In General...**	• Likely exploring or pursuing a variety of AI-enabled employment planning applications generally intended to reduce risk of mission failure.	• Exploring, pursuing, or has implemented AI-enabled approaches to improve guidance and increase prospects of mission success, as well as to support real-time analysis of pol-mil situation.[13]	• Unclear whether/how AI may support PRC employment planning beyond guidance system applications.
	Deliberate and Adaptive Strategic Nuclear Mission Planning	• Likely pursuing CAI for data fusion to enable penetrating low-observable aircraft and cruise missile route planning.[14] • Likely exploring use of ML to enable adaptive route planning for penetrating low-observable aircraft and cruise missiles.[15] • Roughly even odds exploring AI to automate and optimize weapon-target assignments.[16]	• Almost certainly implemented AI-enabled modeling in the National Defense Management Center (NDMC) systems to be able to model real-time and forward-looking analysis of the military political situation (correlation of forces, and courses of action (COAs)).[17] • Very likely using ML/DL-enabled natural language processing to automate translation from several languages, as well as data analysis in NDMC.[18]	
	NSNW Mission Planning	• Likely pursuing CAI for data fusion to enable penetrating aircraft route planning, Suppression / Destruction of Enemy Air Defenses (SEAD/DEAD).[19]	• Very likely implemented CAI in some non-strategic nuclear weapons (NSNW) guidance systems to automate guidance/optimize accuracy using scene-recognition matching.[20]	
	Model Weapons Effects on Targets		• Likely pursuing AI-enabled approaches to nuclear effects modeling.[21]	• Almost certainly pursuing AI-enabled modeling for nuclear effects simulations.[22]
	Mission Execution	• Likely pursuing CAI to automate reprogramming (retargeting) spare missiles to compensate for primary missile failure/ensure strike success.[23] • Likely exploring use of AI-enabled UAS to enable bomber/cruise missile penetration to targets.[24]	• Almost certainly pursuing AI-enabled approaches to automate/optimize guidance and navigation for novel nuclear strike systems, as well as hypersonic systems.[25] • Almost certainly implemented CAI for automation in Strategic Rocket Forces systems, including for potential quick retargeting.[26]	• Almost certainly pursuing AI to optimize accuracy of hypersonic weapons (following release).[27]
Other Possible AI Applications for Targeting		• DL or ML to understand enemy values, identify centers of gravity, and trace enemy systems and networks (broadly understood) in order to tailor deterrent threats. • DL or ML to model effects of various COAs on strategic interaction with adversary (in game theory sense) across spectrum from competition through post nuclear environment. • Classic AI to automate process of building adaptive plans from parts of existing deliberate plans. • Classic AI to enhance capabilities/capacities of smaller survivable battle staffs through automation. • ML/DL for natural language processing to expedite transformation of stated or implied leadership intentions (e.g., in decision conferences) to adaptive plans. • ML for decision support to enable target prioritization, weaponeering, weapon allocation. • DL to model climate effects of war plan execution in order to avoid inadvertently triggering nuclear winter. • DL to understand likelihood of/impact of mass fire on targets for given weapon application(s). • DL to generate local (grid square/city block)-level fallout models. • Combination of DL-enabled mass fire and/or fallout models with decision support function.		

Image N: AI for nuclear employment and planning: Differences in application and technology between the United States, Russia and China

Source: McDonnell Tim, Chesnut Mary, Ditter Tim, Fink Anya, Lewis Larry, with contributions by Westerhaug Annaleah, Artificial Intelligence in Nuclear Operations, Challenges, Opportunities, and Impacts, CNA – National Security Analysis, IRM-2023-U-035284-Final, p. 52, April 14, 2023.

Nuclear Operation	Operational Challenge	UNITED STATES	RUSSIA	PEOPLE'S REPUBLIC OF CHINA
Intelligence/ Surveillance/ Reconnaissance	**In General...**	• Unclear to what extent "advanced" AI approaches such as ML/DL are being used to support Intelligence, Surveillance, Reconnaissance (ISR). • Long-standing use of CAI for automated data fusion.	• Apparent focus on enhancing capabilities of NDMC, including via AI. • Likely implemented unknown AI in avionics systems to automate remote sensing for "automatic detection of submarines and surface, ground, and air targets".[28]	• Almost certainly pursuing the adoption of advanced algorithms and ML for remote sensing and battlefield environmental support.[29]
	Understand Adversary Capabilities, Intentions, Decision-Making	• Probably exploring ML/DL for natural language processing-based data collection to augment human analysts assessing risk of adversary aggression.[30]	• Almost certainly implemented unknown AI-enabled modeling in the NDMC systems to be able to model real-time and forward-looking analysis of the military-political situation (correlation of forces and COAs).[31]	
	Optimize Situational Awareness Within/ Across Command Centers	• Definitely implemented CAI to automate data fusion from multiple sources within/across various nuclear operations centers.[32]	• Almost certainly implemented AI-enabled systems in NDMC systems to be able to fuse data to generate a common operational picture.[33]	• Almost certainly pursuing the use of AI-enabled systems to improve common operational picture.[34]
Other Possible AI Applications for Intelligence/Surveillance/ Reconnaissance		• DL or ML to analyze adversary current and future military capabilities and strategy. • DL or ML to optimize analysis of what specific concepts/capabilities/individuals adversary leaders value to enhance deterrence through superior target identification. • DL or ML decision support via social network analysis of adversary decision-making cadre to identify cliques and cleavages that can be manipulated for deterrent effect.		

Image O: AI for nuclear intelligence, surveillance, reconnaissance: Differences in application and technology between the United States, Russia and China
Source: McDonnell Tim, Chesnut Mary, Ditter Tim, Fink Anya, Lewis Larry, with contributions by Westerhaug Annaleah, Artificial Intelligence in Nuclear Operations, Challenges, Opportunities, and Impacts, CNA – National Security Analysis, IRM-2023-U-035284-Final, p. 52, April 14, 2023.

Early Warning & Attack Characterization	**In General...**	• Unclear whether the US has proceeded beyond use of classic AI for tactical warning.	• Apparent interest in advanced AI applications for early warning and attack characterization.	• Roughly even chance of the PRC pursuing AI for early warning and attack characterization.[35]
	Strategic & Tactical Warning	• Definitely implemented classic AI to automate data fusion from multiple sensors and across command centers for tactical warning.[36]	• Almost certainly implemented AI-enabled modeling in the NDMC systems to be able to model real-time and forward-looking analysis of the military political situation for strategic warning.[37]	
	Attack Characterization		• Very likely pursuing upgrade of radar stations that are part of Russia's missile attack warning system to incorporate AI technology to optimize its ability to measure and assess incoming threats.[38]	
Other Possible AI Applications for Early Warning/Attack Characterization		• Classic AI to automate monitoring of existing sources of indication and warning. • DL or ML to identify new sources of indication and warning for assessment. • ML to automate and optimize trans- or post-attack characterization of adversary targeting strategy in support of response decision-making. • Classic AI to automate impacts of enemy attack on universe of possible response options.		

Image P: AI for nuclear early warning and attack characterization: Differences in application and technology between the United States, Russia and China
Source: McDonnell Tim, Chesnut Mary, Ditter Tim, Fink Anya, Lewis Larry, with contributions by Westerhaug Annaleah, Artificial Intelligence in Nuclear Operations, Challenges, Opportunities, and Impacts, CNA – National Security Analysis, IRM-2023-U-035284-Final, p. 53, April 14, 2023

Active Air & Missile Defense	**In General…**	• Significant interest in uses of various AI approaches to enhance active air and missile defense.	• Likely pursuing for Air and Missile Defense (AMD)[39], particularly given concerns about Western aerospace attack and the evolution of hypersonic systems, interest in the ability to "speed up the task of defining the characteristics and the type of detected objects and the direction of their flight."[40]	• Probably pursuing CAI for sensor fusion to automate and optimize AMD.
	Fuse Data from Multiple Sensors	• Definitely exploring use of ML to optimize analysis of data from multiple Integrated Air and Missile Defense (IAMD) sensors and cue human operators.[41]		
	Sensor, Weapon Allocation/Release	• Very likely exploring use of ML to optimize target discrimination.[42] • Very likely exploring use of ML + CAI to optimize and automate C2 and target engagement.[43]		
	Deny/Degrade Enemy Dissemination of Launch Orders (Left of Launch)	• Likely exploring use of ML/DL to deep fake enemy leadership voiceprints/images to degrade nuclear command, control, communications (NC3) functioning.[44] • Likely exploring use of ML to exploit cyber or other electronic vulnerabilities in adversary NC3 to degrade functioning/prevent launch.[45]		
Other Possible AI Applications for Active Missile Defense		• Classic AI or ML to adaptively optimize interceptor shot doctrine in response to different adversary salvo sizes and targeting strategies. • ML to optimize sensor and interceptor allocation decisions.		

Image Q: AI for nuclear active air and missile defence: Differences in application and technology between the United States, Russia and China
Source: McDonnell Tim, Chesnut Mary, Ditter Tim, Fink Anya, Lewis Larry, with contributions by Westerhaug Annaleah, Artificial Intelligence in Nuclear Operations, Challenges, Opportunities, and Impacts, CNA – National Security Analysis, IRM-2023-U-035284-Final, p. 53, April 14, 2023.

Leadership Decision-Making/C2	**In General…**	• Unclear whether US exploration or pursuit of AI-enabled decision support for non-nuclear operations has or will be applied to nuclear decision support.[46]	• Russia is very interested in the use of AI for decision support. However, many of its AI decision-support systems are likely relatively simple classic AI or expert systems.	• Very likely exploring AI-enabled data fusion processes to improve and accelerate combat guidance and command and control methods.[47]
	Develop & Evaluate COAs for Intended & Incidental Effects		• Almost certainly implemented AI-enabled systems in the NDMC systems to optimize decision-making.[48] • Almost certainly the Russian Strategic Rocket Forces are also exploring AI technologies for decision support systems, intelligent systems and weapons (onboard control systems), and expert systems and automation.[49] • Almost certainly developing system of systems utilizing AI for managing battlefield information in the Automated Control System of the Russian Military.[50] • Very likely implemented CAI for sensor fusion to enable nuclear launch order dissemination via Perimetr system.[51]	
Other Possible AI Applications for Leadership Decision-Making/C2		• ML/DL enabled natural language processing to present leadership with discussion-relevant COAs during decision conferences. • ML/DL enabled natural language processing to identify leadership intentions in support of COA development. • ML or DL to automate analysis of likely adversary responses to potential COA execution, in support of COA evaluation.		

Image R: AI for nuclear leadership decision-making, C2: Differences in application and technology between the United States, Russia and China
Source: McDonnell Tim, Chesnut Mary, Ditter Tim, Fink Anya, Lewis Larry, with contributions by Westerhaug Annaleah, Artificial Intelligence in Nuclear Operations, Challenges, Opportunities, and Impacts, CNA – National Security Analysis, IRM-2023-U-035284-Final, p. 53, April 14, 2023.

Communications	**In General…**	• Long-standing use of classic AI to automate communications being augmented by more sophisticated AI-enabled cyber defenses.	• Some use of unknown AI to automate launch order transmission.	• Very likely the PRC is exploring CAI for cyber network defense. • It is doable that CAI could be applied to to NC3 network monitoring and protection.[52]
	Monitor Comms System/Network Status	• Very likely exploring use of ML to automate NC3 network monitoring and protection.[53]		
	Automated / Adaptive Comms Routing	• Implemented CAI to automate transmission of Emergency Action Messages.[54]	• Almost certainly using AI to automate communications in support of secure transmission of orders for mission launch.[55]	• Almost certainly implemented CAI to automate C2 for land-based missiles, including for transmitting commands, fusing intelligence, and real-time monitoring of launches.[56]
Other Possible AI Applications for Communications		• ML enabled message routing/re-routing around damage to networks.		
Battle Damage Assessment	**In General…**	• Unclear whether/how US envisions using AI for battle damage assessment.	• Unclear whether/how Russia envisions using AI for battle damage assessment.[57]	• Unclear whether/how PRC envisions using AI for battle damage assessment.
	Fuse Data from Multiple Sensors			
	Assess Pol-Mil Implications of Damage		• There is very likely interest within Russia's Ministry of Defense institutes, but capability unclear.[58]	
Other Possible AI Applications for Battle Damage Assessment		• DL or ML to automate analysis of politico-military implications of attack.		

Image S: AI for nuclear communications, battle damage assessment: Differences in application and technology between the United States, Russia and China
Source: McDonnell Tim, Chesnut Mary, Ditter Tim, Fink Anya, Lewis Larry, with contributions by Westerhaug Annaleah, Artificial Intelligence in Nuclear Operations, Challenges, Opportunities, and Impacts, CNA – National Security Analysis, IRM-2023-U-035284-Final, p. 53, April 14, 2023.

As demonstrated by the numerous data about research projects and practical usages, the major nuclear country-players are introducing AI within the nuclear military system to a greater extent within the decision-making process and other systems. Considering the non-addressed risks and technical problems that the incorporation of AI in such nuclear context brings, the direction to be followed at the international level should be either completely banning the research and development of semiautonomous and fully autonomous AI-based nuclear weapons, or limiting their usages, rather than following mere bilateral cooperation and agreements. To this day, there is still no specific legal framework able to limit such usage, but several discussions on the topic have been held at the UN level. Specifically, the UN Group of Governmental Experts on Lethal Autonomous Weapons Systems has met several times since 2014, but unfortunately, it has been able to achieve solely a quite-disappointing result: having member states agree on eleven guiding principles instead of a more general LAWS binding treaty[250], which are also non-binding and possess a revocable nature. Such a stalemate in international talks related to this issue highlights the fact that countries do not consider the usage of AI-based nuclear weapons as a current issue. The recent Russia's invasion of Ukraine re-inflamed such discussions on the nuclear risks due to constant Russian threats of using nuclear weapons in the case of a Western direct involvement into the conflict. In a context where Russia and the US continue to invest in new-age hypersonic missiles capable of evading well-built defensive measures, alongside China investing in the increase of its nuclear arsenal, the addition of an autonomous base would enable such sophisticated nuclear technology to easily and quickly elude defensive systems. The inability to find an agreement on this topic means leaving the door open to a possible future AI-enabled nuclear arms race. This is the reason why international agreements and cooperation on the topic are crucial; on this topic, the future UN World Summit on the Information Society (WSIS)+20 High-Level Event 2025[251] will help in giving a direction to on this topic. Before that, just one year ago in 2024, it has been organized a series of UN initiatives (April 2024 Vienna Conference on Autonomous Weapons Systems) raise the issue of the importance of the ethical, legal and humanitarian attention in the context of the AI introduction to Weapons of Mass Destruction (WMD), specifically on nuclear weapons use. Moreover, a widespread accepted international framework for discussion on this topuc is the REAIM framework, the Summit on Responsible Artificial Intelligence in the Military Domain, with many other new

[250] United Nations, *Lethal Autonomous Weapons Systems (LAWS)*, Office for Disarmament Affairs. https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/

[251] Swiss Confederation, ITU, UNESCO, UNDP, UN Trade & Development, *World Summit on the Information Society (WSIS)+20 High-Level Event 2025*, July 2025. https://www.itu.int/net4/wsis/forum/2025/

agreements (mainly bilateral). Finding an international agreement on the topic is of paramount important necause nowadays, AI could alter the delicate equilibria between the nuclear powers. De facto, "an insecure nuclear-armed state would therefore be more likely to automate nuclear early-warning systems, use unmanned nuclear delivery platforms or, due to fear of rapidly losing a conventional war, adopt nuclear launch postures that are more likely to lead to accidental nuclear use or deliberate escalation"[252] .The result of such a "revolution" would be having a system where minor nuclear powers will start to use AI in a fast way to be sure of catching up with greater nuclear power, inevitably risking of incurring in dangerous incidents. The aforementioned points highlight how artificial intelligence is a double-edged sword when applied within the field of nuclear weapons and nuclear military systems, above all if risks associated to its usage are not considered or worse, ignored to avoid falling behind the nuclear arms race.

To sum up, the main revolution deriving from the AI application in the nuclear military field is inextricably linked to its introduction within the NC3 system. Even in the absence of an international agreement on the matter, it seems to be commonly shared the idea that such AI implementation should be based on:

- Transparency and explainability of the AI reasoning system, to enhance the human trust towards it;
- Human control, as explained with the various countries analysis, even the most advanced automated nuclear AI system still follows the human-in-the-loop principle, so that the recommendations of the AI gets validated by humans who take full responsibility for such decisions;
- Adherence to international humanitarian law, specifically guaranteeing the protection of the civilian population. This point is less widely shared, not considered by Russia and China.

A further step forward would be represented by an enhanced international collaboration, specifically:

- Sharing national nuclear doctrines and technicalities of the AI implementation within the NC3 system;
- Commonly agreed guidelines and rules for the development, testing phase and usage of

[252] SIPRI, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, Volume I Euro-Atlantic Perspectives, edited by Vincent Boulanin, May 2019.

AI for nuclear weapons to diminish any eventual casualty;

- Joint training programs which would help shaping commonly accepted rules and policies to be followed;
- Periodic meetings, on the UN example, to establish the presence of any risk or threat, or either just to discuss about issues;
- Delineation and establishment on an independent intergovernmental body with the ability of foreseeing the respect of the aforementioned points, mainly the national adherence to internationally agreed rules.

AI applied to the nuclear military sector is bringing and it is going to push towards a revolution on countries national research alongside international diplomacy. Abiding to the aforementioned principles could help in maintaining the stability that has been enjoyed until today with the nuclear deterrence, mainly on the importance of maintaining a human supervision of the decision-making process to avoid any inadvert escalation that could arise from the mere rational calculation of the AI. Something that should never be forgotten is that "the only winning move in nuclear war is not to play"[253].

Chapter 4: National and international law concerns

Discussing and analysing the research or development of AI military technologies does not stop solely to the mere technical considerations about the positive improvements deriving from AI integration, but also a series of crucial reflections on the moral and ethical side, besides several juridical considerations. Considering the constantly evolving state of the things, mostly for what concerns the juridical considerations, this chapter will analyse the chronological development of the various trials to establish an international legislation or ban to regulate the usage of such weapons alongside the relative and specific moral and ethical considerations from both experts and civil society organizations.

[253] Kimberly Manuel, *Russia's only winning move*, US Naval Institute, Vol. 148/2/1,428, February 2022. https://www.usni.org/magazines/proceedings/2022/february/russias-only-winning-move#:~:text=In%20the%201983%20movie%20WarGames%2C%20a%20NORAD%20supercomputer%20runs%20through,move%20is%20not%20to%20play.%E2%80%9D

4.1 Chronological examination of international legislation development about LAWS

"I send greetings to everyone attending these important consultations on a defining issue of our time — the threat posed by lethal autonomous weapons systems. Machines that have the power and discretion to take human lives without human control are politically unacceptable, morally repugnant and should be banned by international law. I reiterate my call for the conclusion of a legally binding instrument by 2026. The work being done by you and others around the world — including within the context of the Convention on Certain Conventional Weapons — is moving us in the right direction. And my report of last year on this issue demonstrates widespread agreement on some fundamental principles. First — human control over the use of force is essential. We cannot delegate life-or-death decisions to machines. And second — time is running out to take preventative action. […] From security and human rights to ethics — there are many aspects that require deeper consideration. On an issue of such concern to all of humanity, every voice must be heard. I thank you for providing yours, and helping us draw closer to a world in which lethal autonomous weapons systems have no place"[254]. This is a citation from UN Secretary General Antonio Guterres speech on May 12, 2025, pronounced during an informal consultation about lethal autonomous weapon systems. His preoccupations about the lack of any regulation whatsoever of LAWS integrated by AI are voiced by the hurry expressed through his calls on the UN states to find an agreement on a binding instrument to be reached by 2026, mostly considering that such weapons are being used more and more, which could lead to a greater difficulty in regulating them in the near future. He further stresses the need for a human control on such technologies alongside the urgency of these discussions due to the exponential increase in the usage of such technology. To be able to fully understand Guterres call to action and preoccupations it is important to go back in time and examine, via a chronological analysis, what the international community has done or has tried to do on the regulation of LAWS. The author of this thesis reckons that Roucy-Rouchegonde's analysis of the three laws of Robotics developed by Isaac Asimov in 1972[255] is probably the time frame

[254] United Nations, *Meetings Coverage and Press Release, Lethal Autonomous Weapon System 'Politically Unacceptable, Morally Repugnant and Should Be Banned'*, Secretary-General Says during Informal Consultations on Issue, SG/SM/22643, May 12, 2025. https://press.un.org/en/2025/sgsm22643.doc.htm

[255] Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, p. 11, October 2024.

better fitting the analysis of the juridical developments on the topic. Asimov's three laws of robotics are explained as follows (with a fourth one added later on):

1. First law :"A robot may not injure a human being or, through inaction, allow a human being to come to harm";
2. Second law: "A robot must obey the orders given it by human beings except where such orders would conflict with the First Law";
3. Third law: "A robot must protect its own existence as long as such protection does not conflict with the First or Second Law."
4. Forth law (also defined as zeroth law because of its importance over the other three): "A robot may not harm humanity, or, by inaction, allow humanity to come to harm".[256]

At a time when robots were more associated to a sci-fi sphere, these laws have been considered in a less serious way than today's serious considerations and examinations. In hindsight, by looking at such laws, enounced nearly fifty-three years ago, it is evident how the concerns at the basis of the usage of autonomous technology are similar: the importance of human life and its protection during the usage of such technologies, the maintainence of the human control during the entire research, development and application process of LAWS (both humain in the loop and on the loop, because both guarantee a human supervision of the software's work). The third one, re-analysed in today's context with the development of Generative AI, reveals as crucial because it poses a strict framework to both its development and usage, placing the protection of human life and the obligatoriness of human control at the basis of a system of law. Having recurring basic principles both fifty-three years ago and now, probably highlights that such assumptions should be transformed into fundamentals of international legislation about LAWS. These three main laws (de facto four) have been subsequently re-used, indirectly and probably involuntarily, from 2013 with the first discussions and debates on the topic of LAWS within the international system, at the UN level, mostly when discussing AI application within military weaponery. Specifically, first international discussions have taken place at the Human Rights Council, then framed by the Convention on Certain Conventional Weapons, followed by a more official approach with the establishment of a group of experts within the UN frame,

[256] Encyclopaedia Britannica, *Three Laws of Robotics, Concept by Asimov*, last update April 25, 2025. https://www.britannica.com/topic/Three-Laws-of-Robotics

the GGE, with the aim of analysing if newly researched and implemented weaponery respects all humanitarian international laws alongside the ones strictly linked to war.

Before 2013, there have already been several debates on the topic on a more "unformal" and "unofficial" settings, specifically organized by civil society organisation (with no involvement whatsoever of the international community at the UN level). For example, in 2009 the first civil society organizations started to voice their concerns on the dangerosity of LAWS. One of the most well-know is the International Committee for Robots Arms Control (ICRAC), alongside the ONG who took part to the Campaign to Stop Killer Robots. Roucy-Rouchegonde explains that in this newly forming context, numerous experts and professors have voiced their cincerns, such as Noel Sharkey, who defined a quite interesting set of rules, which reminds of Asimov's laws, published in the Journal of Military Ethics[257]. Two years later, in 2011, even the Nobel Peace Price Jody Williams underlined the need to ban the so-called "killer robots"[258], which some professors, such as Ronald Arkin (Georgia Institute of Technology), warned to not rush within the battlefield[259]. It is exactly with such new fears on the degree of autonomy of these weapons that the attention of international debates shifted towards the importance of maintaining human control over these weapons; as voiced by Human Rights Watch: "Maintaining meaningful human control over the use of force is an ethical imperative, a legal necessity, and a moral obligation"[260].

While reading such reports it comes to the readers' attention that two main definitions are used alternatively to refer to autonomous weapons: "killer robots" and "LAWS". Before continuing the chronological analysis, it is important to understand the difference among them. The first one is less technical, and it is generally employed by civil society organizations and NGOs to cause emotional responses as a result of the negative meaning that it entails, while the second refers to the definition Lethal Autonomous Weapon Systems, and it is more employed in technical and official debates.

---

[257] Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, p. 17, October 2024.

[258] Hughes Stuart, *Campaigners call for international ban on 'killer robots'*, BBC News, April 23, 2013. https://www.bbc.com/news/uk-22250664

[259] Ibidem.

[260] Human Rights Watch, *Stopper les robots tueurs*, Agosto 26, 2020. https://www.hrw.org/fr/report/2020/08/26/stopper-les-robots-tueurs/positions-des-pays-sur-linterdiction-des-armes#:~:text=Tous%20les%20pays%20ont%20le%20devoir%20de%20prot%C3%A9ger%20l'humanit%C3%A9,juridique%20et%20une%20obligation%20morale.

Going back to the chronological analysis, between 2014 and 2016 numrous informal debates and meetings among diplomats and experts have been organized to discuss the subject; one of the most umportant cycle of meetings about LAWS has been the one organized by the French ambassador Jean-Hugues Simon-Michel alongside the German Ambaassador Michael Biontino. The twos have enabled a general agreement between numerous states which led to the creation of the UN GGE in 2016. The latter has hosted numerous meetings on the toic between 2017 and 2021, but whenever it proposed a banning treaty for LAWS, the proposition obtained the disagreement from Russia, South Korea and the United States[261]. During this timeframe, several calls on the importance of reaching an agreement on a regulation or a ban of LAWS have been voiced by figures like Antonio Guterres, who specifically invited the international community to ban LAWS considered as "unacceptable and morally revolting"[262] moralement revoltantes" during the First forum of Peace organized in Paris in November 2018.

These calls have been left unanswered, in fact between 2022 and 2024 the UN GGE meetings have been solely for mere simple discussions, which ended up in no clear result or regulation whatsoever. During all these discussions, from 2013 to 2024, two general tendencies have emerged on the LAWS topic:

- A group of countries support the complete ban of LAWS, extending the ban to every typology and degree of autonomy of such weapons;
- A second group of countries sustains instead the establishment of a set of rules depending on the level of autonomy and characteristucs of the specific weapons under analysis, sustaining regulation mostly for fully-autonomous weapons, considered to be difficult to control, and so truly dangerous for human lives, within a military context.

The problem here is the definition per se of the level of autonomy, which varies from nation to nation depening on the scientific and juridical standards. What has emerged just until now is that nearly all countries agree on the need to define clear rules for the development, research and usage of such weapons, but their opinions about such regulations or ban depends on their national interests. Specifically China sustains the banning campaign while it continues to develop and produce autonomous weapons, France and the United Kingdom sustain the need for clear rules which however should not be too strict because this would hinder technological

[261] Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, p. 20, October 2024.
[262] United Nations News, *When nations work together, hope prevails and collective solutions can be found - UN chief tells Peace Forum, marking World War centenary in Paris*, November 11, 2018. https://news.un.org/en/story/2018/11/1025461

research, alongside the fact that in their opinion international legislation. Other countries which are researching and experimenting such weapons, like Australia, France, India, Israel, Japan, the Netherlands, Russia, South Korea, Turkey, The United States and many others consider that it is still premature to discuss about a complete ban on such technologies because they are still being researched upon[263].

Technically speaking, by looking at war and political science theorists like Clausewitz, such trials to regulate weapons sounds as logical absurdity[264] because the nature per se of the war is instable and total, which in times of crisis and need does not accept any restriction whatsoever on the weapons and other means. This principle has been proved by the usage of chemical weapons in the war in Ukraine[265], which even if prohibited come to be used in unstable and dynamic contexts like war situations. Theoretically, the regulation of war and means used in warfights are regulated by international norms on the appropriate use of force, which however are not always respected because war per se is unstable. Moreover, it has always to be kept in mind that, due to the horizontal nature of the international system, state sovereignty is always maintained which means that every nation interprets laws in the way it prefers the most, leading to a constant instability. The reason why the world does not fall in a chaotic context where everyone is against everyone else, more on the realism theories, is because the Leviathan of the situation is represented by every singular state willingness to respect norms and rules agreed internationally. Moreover, punishment systems like sanctions, help in maintaining the equilibrium within such horizontal international system thanks to a constant reciprocal check on each other compliance to international norms.

The reason why a clear set of rule of norms about LAWS hasn't still been established internationally is also to be researched on the process of law-making per se. Constructivists, specifically K. Sikkink and M. Finnemore, analysed the creation and life cycle of international norms, which led them to define three main steps:

- Emergency of norms: this step refers specifically to the proposal of law, which could derive from ONG, civil society organizations or states per se;
- Cascade of norms: crucial step where persuasion acquires a central role to diffuse the

[263] Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, p. 20, October 2024.
[264] C. von Clausewitz, *De la Guerre*, trad. par L. Murawiec, Paris, perrin (Tempus), 2014. Retrieved from: Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, p. 37-38, October 2024.
[265] Rozei Mina, *OPCW Finds More Chemical Weapons Use in Ukraine*, Arms Control Association, April 2025. https://www.armscontrol.org/act/2025-04/news/opcw-finds-more-chemical-weapons-use-ukraine

proposal of norm so that states start to homologate and act accordingly to it (giving birth to an "imitation game", which leads to indirect pressure in accepting this new norm), or either having a group of states which has an important role or place within the inyternational system, to propose the norm, so that the probability of acceptance of the norm increases;

- Internationalisation of norms: the need to have a growing number of states that accept the norm, so that the others feel pressured to do the same.

The problem with such conception lies always on the horizontal nature of the international system: every state chooses to interpret the law, alongside whether acting in accoirdance to it or not.

However, whenever new military technologies emerged in history, after specific events or catastrophes or other, states came together to regulate or ban them, as it happened with nuclear and chemical weapons. The hope is that something different would happen with LAWS, so they idea is to ban them before they come to be fully employed and used, to avoid bloodbaths. It has nearly happened once with the blinding lasers weapons. The blinding laser weapons have been used from 1960s to target the objective of guided munitions and other tasks, but the usage of such weapons revealed dangerours because, as suggested by the name, they caused blindness of people who were exposed. There have been several propositions to ban such weapon, first by Switzerland and Sweden in 1986, then by France in 1993, but countries like the United States and the United Kingdom did not agree at the beginning with such proposals in fear that their soldiers and commanders would have been accounted as responsible for cases of "accidental" blindness, while Russia sustained that the development of such weapons was a "military necessity"[266]. A further crucial issue in that context was agreeing on a common definition of blindness, specifically on agreeing when considering it to be permanent or total, which resembles enormously today's difficulty for states and experts in agreeing on a common definition and understanding about the degree of autonomy of weapons integrated with AI.

4.2 AI-integrated weapons versus international legislation

Looking at the way nuclear and chemical weapons have been respectively regulated and banned, makes it clear that states within the international system generally act in a reactive way,

[266] Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, p. 55, October 2024.

instead of a preventive manner. Even for the blinding weapons there have been some victims before they came to be banned, and even to reach such a crucial result it was needed a large consensus among states. De facto, thorough history, states have been using legislative and normative gaps in their favour, to use dangerous weapons even for mere experimental purposes, to understand limits, issues, and benefits (if there have ever been any). At the moment in which huge humanitarian issues come to the eye of the civil society and legislators, states move towards the process of searching for an international agreement. It is in fact, generally thanks to civil society protests and movements that states are pushed to act to limit the usage of such weapons causing human tragedies. This process comes to be defined as "humanitarian disarmament", and it generally progresses by campaigns of naming and shaming, respectively the identification of the issue and the responsible actor, then publicly denouncing the scandal to the civil society. The steps followed by the "humanitarian disarmament" are the following:

1. Weapon stigmatization, publicly denouncing the abhorrent effects and consequences deriving from its usage;
2. Weapon delegitimation at the juridical level, to regulate (or directly ban), researches on it, development, construction and usage;
3. Distruction of the weapons from national stocks[267].

In this context, the usage of AI-integrated weapons risks of violating the basic risks of *jus ad bellum*, the principles that regulate the use of force in war, alongside the *jus in bello*, specifically referring to the respect of life during war. To specifically understand the way a state resorts to war (strategies and objectives), alongside its reasons, it is crucial to analyse the *jus ante bellum*. The researchers Davidovic Jovana and Regan Milton affirmed that, to be sure that the war will be performed in the quickest way with the lowest possible number of civilian and military weapons (the reasons why AI-armaments gets deployed) alongside respecting international legislation on the matter, jus ante bellum should respect the following two main criteria:

1. "Not deploying any AI-enabled weapons whatsoever without having previously tested, evaluated, verified and evaluated their capabilities, risks, biases and many others;
2. Not engage in the development or research of these weapons in ways that trigger

[267] Hanson M., *Normalizing Zero Nuclear Weapons: The Humanitarian Road to the Prohibition Treaty, Contemporary Security Policy*, 39/3, pp. 464-486, 2018. Retrieved from : Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, p. 20, October 2024.

insecurity in other states, leading them to deploy AI-enabled systems without engaging in TEVV analysis"[268].

The important suggestion that has been made here is to follow the TEVV principle, so the need to test the weapons for safety reasons, aloingside evaluate eventual risks or issues already-existing or even that could arise during the usage in specific contexts and environments, evaluating the capabilities of the weapons to avoid any disproportionate usage. Generally, a disproportionate use amounts to a usage of force which goes beyond the main four axes: limitation of the aim, space, time and targets, crucial to operate militarly without incurring in civilian or military victims which go beyond the main four pillars of the international humanitarian law[269]:

1. Principle of Distinction, which imposes to the two belligerents to pay attention to not harm civilians, alongside the ban on directly targeting the population via strikes, forced displacements and similar;
2. Principle of Humanity, referring to the importance of safeguarding civilians;
3. Principle of Proportionality, which highlights the need to limit damages to civilians in situations where it is inevitable that they become victims of military strikes;
4. Principle of Military Necessity, specifically underlining the need to avoid any superfluous and unnecessary suffering to the population.

The debates around the usage of force and the ability to control it are crucial in the context of AI-enhanced weapons, specifically referring to the concept of human in (or on) the loop. The problem lies on the definition of human control per se over the autonomy of the AI system: there is no clear definition abour degrees of autonomy, and this is why some states are pushed to refuse any legislation on the topic. For example, the French delegation on the topic affirmed in 2021 that: "this notion does not match fully and adequately the requirements for characterizing LAWS, because it is too vague and it could inglobe in its perimeter even systems which does not respond to the criterion of autonomy"[270]. This is what is pushing states to refuse

[268] Davidovic Jovana, Regan Milton, *AI-enabled weapons and just preparation for war*, Babl AI, Stockdale Center for Ethical Leadership, 2023.
[269] International Committee of the Red Cross, *Fundamental Principles of IHL*, How does law protect in war?, https://casebook.icrc.org/a_to_z/glossary/fundamental-principles-ihl
[270] Research interview done on September 30, 2021 in Geneva. Retrieved from: Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, p. 79, October 2024.
Original version: "Cette notion ne permet pas de repondre de manière adequate aux exigences de caracterisation des Sala car elle est trop vague et pourrait conduir à englober dans le perimetre des discussions des systemes qui ne repondent pas au critere d'autonomie".

fully autonomous weapons, while at the same time continuing to research and develop AI-enabled weapons to avoid "falling behind" the "AI race". If states do not feel pressured to regulate the usage of such technology it means that de facto they do not perceive its dangerouseness, as they did with nuclear and chemical weapons. This all means that they will not plan to regulate them at the international level just until these weapons start to be deployed on the field, so that data on the usage, issues and similar can be collected and analysed to understand how to proceed on the topic. An interesting idea could be to integrate the TEVV principles to the Legal Weapons Review within the Article 36 of the Additional Protocol I of the Geneva Convention[271], so that legal regulations would come to be made alongside the testing and evaluation of systems capabilities and limits, still not fully known.

A final main question lies around the debate on responsibility. Specifically, the main doubts revolve around the question on whether the responsibility of the actions committed by the military system with AI lies on the commander that gives the order, on the developers of the system, on the industry, or other. De facto, it is nearly impossible to account the military personnel as responsible for any mistake or issues within the machine's activities simpky because generally soldiers are not fully aware of the technical functioning of the machines, which means that they do not know the issues related to biases, bugs and other. The unpreparadeness of soldiers in using such systems is truly evident in the cases when while in training, soldiers desactivated the automatic systems because of lacj if trust and knowledge, or even it happened the opposite situation: relying excessively on them. A glaring example on the topic is the Israeli Lavender system, which gives suggestions to military operators on where to strike and who to target. As denounced by the journal +972 on April 2023[272], the Lavender system finds out Hamas fighters within the population of Gaza intersection of the information from people previously accused of being linked to previously performed terrorist attacks, which helps the software in finding out any eventual relation or pattern scheme confronting the past experiences to the current situations. The main issue with such AI system lies at the very "source" of its functioning: the input data. The latter are generally not verified and clearly sleected, resulting in misleading and inexact outputs, than in the military field can amount to humanitarian tragedies. It has been estimated that, within the first phase of the conflict in Gaza, israeli soldiers have identified nearly 37.000 Palestinians using the Lavender system, using its

[271] Davidovic Jovana, Regan Milton, *AI-enabled weapons and just preparation for war*, Babl AI, Stockdale Center for Ethical Leadership, 2023.
[272] Yuval Abraham, *"Lavender": The AI machine directing Israel bombings in Gaza*, +972 Magazine, April 3, 2024. https://www.972mag.com/lavender-ai-israeli-army-gaza/

results not as recommendations of action but more as sure and certain indications, ignoring any consideration whatsoever about the mistakes that the system can commit, which amounts to approximately 10% of the cases analysed. This "blind confidence" onto the software's decisions, deriving from a huge digital illiteracy of soldiers and lack of understanding of its functioning (black box concept), pushes also the issue of the identification of responsibility. This is the reason why it has been proposed by the MEP Mady Delvaux to create a juridical personality specifics to robots, to help with the detection of the responsibility of action from autonomous machines[273]. The latter is a difficult proposal to be put in place for the moment, which could however open the door to a future revolution on the identification of responsibility of LAWS actions, alongiside limitations of the contexts of usage. Alongside the question of responsibility, new proposals are being submitted for the ethical principles that LAWS should abide to, such as the ALTAI guiding document, an Assessment List for Trustworthy AI[274], addressed directly to developers, instead of the users.

Conclusion

This thesis presented the development of AI within the military sector from a historical, somewhat technical, geopolitical, and juridical point of view, with the examination of a specific case study: the application of military-AI within the nuclear field. The benefits brought by the integration of AI within the military field are crystal clear, alongside the still numerous issues related to the high error rate due to recurring biases in the training material, alongside wide flaws in the training system. The aim that brings states to invest in the research of military-AI is to limit the "collateral damage" (civilian victims) alongside preserving more soldiers' lives, while "making sure" that wars last less (there are still no data about that). Numerous issues arise from an indisciplinate usage of such technologies within conflicts, amounting from the risk of targeting and killing innocent people by mistake, havings its own systems misled by adversarial exploiting of flows of AI-military system or directly cyberattack exploiting server vulnerabilities, to direct issues on the responsibility per se. The dangerouseness of such problems highly varies on the degree of autonomy left to the system alongside the human

[273] *Résolution du Parlement Européen du 16 février 2017 contenant des recommandations à la commission concernant des règles de droit civil sur la robotique* (2015/2103 [INL]). Retrieved from: Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, p. 142, October 2024.

[274] Fanni Rosanna, Giancotti Fernando, Taddeo Mariarosaria, *Guerre di macchine, Intelligenza artificiale tra etica ed efficacia*, Guerini e Associati, p. 56, May 2024.

control over its decisions and actions. The mainteinance of the human control over the performing of final actions has been put as the basis for an eventual future international agreement to regulate LAWS, which still finds difficulty in the definition per se of the degree of autonomy alongside what truly means human control and which typologies exist. Furthermore, the responsibility for the committed actions lie on the establishment of the human control that was behind the AI system before acting, however it is not clear whether it should be borne by the military commander who gave the order, the technical team who trained the AI, the technical equip which collected and analysed the material for the AI training or the company that sold it to the government. It should always be kept in mind that the latter, so the private industry, pursues its own economic benefits, which not always correspond to the national ones, exception made for specific cases such as the Chinese one, where industries are kept under the State control. In the case of the responsibility allocated to the military commanders, the issue lies, at least for the moment, on the existent huge gaps in digital literacy, which imply that commanders or even directly officers, are not always aware of the technical functioning of the system, alongside its flaws and risks. These are the main reasons why the international community is still in a situation of substantial impasse to decide whether regulating or banning such weapons. By analysing past experiences with nuclear, chemical and blinding laser weapons one thing comes to be clear: before reaching a widely agreed international legislation there is the need to develop the weapon, fully use it, become aware of the issues, besides the humanitarian and environmental tragedies it entails, to then come and sit at the discussions table at the UN. The road towards the establishment of widely agreed definitions alongside a commong agreement on a limitation, regulation or banning of LAWS is still far and convoluted. The three experts Fanni, Giancottu and Taddeo unveiled a three-point proposal that could for sure help in managing the majority of these issues and doubts, which unfold as follows:

1. To develop national strategies on AI application, with limitations on contexts and applications.
2. To establish a digital literacy program involving universities, schools, reseach centers, private industries alongside common citizens and soldiers to develop a broad digital education on risks and benefits of AI applications;
3. To institute frequent debates with European and extra-EU countries about AI ethics, issues, legislation, needs and new technology, within both the military and civilian sectors.

The hope for the future is that Guterres call made on May 2025 on the need of finding an agreement on LAWS regulation will not remain unlistened, but instead countries will come together to indentify at least some general basic principles to frame the application of LAWS to guarantee the protection and respect of human lives, alongside for the general wellbeing.

Bibliography of Books

Fanni Rosanna, Giancotti Fernando, Taddeo Mariarosaria, *Guerre di macchine, Intelligenza artificiale tra etica ed efficacia*, Guerini e Associati, May 2024.

Galton Francis, *Finger Prints – The Classic 1892 Treatise*, 1892.

Gelin Rodolphe, *Dernières nouvelles de l'Intelligence Artificielle*, Edition Flammarion, March 2022.

Giancotti Fernando, Shaharabani Yakov, *The Relevant Warriors, Leadership and Agility in Complex Environments*, The Industrial College of the Armed Forces, National Defense University, Fort McNair, Washington, D.C 20319-5062, 2008.

Heidegger Martin, *Das Ende des Denkens in der Gestalt der Philosophie*, precedently published under the name of *Zur Frage Nach der Bestimmung der Sache des Denkens,* Erker,St. Gallen, 1984, Translated edition by Adriano Fabris, Edizioni ETS Pisa, 1988.

Laure de Roucy-Rouchegonde, *La guerre à l'ère de l'intelligence artificielle, Quand les machines prennent les armes*, PUF – Presses Universitaires de France/Humensis, October 2024.

Luttwak Edward, *Le Paradoxe de la Stratégie*.

Von Clausewitz C., *De la Guerre*, trad. par L. Murawiec, Paris, perrin (Tempus), 2014.

Wiener Norbert, *The Human Use of Human Beings*, Houghton Mifflin, 1950.

Zanichelli, *Storia delle Macchine Pensanti, Una timeline dell'intelligenza artificiale*, Chapter 1.

Zuboff Shoshana, *The Age of Surveillance Capitalism, The Fight for a Human Future at the New Frontier of Power*, Public Affairs, October 4, 2019.

Bibliography of Institutional and Juridical documents

Министерство обороны Российской Федерации (Минобороны России), *Министр обороны России обратился с приветственным словом к участникам конференции по искусственному интеллекту*, March 14, 2018. https://function.mil.ru/news_page/country/more.htm?id=12166652@egNews

Президент России, *Форум профнавигации «ПроеКТОриЯ»*, События, September 1, 2017. http://kremlin.ru/events/president/news/55492

Assemblée Nationale, *Dissuation nucléaire : audition du délégué général pour l'armement*, Emmanuel Chiva, February 1, 2023. https://www.assemblee-nationale.fr/dyn/16/organes/commissions-permanentes/defense/actualites/dissuasion-

nucleaire-audition-du-delegue-general-pour-l-armement

Chinese State Council, *Notice of the State Council on Issuing the Development Plan for the New Generation of Artificial Intelligence*, Guofa 2017, No 35, 000014349/2017-00142, July 20, 2017. https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm
Chinese State Council, *China's National Defense in the New Era*, July 2019. https://www.gov.cn/zhengce/2019-07/24/content_5414325.htm

Department of Defense of the United States of America, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, Harnessing AI to Advance Our Security and Prosperity.

EU Artificial Intelligence Act, The AI Act explorer.https://artificialintelligenceact.eu/ai-act-explorer/

EUR-Lex Access to European Union Law, Document 32016R0679, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, May 4, 2016. https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=celex%3A32016R0679

European Commission, *Flagship Technical Support Projects – implementing priorities across member states*, 2023. https://reform-support.ec.europa.eu/our-projects/flagship-technical-support- projects_en#ref-2024-flagship-technical-support-projects

European Commission Speech, *Speech by President von der Leyen at the Annual EU Budget Conference 2025*, Brussels, SPEECH/25/1284, 20 May 2025. https://europa.eu/newsroom/ecpc-failover/pdf/speech-25-1284_en.pdf

Future of Life Institute, *Autonomous Weapons Open Letter: AI & Robotics Researchers*, issued at the

International Joint Conferences on Artificial Intelligence 2015, July 28, 2015. https://www.stopkillerrobots.org/wp-content/uploads/2013/03/FLI_LtrJuly2015.pdf

IAEA – International Atomic Energy Agency, *Computer Security in the Nuclear World*, IAEA Bulletin, The IAEA's flagship publication, ISSN 0020-6067, 23-01968, June 2023. Retrieved from: Hewes Mitchell, *How Artificial Intelligence Will Change Information and Computer Security in the Nuclear World*, Vol. 64-2, June 2023. https://www.iaea.org/bulletin/how-artificial-intelligence-will-change-information-and-computer-security-in-the-nuclear-world

Information Office of the State Council of the People's Republic of China, *Government Work Report delivered by Premier Li Keqiang at the Fifth Session of the Twelfth National People's Congress*.

Ministère des Armées, *Comprendre l'IA de défense*, February 10, 2025. https://www.defense.gouv.fr/actualites/comprendre-lia-defense#:~:text=Depuis%202024%2C%20le%20minist%C3%A8re%20des,comprendre%20l'IA%20de%20d%C3%A9fense

Ministry of Foreign Affairs of the Russian Federation, *Комментарий официального представителя МИД России М.В. Захаровой о деятельности Группы правительственных экспертов государств-участников онвенции о «негуманном»оружии по смертоносным автономным системам вооружений*, August 23, 2022. https://www.mid.ru/ru/foreign_policy/news/1827203/.

Ministry of Foreign Affairs of the Russian Federation, *Основы государственной политики Российской Федерации в области ядерного сдерживания*, June 2, 2020. https://archive.mid.ru/en/web/guest/foreign_policy/international_safety/disarmament/-/asset_publisher/rp0fiUBmANaH/content/id/4152094?p_p_id=101_INSTANCE_rp0fiUBmANaH&_101_INSTANCE_rp0fiUBmANaH_languageId=ru_RU

People's Republic of China Congress, March 17, 2017. https://www.scio.gov.cn/m/zxbd/tt/34849/Document/1545199/1545199.html

Pinault Desmoulins Marie Christine, *AI and literature: is it really all for thebest?*, UNESCO, 15th May 2023. https://www.unesco.org/en/articles/ai-and-literature-it-really-all-best-0

President of Russia, *Valdai Discussion Club Meeting*, November 7, 2024. http://kremlin.ru/events/president/news/75521

REAIM – REsponsible AI in the Military domain Summit 2024, 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, September 9 and 10 2024. https://www.reaim2024.kr/home/reaimeng/contents/ctntsDetail.do?encCtntNo=4d734f6c504f714f51657948783678736a6f2f6a44673d3d&encMenuId=4e63794b574d79692f7962726b56685150665a652f673d3d#:~:text=The%20REAIM%20Summit%202024%20will,in%20Seoul%2C%20in%20September%202024.

Senat, *La nécessaire modernisation de la dissuasion nucléaire*, Session Ordinaire de 2016-2017, Rapports d'informations, Rapport d'information n° 650 (2016-2017), May 23, 2017. https://www.senat.fr/rap/r16-560/r16-560.html

The Ministry of Foreign Affairs of the Russian Federation, *On Russian-Chinese consultations regarding the use of AI technologies*, February 2, 2024. https://mid.ru/en/foreign_policy/news/1929203/

The State Council of the People's Republic of China, *New Generation Artificial Intelligence Development Plan*, 20 Jul.2017, http://www.scio.gov.cn/. Retrieved from: Fei Su and Jingdong Yuan, *Chinese thinking on AI integration and interaction with nuclear*

command and control, force structure, and decision-making, European Leadership Network, November 2023.

UK Ministry of Defence, *Defence Artificial Intelligence Strategy*, June 2022. https://assets.publishing.service.gov.uk/media/62a7543ee90e070396c9f7d2/Defence_Artificial_Intelligence_Strategy.pdf

United Nations, Without Adequate Guardrails, *Artificial Intelligence Threatens Global Security in Evolution from Algorithms to Armaments, Speaker Tells First Committee*, Media Coverage and Press Releases, Meetings Coverage, General Assembly, First Committee, SEVENTY-EIGHTH SESSION, 20TH & 21ST MEETING (AM & PM), GA/DIS/3725, October 24, 2023. https://press.un.org/en/2023/gadis3725.doc.htm

United Nations General Assembly, *Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development*, Seventy-eighth session, A/78/L.49, March 11, 2024. https://documents.un.org/access.nsf/get?OpenAgent&DS=A/78/L.49&Lang=E

United Nations, *Lethal Autonomous Weapons Systems (LAWS)*, Office for Disarmament Affairs. https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/

United Nations, *Meetings Coverage and Press Release, Lethal Autonomous Weapon System 'Politically Unacceptable, Morally Repugnant and Should Be Banned'*, Secretary-General Says during Informal Consultations on Issue, SG/SM/22643, May 12, 2025. https://press.un.org/en/2025/sgsm22643.doc.htm.com/military-ai-overhauling-test/

United Nations Office for Disarmament Affairs, *Working Paper of the People's Republic of China on Lethal Autonomous Weapons Systems*, Jul. 2022, https://documents.unoda.org/wp-content/uploads/2022/07/Working-Paper-of-the-Peoples-Republic-of-China-on-Lethal-Autonomous-Weapons-Systems%EF%BC%88English%EF%BC%89.pdf.

United Nations News, *When nations work together, hope prevails and collective solutions can be found - UN chief tells Peace Forum, marking World War centenary in Paris*, November 11, 2018. https://news.un.org/en/story/2018/11/1025461

United Nations Security Council, *Letter dated 8 March 2021 from the Panel of Experts on Libya established pursuant to resolution 1973 (2011) addressed to the President of the Security Council*, S/2021/229, March 8, 2021. https://docs.un.org/en/S/2021/229

US Department of Defense, *2022 National Defense Strategy of the United States of America*, Including the 2022 Nuclear Posture Review and the 2022 Missile Defense Review, October 27th 2022.

US Department of Defense, *Autonomy in Weapon Systems*, DoD Directive 3000.09, January 25th 2023.

US Department of State, Bureau of Arms Control, Deterrence and Stability, *Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy*, February 2023.

US Department of Defense, Military and Security Developments Involving the People's Republic of China 2024, Annual Report to Congress, 2024. https://media.defense.gov/2024/Dec/18/2003615520/-1/-1/0/MILITARY-AND-SECURITY-DEVELOPMENTS-INVOLVING-THE-PEOPLES-REPUBLIC-OF-CHINA-2024.PDF

US Department of Defense, *Nuclear Posture Review*, Office of the Secretary of Defense, February 2, 2018.

US Department of Defense, *Remarks by General John E. Hyten to the Joint Artificial Intelligence Symposium and Exposition*, September 9, 2020. https://www.defense.gov/News/Transcripts/Transcript/Article/2344135/remarks-by-general-john-e-hyten-to-the-joint-artificial-intelligence-symposium/

US Office of the Secretary of Defence, *Nuclear and Missile Defence Policy, The importance of Nuclear Triad*, November 2020.

US Congress, *To prohibit the use of Federal funds to launch a nuclear weapon using an autonomous weapons system that is not subject to meaningful human control, and for other purposes*, 118th Congress, 1st Session, April 25, 2023. Retrieved from: Adams Shawn, *Why Humans, Not AI, Should Control Nuclear Weapons*, Friends Committee on National Legislation, December 16, 2024. https://www.fcnl.org/updates/2024-12/why-humans-not-ai-should-control-nuclear-weapons

U.S. Customs and Borders Protection, *Say Hello to the New face of Efficiency, Security and Safety, Introducing Biometric Facial Comparison Technology*, March 13, 2024. https://www.cbp.gov/travel/biometrics

US Office of the Secretary of Defence, *Nuclear and Missile Defence Policy, The importance of Nuclear Triad*, November 2020.

US Government Publishing Office, *United States Code*, Title 10 - ARMED FORCES, Subtitle A - General Military Law, PART IV - SERVICE, SUPPLY, AND PROCUREMENT, CHAPTER 144 - MAJOR DEFENSE ACQUISITION PROGRAMS, 2011 Edition. https://www.govinfo.gov/content/pkg/USCODE-2011-title10/html/USCODE-2011-title10-subtitleA-partIV-chap144.htm

Bibliography of Academic Articles

Allen Greg, Chan Taniel, *Artificial Intelligence and National Security*, Harvard Kennedy School, Belfer Center for Science and International Affairs, A study on behalf of Dr. Jason

Matheny, Director of the U.S. Intelligence Advanced Research Projects Activity (IARPA), July 2017. https://www.belfercenter.org/publication/artificial-intelligence-and-national-security
Antolini Diego, *A.I. – Il Futuro Appartiene alle Macchine?*, Ecce Machina, 2019.
Armament and Economics, *Fourty-five years in the service of the Motherland*, No.3, 2022.

Bezombes P., *Autonomie et respet de la règle pour les robots militaires. Considérations sémantiques et point de vue de la doctrine interarmées*, Revue défense nationale, pp. 23-35, 2018.

Centro Alti Studi per la Difesa, Istituto di Ricerca e Analisi per la Difesa, *Wargaming e Intelligenca Artificiale – situazione, livello tecnologico internazionale e nazionale, prospettive di sviluppo per la difesa*, Istituto Superiore di Stato Maggiore Interforze, 25° Corso – 1° Sezione – 1° Gruppo di Lavoro, AS-SMD-06, July 2023.

Cheeseman Peter, Stutz John, Self Matthew, Taylor William, Goebel John, Volk Kevin, Walker Helen, *Automatic Classification of Spectra From the Infrared Astronomic Satellite*, NASA Reference Publication 1217, March 1989. https://ntrs.nasa.gov/citations/19900001491

Chernavskikh Vladislav, *Nuclear Weapons and Artificial Intelligence: Technological promises and practical realities*, SIPRI – Stockholm International Peace Research Institute, SIPRI Background Paper, Project funded by the German Federal Foreign Office, September 2024.

CICDE - Centre interarmées de concepts, de doctrines et d'expérimentations, *Commandement et contrôle interarmées en environnement multimilieux-multichamps*, Concept exploratoire interarmées CEIA-3.0_C2IA-M2MC_vision prospective(2022), N°110/ARM/CICDE/NP, July 18, 2022.

CICDE - Centre interarmées de concepts, de doctrines et d'expérimentations, *Emploi opérationnel de l'intelligence artificielle*, Concept exploratoire interarmées CEIA-3.0.1_EMP-OPS-IA(2020), N° 127/ARM/CICDE/NP, December 2, 2020.

CICDE - Centre interarmées de concepts, de doctrines et d'expérimentations, *Emploi de l'intelligence artificielle et des systemes automatisées*, Concept exploratoire interarmées, CEIA-3.0.2_I.A.&SYST-AUT(2018)N 75/ARM/CICDE/NP, July 19, 2018.

Corso Francesco, Maggiore Ingegnere dell'Esercito Italiano, *Intelligenza Artificiale nella Difesa: le attività di ricerca ed innovazione tecnologica*, Artificial Intelligence Hype, AFCEA Roma, Casa dell'Aviatore, May 3, 2023.

Davidovic Jovana, Regan Milton, *AI-enabled weapons and just preparation for war*, Babl AI, Stockdale Center for Ethical Leadership, 2023.

Defence Ministry Board Meeting, December 24, 2019, http://en.kremlin.ru/events/president/news/62401. Retrieved from: Kristensen Hans, Korda Matt, *Russian nuclear forces, 2020,* Bulletin of the Atomic Scientists, vol.77, no.

2, 2021. https://www.tandfonline.com/doi/pdf/10.1080/00963402.2021.1885869?needAccess=true

Dyson Esther, Gilder George, Keyworth George, Toffler Alvin, *Cyberspace and the American Dream: A Magna Charta for the Knowledge Age*, from the journal *The Information Society: An International Journal*, vol 12 n.°3, 1996, August 22, 1994.

Ernst Dieter, *Competing in Artificial Intelligence Chips: China's Challenge amid Technology War*, Centre for International Governance and Innovation, January 1, 2020.

European Defence Agency, *Enhancing interoperability, Train Together, Deploy Together*, European Defence Matters, Issue #19, 2020.

Fei Su and Jingdong Yuan, *Chinese thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.

Future of Life Institute, Strategic Foresight Group, *Framework for Responsible Use of AI in the Nuclear Domain*, Secretary General's designated civil society organization for recommendations on the governance of AI, February 5, 2025.

Geist Edward, Lohn Andrew J., *How Might Artificial Intelligence Affect the Risk of Nuclear War?*, Security 2040, Perspective, Experts insights on a timely policy issue, Rand Corporation, 2018.

Gilli Andrea, de Dreuzy Pierre, *Российское ядерное принуждение в Украине*, NATO Review, Мнения, анализ и обсуждение вопросов безопасности, November 29, 2022. https://www.nato.int/docu/review/ru/articles/2022/11/29/rossijskoe-yadernoe-prinuyodenie-v-ukraine/index.html

Hanson M., *Normalizing Zero Nuclear Weapons: The Humanitarian Road to the Prohibition Treaty, Contemporary Security Policy*, 39/3, pp. 464-486, 2018.

Hawley John K., *Looking Back at 20 Years of MANPRINT on Patriot: Observations and Lessons*, Army Research Laboratory, ARL-SR-0158, Adelphi, MD 20783-1197, September 2007.

Hays Peter, Mineiro Sarah, *Modernizing Space-Based Nuclear Command, Control and Communications*, Atlantic Council, June 2024.

Héloïse Fayet, *French thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.

Horowitz Michael C., Kreps Sarah E., Fuhrmann Matthew; *Separating Fact from Fiction in the Debate over Drone Proliferation*, MIT press Direct, *International Security*, Vol 41, Issue 2, pp. 7-42, 2016. https://doi.org/10.1162/ISEC_a_00257

Hruby Jill, Miller M. Nina, *Assessing and Managing the Benefits and Risks of Artificial Intelligence in Nuclear-Weapon Systems*, NTI Paper, August 2021.

Hynek Nik, Solovyeva Anzhelika, *Militarizing Artificial Intelligence Theory*, Technology, and Regulation, 1st Edition, 2023.

Jeangène-Vilmer J.B., *Introduction : robotisation et transformations de la guerre*, Politique Etrangère, 78/3, pp. 80-89, 2013.

Karniadakis George Em, Kevrekidis Ioannis G, Lu Lu, Perdikaris Paris, Wang Sifan, Yang Liu, *Physics-informed machine learning*, Nature Reviews Physics 3, 2021.

Kimberly Manuel, *Russia's only winning move*, US Naval Institute, Vol. 148/2/1,428, February 2022. https://www.usni.org/magazines/proceedings/2022/february/russias-only-winning-move#:~:text=In%20the%201983%20movie%20WarGames%2C%20a%20NORAD%20supercomputer%20runs%20through,move%20is%20not%20to%20play.%E2%80%9D

Knight Heather – National Science Foundation – MIT, *Early Artificial Intelligence Projects – A Student Perspective*, Part of NSF's Recovering MIT's AI Film History Project, August 2006. https://projects.csail.mit.edu/films/aifilms/AIFilms.html

Kozak Yulia, *Вот уже более полувека непрерывное дежурство несёт эта уникальная система*, Red Star, February 17, 2023.

Lyon David, *State and Surveillance, Governing Cyberspace during a Crisis in Trust: An essay series on the economic potential – and vulnerability – of transformative technologies and cyber security*, Forthcoming, January 1, 2019, pp. 21-25.

Matvienko Yury, Uvarov Alexander, *Science and technology Issues of Using Artificial Intelligence Technologies and Neural Network Technologies of Data Processing in the Automated Systems of Strategic missile Forces*, Military Thought, No. 4, 2021.

McDonnell Tim, Chesnut Mary, Ditter Tim, Fink Anya, Lewis Larry, with contributions by Westerhaug Annaleah, *Artificial Intelligence in Nuclear Operations, Challenges, Opportunities, and Impacts*, CNA – National Security Analysis, IRM-2023-U-035284-Final, April 14, 2023.

NATO Science and Technology Organization, *Science & Technology Trends 2020-2040, Exploring the S&T Edge*, NATO Science & Technology Organization, March 2020. https://www.nato.int/nato_static_fl2014/assets/pdf/2020/4/pdf/190422-ST_Tech_Trends_Report_2020-2040.pdf

Rand Corporation, *How Artificial Intelligence Could Increase the Risk of Nuclear War*, Research and Commentary, April 24, 2018. https://www.rand.org/pubs/articles/2018/how-artificial-intelligence-could-increase-the-risk.html

Rivera Juan-Pablo, Mukobi Gabriel, Reuel Anka, Lamparth Max, Smith Chandler, Schneider Jacquelyn, *Escalation Risks from Language Models in Military and Diplomatic Decision-Making*, Cornell University, January 4, 2024. Retrieved from: arXiv:2401.03408v1 [cs.AI]

Saltini Alice, *AI and nuclear command, control and communications: P5 perspectives*, European Leadership Network, November 2023. Retrieved from: Artificial Intelligence and NC3: P5 Perspectives, Vienna Center for Disarmament and Non-Proliferation, April 3, 2024. https://vcdnp.org/ai-and-nc3/

Saltini Alice, *Applicazione dell'intelligenza artificiale nei sistemi di comando nucleare*, Istituto Affari Internazionali, January 6, 2025. https://www.affarinternazionali.it/applicazione-dellintelligenza-artificiale-nei-sistemi-di-comando-nucleare/

Saltini Alice, *UK thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.

Schalkwyk Johan, Beeferman Doug, Beaufays Françoise, Byrne Bill, Chelba Ciprian, Cohen Mike, Garret Maryam, Strope Brian, *Google Search by Voice: A Cade Study*, Google Inc, Amphiteatre Pkwy (Mountain View, California, USA).

Shakirov Oleg, *Russian thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.

SIPRI, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, Volume I Euro-Atlantic Perspectives, edited by Vincent Boulanin, May 2019.

Stokes Jacob, Kahl Colin H., Kendall-Taylor Andrea, Lokker Nicholas, *Averting AI Armageddon, U.S.-China-Russia Rivalry at the Nexus of Nuclear Weapons and Artificial Intelligence*, CNAS – Center for a New American Security, February 2025.

Tarraf Danielle C., Shelton William, Parker Edward, et al., *The Department of Defense Posture for Artificial Intelligence, Assessment and Recommandations*, Rand Corporation, 2019. https://www.rand.org/pubs/research_reports/RR4229.html

The Centre for Arms Control and Non-Proliferation, Multiple Independently-targetable Reentry Vehicle (MIRV). Retrieved from: https://armscontrolcenter.org/wp-content/uploads/2017/08/MIRV-Factsheet.pdf

The National Security Archive, The George Washington University, Jimmy Carter's Controversial Nuclear Targeting Directive PD-59 Declassified, National Security Archive Electronic Briefing Book No. 390, September 14, 2012. https://nsarchive2.gwu.edu/nukevault/ebb390/

Timothy Wright, *Is China Gliding Toward a FOB Capability?*, International Institute for Strategic Studies, Analysis, October 22, 2021. https://www.iiss.org/sv/online-analysis/online-analysis/2021/10/is-china-gliding-toward-a-fobs-capability/

Ünver Hamid Akin, *Artificial Intelligence, Authoritarianism and the Future of Political Systems*, Centre for Economics and Foreign Policy Studies, Cyber Governance and Digital Democracy 2018/9, July 30, 2018. https://www.ctga.ox.ac.uk/article/artificial-intelligence-authoritarianism-and-future-political-systems

Vienna Center for Disarmament and Non-Proliferation, European Leadership Network, *Artificial Intelligence and NC3: P5 Perspectives*, April 3rd 2024. https://vcdnp.org/ai-and-nc3/

Wiener Norbert, *Cybernetics: Or Control and Communication in the Animal and the Machine*, MIT Press, 1948 https://direct.mit.edu/books/oa-monograph/4581/Cybernetics-or-Control-and-Communication-in-the

Wu X., Shi Z., Xia L., *The Impact of AI Technology on Nuclear Weapons*, Aerodynamic Missile Journal, Vol. 6, 2020.

Young Ken, *Tesla and the First Remote-controlled Boat*, The Model Yacht, Vol 22, No 3, 2021. https://usvmyg.org/history-articles/tesla-and-the-first-remote-controlled-boat/

Bibliography of Journal Articles, Interviews, Websites

Виктор БАРАНЕЦ, *Система «Периметр» или «Мертвая рука»: для чего создавалась машина судного дня и как она отразит ядерную атаку*, Комсомольская Правда, Главная, Интересные факты и события. https://www.kp.ru/putevoditel/interesnye-fakty/sistema-perimetr-ili-mertvaya-ruka/

Газета.РУ, *Война машин: Россия создает группировки боевых роботов*, *В России создадут группировки боевых роботов к 2025 году*, November 23, 2019. https://www.gazeta.ru/army/2019/11/23/12828206.shtml

Риа Новости, *Виталий Давыдов: живых бойцов заменят терминаторы*, April 21, 2020. https://ria.ru/20200421/1570298909.html

Российская Академиа Транспорта, *Перспективы внедрения искусственного интеллекта в транспортной отрасли и логистике*, October 1, 2024. https://rosacademtrans.ru/ct2024/

РБК, *Эксперты объяснили принцип работы российской системы «Мертвая рука»*, Политика , March 29, 2018. https://www.rbc.ru/rbcfreenews/5abd11739a7947155dde1d86

영남일보, *흔들리는 혈맹...중국, 北 향해 "구동존이", 정치, April 16, 2015.* *https://www.yeongnam.com/web/view.php?key=20150416.010160800350001*

ABC News, *U.S. Predator Kills 6 Al Qaeda Suspects*, November 5, 2022. https://abcnews.go.com/WNT/story?id=130027&page=1

Afina Yasmin, Lewis Patricia, *The nuclear governance model won't work for AI*, Chatham House, June 28, 2023, updated July 3, 2023. https://www.chathamhouse.org/2023/06/nuclear-governance-model-wont-work-ai

Aerospace Global News, *AI F-16 pilot beats human in virtual dogfight*, April 2024. https://aerospaceglobalnews.com/news/ai-f-16-pilot-beats-human-in-virtual-dogfight/

Amnesty International France, *Reconnaissance Faciale : Quelles Menaces pour Nos Droits ?*, February 16, 2022. https://www.amnesty.fr/liberte-d-expression/actualites/reconnaissance-faciale-quelles- menaces-pour-nos-droits

Anduril Industries, *Anduril Partners with OpenAI to Avdance U.S. Artificial Intelligence Leadership and Protect U.S. and Allied Forces*, December 4, 2024. https://www.anduril.com/article/anduril-partners-with-openai-to-advance-u-s-artificial-intelligence-leadership-and-protect-u-s/

Arcesati Rebecca, Lofty principles, conflicting incentives: AI ethics and governance in China, Merics – Mercator Institute for China Studies, Report, June 24, 2021. https://merics.org/en/report/lofty-principles-conflicting-incentives-ai-ethics-and-governance-china

Atomic Archive, *Broken Arrows: Nuclear Weapons Accidents*, 2024. https://www.atomicarchive.com/almanac/broken-arrows/index.html#:~:text=Since%201950%2C%20there%20have%20been,been%20lost%20and%20never%20recovered.

Axe David, *An Elite Russian Drone Group Is Hunting Ukrainian Vehicles In Kursk. It'll Take More Than A Few MiG-29 Raids To Stop It*, Forbes, March 8, 2025. https://www.forbes.com/sites/davidaxe/2025/03/08/an-elite-russian-drone-group-is-hunting-ukrainian-vehicles-in-kursk-itll-take-more-than-a-few-mig-29-raids-to-blunt-the-russian-edge/

BBC News, *Donal Trump: US will build up nuclear arsenal*, October 23rd 2018. https://www.bbc.com/news/world-us-canada-45946930

Benanti Paolo, *Algor-Ethics: Developing a language for a human-centred AI*, TEDxRoma, 17, July 2018.

https://www.youtube.com/watch?v=rFzjsHNertc&list=PLh81ro1ZolekYDeI5VHzgdhF0E TjIWr1J&index=18

Benanti Paolo, *Intelligenza Artificiale: chi insegnerà l'etica alle macchine?*, TEDxTreviso, 6, November 2023. https://www.youtube.com/watch?v=6KGVNuOCwxU&list=PLh81ro1ZolekYDeI5VHzgd hF0ETjIWr1J&index=17

Biddle Sam, *OpenAI quietly deletes ban on using ChatGPT for "military and warfare"*, The Intercept, January 12, 2024, updated January 16, 2024. https://theintercept.com/2024/01/12/open-ai-military-ban-chatgpt/

Brennan Centre for Justice, *ICE Extreme Vetting Initiative: A Resource Page*, 16 November 2017, updated 24, May 2018. https://www.brennancenter.org/our-work/research-reports/ice-extreme-vetting-initiative- resource-page

Buolamwini Joy, *How I'm Fighting Bias in Algorithms*, TedX BeaconStreet talk, 29, March 2017. https://www.youtube.com/watch?v=UG_X_7g63rY&list=PLh81ro1ZolekYDeI5VHzgdhF 0ETjIWr1J&index=22

Buolamwini Joy, *The Coded Gaze: Bias in Artificial Intelligence*, Equality Summit, Bloomberg, March 29, 2019. https://www.youtube.com/watch?v=eRUEVYndh9c&list=PLh81ro1ZolekYDeI5VHzgdhF0ETjIWr1J&index=23

Buolamwini Joy, *Unmasking the bias in facial recognition algorithms*, MIT Management – Sloan School, Ideas to matter – artificial intelligence, December 13, 2023. https://mitsloan.mit.edu/ideas-made-to-matter/unmasking-bias-facial-recognition-algorithms

CBS News, *Army drone missing from Arizona found in Colorado*, February 10, 2017. https://www.cbsnews.com/news/army-drone-missing-from-arizona-found-in-colorado/

CBS News, *U.S. Official: Iran does have our drone*, December 8, 2011. https://www.cbsnews.com/news/us-official-iran-does-have-our-drone/

Cheung Sunny, *AI frontlines: China's PLA evolution: vs NATO's counterplay, Friedrich Naumann Foundation for Freedom*, Digital Transformation, November 24, 2023. https://www.freiheit.org/taiwan/ai-frontlines-chinas-pla-evolution-vs-natos-counterplay#:~:text=In%20essence%2C%20while%20China%27s%20AI,amidst%20this%20turbulent%20geopolitical%20landscape

Csernatoni Raluca, *Governing Military AI Amid a Geopolitical Minefield*, Carnegie Europe, July 17, 2024. https://carnegieendowment.org/research/2024/07/governing-military-ai-amid-a-geopolitical-minefield?lang=en&center=europe

Defence Industry Europe, *Epirus: high-power microwave system delivered to U.S. Navy for ground-based drone defence*, April 30, 2025. https://defence-industry.eu/epirus-high-power-microwave-system-delivered-to-u-s-navy-for-ground-based-drone-defence/

Defense Innovation Unit, *Replicator*, 2024. https://www.diu.mil/replicator

Depp Michael, Scharre Paul, *Artificial Intelligence and Nuclear Stability*, War on the Rocks, January 16, 2024. https://warontherocks.com/2024/01/artificial-intelligence-and-nuclear-stability/

Duboust Oceane, *AI models chose violence and escalated to nuclear strikes in simulated wargames*, EuroNews, February 22, 2024, updated February 23, 2024. https://www.euronews.com/next/2024/02/22/ai-models-chose-violence-and-escalated-to-nuclear-strikes-in-simulated-wargames

Encyclopaedia Britannica, *Defense Advanced Research Projects Agency*, United States government. https://www.britannica.com/topic/Defense-Advanced-Research-Projects-Agency

Encyclopaedia Britannica, *Three Laws of Robotics, Concept by Asimov*, last update April 25, 2025. https://www.britannica.com/topic/Three-Laws-of-Robotics

Facia, *AI-Generated Identity Fraud: Bringing Real-Time Identity Intelligence for Fraud Prevention*, March 7, 2024. https://facia.ai/blog/ai-generated-identity-fraud-bringing-real-time-identity-intelligence- for-fraud-prevention/

Faulconbridge Guy, *Russia warns of nuclear, hypersonic deployment if Sweden and Finland join NATO*, Reuters, April 14, 2022. https://www.reuters.com/world/europe/russia-warns-baltic-nuclear-deployment-if-nato-admits-sweden-finland-2022-04-14/

Federation of American Scientists, *Chinese Nuclear Weapons, 2025: Federation of American Scientists Reveals Latest Facts On Beijing's Nuclear Buildup*, March 12, 2025. https://fas.org/publication/nuclear-notebook-china-2025/

Field Hayden, OpenAI launches ChatGPT Gov for U.S. government agencies, CNBC, Tech, January 28, 2025. https://www.cnbc.com/2025/01/28/openai-launches-chatgpt-gov-for-us-government-agencies.html

Field Hayden, *OpenAI partners with U.S National Laboratories on scientific research, nuclear weapons security*, NBC News, January 30, 2025. https://www.nbcnews.com/tech/tech-news/openai-partners-us-national-laboratories-research-nuclear-weapons-secu-rcna190008

Fink Anya L., Russia's Nuclear Weapons: Doctrine, Forces, and Modernization, US Congressional Research Service, R45861 April 21, 2022. Retrived at : https://www.congress.gov/crs-product/R45861

Flatoff Libby, Kimball Daryl G., *Russia Rejects New Nuclear Arms Talks*, Arms Control Association, March 2024. https://www.armscontrol.org/act/2024-03/news/russia-rejects-new-nuclear-arms-talks

Frantzman Seth J., *How Israel Became a Leader in Drone Technology*, Middle East Forum, The Jerusalem Post, July 13, 2019. https://www.meforum.org/how-israel-became-a-leader-in-drone-technology

Gardner Frank, *Mohsen Fakhrizadeh: Iran scientist "killed by remote-controlled weapon"*, BBC, November 30, 2020. https://www.bbc.com/news/world-middle-east-55128970

Gebru Timnit, *How To Stop Artificial Intelligence From Marginalizing Communities?*, Timnit Gebru, TEDxCollegePark, October 18, 2018. https://www.youtube.com/watch?v=PWCtoVt1CJM

Gebru Timnit, Jamie Morgenstern, Briana Vecchione, Vaughan Jennifer Wortman, Hanna Wallach, Daumé III Hal, Kate Crawford, *Datasheets for Datasets*, December 2021. https://arxiv.org/pdf/1803.09010

Georgetown Law, Centre on Privacy and Technology, *America Under Watch: Face Surveillance in the United States*, May 16, 2019. https://www.law.georgetown.edu/privacy-technology-center/publications/america-under- watch-face-surveillance-in-the-united-states/ & https://www.americaunderwatch.com/

Georgetown Law, Centre on Privacy and Technology, *The Perpetual Line-Up: Unregulated Police Face Recognition in America*, October 18, 2016. https://www.law.georgetown.edu/privacy-technology-center/publications/the-perpetual-line-up/ & https://www.perpetuallineup.org/

Greene Noah, *AI Nuclear Weapons Catastrophe Can Be Avoided*, March 1, 2023, Lawfare, Armed Conflict, Cybersecurity & Tech, Surveillance & Privacy. https://www.lawfaremedia.org/article/ai-nuclear-weapons-catastrophe-can-be-avoided

Google Cloud, *Unsupervised Learning*. https://cloud.google.com/discover/what-is-unsupervised-

learning#:~:text=Unsupervised%20learning%20in%20artificial%20intelligence,any%20explicit%20guidance%20or%20instruction.

Government of Canada, *What is AI*, The Department of National Defence and Canadian Armed Forces, Artificial Intelligence Strategy, March 11, 2024. https://www.canada.ca/en/department-national-defence/corporate/reports-

publications/dnd-caf-artificial-intelligence-strategy/what-is-ai.html

Hawking Stephen, *Stephen Hawking warns Artificial Intelligence could end mankind*, BBC News Tech, December 2, 2014. https://www.bbc.com/news/technology-30290540

Helbing Dirk, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Henely John, Borger Julian, *Russia warns of nuclear weapons in Baltic if Sweden and Finland join Nato*, The Guardian, April 14, 2022. https://www.theguardian.com/world/2022/apr/14/russia-says-it-will-reinforce-borders-if-sweden-and-finland-join-nato?CMP=share_btn_tw

Hofstetter, Jeroen Van Den Hoven, Roberto V. Zicari, Andrej Zwitter, Scientific American, *Will Democracy Survive Big Data and Artificial Intelligence*, February 25, 2017. https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/

Hooker Lucy, Vallance Chris, *Concens over Google ending ban on AI weapons*, BBC, February 5, 2025. https://www.bbc.com/news/articles/cy081nqx2zjo

Human Rights Watch, *Killer Robots and the Concept of Meaningful Human Control, Memorandum to Convention on Cinventional Weapons (CCW) Delegates*, Background Briefing, April 11, 2026. https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control#:~:text=Meaningful%20human%20control%20guarantees%20that,Martens%20Clause%20is%20also%20relevant

Human Rights Watch, *Stopper les robots tueurs*, Agosto 26, 2020. https://www.hrw.org/fr/report/2020/08/26/stopper-les-robots-tueurs/positions-des-pays-sur-linterdiction-des-armes#:~:text=Tous%20les%20pays%20ont%20le%20devoir%20de%20prot%C3%A9ger%20l'humanit%C3%A9,juridique%20et%20une%20obligation%20morale

Hughes Stuart, *Campaigners call for international ban on 'killer robots'*, BBC News, April 23, 2013. https://www.bbc.com/news/uk-22250664

Hunter Christie Edward, *Artificial Intelligence at NATO: dynamic adoption, responsible use*, NATO Review, November 24, 2020. https://www.nato.int/docu/review/articles/2020/11/24/artificial-intelligence-at-nato-dynamic-adoption-responsible-use/index.html

IAEA – International Atomic Energy Agency, *Nuclear Data Services*, Nuclear Data Section. https://www-nds.iaea.org/

IBM – International Business Machine Corporation, *What are Convolutional neural networks?* https://www.ibm.com/topics/convolutional-neural-networks

IBM – International Business Machine Corporation, *What is Supervised Learning*. https://www.ibm.com/topics/supervised-learning

IDC, *IDC's Worldwide AI and Generative AI Spending – Industry Outlook*, Artificial Intelligence and DAAS, August 21, 2024. https://blogs.idc.com/2024/08/21/idcs-worldwide-ai-and-generative-ai-spending-industry-outlook/

International Committee of the Red Cross, *Fundamental Principles of IHL*, How does law protect in war?, https://casebook.icrc.org/a_to_z/glossary/fundamental-principles-ihl

Jaffri Afraz, Wiles Jackie, *What's new in Artificial Intelligence from the 2022 Gartner Hype Cycle*, Information Technology, September 15, 2022. https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle

Ji Jessica, *What does AI-Red Teaming actually mean*, CENTRE for SECURITY and EMERGING TECHNOLOGY, October 24, 2024. https://cset.georgetown.edu/article/what-does-ai-red-teaming-actually-mean/

Jordan John M., *The Czech Play That Gave Us the Word "Robot"*, The MIT Press Reader, July 29, 2019. https://thereader.mitpress.mit.edu/origin-word-robot-rur/

Joseph Clark, *Hicks announces the Delivery of Initial CJADC2 Capability*, US Department of Defense, February 21, 2024. https://www.defense.gov/News/News-Stories/Article/Article/3683482/hicks-announces-delivery-of-initial-cjadc2-capability/

Konaev Margarita, Imbrie Andrew, Fedasiuk Ryan, Weinstein Emily S., Sedova Katerina, Dunham James, *Headline or Trend Line?*, Evaluating Chinese-Russian Collaboration in AI, Center for Security and Emerging Technology, August 2021. https://cset.georgetown.edu/publication/headline-or-trend-line/

Kozak Yulia, *Её называют восьмым чудом света*, Red Star, January 24, 2022, http://redstar.ru/eyo-nazyvayut-vosmymchudom-sveta/

Kristensen Hans M., Korda Matt, Johns Eliana, Knight Mackenzie, *Estimated Global Nuclear Warhead Inventories 2025, Federation of American Scientists*, March 12, 2025. https://fas.org/initiative/status-world-nuclear-forces/

Kuipers Benjamin, Feigenbaum Edward A., Hart Peter E., Nilsson Nils J., *Shakey: From Conception to History*, p. 92, 2017. https://ai.stanford.edu/~nilsson/OnlinePubs-Nils/General%20Essays/Shakey-aimag-17.pdf

La Robotique aux cycles 3 et 4, *1914-1953 Robots de déuxieme génération*. https://atelier-canope-95.canoprof.fr/eleve/Automates%20et%20robots/res/robot.dossierHtml/co/1914_50robotsCapteurs.html

Litovkin, Dmitry, *"Onyx" and "Granit": how smart missiles choose their targets*, Gazeta.Ru, Novemebr 18, 2021, https://www.gazeta.ru/army/2021/11/18/14218657.shtml & Litovkin, Dmitry, *Missile swarm: how artificial intelligence controls target engagement*, TASS, March 15, 2021, https://tass.ru/opinions/10891627
Lowther Adam, McGiffin Curtis, "America Needs a 'Dead Hand'," War on the Rocks, Aug. 16, 2019. https://warontherocks.com/2019/08/america-needs-a-dead-hand/
Mathison Turing Alan, *Computing Machinery and Intelligence*, Journal *Mind 49*, 1950, pages 433-460. https://mitsloan.mit.edu/ideas-made-to-matter/unmasking-bias-facial-recognition-algorithms
Matteo Barbarino, *New CRP: Artificial Intelligence for Accelerating Fusion R&D (F13022)*, IAEA – International Atomic Energy Agency, New Coordinated Research Process, June 9, 2022. https://www.iaea.org/newscenter/news/new-crp-artificial-intelligence-for-accelerating-fusion-rd-f13022
Meier Olivier, *The fast and the deadly: When Artificial Intelligence meets Weapons of Mass Destruction*, European Leadership Network, June 27, 2024. https://europeanleadershipnetwork.org/commentary/the-fast-and-the-deadly-when-artificial-intelligence-meets-weapons-of-mass-destruction/

Military Embedded Systems, *Artificial Intelligence Timeline*, January 24, 2019. https://militaryembedded.com/ai/machine-learning/artificial-intelligence-timeline

Military The People, *Experts: Military Intelligentisation is Not Only About Artificial Intelligence*, December 6th 2017, https://military.people.com.cn/n1/2017/1206/c1011-29689750.html. Retrieved from: Fei Su and Jingdong Yuan, *Chinese thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making*, European Leadership Network, November 2023.

Najibi Alex, *Racial discrimination in Face Recognition Technology*, Blog, Science Policy, Special Edition: Science Policy and Social Justice, Harvard Kenneth C. Griffin, Graduate School of Arts and Sciences, 24, October 2020. https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/

National Museum of American History, *"Stanley" Robot Car*, Behring Center, Smithsonian. https://americanhistory.si.edu/collections/object/nmah_1377824

National Air and Space Museum, Smithsonian, *Pioneer RQ-2A UAV*. https://airandspace.si.edu/collection-objects/pioneer-rq-2a-uav/nasm_A20000794000

NATO, *NATO acquires AI-enabled Warfighting System*, Apr 14, 2025. https://shape.nato.int/news-releases/nato-acquires-aienabled-warfighting-system-
Nichols Michelle, *UN chief backs idea of global AI watchdog like nuclear agency*, Reuters, June 12, 2023. https://www.reuters.com/technology/un-chief-backs-idea-global-ai-watchdog-like-nuclear-agency-2023-06-12/
Noel Sharkey, *Robot wars are a reality*, The Guardian, August 18, 2007. https://www.theguardian.com/commentisfree/2007/aug/18/comment.military

Office of the Victorian Information Commissioner, *Biometrics and Privacy – Issues and Challenges*, July 2019. https://ovic.vic.gov.au/privacy/resources-for-organisations/biometrics-and-privacy-issues-and-challenges/
O'Hanlon Michael E., *How unchecked AI could trigger a nuclear war*, Brookings, February 28, 2025. https://www.brookings.edu/articles/how-unchecked-ai-could-trigger-a-nuclear-war/
OpenAI, *Governance of Superintelligence*, May 22, 2023. https://openai.com/index/governance-of-superintelligence/
Pellerin Cheryl, *Project Maven to Deploy Computer Algorithms to War Zone by Year's End*, US Department of defense, July 21, 2017. https://www.defense.gov/News/News-Stories/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/
Palantir, *Palantir Expands Maven Smart System AI/ML Capabilities to Military Services*, September 9, 2024. https://investors.palantir.com/news-details/2024/Palantir-Expands-Maven-Smart-System-AIML-Capabilities-to-Military-Services/
Paleja Ameya, *Ground Warden: New AI-based system enables forces to see hidden targets*, Interesting Engineer, Military, June 18, 2024. https://interestingengineering.com/military/ai-ground-warden-finds-hidden-targets
Pistilli Giada, *AI Generativa: Uno sviluppo etico è possibile?*, TEDxVerona, September 11, 2023 https://www.youtube.com/watch?v=NreFQFKahxw&t=208s
Puwal Steffan, *Should artificial intelligence be banned from nuclear weapons systems?*, NATO Review, Opinion, Analysis and Debate on Security Issues, April 12, 2024. https://www.nato.int/docu/review/articles/2024/04/12/should-artificial-intelligence-be-banned-from-nuclear-weapons-systems/index.html
Rai News, *Timnit Gebru, la ricercatrice su etica dell'AI che non si fida di Google, Costretta a licenziarsi da Mountain view un anno fa, l'ingegnere informatico ha creato un isituto per lo studio dei pregiudizi nell'uso dell'Intelligenza Artificiale*, Sezione *Scienza e Tecnologia, Tech*, March 26, 2022. https://www.rainews.it/articoli/2022/03/timnit-gebru-la-ricercatrice-su-etica-dellai-che-non-si-fida-di-google-e179d072-dcc5-4383-ba36-077ccd9ac40b.html
Rautenbach Peter, *On Integrating Artificial Intelligence with Nuclear Control*, Arms Control, September 2022. https://www.armscontrol.org/act/2022-09/features/integrating-artificial-intelligence-nuclear-control
RBC, *Эксперты объяснили принцип работы российской системы «Мертвая рука»*, Политика, March 29, 2018. https://www.rbc.ru/rbcfreenews/5abd11739a7947155dde1d86
Réal Pierre, *L'intelligence artificielle et ses applications: un défi stratégique pour la France*, Les Cahiers de la Revue Défense Nationale, May 2019. https://www.defnat.com/e-RDN/vue-article-cahier.php?carticle=148&cidcahier=1188

Rozei Mina, *OPCW Finds More Chemical Weapons Use in Ukraine*, Arms Control Association, April 2025. https://www.armscontrol.org/act/2025-04/news/opcw-finds-more-chemical-weapons-use-ukraine

Ryan Missy, Demirjian Karoun, Hudson John, Harris Shane, *With Russian nuclear forces on alert, Ukraine crisis enters more dangerous phase*, Washington Post, February 27, 2022. https://www.washingtonpost.com/national-security/2022/02/27/ukraine-russia-nuclear-alert/

Rogoway Tyler, *Meet Israel's "Suicide Squad" of Self-Sacrificing Drones*, The War Zone, September 29, 2021. https://www.twz.com/4760/meet-israels-suicide-squad-of-self-sacrificing-drones#:~:text=In%201982%2C%20when%20open%20hostilities,hours%2C%20it%20did%20just%20that.

Root Brian, *US Immigration Officials Pull Plug on High-Tech'Extreme Vetting'*, Human Rights Watch, 18, May 2018. https://www.hrw.org/news/2018/05/18/us-immigration-officials-pull-plug-high-tech-extreme-vetting

Salles Arleen, Evers Kathinka, Farisco Michele, *Anthropomorphism in AI*, Journal *PubMed* of the National Library of Medicine – National Centre for Biotechnology Information, 2020. https://pubmed.ncbi.nlm.nih.gov/32228388/

Seibt Sébastian, *From the A bomb to the AI bomb, nuclear weapons' problematic evolution*, France24, Technology, May 10, 2019. https://www.france24.com/en/20190510-nuclear-weapons-artificial-intelligence-ai-missiles-bombs-technology-military

Simpson Kurtis H., Paquette Samuel, Racicot Raphael, Villanove Samuel, *Militarizing AI: How to Catch the Digital Dragon?, AI holds the potential to fundamentally redefine modern warfare*, Center for International Governance and Innovation, February 26, 2025. https://www.cigionline.org/articles/militarizing-ai-how-to-catch-the-digital-dragon/

Stanford University, *Stanford's Robot Legacy*, Stanford Report, Science & Engineering, January 16, 2019 https://news.stanford.edu/stories/2019/01/stanfords-robotics-legacy

Stilwell Blake, Military.com, *American Military History, Russia's 'Dead Hand' Is a Soviet-Built Nuclear Doomsday Device*, March 9, 2022. https://www.military.com/history/russias-dead-hand-soviet-built-nuclear-doomsday-device.html

Stokes Jacob, Military Artificial Intelligence, *the People's Liberation Army, and U.S.-China Strategic Competition*, CNAS, February 1, 2024. https://www.cnas.org/publications/congressional-testimony/military-artificial-intelligence-the-peoples-liberation-army-and-u-s-china-strategic-competition

Sutton H., *New North Korean Atomic Submarine Escalates Nuclear Arms Race*, NavalNews, April 2, 2025. https://www.navalnews.com/naval-news/2025/04/new-north-korean-atomic-submarine-escalates-nuclear-arms-race/

Swiss Confederation, ITU, UNESCO, UNDP, UN Trade & Development, *World Summit on the Information Society (WSIS)+20 High-Level Event 2025*, July 2025. https://www.itu.int/net4/wsis/forum/2025/

Tass, *В Минобороны создали центр перспективных беспилотных технологий "Рубикон"*, October 11, 2024. https://tass.ru/armiya-i-opk/22099195

Thales, *Biometrics: definition, use cases, latest news, Digital identity and security*, May 20, 2023. https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/inspired/biometrics
The Washington Post, *Google drops pledge not to use AI for weapons or surveillance*, February 4, 2025 https://www.washingtonpost.com/technology/2025/02/04/google-ai-policies-weapons-harm/
T-Invariant, "*Sovereign" Means Military: How Russia Militarized AI, Drone and Cryptography Industries*, March 20, 2025. https://t-invariant.org/2025/03/sovereign-means-military-how-russia-militarized-ai-drone-and-cryptography-industries/
United States Air Force Website, B-21 Raider. https://www.af.mil/About-Us/Fact-Sheets/Display/Article/2682973/b-21-raider/
United States Air Force Website, *MQ-1B Predator*. https://www.af.mil/About-Us/Fact-Sheets/Display/Article/104469/mq-1b-predator/
Vlasov Artem, Barbarino Matteo, *Seven Ways AI Will Change Nuclear Science and Technology*, IAEA – International Atomic Energy Agency, September 22, 2022. https://www.iaea.org/newscenter/news/seven-ways-ai-will-change-nuclear-science-and-technology
Waterman Shawn, *Military AI Will Mean Overhauling Test as Well as Tactics: DOD's first AI Chief*, Air Space and Forces Magazine, April 9, 2025. https://www.airandspaceforces
Wong Leung Jenniferm Robin Stephan, Cave Danielle, *ASPI's two-decade Critical Technology Tracker: The rewards of long-term research investment*, Australian Strategic Policy Institute, August 28, 2024. https://www.aspi.org.au/report/aspis-two-decade-critical-technology-tracker-the-rewards-of-long-term-research-investment/
Yao Deborah, *25 Years Ago Today: How Deep Blue vs. Kasparov Changed AI Forever*, Interview with Deep Blue co-creator MurrayCampbell, May 11, 2022. https://aibusiness.com/ml/25-years-ago-today-how-deep-blue-vs-kasparov-changed-ai-forever
Yuval Abraham, *"Lavender": The AI machine directing Israel bombings in Gaza*, +972 Magazine, April 3, 2024. https://www.972mag.com/lavender-ai-israeli-army-gaza/

Conferences attended by the author of this thesis

Brando Menifei during the panel discussion in celebration of the Europe Day, *European Values: Challenges from AI to elections*, organized by the EU Delegation to the Holy See, the Sovereign Order of Malta and the UN Organizations in Rome to the Republic of San Marino, and by the Embassy of the Kingdom of Belgium to the Holy See, May 16, 2024.

Conference *Vauban Sessions*, organized by Fordward Global at the CRR-FR Lille, May 14 and 15, 2025. Event under Chatham House Rules.

Conference *Paris Defence and Strategy Forum*, Europe at a crossroad, organized by the

Académie de Défense de l'Ecole Militaire, March 11, 12 and 13, 2025. Event under Chatham House Rules.

Professor Blerina Sinamieri during the conference *Intelligenza Artificiale e Intelligenza Umana a Confronto*, organized by the Embassy of the Republic of Albania to the Holy See and the Sovereign Order of Malta, May 21, 2024.

Professor Paolo Benanti during the conference *Intelligenza Artificiale e Intelligenza Umana a Confronto*, organized by the Embassy of the Republic of Albania to the Holy See and the Sovereign Order of Malta, May 21, 2024.