American University in Cairo

AUC Knowledge Fountain

Theses and Dissertations

6-1-2019

The weaponization of artificial intelligence (AI) and its implications on the security dilemma between states: could it create a situation similar to "mutually assured destruction" (MAD)

Gilan Osama Dahab

The American University in Cairo

School of Global Affairs and Public Policy

THE WEAPONIIZATION OF ARTIFICIAL INTELLIGENCE (AI) AND ITS IMPLICATIONS ON THE SECURITY DILEMMA BETWEEN STATES: COULD IT CREATE A SITUATION SIMILAR TO "MUTUALLY ASSURED DESTRUCTION" (MAD)?

A Masters Project Submitted to the

Public Policy and Administration Department

in partial fulfillment of the requirements for Global Affairs

By

GILAN OSAMA KAMAL MAHMOUD MOHAMED DAHAB

Fall 2018

The American University in Cairo

School of Global Affairs and Public Policy

THE WEAPONIIZATION OF ARTIFICIAL INTELLIGENCE (AI) AND ITS IMPLICATIONS ON THE SECURITY DILEMMA BETWEEN STATES: COULD IT CREATE A SITUATION SIMILAR TO “MUTUALLY ASSURED DESTRUCTION” (MAD)?

A Master’s Project by

Gilan Osama Kamal Mahmoud Mohamed Dahab

to the Department of Public Policy and Administration

Fall 2018

in partial fulfillment of the requirements for the
Master of Global Affairs
has been evaluated by

Evaluated by: (*Prof. Allison Beth Hodgkins*)

For consideration by: (*Dr. Waleed Rashad Zaky Omar; The National Center for Social and Criminological Research)*

The American University in Cairo
School of Global Affairs and Public Policy
Department of Public Policy and Administration
Global Affairs Master's Project

THE WEAPONIZATION OF ARTIFICIAL INTELLIGENCE (AI) AND ITS IMPLICATIONS ON THE SECURITY DILEMMA BETWEEN STATES: COULD IT CREATE A SITUATION SIMILAR TO "MUTUALLY ASSURED DESTRUCTION" (MAD)?

Gilan Osama Kamal Mahmoud Mohamed Dahab

Supervised by: Prof. Allison Beth Hodgkins

ABSTRACT

There is no a consensus in the IR literature on the possible implications of AI for cyber or nuclear capabilities, and whether AI would exacerbate, or potentially mitigate, the security dilemma between actors with varying capabilities. This paper explores these questions, using experts' interviews and secondary data. It has tackled the issue under study by using the most-similar method in which most of the variables are similar.

The paper argues the weaponization of AI exacerbates the security dilemma between states since it increases uncertainty. What is actually problematic about the military AI applications, as opposed to other military capabilities, is the declining role of humans. AI could be productive and counterproductive when it comes to policy making, implying the necessity of keeping humans over-the-loop. Neutralization makes AI deterrence reasonable for avoiding destructive, disruptive and manipulative outcomes. Like nuclear capabilities, establishing an AI-MAD structure, regulating the uses of AI and establishing a governing regime for AI arms race are the best possible policies.

Keywords: Artificial Intelligence, Deterrence, Mutually Assured Destruction, Arms Control

TABLE OF CONTENTS

Introduction:

The weaponization of Artificial Intelligence (AI) has the potential to change the nature and the character of warfare. As seen in cyber warfare, AI is expected to be the future battlefield of warfare since it can be used as an enabler of a weapon or it could be weaponized as it was the case with nuclear capabilities. To put it simple, it would allow states to employ both kinetic and non-kinetic capabilities, separately or altogether.

The chief objective of this paper is to investigate the worrisome phenomenon of weaponized Artificial Intelligence and how it exacerbates the security dilemma between states in both symmetric and asymmetric settings. This in turn accelerates AI arms race, which eventually invokes crisis instability and arms race instability. In respect, this paper is to explore the potential implications of AI on national polices, interstate relations, and the foundations of the international regime governing relations between states. This piece is to suggest the formation of an international regime for governing AI.

The weaponization of AI might put the world order and the foundations of international peace and security at a shaky ground. Thus, state actors, non-state actors, such as the IAEA and the UN Office for Disarmament Affairs, international lawyers and private companies should consider either of the two policy options: regulating AI and ensuring that its uses are in conformity with the parameters of the current global regime, or establishing a novel global system, in which AI replaces states, for maintaining international peace and security. From a utopian angle, the establishment of a new global system looks awesome at first glance. But, the first policy option is the most doable one given that it just requires the establishment of a regime similar to the one which was established for regulating nuclear arms race.

- *Problem Statement:*

There is an excessive use of Artificial Intelligence-enabled applications in the military realm coupled with the unprecedented advancement in killer robots and the massive production of drones. This mirrors the hasty inclination to possess the most advanced AI military applications, so as to intensify AI race. The IR literature has narrowed down the focus to the possible implications of AI on nuclear capabilities without investigating how AI will alter the nature of a weapon technology. The IR literature has failed to see the other side of coin and did not investigate how AI could be employed for confidence-building and for enhancing security. It is illogical to assume the uselessness of AI; a developed version of cyber technology, in cyber defense.

Besides to investigating the potentials of AI from a technical point of view, there is a need to explore the implications of AI on interstate relations and how the AI race could be addressed. Also, ushering for international legal instruments and the call for coherent policies at the national and international levels are essential for regulating the uses of AI. The international community should accept the fact that AI race is irreversible, but regulating it is the best possible choice.

The IR literature has largely overlooked the possible implications of coupling other weapons, including nuclear and cyber, with AI capabilities and has lamentably disregarded to investigate how that might increase their destructive potentials, and uncertainty. This paper explores ***whether the weaponization of AI could create a MAD-like structure since it exacerbates the security dilemma between states?*** The paramount objective of this paper is to investigate the efficacy of AI, which aggravates the security dilemma between states, as a deterrent tool. In the context of offense-defense theory, it investigates how AI, either as an enabler of a weapon or a weapon, would dictate future wars, and it also examines would it makes offense or defense dominant. Along with exploring the implications of AI on other weapon systems, this paper raises a question: ***would signaling an AI second-strike capability or establishing an AI equivalent of Mutually Assured Destruction reduce the probability of a cyber, conventional or even a nuclear war?*** To answer this, this piece introduces ***two hypotheses: (i) Nuclear MAD could create an AI-MAD even if the first-strike capability is advantageous in the cyber realm; (ii) AI capabilities could strengthen cyber defense, thereby AI-MAD could be feasible***.

- *Argument:*

Since the civilian AI applications have been weaponized with their potentials to revolutionize and change the nature and the character of future wars, there is no doubt that AI with its dual-use nature does exacerbate the security dilemma between states in both symmetric and asymmetric settings. AI strengthens cyber deterrence by preemption and by demonstrating the ability to retaliate in the cyber realm. AI, like nuclear capabilities, has its advantages and disadvantages since it can enhance cyber defense and nuclear safety, on the one hand, and can be employed wrongfully and in a manipulative way, on the other. This further indicates that AI has the second-strike capability amid the growing uncertainty over its potentials, and the intentions of rivalry states. AI deterrence can exist in tandem with nuclear deterrence due to its destructive, disruptive and manipulative potentials. Based on that, an AI MAD like structure is the optimal policy option for mitigating the security dilemma between states and maintaining international peace and security. This requires finding out ways for diluting the pace of AI arms race, which has been deviated from the commercial sphere to the military one. Accordingly, the international community should make all efforts to regulate the uses of AI and control AI proliferation since we cannot reverse it. This indicates that regulating the uses of AI through the establishment of legal instruments will not be sufficient, notably with the involvement of private companies. Regulating AI requires both a political will and a consensus, otherwise the outcomes would be disastrous. A comprehensive framework, incorporating the legal, political, ethical, economic and security aspects, is highly recommended for maintaining international peace and security. More important, the international community should not allow, under any circumstances, militaries to be governed by machines since the psychological factor is crucial in military's decision-making.

Client Description:

Dr Waleed Rashad:
Brief Bio:

Dr. Walled Rashad is an assistant professor of sociology at the National Center for Social and Criminological Research. In parallel with his career of over 15 years, he has contributed to academic research in the area of cyber security. He publishes his studies and findings at many periodical journals, including the Democracy Journal (Al-Democrateya) and The Contemporary Thinking Journal (Al-Fakr Al-Mo'aser). Some of his contributions were published by Egypt Police Research Center. His academic contributions have included: two studies on "Cyber/Internet Cafés: under the titles of "Internet Cafes as a Public Sphere" and "Virtual Actors' Interactions"; two book chapters under the titles of "*Social Mobilization in the Cyber Domain*" and "*Social Strata and the Transformations of the Virtual Community: National and*

International Debate"; a study on the interactions that take place in the cyber domain. He also co-authored a study under the title of "*The Internet as An Alternative Media Platform*". He also wrote a book review entitled as "*Cyber Culture*" and a journal article entitled as "*The Internet of Things (IoTs): A Sociological Approach*". Moreover, he participated in many symposiums and conferences. He also has earned a Ph.D. of Arts from Ain Shams University and three master's degrees of sociology. One of his master thesis tackled the multiple actors in the cyber sphere.

His recognizable experience in the field of research will certainly provide a thorough insight into the analogy between AI and cyber capabilities, in addition to nuclear ones. Based on his academic experience in the field, he will provide a thorough analysis and a convincible evaluation of the research findings. He could also validate or nullify the findings of this research when it comes to practicality. Also, he could assess the applicability of the policy recommendations suggested in this research or even add more recommendations. Adding to this, he might direct the academic community to the negative implications of AI weaponization coupled with the weaponization of the Internet of Things (IoTs) or even usher for, at the sidelines of symposiums and conferences, any of the policy recommendations mentioned in this paper.

Background:

The use of Artificial Intelligence in militaries is not a new phenomenon, but the inclination to upgrade AI military applications and semi-autonomous drones to those that can operate autonomously and without humans' intervention is the eye-catching phenomenon that raises concerns among scholars and experts. For the time being, AI is bolted into arrays of weapon systems, such as aircrafts, submarines, and is also installed in command and control systems (C2), and critical logistical infrastructure, (Meserole, 2018). Today's AI is somehow limited in its capacities but with the possible progress in the dreamy one-shot learning and quantum computing, the security dilemma will become irreducible.

Over the past years, militaries have proliferated and have produced armed drones that can serve at both the tactical and operational levels, in an effort to reduce human causalities and gain a military advantage. Donald Rumasfeld; the former Secretary of Defense, had introduced the concept of "*the mechanization of war*" by which the US army is made up of half robots and half humans, (Soliaman, 2019). The US Department of Defense has made AI, besides to human soldiers and manned personnel, as an integral element of its Third Offset Strategy. In an effort to slash costs, it was reported that Japan's AI-enabled rockets is underway, (Nausca, 2011). Due to the effectiveness of semi-autonomous weapons systems, both "*killer robots*" and unmanned air vehicles (UAVs) have become very appealing to many actors.

AI was supposed to be used for surveillance, reconnaissance and military tactical operations, but its uses have broadened when states employed AI applications in their information and cyber warfare. In 2010, Israel and US launched Stuxnet against Iran's nuclear facilities in lieu of a conventional military attack, marking a paradigm shift in warfare. Since then, the cyber domain, coupled with the growing reliance on UAVs and drones, has become the new battleground. The most recent example is the Russian information warfare by which it took measures to overtly or covertly influence, (Polyakova, 2018) the American public opinion during the latest US presidential elections. There are many other examples of an AI-enabled cyber warfare, of which the deviation of a civilian flight from its destination, (Rashad, 2019). All of these examples signify the protraction in the uses of AI applications and cyber capabilities. Such an observation is critically important since a number of experts expressed their concerns over the possibility of misusing AI capabilities for subverting those of an adversary, (Giest et al, 2018).

As a result, the term of "*Algorithmic Warfare*" has dominated the IR literature since it will change the battlefield we know with the primacy of intelligence warfare and will dictate future wars. 30 scientists, technologists and military experts pointed out that there will be three new elements that will define and will shape the future battlefield by 2050. These are cyber capabilities and technologies; a complex, highly disputed information sphere, as well as a human force with advanced physical and cognitive skills, (Kott et al, 2015). These elements have stimulated an AI arms race, with China attempting to surpass the US and to become a key player in the AI plane, (China May Match, 2017). More than 30 countries possess or are developing drones for military uses, such as intercepting high-speed rockets, (Scharre, 2017).

Ongoing speculations about the weaponization of AI have generated contradictory opinions over the potentials of AI on nuclear, cyber and even conventional weapons.

Each camp holds differing views over the promising and negative potentials of AI over nuclear and cyber capabilities. For nuclear capabilities, Theorists argued that exploiting new technologies makes nuclear stalemate reversible and reduces nuclear survivability which is actually based on concealment, hardening and redundancy, (Lieber et al, 2017). Meanwhile, it can bolster nuclear counterforce. Concerning cyber capabilities, the automation of data analysis and targets prioritization trigger data poisoning, (Brundage et al, 2018). However, AI-enabled detecting software will embolden cyber defense with their abilities to detect code vulnerabilities.

Scholars have also observed that a mass of AI-enabled applications could threaten both combatants and non-combatants, and it would make "algorithmic warfare", (Layton, 2018) in contravention to international law. The inherent hazards of unregulated "algorithmic warfare", coupled with the absence of humans, could entail unintended engagement, causing fratricide and civilian causalities or triggering inadvertent escalation, (Layton, 2018). From a security perspective, fully autonomous weapons and unmanned vehicles seem impractical for matters of life and death, unless humans are over-of-the loop, since they cannot operate in or effectively adapt to highly changeable and complex environments, (Layton, 2018). Also, in the context of cyber warfare, unregulated algorithms, coupled with the weaponization of the Internet of Things (IoTs), could induce cost on adversaries by attacking critical infrastructure and networks, (Liff, 2012) thereby triggering causalities among non-combatants and civilians. The literature has suggested various policy options for regulating and mitigating the uses of AI, and for avoiding future intelligence and algorithmic warfare. Some of these policy options are useful, such as the synergy between human cognition and machines intelligent computation. Arguably, the human-machine teaming would complement the missing piece of the puzzle through advanced, speedy data analysis and human cognition. Adding to this suggestion, there are calls for drafting a Digital Geneva Convention, (Why We Urgently Need, 2017) and preventive arms control. Further, the weaponization of AI has raised concerns among states' leaders, CEOs of private companies, and over 60 NGOs, (Scharre, 2017) which called for the banning of AI, (Autonomous Weapons: An Open Letter, 2015).

In parallel, several steps and endeavors were taken, including the announcement of an *International Panel on Artificial Intelligence* by Canada and France at the sideline of a G7 conference. The aims have been providing support, embracing the responsible adoption of a human-oriented AI and facilitating international cooperation, (Shead, 2018). Also, the Berkman Klein Center launched the “Ethics and Governance of Artificial Intelligence Initiative” for bolstering the proper use of AI, (Ethics and Governance of AI). All of these mesmerizing endeavors imply how pressuring the AI weaponization and denote the urgency of taking a global collective action.

Therefore, experts and policy-makers should be far-sighted while assessing AI as a weapon/an enabler by investigating how it would spark arms race, and should hypothesize its implications when it is either nascent or advanced and when humans are over- and out-of-the loop. They should also consider the malignant and the harmless uses of AI while investigating its implications on the security dilemma, which usually exacerbates because of the AI’s effectiveness, scalability, rapid diffusion, speedy potentials, and its dual-use nature, (Brundage, 2018). They should also consider how to ameliorate and reduce uncertainty which arises because of the manipulative and disruptive potentials of AI and because of the emergence of new threats and vulnerabilities, such as impersonation, (Brundage, 2018) redirection of flights, amid the absence of punitive and attributive measures.

Literature Review:

Technological advances are a double-edged sword for a state’s national security. On the one hand, new technology can enhance a state’s defensive capacity, and can enhance a state’s ability to deter potential hostile acts by adversaries. On the other hand, technological advances can also exacerbate the security dilemma. Likewise, technological advances spur potentially destabilizing arms races between rival powers when they are neither certain over the sort of capabilities developed nor their implications on the balance of power. When a rival state increases its defensive capability, the security dilemma exacerbates in this respect. With weapons of mass destruction (WMDs), states may launch pre-emptive or even preventive strikes, when they suspect an adversary of developing new and potentially dangerous capabilities as it was the case when Israel attacked Iraq in 1981. Paradoxically, such strikes or rhetoric threatening such acts are often seen as justifications for acquiring more advanced and destructive weapons.

Technological advances lead to the escalation of the security dilemma and the emergence of new military revolutions. Based on the history of military development, ten ***military revolutions*** took place as a result of technological advances, (Krepinevich, 1994). The weapons of mass destruction and nuclear weapons are perfect examples of technological advances that exacerbated the security dilemma between rival states during the Cold War era. Based on Krepinevich's argument, further military revolutions will occur inasmuch as technological advances are steady, thereby exacerbating the security dilemma since an adversary maintains a competitive advantage, (Krepinevich, 1994). Like other technologies, Artificial Intelligence (AI), the latest innovative technology which is currently used in daily life routines, might heighten the security dilemma and might underwrite a new military revolution amid the increasing tendency to use it in the military sphere. Based on Krepinevich's argument, the application of Artificial Intelligence to military sphere would result in a military revolution that requires organizational innovation, the production of a new system, organizational adaptation and technological change within military organizations, (Krepinevich, 1994).

Though ***Artificial Intelligence (AI)*** is not a newly invented technology, it has lately gained momentum due to recent cataclysms over its potential implications over both national and international security amid the increasing tendency to weaponize it and to use it in military applications. AI has dozens of definitions which mirror the developments and advances in such a kind of technology throughout the past decades. The definition of AI has broadened from being merely termed as the automation and the computation of intelligent behavior to be defined as the ability of computerized systems to implement tasks which are used to be performed by humans only and to replicate mental skills, including the perception of natural languages, pattern recognition and adaptive learning, which have been monopolized by humans, (De Spiegeleire et al, 2017). As a result of steady progress in AI, the AI literature has laid out four approaches of artificial intelligence: (1) computerized systems that think humanly, (2) computerized models which are designed to think rationally, (3) machines that act like human beings, and (4) the creations of automated systems that act and behave rationally, (De Spiegeleire et al, 2017). These four approaches can be categorized under two dimensions: (1) thought process and (2) rationality, (Russell, 2009). This classification highlights various orientations and paradigms of AI.

Such a progress in AI sheds the light on the plausible implications of AI upon the role of human in the military sphere in the aftermath of using machine learning and deep reinforcement learning. This implies that AI could pose a threat to a state's security since there are aspirations for making human-out of the loop, thus machines will surpass human intelligence after they have been used either for carrying out certain tasks in alignment with human intelligence or performing a full range of tasks with a human supervision, (De Spiegeleire et al, 2017). AI meanwhile enhances the capabilities of a state. Therefore, uncertainty over the military, legal, and humanitarian impacts of AI weapons looms over the horizon.

Artificial Intelligence resembles nuclear weapons in terms of being initially invented for peaceful and civilian purposes, and for being eventually used for military purposes. Artificial intelligence has a great potential for being used in diverse sectors ranging from medicine, education, business, finance, cybersecurity to marketing, (De Spiegeleire et al, 2017). However, both AI researchers and IR specialists are concerned with the potential weaponization of AI, and they highlight the need to avoid errors and regulate AI for peaceful purposes. There have been military applications of Artificial Intelligence such as drones, including robots and anti-missile systems, (Bates, 2017). The commonality between nuclear weapons and artificial intelligence also includes the probability of spurring AI arms race to mitigate the security dilemma and restore strategic stability, (Geist et al, 2018). Equivalent to the nuclear weapons, it is hard to define the nature of AI as weapon and how it would affect states' behavior, as per Mohan who highlighted the problematic nature of good and bad weapons/technologies. He stressed that the differentiation between good and bad weapons/technologies is challenging given that certain weapons/technologies could be a stabilizing factor at some point due to targeting accuracy and their efficacy in a second-strike capability. However, they could eventually be regarded as destabilizing weapons, if other sorts of anti-weapons technologies such as Anti-Ballistic Missiles are developed, (Mohen, 1986).

With this analogy between AI and nuclear weapons and their destructive potentials, and with the rapid advances in AI, it is worth considering how they might exacerbate the security dilemma since there is a consensus in the IR literature over the undeniable inclination to militarize the AI technology. However, the AI literature in itself is polarized over the use of AI in the military realm. Proponents claimed that AI militarization has its own advantages. Such advantages range from decreasing the number of human combatants which would definitely reduce causalities to the accessibility to dangerous areas through the employment of unmanned vehicles and robots, (Etzioni et al, 2017). They additionally argued that AI would be of great help in the decision-making process since robots are equipped to carry out and coordinate multitasks, (Etzioni et al, 2017). Consequently, around of 30 states, (Autonomous Weapons, 2016) are pursuing AI capabilities including the United States which develops killer robots for integrating them in its third offset strategy.

Proponents also supported automation in weapons since they mistakenly assume that autonomous weapons could put an end to the legal dilemma over civilian causalities, (Autonomous Weapons, 2016) thereby precluding a state's responsibility. On the contrary, as opponents always emphasize on both ethical and legal dilemmas of using autonomous weapons and the negative repercussions of the declining humans control over the course of war, autonomous weapons would increase civilian causalities. With the development of AI-enabled weapons, humans might not be the essential operators, (Autonomous Weapons, 2016). Hence, this nullifies the view point of proponents, arguing that autonomous weapons could have better abilities in targeting and discriminating military objects from civilian ones, as well as performing tasks with greater precision and reliability. Proponents' assumption is dubious given that humans have better judging abilities and can act as either moral agents or human as fail-safe when autonomous weapons fail to judge the situation correctly, to adapt to changing circumstances or to perform tasks effectively, (Autonomous Weapons, 2016). More importantly, the utilization of AI in armed conflicts sets off alarm bells over the applicability of International Humanitarian Law (IHL), (Kreps et al, 2012). Defenders of autonomous weapons see that AI would fulfil the requirements of Article 48 of the 1977 additional Protocol, (Kreps, 2012). However, such an assumption is unreasonable because AI could instigate collateral damage due to fallacious distinction. Thus, the ongoing controversy over AI highlights the dichotomy between autonomy in weapons and human control.

This transformation in the usage of AI technology accentuates that the security dilemma will be intensifying. Today's conflicts accelerate vicious races in technology for the purpose of enhancing a state's defensive power to avoid annihilation, as John Hers argued. Though this argument is short-sighted given its disregarded the fact that the security dilemma is mutual. Robert Jervis speculates that the security dilemma arises when a state accumulates more capabilities, such as AI, in an effort to strengthen its security, thereby endangering the security of the other state, (Tang, 2009).

Tang argued that a genuine security dilemma exists when anarchy prevails, and defensive measures are taken without malicious intentions. (Tang, 2009). This classification helps rivals differentiate between accidental escalation under the security dilemma and a measured response to possible aggression. Thus, Jervis and Tang's definitions incorporate of both objective sense which assesses the lack of threats to a state's acquired values, and subjective sense which represents the freedom from fear over the loss of a state's values, (Buzan, 1991). In other words, their definitions are based on Arnold Wolfers' definition of security that entails both the material aspect and the psychological factor. They likewise coincide with Kenneth Waltz's definition of security that sees world as anarchic due to the lack of upper-hand authority, resulting in the emergence of self-help system where competition exists, (Williams et al, 2008).

Thus, in the current anarchic system, (Waltz 1959), AI could tighten the security dilemma due to the uncertainty over its destructive capabilities and adversaries' intentions even if they are merely security-seekers. Thus, Tang's contribution to the literature would help states in measuring the severity of the security dilemma when it applies to the AI realm. To assess the severity of the security dilemma in the AI realm, there is a need to decide whether offense or defense is dominant. Given that the AI literature is still undeveloped, there is no a clear-cut assumption about the nature of AI as a weapon. The AI literature has mistreated the offense-defense balance and it has not thoroughly tackled the relation between AI and other types of weapons.

The AI literature has unduly covered the mismatch between AI and conventional weapons in spite of striking advancement in killer robots and unmanned weapons that are expected to replace manned soldiers in the foreseeable future. Accordingly, what stands out is the profound investigation of the relation between AI, on one hand and nuclear, cyber and conventional weapons, on the other hand. This further illustrates that these weapons should not be investigated separately when their implications upon AI deterrence are investigated. The rationale behind this is to hypothesize probable scenarios of the security dilemma based on types of weapons developed and possessed in tandem with AI capabilities.

As the security dilemma exists when a state develops weapons or technologies that enhance its ability to attack and when a defender finds itself in status where the strategic balance has shifted, the theory of the offense-defense balance should be considered to determine the severity of the security dilemma. Jervis argued that defense is dominant when a defender has no willingness to launch preemptive strikes or to carry out preventive attacks to avoid depletion of resources and enormous costs of war, (Jervis, 2009). In other words, a state's perception about the severity of the security dilemma is partially based on its relative ease and the shift in the balance of power, (Lieber et al, 2017). Furthermore, Jervis claimed that restoring the balance of power is feasible by catching up capabilities, so as to increase the chances of cooperation, (Jervis, 2009). While, offense becomes dominant when both sides have equal defense budgets; the benefits of preemptive attacks are much higher than inaction; the first-strike is advantageous, and when the loser lacks concrete evidences about the winner/adversary's intentions, (Jervis, 2009). By applying this to AI, it is worthy to consider the effect of such a novel technology on the mobility of weapons (killer robots, cyber weapons) and their destructive power, (Lieber et al, 2017) to decide whether AI weapons favor offense or defense. Beside to evaluating the strategic, operational and the tactical aspects of an AI strike, empirical logic says that both the psychological aspect, and the reconcilability and irreconcilability of interests should be studied, (Tang, 2009). Given that AI specialists and researchers have disregarded the possibility of using civilian AI capabilities in the military sphere, in spite of plentiful incidents in the history of warfare and the noticeable reliance on drones and unmanned vehicles, it is highly significant to take conflict of interests, the offense-defense variables, and technological advances into consideration when measuring the severity of the security dilemma.

Based on the severity of the security dilemma that could be exacerbated by the development of AI capabilities, states' behavior, as Jevris noted, will be influenced either by reciprocal fears of retaliation, reciprocal malign intention as rivals develop capabilities to intentionally deter each other, (Tang, 2009) and the enormous implications of exhausting military resources if the security dilemma is genuine, (Jervis, 2009).

The novelty of such a kind of technology mounted an intense debate over the development of AI capabilities since peaceful applications could trigger AI machine-led wars. Some experts concluded that artificial intelligence could put personal privacy at stake through surveillance monitoring; it also could be used as a coercive weapon since it can explore points of weakness in a business organization, (The New Dogs of War, 2017). Since it could threaten a business organization, a state's security could be threatened as well. More importantly, tracking AI weaponry suppliers would be problematic since AI factories are just integrated networks of virtual facilities. Further, it would be challenging to identify the types of AI capabilities whether for peaceful or military and subversive purposes, (The New Dogs of War, 2017). Though the security experts who participated in Threatcasting Workshop accurately identified threats of AI, they disregarded other possible threats of weaponized artificial intelligence. A different group of scholars, on the other hand, see that AI could heighten the security dilemma through the utilization of malicious cyber capabilities and disinformation, as well as surveillance for data mining, (Osoba et al, 2017). This raises a question about the difficulty of attribution

There are other factors that could exacerbate the security dilemma in a dyadic relationship even when AI capabilities are developed for peaceful, commercial and civilian purposes, including the Research and Development (R&D) expenditure, progress in education, economic prosperity, as well as surveillance and reconnaissance. The AI medical applications, for instance, could be weaponized through the exploitation of or the hacking of medical data attached to the internet by attackers/states to inflict damage upon defenders. The production of AI intelligent machines for the sake of profit could trigger arms race at the regional level, (Layton, 2018).

Adding to this, cyber capabilities which are linked to AI software could be destructively exploited to launch offense strikes and to disrupt a state's infrastructure, (Eckersley et al, 2018). Cyber capabilities coupled with AI ones could worsen the security dilemma since it enhances both offensive and defensive powers of a state vis-a-vis its neighboring country. This reflects uncertainty over adversaries' intentions and vulnerability of a states' security system since AI software could stalk on opponent’s security system to attack its weakest point and make it inoperative, (Allen et al, 2017).

Concerning information security, cyber-enabled software, along with social media botnets would aggravate the security dilemma and would menace a state’s economy and its regime through the spread of fake news and data poisoning, (Allen et al, 2017). Peaceful applications of AI could be used as a sabotage to inflict grave economic loss, (Allen et al, 2017).

In the conventional domain, the diffusion of killer robots into real militaries poses a threat to a state’s security since its territory is prone to attacks by robots, (Eckersley et al, 2018). So, peaceful AI applications have their own pros and cons since they make individuals’ life easier but endangering their privacy in the light of individualized and intelligentized warfare.

Adding to this, the heated debate over the legality of AI-enabled weapons with the difficulty of attribution also reflects international lawyers overwhelming perplexity. Around of five arguments have emerged in the international law literature. Of which, bestowing a legal personality for AI entities which in turn raises a question about liability and accountability in case of non-compliance to international legal instruments or the commissioning of illegal acts, (Burri, 2017). This further raises the alarm bells over the inability of international lawyers to define liability in the AI realm and to outline the cases where a state would be legally responsible for using AI applications in the military sphere. While, another argument articulates, the banning of fully autonomous weapons under a new set of international legal instruments. It further suggests that low level/semi-autonomous weapons could be lawful and could be regulated under the international law, thereby lessening the security dilemma. Thus, a precise legal definition of a meaningful human control, where the symbiosis between humans and machines is defined, should be drafted for avoiding future conflicts between AI possessing countries and AI not possessing countries, (Burri, 2017). Since retaining a degree of control over machines and autonomous weapons seems challenging in the age of algorisms-based warfare, a new set of international legal instruments could regulate the utilization of AI weapons in conformity with the Law of Armed Conflict and the International Criminal Law, (Burri, 2017). The problem of attribution, notably when humans are out of the loop, provides an illustration of how states are subject to the will of machines and are also might be legally responsible according to public international law, (Burri, 2017). This further implies that public international law should lay out the decisions that should not be delegated, under any circumstances, to autonomous machines, (Burri, 2017), in addition to outlining the situations where humans should be in/over the loop to master the course of war. Another group of international lawyers have introduced a supposition suggesting the emergence a super-soft law through the creation of international ethical and moral standards, (Burri, 2017). As per this argument, such a bottom-up law-making process could be binding at the state level. Janet Koven's counterargument, refuting ethical and moral standardization and their inapplicability to the international landscape, (Burri, 2017) was factual and logical amid ongoing AI arms race. Further, such a bottom-up lawmaking raises a question about the political will and the essentiality of incorporating states in the lawmaking process.

Since AI has triggered an arms race in the commercial sphere which in turn has been shifted to the military one (Research and Development in automotive, information and communication; aerospace and defense constituted were immense throughout 2014-2016), (Cuminings, 2017), the AI literature anticipated an array of AI future scenarios. One of those scenarios is a "Sputnik Event" triggering a sharp AI race between states since maintaining an AI superiority could enhance economic, military, defensive, scientific and geopolitical powers of a state, (De Spiegeleire et al, 2017). This implies that a Sputnik-like incident for AI is not improbable amid the ongoing space warfare. Thus, AI race could pose a threat to a state's security, notably the weaker one. But it could be a stabilizing factor in case of parity.

This raises a plethora of questions about the validity of using AI as a deterrent tool and the probable implications of developing AI upon the relations between rivals. One possibility, as previously discussed in this paper, is the exacerbation of the security dilemma, which could potentially lead to pre-emptive strikes. Another possibility is the operation of deterrence -– much like MAD with nuclear weapons - when a state is being informed of a rival's capacity to launch its own destructive strike.

Based on the above, AI could exacerbate the security dilemma, thus the strategy of ***deterrence*** which has gained momentum among IR scholars during the Cold War era, reintroduces itself as a possible solution for the underlying dilemma. However, it could be problematic when it is applied to AI. Firstly, the elements of deterrence should be investigated to assess the soundness of AI as a deterrent tool. The elements of the classical deterrence theory, coined by Hobbes, include self-interests, material gains, unavoidable conflict and rationality to the international realm. As per the findings of other scholars, namely Cesare Beccaria, the strategy of deterrence pertains the threat of inflicting high costs on perpetrators to dissuade them from committing crimes, (Dilulio).

In other words, the rational theory of deterrence revolves around a state's ability to dissuade its adversary from carrying out certain actions through latent force. According to classical theory, deterrence operates when an adversary assumes that its rival has considerable military capabilities, threats are credible, and costs would be undesirable should provocative actions be taken, (Quackenbush, 2011). Therefore, credible ultimatums and the threat of use force are fundamental for effective deterrence. Secondly, the level of technological advancement and the dominant trend of weaponization should be studied to determine the severity of the security dilemma, and to question if AI would deter a state from attack, thus the security dilemma will no longer operate since none of the rival states would be defensive, (Jervis, 2009).

By tracing rapid advancement in technologies and weapons and how it has altered the art of war throughout the past decades, the term "killer robots" was invented in response to the excessive use of drones and robots in military. This term underlines the salient apprehension over the ability of "killer robots", as per Sharkey's argument, to act like humans since they lack human capabilities and human intelligence that are required for making military decisions, (Sharkey, 2012). This illustrates that killer robots have their own limitations when it comes to war. Since AI weapons would not be able to differentiate between civilian and military targets and can cause collateral damage, autonomous weapon targeting is worrisome, (Etizioni et al, 2017).

The Israeli Harpy is a perfect example of this problem since it cannot distinguish whether the radar is located on an anti-aircraft station or on a civilian facility, (Sharkey, 2012). This raises a question about the ability of lethal artificial weapons and killer robots to cope the pace of strategic decision-making in combat, especially in densely populated areas. Moreover, Garcia's argument about the inevitability of disruptive change in the domains of international peace and security is convincing given that the weaponization of AI signifies the erosion of fundamental international norms that regulate the use of force, (Garcia, 2018).

If superiority in AI, as Garcia pointed out, would come in favor of the superior, (Garcia, 2018), should it wipe out a state's ability to respond. If the superior state has the ability to launch a first-strike, deterrence will not work and offense will dominate. As nuclear weapons have changed the calculus of war during the Cold War, AI, as Randolph claimed, could tighten cyberspace and outer space warfare in the wake of unprecedented reliance on easily disrupted cyber capabilities, (Kent, 2015). In spite of this, scholars are looking forward to tailoring a new doctrine for regulating warfare in cyberspace, some of them argued that the first-strike is advantageous in cyberwarfare because it is cheaper, and attribution will be challenging since it is hard to track perpetrators. Thus, the IR literature should devote more focus on the influence of AI capabilities coupled with either cyber capabilities, nuclear capabilities or even both capabilities on the second-strike capability. The literature has tackled the first-strike capability in the cyber sphere, but with the weaponization of AI, there is a pressing need to reassess this argument given that "killer robots" could make the second-strike capability a preferable option since the extent of destruction is still obscure and the immunity of noncombatants, (Crosston, 2011) a fundamental criterion of Jus ad bello, is still unsettled.

By the same token, there is a strong debate over the possibility of a nuclear war in the light of robust advantages in both cyber and AI capabilities. Subversionist scholars purported that AI could trigger nuclear warfare since adversaries could mislead or alter AI capabilities, (Geist et al, 2018). Subervisionists' view point concurs with the alarmists who conceive that advanced Artificial Intelligent capabilities would render nuclear arsenals vulnerable, thereby diminishing the strategic balance, (Geist et al, 2018). Accordingly**,** AI could be destabilizing given that it could make the second-strike capability ineffective, (Geist et al, 2018). However, the literature has dismissed the fact that nuclear weapons have been used for deterrence even with the occurrence of disinformation and cyberattacks. This could tell that nuclear weapons could deter an AI first-strike capability since city-sparing, cyberspace-sparing and machine-sparing dictate leaders' decisions.

Since AI could sharpen the security dilemma through uncertain technological asymmetries between great powers and small states, a question raises itself about the efficacy of small arsenals as a means of deterrent. In the nuclear realm, small arsenals have been successful deterrent as the literature demonstrated it is a matter of possession such a kind of destructive weapons. As Jervis stated small arsenals and moderate military expenditure could neutralize disparity and high military expenditure, so as to restore the second-strike capability. This has been the situation with the nuclear weapons. Weak and small states, as Jervis argued, usually prefer defense and seek cooperation, but because they might resort to preemptive or preventive strikes due to their undesirable position, (Jervis, 2009). This, consequently, reduces chances of cooperation. If it is true that the first-strike is advantageous in the cyber realm, small and weak states could launch cyberattacks as preemptive strikes. This could generate two scenarios: (1) a retaliatory attack by using AI capabilities, causing collateral damage and making defense dominant or (2) inaction since the defender has no other options to retaliate, therefore making offense dominant. If the defender does not possess a nuclear arsenal, the weaker state could launch AI-enabled cyberattacks. Therefore, it is highly possible to carry out a nuclear or an AI strike. However, the situation would be quite different when a state possesses cyber, nuclear and AI capabilities given that both nuclear capabilities and AI applications, which enhance cyber defense, could make defense dominant. More importantly, geography and the location of weapons could be determinant factors in the strategy of deterrence. As Jervis allured both conventional weapons and nuclear weapons are defense-oriented based on geostrategic position and the location of nuclear weapons. The same can be applied to the AI realm, though it instantly favors offense, due to tactical and operational considerations; the vulnerability of both nuclear and AI weapons, in addition to high exposure of critical infrastructure through cyberattacks. In conclusion, deterrence could be effective in today’s world.

In addition to technology, geography and various capabilities, the power to hurt is an integral element of deterrence as Thomas Schelling elaborated that the power to hurt is a sort of diplomacy that makes threats credible since it is measured by the degree of suffering and pain that could be inflicted upon a rival, (Schelling, 2008). It basically rests on the use of latent violence and the infliction or the withholding of pain, (Schelling, 2008). This further indicates that deterrence requires the defender to communicate with the defector about possible course of actions in case of noncompliance while not necessitating to haphazardly leave the course of war to chance, otherwise, destructive war will erupt. Comparable to latent nuclear deterrence which is grounded on a state's intention to reduce the time required for producing a nuclear bomb, and nuclear latency which is based on the capabilities, (Fuhrmann, 2018) AI could be latent since AI proliferation is expected not to end and rivals would seek more capabilities, as well as the fact that cyber capabilities could inflict pain upon the defender and could also be a credible threat. In addition to that, R&D expenditure and the production of enormous commercial and medical applications could be signs of latent violence since rival states can convey ultimatums through steady progress in AI technologies.

This argument nullifies the IR literature's suggestion of "deterrence by denial" which deters the adversary from acquiring further capabilities. It is almost impossible to deter a state from possessing AI or cyber capabilities amid the ongoing arms race and the increasing asymmetry. Past incidents in nuclear deterrence accentuate the efficacy of deterrence by punishment as opposed to deterrence by denial, as it was the case in the Israeli strike against Iraq's nuclear arsenal. While deterrence by denial is the favored option of small states, it failed to dissuade great powers from developing more weapons. That is why, Paul Davis introduced dissuasion by denial as a replacement of deterrence by denial. He claimed that dissuasion by denial pertains the calculation of potential repercussions of carrying out an attack based on expected value and worst/best-case scenario, (Davis, 2014). To this end, the defector should be informed of the positive outcomes of de-escalation and vice versa. Therefore, AI deterrence could be a mosaic of latent violence and dissuasion by denial. This retells the Cuban Missile Crisis when deterrence by punishment along with concessions and assurances prevented the outbreak of a destructive war. However, latent violence and punishment will be the core of AI deterrence.

The cognitive theory/prospect theory of deterrence, which was coined by Jeffrey Bekejikian, provides a suitable model for evaluating threats from highly credible to highly incredible. Hence, this scale would definitely guide decision-makers to make the right decisions based on actual capabilities and accurate calculation of costs and precise assessment of threats' credibility, (Bekejikian, 2002). Such a scaling of threats that is based on the variants of coercive diplomacy, presented by Alexander George, which includes classic ultimatums that integrate three main elements: a demand, sense of urgency and threat of punishment; tacit ultimatums that succeed when conveyed deliberately and effectively, or positive assurances/concessions, (George, 2009), would help the defector in calculating the credibility of threats. When it applies to the AI sphere, rivals would mutually deter each other not only because of uncertainty and credibility of threats, but also the fact that neither of them would gamble the status quo even when the estimated outcomes of defection are higher than the status quo, (Bekejikian, 2002).

It is also significant to study the psychological factor in the decision-making process since AI requires humans to be out of the loop. The psychological aspect contributed to the effectiveness of deterrence as it was the case in nuclear deterrence, precisely the Cuban Missile Crisis which was an ideal example of general deterrence that exemplifies rivals' satisfactions with the status quo. The psychological aspect would be non-existent in AI to AI interactions. This tells that AI could be disadvantageous in crisis management as some crisis require more time to be resolved diplomatically, (AI and the Military, 2019). This further implies AI to AI interactions could increase the probability of war that might produce unexpected outcomes, thereby increasing uncertainty and yielding strategic surprises, (AI and the Military, 2019). Such an observation is based on the embryonic capabilities of AI military applications and the mainstreamed assumption about the impossibility of developing AI applications capable of analyzing and reporting all diplomatic endeavors and efforts, (AI and the Military, 2019). Theoretically speaking, this conclusion is convincing, but practically speaking, it tackled the issue from one angle and overlooked the other angle which is the participation of humans in strategic decision-making. When humans are over-the-loop, in spite of the high potential of data manipulation and errors, the psychological factor would be prominent in the anticipated AI deterrence. The IR literature regrettably overlooked all possibilities and scenarios while investigating the effects of technological advancement on security.

Hence, AI could exacerbate the security dilemma amid the ongoing arms race in the commercial sphere, banning AI, as some scholars suggested, is highly unlikely. The literature has debated over the legality of AI weapons and the potentiality of banning AI itself or solely banning AI in military applications. As Glaser claimed cooperation is possible under the security dilemma when offense-defense variables are segregated and when states have knowledge about motives and intentions of an adversary, (Glaser, 1997). The IR literature should cooperate with the AI literature to regulate R&D in AI, on one hand and to lay out a legal framework for governing AI in military applications, on the other. Although, drafting an NPT-like agreement or regulating it is very hard to achieve, Glaser argued that drafting arms control agreements is the best desirable solution, (Glaser, 1997) given that they would promote mutual restraints, (Glaser, 1997).

Since AI is currently seen as a new frontier for Weapons of Mass Destruction, it is worth considering whether a scenario of Mutually Assured Destruction could evolve, and whether, like during the Cold War and the post-Cold War, mutual kill via AI (Jervis, 2009), would lead to deterrence or intensify tensions between superpowers, (Lebow et al, 1995).

Kenneth Waltz’s argument, which sees that nuclear deterrence focuses on the ability to cause damage to the aggressor rather completely defeating it concurs with Jervis and Schelling’s views of nuclear deterrence, (Waltz, 2009). Despite Kenneth’s viewpoint regarding the elimination of the essentials of war-fighting on account of nuclear deterrence is worthy of consideration, (Waltz, 2009) his argument about the elimination of the elements of defense was misleading given that nuclear weapons have made mutual fear intense. As Jervis noted nuclear deterrence has created general stability due to the alterations in political values of wars and the advert changes in states’ perceptions, intentions and motivations, (Jervis, 2009). General stability, therefore, ascertains that nuclear weapons may help in maintaining peace between rivals by dissuading them to overturn the status quo even when they have the motivation, (Jervis, 2009). In addition, general stability has negated the view saying that nuclear weapons did not preclude non-nuclear states to carry out escalatory acts, (Quackenbush, 2011). The opponents of nuclear deterrence have disregarded how the imbalance of power exacerbates the security dilemma. This additionally manifests that nuclear superiority is a destabilizing factor and does not guarantee a decisive military victory, (Mohan, 1989). Hence, nuclear deterrence, in contrast to the views of staunch opponents of deterrence, has proved to be empirically fruitful because it has precluded enormously destructive wars and has maintained stability in times of conflicts and peacetime.

Given that ***Mutually Assured Destruction***, an offshoot of nuclear deterrence, was acclaimed by IR scholars, AI-MAD could be a workable strategy since the literature debates over the applicability of MAD in the cyber sphere.

The chief essence of MAD is the vulnerability of both sides to retaliation with the possibility of launching a second-strike capability, (Mohan, 1989). Therefore, such mutual vulnerability and mutual fears had contributed to the success of general deterrence when the Cuban Missile Crisis erupted by pushing the leaders of the two superpowers towards a settlement rather than pushing them to a severe confrontation, (Lebow, 1995). Accordingly, the more nuclear capabilities, the higher possibility of effective deterrence since each side will be deterred due to the uncertainty over the devastating consequences of a second-strike, (Mohan, 1986). The use of general deterrence at the peak of the Cuban Missile Crisis attributed to the prevention of a catastrophe as it influenced the risk of war. This further implies that deterrence was rather effective because of the asymmetry of interests and nuclear parity rather than nuclear superiority, (Lebow, 1995). This likewise proves that deterrence is a viable strategy since it promotes leaders to refrain from war and to accept the status quo when it is proved to be the best-case scenario.

Therefore, the AI literature should posit how a Cuban Missile Crisis similar incident in the AI could happen and what could generate mutual fears: would it be mutual disruption of cities, machines, cyber systems or overkill? The Cuban Missile Crisis was between two superpowers but if a Cuban Missile Crisis incident took place at the regional level, would asymmetry in technology promote regional adversaries gone AI-MAD? Since the psychological factor played a crucial role in the Cuban Missile Crisis, could human intervene in a state’s offset strategy to assist AI weapons and killer robots to pinpoint the right targets when it applies to the AI realm? The AI literature should also investigate the implications of AI when humans are out-of-the-loop and when they are over-the-loop, as well as identifying which scenario would be the most destructive since MAD is centered on “the indivisibility of control”, (Fairbanks, 2004). In addition to that, the AI literature should make a comparable study on the implications of AI and the severity of the security dilemma based on a state’s dependence on technology and a state’s military capabilities.

While the term “Mutually Assured Deletion/Delibitation” has become trendy in the IR literature, some argued that the first-strike is favorable in the cyber realm. So, if the first-strike becomes a preferable option and offense is dominant in the cyber sphere, AI-MAD could be a substitute for the so-called “Mutually Assured Deletion/Delebitation” since massive destruction could be the logical outcome either through the eruption of a conventional war, cyber warfare or even a nuclear war. Though Fairbanks proclaimed that “damage limitation” was not the main goal of nuclear MAD, (Fairbanks, 2004), today’s MAD could be overwhelmed by “damage limitation” since current capabilities have surpassed human control. However, this does not necessarily mean that AI regulations should not be merely concerned with the damage limitation since the severity of damage could be grimmer, as opposed to other types of warfare.

The superiority of AI capabilities over cyber ones is also debatable. Some argued that AI capabilities could overturn cyber ones since they could discover vulnerabilities in other cyber defense systems and exploit them, (Horowitz et al, 2018). The flipside of utilizing AI capabilities is enhancing a state’s cyber defense system by patching vulnerabilities in its own cybersecurity systems, thereby protecting its system from AI-enabled cyberattacks, (Horowitz et al, 2018). By the same token, AI could tighten disinformation by disseminating fake propaganda at a large scale, and could also counter disinformation through the utilization of bots and algorithms for detecting, analyzing, disrupting, vetting, blocking and filtering false/unauthentic data, (Horowitz et al, 2018). AI would be an effective tool for intelligence by gathering a tremendous amount of data, albeit it could be vulnerable to counter AI-spoofing, (Horowitz et al, 2018). Thus, AI triggers the security dilemma and demonstrates “mutual vulnerability”, which further implies that defense could be dominant. This further illustrates that self-deterrence would be successful since weaker actors can circumvent disparities by using other capabilities and inflicting a political pain, (Wasser et al, 2018). This, additionally, demonstrates that offense could be dominant in case of disparity and the lack of nuclear capabilities.

Besides to mutual vulnerability and collateral damage, AI-MAD is highly plausible since AI applications could have strategic implications over a state’s military, economic and information superiority, as well as its nuclear superiority. AI-enabled applications would change the balance of power between developed and developing countries, notably with the mammoth utilization of the 3D printing technology (which is also known as additive manufacturing {AM}) that will facilitate the development of highly disruptive and speedy technologies, and will accelerate weapons proliferation, (Johnston et al, 2018). Since Additive Manufacturing is a cheap technology and has the ability to replicate its applications, (Johnston et al, 2018), AI arms race would be accelerated. AI MAD is not improbable since machine takeover with its four possible scenarios could aggravate the security dilemma particularly for networked societies, (Bouskill et al, 2018).

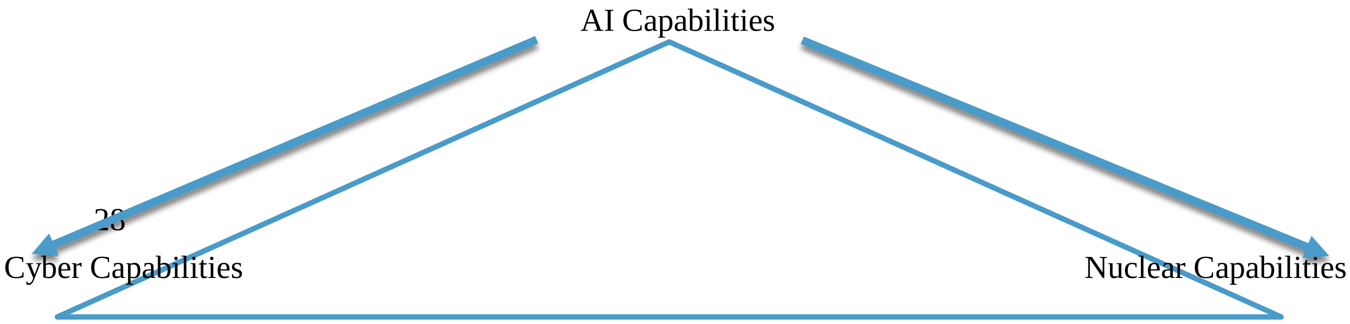
Resembling to nuclear MAD, the foundations of the anticipated AI-MAD could include: (i) the indivisibility of control, (ii) mutual fears of retaliation, (iii) severe destruction, (iv) the psychological factor, (v) parity/disparity, (vi) sparing, (vii) latent force and (viii) error. Corresponding to cyber MAD, the cores of the propositioned AI MAD could include: (i) attribution, (ii) costs and (iv) degree of dependence on technology. Opposed to nuclear and cyber MAD, AI MAD could also investigate the roles of humans and machines in a military's command and control.

To sum up, the literature should focus on the potential destructiveness of AI amid the massive use of cyber capabilities; find ways to ameliorate the security dilemma, and it should also consider if MAD applies to AI technology.

Conceptual Framework:

Since both cyber and IR literature claim that cyber threats could overcome AI capabilities, this paper will build upon the argument supporting the overpowering potential and future prospects of AI while considering the arguments that discredit the potentialities of AI applications versus cyber capabilities. This paper will explore the possible implications of AI on both cyber and nuclear capabilities. It will also tackle the direct proportion between both cyber-offensive and cyber-defensive capabilities, and nuclear capabilities in relation to AI capabilities.

The Security Dilemma Triad:



Towards this end, this paper develops a "***Security Dilemma Triad***" composing of three main elements: cyber, AI, and nuclear capabilities. Based on the security dilemma triad, this paper will address the relationship between AI and cyber capabilities on the one hand, and the relationship between AI and nuclear capabilities on the other, as well as the relationship between nuclear and cyber capabilities with the presence of AI capabilities.

Therefore, four possible scenarios will be developed to investigate whether offense or defense will be dominant as follows:

(A) When a state possesses nuclear capabilities + AI capabilities + cyber capabilities = defense is dominant;

(B) When a state does not possess nuclear capabilities, but possesses AI capabilities + cyber capabilities = offense is dominant;

(C) When a state possesses nuclear capabilities + cyber capabilities but does not possess AI capabilities = defense is dominant;

(D) When a state possesses nuclear capabilities + AI capabilities but lacks cyber capabilities = defense is dominant.

This paper presumes that defense is dominant under the first scenario given that AI capabilities can overcome cyber capabilities, thereby disavowing the argument of nuclear vulnerability against cyberattacks. Following the second scenario, offense is dominant given that nuclear deterrence is ineffective or absent, while AI capabilities could empower cyber capabilities by attacking points of weakness in the cyber system. Under the third scenario, defense is also dominant, in spite of the lack of AI capabilities, due to nuclear deterrence. Finally, for the fourth scenario, defense is dominant because a state possesses nuclear weapons that maintain a second-strike capability.

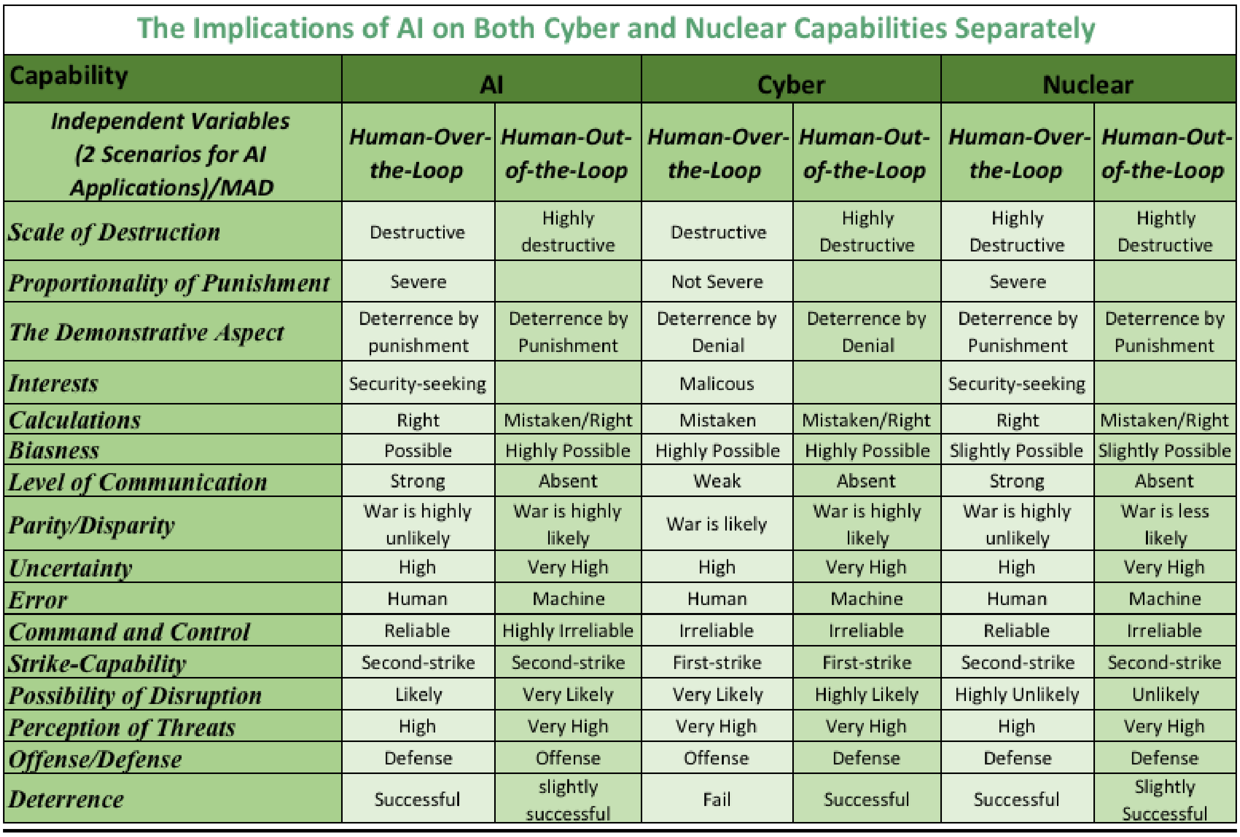
The rationale behind developing the abovementioned scenarios and the "*Security Dilemma Triangle/Triad*" is questioning how AI could change states' perceptions in terms of cyber and nuclear capabilities/doctrines. It will, initially, investigate the AI offense-defense dominance. Then, the implications of AI on both the cyber/nuclear offense-defense dominance will be covered. It will subsequently borrow the elements of nuclear Mutually Assured Destruction (MAD) (*please see annex 1*) as variables.

AI Offense Defense Dominance:

It is logic to start with exploring the dominance of either offense or defense in the AI realm before investigating the prospects of an AI MAD. By virtue of the declining role of humans in the AI sphere, it is prudent to hypothesize two scenarios for the application of AI in the military sphere: (i) humans have a minimal control over AI applications; (ii) human supervision over AI applications is absent. Based on that, this piece presumes that AI favors a second-strike capability, (Schneider, 2018). Notwithstanding, the impossibility of defining a machine's accountability in violation of the Law of Armed Conflicts and the Geneva Conventions makes an offensive AI strike advantageous. This does not necessarily mean that offense is dominant in AI. On the contrary, **defense is dominant in AI when humans maintain control over machines. On the other hand, AI could favor offense when human control is absent and when a military's command and control is digitally-dependent on cyber capabilities.**

The Implications of AI in Terms of the Element of MAD:

Based on that conclusion, the following section will cover the implications of AI on both nuclear and cyber capabilities in order to investigate the offense-defense dominance in the abovementioned scenarios:



The Implications of AI on Both Cyber and Nuclear Capabilities Separately

Capability	AI		Cyber		Nuclear	
Independent Variables (2 Scenarios for AI Applications)/MAD	***Human-Over-the-Loop***	***Human-Out-of-the-Loop***	***Human-Over-the-Loop***	***Human-Out-of-the-Loop***	***Human-Over-the-Loop***	***Human-Out-of-the-Loop***
Scale of Destruction	Destructive	Highly destructive	Destructive	Highly Destructive	Highly Destructive	Hightly Destructive
Proportionality of Punishment	Severe		Not Severe		Severe	
The Demonstrative Aspect	Deterrence by punishment	Deterrence by Punishment	Deterrence by Denial	Deterrence by Denial	Deterrence by Punishment	Deterrence by Punishment
Interests	Security-seeking		Malicous		Security-seeking	
Calculations	Right	Mistaken/Right	Mistaken	Mistaken/Right	Right	Mistaken/Right
Biasness	Possible	Highly Possible	Highly Possible	Highly Possible	Slightly Possible	Slightly Possible
Level of Communication	Strong	Absent	Weak	Absent	Strong	Absent
Parity/Disparity	War is highly unlikely	War is highly likely	War is likely	War is highly likely	War is highly unlikely	War is less likely
Uncertainty	High	Very High	High	Very High	High	Very High
Error	Human	Machine	Human	Machine	Human	Machine
Command and Control	Reliable	Highly Irreliable	Irreliable	Irreliable	Reliable	Irreliable
Strike-Capability	Second-strike	Second-strike	First-strike	First-strike	Second-strike	Second-strike
Possibility of Disruption	Likely	Very Likely	Very Likely	Highly Likely	Highly Unlikely	Unlikely
Perception of Threats	High	Very High	Very High	Very High	High	Very High
Offense/Defense	Defense	Offense	Offense	Defense	Defense	Defense
Deterrence	Successful	slightly successful	Fail	Successful	Successful	Slightly Successful

Drawing on the above mentioned, the possession of other sorts of military capabilities, such as conventional, nuclear and cyber, in tandem with AI ones helps states to take the situation from different angles. These angles could be: (i) AI capabilities have no implications over other capabilities and vice versa, (ii) other capabilities could bolster a state's position when it possesses highly advanced AI software and applications, (iii) other types of military capabilities are valuable, if a state possesses amateur AI software, or (iv) other military capabilities are invaluable, if a state possesses advanced AI applications. This in turn helps states to see if small AI arsenals could create a MAD-like structure with the possession of other military capabilities. Perhaps, small AI arsenals could deter states from launching a preemptive or preventive strike.

Hereafter, the paper supposes that mutual AI deterrence could be established between two nuclear states. While, asymmetric deterrence might operate between a non-nuclear state and a nuclear state since the non-nuclear state would be deterred from launching a first military strike because of the adversary's superiority with the possession of nuclear and fully autonomous AI applications. Though, it might employ asymmetric capabilities instead to deter its adversary from launching a preemptive AI or a conventional strike. For the superior state, it could resort to offensive warfighting rather than depleting its nuclear arsenal that could be neutralized by AI.

Methodology:

To explore the potential impact of AI on the security dilemma, I interviewed policy makers, and experts in the field, in addition to an extensive review of the secondary literature on the weaponization of AI.

Interviews included: personal and phone interviews with a security expert/military advisor, two university professors, two ambassadors and a researcher as follows:

(i) Dr/General Mahmoud Khalaf, advisor at Nasser Military Academy;

(ii) Dr Dalal Mahmoud Al-Sayed, a professor of political science at Faculty of economic and Political Science at Cairo University and Nasser Military Academy;

(iii) Dr Waleed Rashad, assistant professor at the National Center for Social and Criminological Research;

(iv) Ambassador Karim Haggag, professor of practice at the American University in Cairo;

(v) Ambassador Aly Erfan; Program Director at the School of Global Affairs and Public Policy at the American University in Cairo;

(vi) Mona Soliman, doctoral candidate at the Faculty of economic and Political Science at Cairo University and a researcher at International Politics Journal (Al-Siyasa Al-Dawleeya).

Variables and Investigation Methods: (*please see annex2*)

To investigate the four scenarios mentioned in the conceptual framework, the dependent variables include: dependence on technology, sparing, latent violence and the demonstrative aspect, expected utility and cost-benefit analysis, calculus of war, balance of power, relatively of power and comparison of military and non-military capabilities, parity/disparity, margin of error, levels of communication, intentions, scope of human role in the decision-making process, degree of control over machines, indivisibility of command and control systems, attribution and counterforce. In addition, geography and population will be considered to see their impacts on a state's strategic depth. These dependent variables will help in observing variations based on the two independent variables which are: (i) human-over-the-loop, (De Spiegeleire, S et al) and (ii) human-out-of-the-loop, (Russell S. J et al, 2010). Accordingly, it hypothesized two scenarios for AI deterrence, as either successful or failed, based on the degree of human control over machines, the degree of dependence on technology and the impacts of AI applications on nuclear and cyber policies.

Findings:

What is Artificial Intelligence?

The definition of Artificial Intelligence is originated from the definition of intelligence which is defined as an agent's computational ability to perform tasks and achieve goals in different environment. Based on that, Artificial Intelligence is defined as a machine's ability to replicate humans' mental skills and behaviors, namely pattern recognition, reasoning and neuro-linguistic programming (NLP), and to learn by experience, as well as being able to adapt to environment and changes, (De Spiegeleire, S, et al). The US Defense Science Board defined AI as the computation of tasks such as decision-making, perception and conversation, which are used to be exclusively done by humans, (De Spiegeleire, S, et al). Such a definition of AI illustrates that computation and automation are associated with thought processes, reasoning, behaviors, ideals, and fidelity and dependability of human performance, (Russell et al, 2010).

Thus, the core of AI technology is the mimicry of human characteristics autonomously, (Tweedie, 2017). AI technology entails (i) expert systems, (ii) machine learning, (iii) natural-language processing, and (iv) AI planning, (Tweedie, 2018).

The AI literature has generated three types of AI, mirroring the evolution of AI throughout the past decades: (i) Artificial Narrow Intelligence (ANI): It is a sort of technology that mimics a narrow range of human behavior/intelligence. It is a sophisticated technology, albeit it cannot develop codes; (ii) Artificial General Intelligence (AGI): It is a more sophisticated technology as opposed to Narrow Artificial Intelligence since it emulates a wider range of human behaviors. It is a type of technology that mimics human intelligence as if they are made by humans; (iii) Artificial Super Intelligence (ASI): It transcends human intelligence, (Tweedie, 2017). Artificial Super Intelligence, as per AI developers' speculations, is expected to nullify and end the exclusivity of human intelligence, (Tweedie, 2018).

The Possible Implications of AI on Other Military Capabilities (Nuclear and Cyber):

A. Cyber Capabilities:

Cyber capabilities are a sort of capabilities and assets that a state can possess to use them in the conventional, commercial, nuclear, logistical, military and etc to resist possible attacks or project influence in cyberspace, (Craig, 2018). Both defensive and offensive capabilities shape a state's influence since they can be employed as active or latent, (Craig, 2018).

Today's cybersecurity systems' challenges and vulnerabilities are manifold. Cybersecurity systems are usually attacked through a chain of attacks starting with the reconnaissance, weaponizing, the delivery phase and ending with the exploit phase, (Wirkuttis et al, 2017). What is more important, the challenges associated with gathering cyber intelligence, inter alia, the need to constant adaptation with the massive amount of heterogeneous data that flows exponentially; the inadequacy of intrusion detection prevention systems that either defines malware by detecting abnormal patterns or outlines patterns of normal and recognized networks, (Wirkuttis et al, 2017).

Based on such a cursory investigation, offense seems dominant, according to the tenets of the classical offense-defense theory, for plenteous reasons, of which, the constant progress in offensive capabilities over defensive ones and the increasing defensive vulnerabilities, (Locatelli, 2013), including the confidentiality, integrity and availability of data, (Abel Moneim, 2018) as well as the asymmetric nature of cyberwarfare, (Lindsay, 2013). Resembling to nuclear ambiguity, constructive ambiguity is a chief essence of cyber warfare, (Al-Daweek, 2018). But with the massive production of AI applications, such a conclusion needs further investigation since AI could sharpen or mitigate the cybersecurity dilemma which refers to the use of offensive, defensive or commingled cyber tools by states amid the absence of shared cyber norms, (Hennessey, 2017). By the same token, the weaponization of Big Data and the usage of off-the-shelf technology also tighten the cybersecurity dilemma since governments have opportunity to create databases of every single member in the opponents' militaries, (Layton, 2018).

The utilization of AI capabilities in the cyber realm has two poles. AI with its predictability and automation, could mitigate the cyber security dilemma and could enhance cyber defense by addressing underlying challenges and vulnerabilities in the cyber ecosystem. Thus, AI capabilities would enhance the effectiveness of the Integrated Security Approach" (ISA); a holistic approach encompasses early-warnings; the selection and the adoption of the most adequate countermeasures to deter possible cyberattacks; the detection of potential attacks in case of failing to prevent a cyberattack, and adequate responses, (Wirkuttis et al, 2017). AI, with its offensive and defensive capabilities, has exhibited its ability to enhance cybersecurity by pinpointing and patching inherent vulnerabilities in cyber defense systems, while probing, manipulating and spoofing those of adversaries, (King et al, 2018), as well as detecting software bugs and performing responsive and defensive actions such as self-patching, thereby deterring cyberattacks at early stages, (Artificial Intelligence (AI) Enabled Cyber Defense).

While, the negative pole of AI is exemplified in a new bunch of AI applications capable of evading cyber defense systems and remaining dormant till detecting their targets, such as the Stuxnet, (Menn, 2018), as well as masking the identity of a malware after observing and figuring out how adversarial defense systems detect malware and malicious codes and what they are detecting, (Goosen et al, 2018). Also, data diet and algorithms biasness are archetypically the Cassandra of misbehaving algorithms, (Osoba et al, 2017). Heavy reliance on robots and technology increase warriors' vulnerability to information attacks by spoofing, denial-of-service, eavesdropping and exploitation, (Kott et al, 2015).

Ostensibly, the use of AI in the cyber realm is a double-edged sword. It enhances cyber security and cyber deterrence, at the meantime it intensifies cyber proliferation. The proliferation of advanced cyber capabilities could serve a state's strategic purposes through coercion, and could be useful for employing brute force which helps a state to achieve its purposes at the tactical level through kinetic or non-kinetic cyberattacks, (Liff, 2012). Cyberattacks allow states to extract meaningful concessions from adversaries, undermining their abilities to retaliate or defend themselves with conventional or cyber capabilities, (Liff, 2012). On the backdrop of this vignette, the threat of cyberwarfare, coupled with AI capabilities, could be an expedient deterrent tool and a practical brute force measure against superior adversaries that possess highly advanced conventional weapons, (Liff, 2012). Also, AI proliferation deter states in the cyber plane since every single application has its counter application, (Rashad, 2019) thereby making it useless.

AI, despite boosting cyber defense, states employs AI with varying degrees for cyber deterrence, (Rashad, 2019). Hence, the aim of preemptive cyber deterrence, in certain cases, is demonstrating the ability to disrupt or penetrate security systems rather than inflicting complete destruction, and having access to sensitive data, (Rashad, 2019). This illustrates that cyber deterrence is usually based on calculus.

b) Nuclear Capabilities:

The rapid advancement in AI raises a question over the survivability and the resilience of nuclear systems; the ability to resist or circumvent attacks and the aptitude to penetrate defenses of nuclear arsenals, (Payne et al, 2017) The mundane marriage between AI and nuclear weapons coupled with full autonomy and the absence of human from nuclear decision-making is two-folded. It might upend the subtle strategic balance among nuclear states, (Groll, 2018), triggering catastrophic repercussions and cascading tensions between nuclear states on one hand and a non-nuclear state and a nuclear one, on the other.

The cons of AI capabilities on nuclear deterrence involve the vulnerability of nuclear weapons to robust models of cyber-enabled attacks aimed at disrupting machine learning, thereby undermining their survivability, (Brown, 2018). Such a tragic flaw in the AI system, while providing opportunities for mitigating cyber vulnerabilities, could also undermine nuclear safety and reliability since nuclear weapons depend on real-time information exchange for targeting, (Unal, 2018). Further, full automation wherein humans are out-of-the-loop would definitely have knock-on implications on strategic stability, cascading escalatory acts and triggering arms race, (Unal, 2018). On account of automation and the digitalization of militaries, false assessments and responses by algorithms inserted in nuclear weapons systems, which could be labelled as machine error, could create operational hazards notably for digitally-independent states that would make wrongful decisions based on unreliable and inaccurate data, (Unal, 2018). Critically important, AI could make the "no first use" policy of less merit because of accidental errors, (Boulanin, 2018).

There are arrays of risks associated with the digitalization of nuclear command and control systems (C2) which include the possibility of disrupting means of communication, thus putting the reliability of data assessment on a shaky ground, (Unal, 2018). In a similar way, the Integrated Threat warning/Assessment structure which depends on a number of nodes, namely intelligence centers, the missile warning center, ground-and-space-based assets, could be irreliable since the means of communication could be compromised and manipulated, (Unal, 2018). "*AI could undermine system stability (C2 and early-warning)*", (Haggag, 2019). It could also "*undermine nuclear strategic stability because of its asymmetrical way*" by undermining its physical system that supports a nuclear command and control system, (Haggag, 2019).

A striking claim forestalls that AI could sharpen the nuclear second-strike capability. Despite its peculiarity, it could be true, according to an expert on general adversarial networks, when states resort to adversarial manipulation attacks for dissuading adversaries from tracking their nuclear arsenals, (Giest et al, 2018).

Correspondingly to the pros of AI to cyber deterrence, AI could tighten nuclear weapon systems by boosting detection capabilities, improving early-warning systems, empowering humans to carry out a precise cross-analysis of data, as well as protecting the nuclear command and control architecture, (Boulanin, 2018). In line with this, a group of participants in a workshop organized by RAND argued that AI might address underlying frailties in the nuclear arms control regime and might lay out novel foundations of arms control, (Giest et al, 2018).

Paradoxically, it could intensify arms race and could push nuclear states to modernize their nuclear arsenals due to escalatory acts by nuclear and non-nuclear states, (Boulanin, 2018).

C) Nuclear Versus Cyber:

From a technical perspective, cyber capabilities menace nuclear weapons since cyberattacks could agile nuclear command and control when it is unprotected and when cyber resilience is not effective, coupled with human error and fallibility. In addition, AI, through adversarial manipulation, could send false signals or transfer fake information to counter cyberattacks on nuclear facilities. Still, AI helps improve defense systems, including the nuclear ones.

The paramount argument saying that AI might undermine nuclear deterrence and trigger nuclear war needs to be revisited. On contrary, from a purely technical view, AI capabilities could overturn cyber vulnerabilities and mitigate their negative side effects on nuclear capabilities, if they are well-protected, (Al-Sayed, 2019) and highly advanced. Dr/General Mahmoud Khalaf; Nasser Academy Military advisor, asserted that AI has nothing to do with nuclear deterrence. Yet, he acknowledged the negative impacts of AI on nuclear command and control systems, counterforce and survivability from a technical angle. Thus, based on political realities, cyber capabilities have failed to revoke nuclear deterrence.

Nuclear vulnerabilities put states under a dilemma of pursuing cyber offense or cyber defense. Such a dilemma is a normal byproduct of the inherent uncertainty over the survivability and the reliability of nuclear systems that could be silently compromised and infiltrated through dormant and stealth campaigns. Therefore, a state may be incognizant of, in times of peace, the infiltration of its nuclear system for days, months or years which in turn deleteriously affects its military decision-making, deterrence policy, security doctrine, (Unal, 2018) and nuclear posture. In times of war, the situation is quite different given that it may result in information asymmetry, thereby triggering a retaliatory attack based on faulty calculations, (Unal, 2018).

In response, the emergence of AI technology could ameliorate such a dilemma by its detective and predicative capabilities. AI, as an assistive tool, could help leaders to make righteous decisions. Also, cyber intrusion, hacking of critical nuclear facilities and system failures are very common in nuclear weapons systems, (Unal, 2018). As a result, states would be dissuaded from using cyber capabilities due to the uncertainty over the degree of advancement as opposed to their adversaries. Correspondingly, attackers would be dissuaded from attacking adversarial nuclear arsenals. This illustrates, as Dr Waleed Rashad; assistant professor at the National Center for Social and Criminological Research, humans are critical in nuclear policy as they act as rational beings, (Rashad, 2019).

AI + Cyber + Nuclear:

To sum up, technical-wise, AI could undermine nuclear deterrence, especially when humans are absent, but politically speaking, the AI technology cannot overturn nuclear deterrence as long as there is a meaningful degree of human control.

Analysis:

Even with the lack of empirical evidences of the destructive potentials of AI military applications, defining AI is a requisite for investigating how it would reshape interstate relations and how it would alter the foundations of the international peace system.

The IR scholars and international lawyers have narrowly focused on the destructive potentials of AI and its autonomous potentials without defining its nature. They, regrettably, mixed up between AI as a technology, precisely as an enabler of a weapon, and AI as a weapon system per se. They mistakenly assumed that AI can serve *"as a state weapon*", (Haggag, 2019). In fact, "*AI is not a weapon*", (Erfan, 2019), but a technology that can be bolted into a weapon system and that "*can serve as an enabler for cyber and conventional weapons, as well as weapons of mass destruction*", (Haggag, 2019). It will be very problematic to categorize AI as a weapon given that equating AI with other weapon systems, such as conventional and nuclear weapons, would definitely direct the literature to exploring the impacts of AI per se, while disregarding the possible impacts of AI military applications on other weapon systems. Hypothetically speaking, if AI had been classified as a weapon, not a technology, states would have heavily relied on AI, with its highly destructive and disruptive potentials and its cost-benefit effects, for achieving military targets. Therefore, the weaponization of AI refers to "*the development in the uses of weapons*", (Khalaf, 2019).

Based on this definition, the weaponization of AI "*exacerbates the security dilemma because it can enhance the military capability of a state in symmetric relations or it can increase the military disparity between states not only in terms of new capability, but also in terms of attribution*" coupled with the potential of being weaponized by non-state actors, (Haggag, 2019). Apart from the legal aspect, the security dilemma does not only augment because of parity/disparity in capabilities, but also the uncertainty over "*the impacts of using AI on military's decision-making and the calculus of war*", (Erfan, 2019). Adding to this, uncertainty over AI applications' ability to counter-react and respond in the event of sudden attacks, regardless of being intentional or unintentional, is the core of the security dilemma in the age of technology.

There is no doubt that the use of AI for military purposes will dramatically change the calculus of war. Referring to the excessive reliance on drones in lieu of humans for reducing the number of causalities, Ambassador Aly Erfan sees that AI or any technological advancement "*would make the decision to go to war easy*", (Erfan, 2019). Though, such a view point is partially true at first glance, it omits that technological advancement could make causalities higher and could also make the outcomes graver. The use of nuclear weapons during the second World War in 1945 was a perfect example illustrating how technological advancement could be highly destructive and could trigger high death tolls. This tells destructive outcomes always dissuade states from rushing into war. And, the whole issue is not only about causalities, but also cost-benefit effects, interstate relations, legal considerations, state responsibility, military strength, degree of advancement in technology, geography, parity/disparity in capabilities, strategic climate, etc. More importantly, mutual vulnerability, indecisive victory, (Khalaf, 2019) escalatory acts and retaliation are also foundational in war calculus. The use of AI in militaries adds a new criterion to war calculus which is the utility of using AI as an enabler of a certain weapon system.

AI in air defense systems is one area to consider how AI could enhance or undermine the effectiveness of a weapons system. From a purely military perspective, commanders could assess how would AI allow them to employ air defense systems effectively and how would it allow them to maneuver and respond in a timely fashion. AI, for instance, minimizes the time needed for a response from 2 minutes, when humans are on-the-loop, to 10-20 seconds, when humans are no longer on the loop, (Khalaf, 2019). As Dr./General Mahmoud Khalaf said, the whole issue is about choosing and using the most adequate weapon for ensuring a speedy response. In other words, a states' commanders should know the type of the weapon used by the adversary, and should use the most appropriate weapon to respond within no time, (Khalaf, 2019). Therefore, a dichotomy does exist between the ability to identify and detect the advanced weapon used by an adversary, and the ability to respond effectively and in no time by using the appropriate means/weapons, (Khalaf, 2019). Therefore, AI and emerging technologies would not make the decision to war easy.

The absence of a threshold for incidents that could be seen as an act of aggression in the cyber domain, (Erfan, 2019) coupled with the AI's "*dual-use nature and the potential of weaponizing AI civilian applications*", (Haggag, 2019) further exacerbates the cyber security dilemma. The dual-use nature of AI could enable an adversary to manipulate a civilian AI application and change its nature, so as to be employed for military purposes. As a consequence, the AI security dilemma would be exacerbated since a state's commanders and soldiers should be aware of a weapon's capability and technology in order to be able to respond effectively. In that case, the problematic issue of attribution looms over since the defender might be unable to recognize the real nature of an AI application.

Since AI is typically a development in the use of technology in the military realm, one could say that it is two-folded given it could enhance both cyber defense and cyber deterrence, (Rashad, 2019) and could undermine the nuclear policy. Meanwhile, there is no a determinant proof. Hereafter, as Ambassador Erfan implied, the degree through which an AI application controls a weapon system is critical in a state's calculations.

Most of scholarly debate assumed the inapplicability of cyber deterrence for ample reasons: (i) cyber space is an open battlefield, thereby it does not exacerbate the security dilemma, (Al-Sayed, 2019); (ii) the absence of internet governance, (Erfan, 2019). It is true that the absence of internet governance and the difficulty of establishing attribution, notably when the attacker wants to keep his/her identity hidden, hinder the efficacy of cyber deterrence in its classical form. But, by enabling cyber defense systems with AI applications, cyber deterrence will be effectual. From a technical point of view, AI enhances cyber defense by detecting vulnerabilities in one's system and spoofing an adversary's system. From a political point of view, AI exacerbates the security dilemma because it is an advanced version of cyber capabilities, allowing states to mutually penetrate sensitive systems, such as military, intelligence and critical infrastructure, (Rashad, 2019); collect accurate data, and to "have a vivid picture of an adversary's capabilities", (Khalaf, 2019). It also reshapes the balance of power. In an effort to mitigate the cyber security dilemma, states employ AI, drones and robots in preemptive cyber deterrence, (Rashad, 2019) since the victim will unilaterally deter itself, (Lonsdale, 2017) in the event of a widespread disruptive cyberattack that could trigger civilian causalities.

Regarding nuclear capabilities, alarmists' view point, arguing that AI would threaten nuclear weapons and would undermine nuclear deterrence, dominates the IR literature. Theoretically speaking, this view looks awesome because "*the nuclear C2 can be violated by cyber capabilities since hackers can hack the typical system of air mines*", (Erfan, 2019). But, when it comes to nuclear deterrence, it needs further investigation. First of all, "*AI could be used as an enabler in terms of nuclear policy which includes: targeting, command and control, early-warning, potential battle damage assessment and the scenarios for establishing attribution*", (Haggag, 2019). Therefore, from a technical angle, AI can protect nuclear weapons since some of the AI applications are designed for early-warning and detecting any nuclear proliferation. Based on that, AI applications can assist humans and decision-makers, who use skills-based behaviors, in outlining the courses of action in a nuclear policy. In that case, nuclear deterrence will not be threatened provided that the nuclear command and control system is well-structured, well-protected and well-defended, (Erfan, 2019) as well as "defensive measures, including data encryption, are taken", (Al-Sayed, 2019). Hence, it is unexpected that AI would change the defensive nuclear doctrine to a "*preemptive*" one, as Ambassador Haggag claimed, as long as humans are over-the-loop. Such an argument could be valid only when fully autonomous applications are bolted into the nuclear weapon system and when humans maintain no control over machines.

Though, the AI technology cannot equate any of the known weapon systems, the devastating potentials of the AI technology can equate those of the nuclear weapons, (Erfan, 2019). Assuming that an AI application controls a nuclear weapon system, the scale of destruction will surpass the destructiveness of AI-enabled conventional weapons, (Erfan, 2019). This argument is convincing when humans are out-of-the-loop.

As politics speak louder than technicalities, the dichotomy between nuclear weapons safety and highly advanced cyber capabilities could somehow be mitigated by the use of AI early-warning applications and a meaningful humans' supervision.

Offense Versus Defense and the Efficacy of Deterrence in the AI realm:

Theoretically speaking, the malicious use of AI makes offense dominant in the cyber realm when state A has strong cyber defense systems as opposed to state B which has weak defense systems. There is no a unified position over the offensive/defensive nature of AI. Ambassador Erfan, for instance, maintained that AI deterrence could be feasible, though he implied the difficulty of determining whether offense or defense will be dominant, (Erfan, 2019). Ambassador Haggag, on the other hand, sees that "*establishing AI deterrence will be more difficult, if not impossible*", because he sees that deterrence is already difficult in nuclear weapons, (Haggag, 2019). Likewise, Dr. Dalal Al-Sayed argued that AI deterrence is impossible because of the openness of the cyber realm, (Al-Sayed, 2019). Ambassador Haggag's argument about the complex nature of deterrence shall be spotted-on given that deterrence is based on assumptions and hypothetical scenarios. Nonetheless, this does not necessarily mean that deterrence is impossible in other weapon systems and emerging technologies since deterrence is a policy/strategy through which states devise scenarios based on the strategic climate for enhancing their defense. The whole issue of deterrence is "*the political will to deter and having the ability to establish deterrence*", (Khalaf, 2019).

As per the foundations of cyber deterrence, defense would be dominant in the AI sphere, owing to its penetrative, manipulative and disruptive potentials, (Rashad, 2019). The ability to show muscles in the cyber/AI sphere and the ability to retaliate and respond in a timely manner make defense dominant. In some cases, states resort to the cyber sphere and weaponize the Internet of Things (IoT) just for signaling the vulnerability of adversarial cyber defense systems which in turn deter victims from launching offensive cyberattacks. This demonstrates that signaling cyber vulnerabilities is deterrent in and of itself, (Rashad, 2019).

AI MAD is Feasible: (Please see annex 3)

The 20th century Cold War provoked nuclear deterrence and Mutually Assured Destruction. By the same token, the 21th century Cold war and the intense AI race could make an AI MAD-like structure probable. However, such a supposition should not be taken for granted given that an AI MAD could be a workable strategy only when humans are having a degree of control over AI applications and when they participate in the decision-making process, especially at the strategic level.

Scenario One: Humans are out-of-the-loop:

Such a scenario is highly implausible in the foreseeable future, but it should be considered since AI warfare will be the next war due to its little cost and its potentiality to trigger few physical causalities, (Soliman, 2019). Accordingly, with the mechanization of war, this scenario could generate graver outcomes comparable to the second scenario, to be discussed later. Under this scenario, humans would have no control over machines and they would also relinquish their monopoly over the military decision-making process to machines and AI applications. Thus, as Mona Soliman noted, machines/robots and drones would have a powerful role as opposed to humans in future wars, (Soliman, 2019). And, human role would be confined to counting physical and human causalities, (Soliman, 2019). Furthermore, the scale of destruction could not be estimated and could not be mitigated or even controlled in case of wrongful attacks or miscalculations. There is no doubt that the use of fully autonomous AI applications with their high destructive capabilities and the irritability of C2 systems will definitely change the nature and "*the purpose of war in the cyber sphere from trying to influence an adversary's calculus to destroying it*", (Al-Sayed, 2019). Thus, offense would be dominant with the absence of the psychological factor. This further illustrates that a vicious circle of retaliatory attacks (first- and second-strikes) would be highly probable. Adding to this, the lack of accountability would further aggravate the situation amid the strict rejection of states to define a cyber threshold, (Erfan, 2019). Therefore, it would be hard to punish a machine or even preclude a state responsibility, which also means the failure of deterrence. The failure of deterrence and indecisive victory would be the logical outcomes since fully autonomous weapons would take-over other capabilities, causing severe destruction and disruption.

The difficulty of ensuring machines' compliance with international law and international legal norms, the impossibility of fathoming in advance the outcomes of machine-machine interactions, (Altmann et al, 2017) and the dilemma of attribution make deterrence more complex and spark crisis instability. Adding to such a gloomy scenario, an AI system could preserve itself should it suspected that its halt was imminent, and could retaliate by launching a nuclear strike, thereby undermining the doctrine of mutually assured destruction, (Klare, 2019). Also, the deployment of undersea drones might threaten the second-strike capability, (Klare, 2019).

Hypothetically, the only possible way to make machine-based deterrence effective under such an extreme scenario is the regular updates of data and occasional oversight by humans. Ergo, machines are not immune from miscalculations. Under such a very hypothetical and far-fetched scenario, where machines are in control of fire power and other weapon systems, ample forms of latent violence could be used as follows: (i) when AI has been bolted into a nuclear weapon or a WMD, deterrence by punishment or retaliation would have been effectual, (Erfan, 2019); (ii) when AI has been inserted into a cyber defense system, deterrence by disruption would have been effective; (iii) when AI has been used through a conventional weapon, deterrence by punishment would have been plausible. If such a scenario occurred, would states' leaders intervene at the end of the day? There is no a definite answer for such a question since we are unsure to what extent would machines be able to act like humans.

Scenario Two: Humans are over-the-loop:

Under this scenario, states would remain the main actor given such a highly advanced AI technology, especially those which are usually developed for military purposes, cannot be produced or even used by individuals and non-state actors, (Erfan, 2019). It is true that states would be the main actor under this scenario, but non-state actor, including companies and terrorist groups, and individuals could use and could produce AI applications with the technique of addictive manufacturing, as well as they could "*weaponize*" AI applications, (Haggag, 2019). Further, the potential of eclipsing humans' role would be far-fetched, (Al-Sayed, 2019) since the decision to go to war would be under the discretion of humans. In the context of human-machine teaming, AI applications would be active at the tactical and operational levels and humans would be responsible for strategic decision-making.

From a military perspective, AI deterrence is feasible, with or without the possession of nuclear weapons, given that states' leaders will be reluctant to launch a first-strike because of the fear of unknown. Besides, the ever-intensifying AI race in the commercial and military spheres aggravates the inherent dilemma of keeping up pace by possessing the most advanced AI applications to deter and penetrate adversaries, and the ability to develop national AI applications. In other words, each state should possess the most advanced AI applications vis-a-vis its adversary, (Khalaf, 2019). Unlike other conventional and unconventional military capabilities, AI applications should be domestically developed, thereby enhancing states' power and influence, (Khalaf, 2019). AI warfare is a sort of information warfare whereby triumph always goes to the one who possesses more data and information, (Khalaf, 2019). AI warfare is new form of struggle wherewithal competing parties seek to "*destroy data*", (Khalaf, 2019) to paralyze each other and to undermine their choices to respond. However, this reflects the inherent dilemma in AI-enabled warfare which requires possessing and collecting more data without being detected to avoid retaliatory acts that could take place to collect massive data in return, (Khalaf, 2019). In the event of reciprocated penetration and manipulation of data, victory will be indecisive due to data neutralization, (Khalaf, 2019). This tells that data neutralization, coupled with the weaponization of Big Data, triggers neutralization at the battlefield inasmuch as military commanders are uncertain about the reliability of their weapon systems and are also unsure of weapons capabilities. This further implies that data neutralization can also pave the way for weapons neutralization. Therefore, weapons neutralization can pose a problem at the operational and tactical levels given that the defender should "respond effectively and in a timely fashion, as well as should choose the most appropriate weapon to respond", (Khalaf, 2019).

Because of neutralization and mutual vulnerability, the weaponization of AI could create deterrence and could maintain strategic stability in symmetric struggles. In asymmetric conflicts which are usually associated with crisis instability, AI deterrence could also be viable since cyber force and conventional military force are not alike, (Rashad, 2019). Thus, by separating cyber force from other sorts of military force, "*AI could make preemptive deterrence and defense more effective*", (Erfan, 2019). However, such a classification should not disregard the efficacy of other sorts of force. One could argue that the efficacy of cyber force could equate and could go hand in hand with conventional force. It is illogical to confine asymmetric calculus to the cyber sphere since states are rational. The purpose of deterrence in asymmetric struggles is usually demonstrating the ability to attack or retaliate without inflicting massive destruction. Coupled with traditional war calculus, a superior state could adopt AI preemptive deterrence to dissuade an adversary from using AI and such-like capabilities maliciously, whereas, a weak state could adopt cyber deterrence and develop more cyber capabilities to demonstrate its ability to attack a superior state. To that end, the defensive doctrine would be complemented with preemption. Asymmetric deterrence resembles the cat and mouse game where neither the cat nor the mouse would be able to claim victory.

In short, cost neutralization pushes states to think twice. Accordingly, AI mutually assured destruction-like structure is feasible since "*the purpose is not destruction, but gaining a political benefit by making the costs of offense very high and intolerable*", (Khalaf, 2019).

In the context of symmetric and asymmetric conflicts, the defensive doctrine would be dominant in the AI realm as long as humans could reduce uncertainty and they, more or less, could open channels of communication to avoid grave destruction of spared cities and avoid the total disruption of cyber systems and AI-enabled machines.

By applying this to other weapon systems which can be enabled by AI capabilities, AI could maintain a second-strike capability amid the growing uncertainty over the collateral damage that might be triggered by the uncontrollable use of nuclear and conventional weapons. This further illustrates that AI deterrence would be successful since states' leaders are usually driven by security-seeking interests. This also implies that the foreseen AI MAD structure would go in parallel with nuclear MAD, thereby a defensive doctrine would be adopted.

To ensure a successful AI deterrence, states should use latent violence and credible threats to compel and deter adversaries from doing unwanted actions. As Ambassadors Erfan and Haggag argued, attribution and accountability are foundational in deterrence, (Erfan, Haggag, 2019). Like nuclear and cyber deterrence, AI deterrence per se could entail "*deterrence by punishment*" through the execution of an AI retaliatory attack, and "*deterrence by denial*" by the development of more AI capabilities. On contrary to other deterrence postures, the deterring state could invoke credible threats by "*the threat of disruption to a state's political, fiscal, power, weapon, financial, electoral systems*," (Haggag, 2019). As Ambassador Haggag noted, AI can disrupt thing of value for punishment or it can deny the use of AI applications and other capabilities", (Haggag, 2019). Based on that, nuclear weapons are not the sole agent of destruction, as Ambassador Haggag claimed, since the disruption of critical infrastructure could result in complete destruction. AI deterrence could also include the threat of mass manipulation or penetration, thereby paralyzing and neutralizing critical systems, especially the weapon systems. Thus, AI deterrence could be a standalone policy.

However, there are possible scenarios for using latent violence based the type of weapons and capabilities possessed besides to AI, as follows: (i) when two nuclear states possess AI capabilities, deterrence by punishment will be employed not only because of the possession of nuclear weapons, but also the parity in AI capabilities; (ii) when a nuclear state and a non-nuclear state possess AI capabilities, deterrence by preempt ion and denial will be effective; (iii)when two non-nuclear states possess AI and cyber capabilities, deterrence by preemption and denial will be used.

Conclusion:

Based on the foregoing, AI, as a weapon enabler, tightens the security dilemma between states in symmetric and asymmetric conflicts. After the in-depth investigation, deterrence could be effective and a MAD like structure is probable in the AI realm because of neutralization and mutual vulnerability. Notwithstanding, there is no a 100 percent guarantee that leaders won't miscalculate situations amid the growing uncertainty and their great reliance on machines that can be manipulated or neutralized when AI and cyber defense systems are not shielded or amateur. So, human-machine teaming is essential for having a successful deterrence and minimizing errors as much as possible. As Dr/General Khalaf suggested that human intervention would be needed, should a technical error or an intentional error happened. In that regard, he referred to a well-known western saying "*Don't trust too much in technology*." He envisions that as long as AI applications are updated and are scrutinized by humans, on a regular basis, besides to military simulations, wrong war decisions and miscalculations won't take place, (Khalaf, 2019).

To conclude, the second scenario is the most possible scenario since deterrence requires the psychological factor along with rational thinking in war calculus which entails military, political and economic aspects. To that end, states will never relinquish its monopoly over fire power to machines or AI applications.

On the backdrop, the anticipated AI MAD, which could be coined as "*Mutually Assured Manipulation*", could operate in parallel with nuclear MAD. Also, AI MAD could embolden nuclear MAD when humans are over-the-loop.

Finally, further research should be made to tackle the implications of AI on the relations between state actors and non-state actors and such asymmetric struggles which cannot be mitigated amid crisis instability. It is also suggested to do further research on how the weaponization of outer space, coupled with the possession of AI capabilities, would threaten deterrence. By the same token, further research should be done to investigate how AI could shuffle the foundations of international peace and security, such as the concept of collective security.

Policy Implications:

AI vertical proliferation and hasty AI race instigate instability, thereby exacerbating the security dilemma and increasing military expenditure with the aim of catching up capabilities and ensuring arms race stability, (Altmann et al, 2017). AI race has been augmented for maintaining strategic stability and for preventing the adversary from being ahead. However, the proliferation of AI should be regulated for maintaining arms race stability which requires the planned deployments of arms in terms of scope and pace, (Altmann et al, 2017). Maintaining strategic stability rests on ensuring the planned development and proliferation of such novel asymmetric capabilities in age of information and economic warfare.

Though AI exacerbates the security dilemma and accelerates proliferation, AI provides a potential for confidence-building through the formation of a regime for arms control and the promotion of disarmament, (Haggag, 2019). Such an anticipated regime could pave the way for regulating the unplanned deployment of such novel technologies and AI which in turn spark crisis instability and stimulate arms race, (Altmann et al, 2017).

Unsurprisingly, such a fierce commercial competition has been defused to the military sphere, rendering the development of AI applications that meet the requirements of the military uses (Altmann et al, 2017). Such a paradigm shift in the rapid proliferation of autonomous weapons systems (AWS) and AI applications, which do not require Herculean efforts or exotic materials as opposed to nuclear and conventional weapons, demonstrates the urgency of regulating the uses of AI and AWS in the context of the ongoing information warfare, and also indicates the necessity of controlling AI race in the context of the current economic warfare.

Since Big Data and the Internet of Things have been weaponized, the suggested regime should put limitations on the weaponization of Big Data which threatens not only states, but also institutions and individuals, (Rashad, 2019). Also, the 3D printing or Addictive Manufacturing (AM) technology that allows second-, third-tier states and non-state actors to develop AI or AWS raises the alarm over the possible irrational use of AI by non-state actors or individuals. This means that any AI arms control regime should take all necessary measures and steps to ensure the inaccessibility of both the 3D printing technology and AI applications to non-state actors.

Since AI race has evolved in the context of economic rivalries and economic warfare before being diluted to the military sphere, state actors will no longer have a monopoly over the ongoing AI race given that the private sector has become a part of the game. This also means that establishing a regime for regulating the uses of AI and controlling its race requires the incorporation of multi-stakeholders, including the private companies which are implicitly competing with state actors and are thriving for promoting human security. This mirrors the clash between maintaining strategic stability and a state's national security on one hand, and promoting human security and gaining profit on the other. Such ever-intensifying commercial competition illustrates the underlying dilemma between promoting free-market economy and maintaining strategic stability, implying the impossibility of regulating AI, (Khalaf, 2019). Dr/General Khalaf was absolutely right when he articulated that regulating competition is impossible from an economic point of view, but that does not necessarily mean that regulating AI uses in the commercial, cyber and military spheres is improbable too, otherwise militaries will always be under the threat of being neutralized since private companies have the know-how of such applications and are aware of their inherent vulnerabilities. The suggested regime could settle this by promoting the sense of ownership among stakeholders.

Besides to the arms control point, the nexus between maintaining a meaningful human control and eclipsing humans control could trigger states to alter their military doctrines and policies. As per the Bob Work; the US deputy Secretary of Defense, the full delegation of authority to AI and algorithms is highly improbable except for the cyber realm, (Altmann et al, 2017). However, such an option could not be sustained, should an adversarial state signaled it willingness to delegate more authority to AI-enabled machines, ((Altmann et al, 2017). Consequently, the AI race could be protracted to the extent of triggering collateral damage. Though, such a signaling to delegate military's decision-making to fully autonomous applications deemed improbable, it is worrisome since AI and AWS cannot act in conformity with the principles and foundations of international law and the international legal norms, particularly the International Humanitarian Law and the Law of Armed Conflicts, as well as they could increase the incidences of speedy and mechanized wars that cannot be fathomed or controlled. The mere thinking of a swarm combat triggers crisis instability since the assumption of high chances of war will takeover, (Altmann et al, 2017). It further increases the likelihoods of escalation, as Paul Scharre implied, there is no a guarantee for winning a swarm war, unless well-programmed algorithms are developed and are used, otherwise the outcomes will be disastrous because of timely counterattacks, (Altmann et al, 2017). If machines have been delegated to make war decisions, there would have been no chances for practicing restraints or double-checking, (Altmann et al, 2017). The suggested regime, coupled with international legal instruments, could address this point by ushering for a meaningful human control.

The intractability of such a kind of technology makes attribution difficult and problematic. The inherent difficulty of establishing attribution rests on the inability to know the attributor since the attributor could be a state, non-state actor or even a "*third party who has interest in the outcomes of any potential crisis, confrontation with the use of a certain weapon system*". The only possible way for establishing attribution, apart from those suggestions focusing on the legal perspective, is human intelligence by which humans can collect data and process them according to the strategic climate, (Khalaf, 2019). With the establishment of an arms control regime, the issue of attribution could be resolved by the development of legally binding instruments, and the development of political, security and economic frameworks.

Surely, AI arms control does not only pave the way for creating a regime that would maintain strategic stability within the AI sphere, but also preventing the fall of the AI technology in the wrong hands by laying out parameters for AI production and AI arms trade without hindering competition.

Policy Recommendations:

There is no doubt that the AI technology, similarly to nuclear capabilities, has been weaponized. Therefore, the stealthy potentials of AI could pose high security concerns that might reshuffle the world order and might make the parameters of international peace and security at a shaky ground. In the era of globalization, the weaponization of AI, without being regulated, would definitely add further hurdles to strategic stability.

Much as, there is no empirical evidences of destruction triggered by the use of AI in the military domain, the international community should not wait till an AI Pearl Harbor, AI Hiroshima and Nagasaki or such-like incidents take place. Is the history repeating itself? There is no a unified stance on how to manage and regulate the uses of AI for civilian and particularly military purposes amid the new Cold War.

There are two possible scenarios for regulating AI. Each of those scenarios has its own parameters and regulatory agenda:

(I) AI is not a weapon, but a technology that can alter a weapon's technology, (Erfan, 2019) and can be integrated into numerous military systems, (Concluding Report, 2018). Thus, the supposition of drafting additional protocol to the Convention on Conventional Weapons for banning AI seems irrelevant. In such a scenario, it is worthy of consideration to see how the foreseen AI arms control regime would shape the nuclear and cyber arms control regimes. More important, the issue of accountability and state responsibility should be considered in that regard;

(II) AI and such kinds of lethal autonomous weapons systems (LAWS), such as submarines drones, are coined as weapons. In that regard, they should be prohibited, (Geist, 2016). In that case, an additional protocol to the Conventional on Conventional Weapons should be drafted for banning AI and LAWS.

The first policy option is the doable one. Therefore, the international community should take the following measures to incrementally formulate a multilateral regime for regulating AI, as it was the case with nuclear weapons:

1) **National AI and Cyber Policies:** According to the "routine activity" theory which articulates that individuals, institutions and states unilaterally deter themselves/itself when the threats associated with technological advancement are growing, (Rashad, 2019), states should draft national laws for regulating the AI and cyber activities based on the degree of advancement and the degree of dependence on technology, (Rashad, 2019).
2) **Drafting Bilateral Agreements:** Resembling to nuclear weapons, states are recommended to sign such-like START agreements for managing the uses of AI applications; defining a threshold for cyber and AI attacks, and information and technology sharing, as well as strengthening cyber and AI defensive measures at the bilateral level.

   Such bilateral agreements could open the room for the evolvement of a legal norm.
3) **Super-soft Law for AI:** Similar to nuclear restraint, AI restraint could pave the way for managing, regulating or containing the development of AI for military uses, (Maas, 2019). Such a bottom-up law-making approach, which necessitates the incorporation all actors and stakeholders (INGOs, scientists, academia, security experts, developers and individuals), could come out with non-binding speculative rules and regulations, (Burri, 2017). However, such non-binding speculative rules and regulations could be the stepping stone for legally binding rules. They could also set redlines for AI-enabled attacks, such as AI-enabled nuclear strikes, thus establishing an AI taboo.
4) **Promoting AI Arms Control Rather Than Non-Proliferating It:** We cannot reverse or ban the AI technology and lethal autonomous weapons systems, (LAWS) as Schultz articulated "*Proliferation begets proliferation*",

(Maas, 2019). Since the AI technology is not unlawful but its malicious uses, (Cavelty, et al, 2017) all we can do is regulating its uses and circumscribing its lethality through the drafting of a multilateral agreement. Reaching an agreement regulating the uses of AI, more or less, illustrates states' acceptance to regulate AI and its uses inasmuch as they will hold a monopoly over the use of AI for military purposes, (Erfan, 2019).

Mindful that, vague legal terms, such as the term "*control*" could be interpreted differently and loosely by states based on their preferences, (Burri, 2017). This illustrates that tight and precise legal terms should be used.

Further, preventive prohibition seems convincible since it would neither prohibit the technology itself nor add restrictions on quantitative proliferation of AI applications, but the prohibition of certain military practices, (Altmann et al, 2017). Thus, a legally-binding multilateral agreement, comprehensively outlawing certain uses of AI, is highly recommended in that regard.

5) **Drafting a Multilateral Agreement for Regulating AI:** Such an agreement shall be drafted based on the foreseeable AI norms and in conformity with international legal instruments. In addition, it shall include clause(s) on:

   a) **Meaningful Degree of Humans' Control and Keeping Humans Over-the-loop:** Based on the foregoing analysis, a degree of a human control over a machine is essential for commanding and controlling the course of war, otherwise the outcomes will be disastrous. Humans can act as operators, (Autonomous Weapons & Human Control, 2016) under the context of human-machine teaming, so as to manage the course of war at the operational, tactical and strategic levels. They can also be moral agents by weighting the degree of collateral damage that might be trigged by the excessive or inadequate use of force, (Autonomous Weapons & Human Control, 2016). The whole issue is not only about maintaining a meaningful degree of human control, but also making human control on par with and in conformity with the principles of military necessity, proportionality, distinction, etc, and addressing the issues of controllability, moral responsibility and accountability, (Horowitz et al, 2015). To ensure a meaningful human control, it essential to meet three core requirements: (i) making informed decisions about the usage of weapons, (ii) having sufficient information and maintaining a situational awareness of the course of war, so as to ensure the legality of actions, and (iii) training humans on how to control and use weapons effectively after being tested, (Horowitz et al, 2015). Adding to this, conducting regular updates of AI applications, (Rashad, Khalaf, 2019) is a pre-requisite for maintaining a meaningful degree of human control. The suggested clause(s) should also stipulate for defining "*a meaningful control*" as: "control by design" by which the operator has the ability to monitor information about the context and system, and "control in use" through which the operator monitors the operational environment and the system to ensure compliance with IHL, (Concluding Report, 2018).

b) **The Uses of AI:** Resembling to nuclear weapons, we cannot stop or reverse the development of AI. Then, AI should be regulated and humans should be hold accountable in the AI domain, (Erfan, 2019). By regulating AI, it means the regulation of its uses and regulating the conducts of states, individuals, companies and the international community in the AI sphere, (Erfan, 2019). Lucas argued that the use of LAWS in uninhabited areas and against unmanned targets makes it lawful, (Cavelty, 2017). Needless to say, AI regulations should entail the prohibition of certain applications and the permitting of others, (Erfan, 2019). Further, AI regulations should outline what humans can do and what they cannot do in the AI domain.

Since states, according to Article 36 of the Additional Protocol of the 1977 Geneva Convention, are obliged to determine whether a certain use of a weapon be seen as a violation by international law or not, (Cavelty, 2017), it is highly suggested to add a clause stressing on that obligation. To this end, the suggested clause should require every state to take the following into consideration: (i) the characteristics of a weapon and its technology, (ii) the context in which LAWS are used i.e: remote or populated areas, (Lewis, 2013), (iii) the military targets, (iv) the level and degree of residual human control over the LAWS, (Cavelty, et al, 2017).

c) **Accountability and Moral Responsibility:** In the event of malfunction, hacking, miscalculations or inadequate use of force in violation of IHL and the Law of Armed Conflicts, the issues of accountability and liability loom over given that it is hard to hold machines liable and it will be unfair to inflict liability upon commanders or programmers in that case, (Fournier, 2018). It will also be impossible to hold a manufacturer accountable given he/she is not a subject of the International Criminal Law which only prosecutes individuals, particularly states' leaders. Adding to the further muddied situation, states cannot be prosecuted according to the "doctrine of sovereign immunity" even it has been proved that states were responsible for using autonomous weapon systems, (Fournier, 2018). Because of sovereign immunity, certain states have extended sovereignty to manufacturer, (Fournier, 2018), thereby prosecuting manufacturers will be almost impossible. Thus, the international community should not afford machines to make war decisions without holding someone accountable, (Erfan, Haggag, Rashad, 2019). This illustrates that when humans are over-the-loop, perpetrators and programmers should be held accountable according to international law and a state responsibility shall be claimed.

d) **AI/Cyber Red Lines:** All stakeholders should develop a threshold, outlining and defining what constitutes an offensive/defensive AI-enabled attack in the cyber plane, (Rashad, 2019). For instance, AI attacks conducted by fully-autonomous applications should be regarded as offensive.

e) **AI as a Technology of Mass Destruction:** It is intriguing to classify the malicious AI technology as a Technology of Mass Destruction.

f) **The Protection of Critical Infrastructure and Non-Combatants:** Amid the intense inclination to weaponize AI, coupled with the absence of internet governance, a clause for protecting noncombatants in cyberspace should be taken as a priority over other issues, (Guay et al, 2017).

6) **Establishing an IAEA-like Agency for AI Arms Control:** "*It is possible to create an arms control regime by the establishment of an international authority for regulating the usage of AI in the military realm*", (Al-Sayed, 2019). It is highly suggested to establish a supranational agency, referred to as

the "*International Agency for Regulating AI and Newly Emerging Technologies*". The objectives of this Agency are: regulating the uses of AI and curbing its malicious uses; ensuring a state's compliance with AI peaceful safeguards; slowing down AI proliferation. The competences of the Agency include: overseeing the development of AI applications for military purposes through the deployment of inspection missions, on a regular basis; ensuring a state's compliance with international AI safeguards and verification methods, as well as encouraging and overseeing AI research and development in member states. Further, the Agency, with the help of its technical staff, is responsible for providing technical assistance and submitting technical recommendations/reports to the UNSC, UNGA and the UN Office of Disarmament Affairs. Furthermore, the Agency should cooperate with any OIs to be created in the future or other like-minded IOs, which are responsible ensuring nuclear safeguards and verifications, and promoting cyber safety and security.

More important, it shall refer/file a case, when the pace of AI development/race endangers international peace and security, to the UN General Assembly or the UN Security Council.

The organizational structure of the anticipated *Agency* shall be composed of:

a. The *General Forum*; an international forum for discussing technicalities and security implications of AI and emerging technologies. Each member either a state, IO, INGO, academia, developer, technician or private company has one vote. This Forum shall submit its recommendations and suggestions, including multilateral agreements, to the *Supreme Council;*

b. The *Supreme Council* which shall be composed of 20 member-states and 5 miscellaneous members representing the academia, private sector and competent IOs/INGOs, with equitable representation. Its resolutions are binding. Those 20 members shall be elected every two years.

   The competences of the Council shall include, inter alia,

   I. Discussing substantial matters;

   II. Determining if a certain act or step threatens international peace and security. Should an action be proven to be a severe violation of international legal instruments, the *Council* shall refer the issue/case to the UN Security Council or competent IOs;

III. Taking all measures, including, but not limited to, punitive measures, should a member state violated the Charter, international legal instruments regulating AI and other emerging technologies, or have shown non-compliance with the Agency Safeguards;

IV. Cooperating with other IOs and INGOs, to mention but few, the International Atomic energy Agency and the International Telecommunication Union, for discussing and coming out with solutions for any issue that threatens international peace and security;

V. Sponsoring bilateral agreements for AI software control.

c. *The Research and Development (R&D) Department:* This Department shall be a global hub for R&D in AI and other emerging technologies. It shall coordinate and compile all research and endeavors; call for further research; submit reports/compiled recommendations to *the General Forum;*

d. Technical Assistance Task Force and Inspection Missions: This body shall provide technical assistance, if deems necessary or upon a state's request, to ensure a state's compliance with the Agency Safeguards. The Task Force shall be primarily composed of inspectors from the *Agency*. Also, inspectors from like-minded IOs or Agencies, namely the IAEA, can participate in the inspection missions, on a voluntarily basis;

e. M&E mechanisms, AI safeguards and Verifications: It shall ensure members' full compliance with the Agency Safeguards and Verification Measures. It shall also develop new safeguards and verifications, when it is deemed necessary.

Corresponding to nuclear safeguards, of which nuclear material and facilities cannot be upgraded to a weapon-grade and are not used for military purposes, (Safeguards Agreements), AI safeguards are recommended for verifying the peaceful applications of AI and ensuring a state's compliance with the foresseable internationally-recognized AI threshold. The AI Safeguards could include regular weapons and data reviews; regular updates for AI applications; AI applications are not upgraded to a weapon-grade; a meaningful human control in the military sphere; the disaggregation of civilian and military AI applications;

f. *Department for Promoting the Rational Use of Weapons*: This *Department* shall be composed of *sub-departments: nuclear, cyber and conventional.* It shall, in conjunction with the IAEA, ITU or state parties to the United Nations Convention on Conventional Weapons, ensure the proper usage of AI and other emerging technologies when they are bolted into other weapons. It shall also curb or mitigate the misuse of AI and other emerging technologies in the military realm.

g. *The Dispute Settlement Mechanism:* The Dispute Settlement Mechanism shall settle any dispute that may arise between member states or a member-state and a non-member state.

h. The *Attributive Mechanism: The Mechanism* shall provide advisory opinions on attributive measures and shall develop a framework for attribution and accountability by developing AI-enabled thresholds based on the type of weapons used or the degree of destruction.

i. *The Mitigation Mechanism:* The *Mechanism* shall assist states in remediating the unwanted impacts of wrongful use of AI application or unintentional error.

7) **Revising the Nuclear Arms Control Regime:** With the growing challenges of emerging technologies and AI, there is a need to revise the nuclear arms control regime and add clause(s) regulating the uses of AI in the nuclear domain.

A journey of a thousand miles begins with a single step, internet governance is seen as the stepping stone for AI regulations. Thereupon, a revolutionary paradigm-shift, incorporating technical, ethical, moral and political dimensions in the standardization process, (Burri, 2017) is a requisite for internet governance. Microsoft manager's suggestion of the formation of a neutral digital Switzerland is welcome since it will harness the private companies to be detached from developing offensive tech/applications; to combat state-sponsored cyberattacks, as well as establishing attribution for state-sponsored cyberattacks and taking necessary measures to remediate the repercussions of such large-scale attacks, (Smith, 2017).

Reference List/Works Cited:

1. Abdel Moneim, A. (2018, July). Information Security and National Security. *Al-Siyassa Al-Dawleya*, *53*(213), 202-207.

2. AI and the Military: Forever Altering Strategic Stability. (2019). Retrieved from https://www.tech4gs.org/

3. Al-Doweek, A. A. (2018, July). Cyber Deterrence Strategy. *Al-Siyassa Al-Dawleya*, *53*(213), 196-201.

4. Allen, G., & Chan, T. (2017). Artificial Intelligence and National Security. Retrieved from https://www.belfercenter.org/publication/artificial-intelligence-and-national-security

5. Allison, G. (2016). English Translation of the Official Strategy of the Israel ... Retrieved from https://www.belfercenter.org/sites/default/files/legacy/files/IDF doctrine translation - web final2.pdf

6. Altmann & Frank Sauer (2017) Autonomous Weapon Systems and Strategic Stability, Survival, 59:5, 117-142, DOI:10.1080/00396338.2017.1375263

7. Craig, A. (2018). Understanding the Proliferation of Cyber Capabilities. Retrieved from https://www.cfr.org/blog/understanding-proliferation-cyber-capabilities

8. Artificial Intelligence (AI) Enabled Cyber Defense. (n.d.). Retrieved from https://www.eda.europa.eu/webzine/issue14/cover-story/artificial-intelligence-(ai)-enabled-cyber-defence

9. Autonomous Weapons and Human Control. (2016). Retrieved from https://www.cnas.org/publications/reports/autonomous-weapons-and-human-control

10. Bates, S. J. (2017). *Artificial intelligence: A revolution waiting to happen*(Unpublished master's thesis). Thesis / Dissertation ETD. Retrieved 2018, from https://apps.dtic.mil/dtic/tr/fulltext/u2/1041675.pdf

11. BEREJIKIAN, J. D. (2002). A Cognitive Theory of Deterrence*. Retrieved from http://web.mit.edu/sabrevln/Public/GameTheory/Journal of Peace Research/A Cognitive Theory of Deterrence.pdf

12. Boulanin, V. (2018). AI & Global Governance: AI and Nuclear Weapons - Promise and Perils of AI for Nuclear Stability - Centre for Policy Research at United Nations University. Retrieved from https://cpr.unu.edu/ai-global-governance-ai-and-nuclear-weapons-promise-and-perils-of-ai-for-nuclear-stability.html

13. Bouskill, K., Chonde, S., & IV, W. W. (2018, May 01). Speed and Security: Promises, Perils, and Paradoxes of Accelerating Everything. Retrieved from https://www.rand.org/pubs/perspectives/PE274.html

14. Brundage, et al. (2018). The Malicious Use of Artificial Intelligence. Retrieved from https://maliciousaireport.com/

15. Burri, T. (2017). International Law and Artificial Intelligence. Retrieved from https://www.researchgate.net/publication/320938178_International_Law_and_Artificial_Intelligence

16. Buzan, B. (1984). The national security problem in international relations. International Affairs,60(2), 289-290. doi:10.2307/2619056

17. Cavelty, M. D., Fischer, S., & Balzacq, T. (2017). "Killer Robots" and Preventive Arms Control, in: The Routledge Handbook of Security Studies, 2nd Edition (Routledge Hardback 2016), pp. 457-468. Retrieved from https://www.academia.edu/28785811/_Killer_Robots_and_Preventive_Arms_Control_in_The_Routledge_Handbook_of_Security_Studies_2nd_Edition_Routledge_Hardback_2016_pp._457-468

18. China may match or beat America in AI. (2017, July 15). Retrieved from https://www.economist.com/business/2017/07/15/china-may-match-or-beat-america-in-ai

19. Concluding Report: Recommendations to the GGE. (2018). Retrieved from https://www.ipraw.org/recommendations/

20. Crosston, M. D. (2011). World gone cyber MAD: How "Mutually assured debilitation" is the best hope for cyber deterrence. Strategic Studies Quarterly, 5(1), 100-116.

21. Davis, P. K. (2014). Toward Theory for Dissuasion (or Deterrence) by Denial: Using Simple Cognitive Models of the Adversary to inform strategy. Retrieved from https://www.rand.org/content/dam/rand/pubs/working_papers/WR1000/WR1027/RAND_WR1027.pdf

22. De Spiegeleire, S., Maas, M., & Swejis, T. (n.d.). ARTIFICIAL INTELLIGENCE AND THE FUTURE OF DEFENSE. Retrieved from https://www.hcss.nl/sites/default/files/files/reports/Artificial Intelligence and the Future of Defense.pdf

23. Dilulio, J. J. (n.d.). Deterrence Theory. Retrieved from https://marisluste.files.wordpress.com/2010/11/deterrence-theory.pdf

24. Eckersley, et al, (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. Retrieved from https://www.eff.org/deeplinks/2018/02/malicious-use-artificial-intelligence-forecasting-prevention-and-mitigation

25. Ethics and Governance of AI. (n.d.). Retrieved from https://cyber.harvard.edu/topics/ethics-and-governance-ai

26. Etzioni, A., et al (2017). Pros and Cons of Autonomous Weapons Systems. Retrieved from https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/May-June-2017/Pros-and-Cons-of-Autonomous-Weapons-Systems

27. Fairbanks, C. H. (2004). MAD and US Strategy. In In Getting MAD: Nuclear Mutually Assured Destruction, Is Origin and Practice. Strategic Studies Institute. doi:http://ssi.armywarcollege.edu/pdffiles/pub585.pdf

28. Fournier, G. (2018, June). The liability issues related to the use of Lethal Autonomous Weapons Systems (LAWS). UN Special, (750), 12-13.

29. Fuhrmann, M. (2018). The Logic of Latent Nuclear Deterrence. Retrieved from http://www.iserp.columbia.edu/sites/default/files/Deterrence without Bombs 2018-0129.pdf

30. Garcia, D. (2018, May 29). Lethal Artificial Intelligence and Change: The Future of International Peace and Security | International Studies Review | Oxford Academic. Retrieved from https://academic.oup.com/isr/article/20/2/334/5018660

31. Geist, E., & Lohn, A. J. (2018, April 24). How Artificial Intelligence Might Affect the Risk of Nuclear War?. Retrieved from https://www.rand.org/blog/articles/2018/04/how-artificial-intelligence-could-increase-the-risk.html

32. Geist, E. M. (2016). It's already too late to stop the AI arms race-We must manage it instead. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/00963402.2016.1216672

33. George, A (2009). Coercive Diplomacy. In The Use of Force (7th ed., pp. 72-79). Rowman and Littlefield.

34. Glaser, C. L. (1997). The security dilemma revisited. World Politics, 50(1), 171-201. doi:10.1017/S0043887100014763

35. Goosen, R. et al. (2018). Artificial Intelligence Is a Threat to Cybersecurity. It's Also a Solution. Retrieved from https://www.bcg.com/publications/2018/artificial-intelligence-threat-cybersecurity-solution.aspx

36. Groll, E. (2018, April 24). How AI Could Destabilize Nuclear Deterrence. Retrieved from https://foreignpolicy.com/2018/04/24/how-ai-could-destabilize-nuclear-deterrence/

37. Guay, J., & Rudnick, L. (2017, August 02). What the Digital Geneva Convention means for the future of humanitarian action. Retrieved from https://www.unhcr.org/innovation/digital-geneva-convention-mean-future-humanitarian-action/

38. Hennessey, S. (2017, December 04). Deterring Cyberattacks. Retrieved from https://www.foreignaffairs.com/reviews/review-essay/2017-10-16/deterring-cyberattacks

39. Horowitz, M. C., & Allen, G. C. (2018). Artificial Intelligence and International Security. Retrieved from https://www.cnas.org/publications/reports/artificial-intelligence-and-international-security

40. Horowitz, M., & Scharre, P. (2015). Meaningful Human Control in Weapon Systems: A Primer. Retrieved from https://s3.amazonaws.com/files.cnas.org/documents/Ethical_Autonomy_Working_Paper_031315.pdf?mtime=20160906082316

41. Jervis, R.(2009). *The Use of Force* (7th ed.). Rowman and Littlefield.

42. Jon R. Lindsay (2013) Stuxnet and the Limits of Cyber Warfare, Security Studies, 22:3, 365-404, DOI: 10.1080/09636412.2013.816122

43. Johnston, T., Smith, T. D., & Irwin, J. L. (2018, May 08). Additive Manufacturing in 2040: Powerful Enabler, Disruptive Threat. Retrieved from https://www.rand.org/pubs/perspectives/PE283.html

44. Kent, R. (2015). Are We Ready for the Future of Warfare? Retrieved from https://blogs.scientificamerican.com/observations/are-we-ready-for-the-future-of-warfare/

45. King, T. C., & Aggarwal, N. (2018). Artificial Intelligence Crime: An Interdisciplinary ... Retrieved from https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3183238_code2792915.pdf?abstractid=3183238&mirid=1

46. Klare, M. (2019). Arms Control Today. Retrieved from https://www.armscontrol.org/act/2019-03/features/autonomous-weapons-systems-laws-war

47. Kott, A., Alberts, D. S., & Wang, C. (2015). Will Cybersecurity Dictate the Outcome of Future Wars ... Retrieved from https://www.researchgate.net/publication/288857290_Will_Cybersecurity_Dictate_the_Outcome_of_Future_Wars

48. Kreipinevich, A. F. (1994). Cavalry to computer; the pattern of military revolutions. Retrieved from http://users.clas.ufl.edu/zselden/CourseReadings/Krepinevitch.pdf

49. Kreps, S. E., & Kaag, J. (2012). The Use of Unmanned Aerial Vehicles in Contemporary Conflict: A Legal and Ethical Analysis. *SSRN Electronic Journal*. doi:10.2139/ssrn.2023202

50. Layton, P. (2018). Algorithmic Warfare: Applying Artificial Intelligence to Warfighting. Retrieved from http://www.academia.edu/36620913/Algorithmic_Warfare_Applying_Artificial_Intelligence_to_Warfighting

51. Lebow, R. N., & Stein, J. G. (1995). Deterrence and the cold war. Political Science Quarterly, 110(2), 157-181. doi:10.2307/2152358

52. Lewis, J. (2013). The Case for Regulating Fully Autonomous Weapons. Retrieved from https://www.yalelawjournal.org/comment/the-case-for-regulating-fully-autonomous-weapons

53. Lieber, K. A. (2017). Grasping the Keir A. Lieber Technological Peace. Retrieved from http://web.stanford.edu/class/polisci211z/2.3/Lieber IS 2000.pdf

54. Liff A. (2012) Cyberwar: A New ‘Absolute Weapon’? The Proliferation of Cyberwarfare Capabilities and Interstate War, Journal of Strategic

Studies, 35:3, 401-428, DOI: 10.1080/01402390.2012.663252

55. Locatelli, A. (2013). THE OFFENSE/DEFENSE BALANCE IN CYBERSPACE. Retrieved from https://www.ispionline.it/sites/default/files/pubblicazioni/analysis_203_2013.pdf

56. Lonsdale, D. J. (2017). Warfighting for Cyber Deterrence: A Strategic and Moral ... Retrieved from https://link.springer.com/content/pdf/10.1007/s13347-017-0252-8.pdf

57. Maas (2019): How viable is international arms control for military artificial intelligence? Three lessons from nuclear weapons, Contemporary Security Policy, DOI: 10.1080/13523260.2019.1576464

58. Menn, J. (2018). New AI Programs Can Learn How to Beat Best Cyber Defenses. Retrieved from https://www.insurancejournal.com/news/national/2018/08/10/497443.htm

59. Meserole, C. (2018). Artificial Intelligence and the Security Dilemma - Lawfare. Retrieved from https://www.lawfareblog.com/artificial-intelligence-and-security-dilemma

60. Mohan, C. R. (1986). The tragedy of nuclear deterrence. Social Scientist, 14(4), 3-19. doi:10.2307/351717

61. Nusca, A. (2011, March 21). Japan developing rockets with artificial intelligence. Retrieved from https://www.zdnet.com/article/japan-developing-rockets-with-artificial-intelligence/

62. Open Letter on Autonomous Weapons. (2015). Retrieved from https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1

63. Osoba, O., & IV, W. (2017). The Risks of AI to Security and the Future of Work. Retrieved from https://www.rand.org/pubs/perspectives/PE237.html

64. Williams, P., & McDonald, M. (2008). *Security studies: An introduction*. London: Routledge, Taylor & Francis Group.

65. Payne, K. et al. (2017). A New Nuclear Review for a New Age - nipp.org. Retrieved from http://www.nipp.org/wp-content/uploads/2017/06/A-New-Nuclear-Review-final.pdf

66. Polyakova, A. (2018, April 24). Weapons of the weak: Russia and AI-driven asymmetric warfare. Retrieved from

https://www.brookings.edu/research/weapons-of-the-weak-russia-and-ai-driven-asymmetric-warfare/

67. QUACKENBUSH, S. L. (2011). Deterrence theory: Where do we stand? Review of International Studies, 37(2), 741-762. doi:10.1017/S0260210510000896

68. Russell, S. J., & Norvig, P. (2010). Artificial Intelligence: A Modern Approach. Retrieved from https://www.cin.ufpe.br/~tfl2/artificial-intelligence-modern-approach.9780131038059.25368.pdf

69. Safeguards agreements. (2016, June 08). Retrieved from https://www.iaea.org/topics/safeguards-agreements

70. Scharre, P. (2017, September 25). Robots and Artificial Intelligence Could Change War. Retrieved from http://time.com/4948633/robots-artificial-intelligence-war/

71. Schelling, T. C. (2008). Arms and influence. New Haven: Yale University Press.

72. Schneier, B. (2018). Schneier on Security: Artificial Intelligence and the Attack/Defense Balance. Retrieved from https://www.schneier.com/blog/archives/2018/03/artificial_inte.html

73. Sharkey, N. E. (2016, October 13). The evitability of autonomous robot warfare. Retrieved from https://www.icrc.org/en/inte,rnational-review/article/evitability-autonomous-robot-warfare

74. Shead, S. (2018, December 12). Canada And France Create New 'International Panel On AI'. Retrieved from https://www.forbes.com/sites/samshead/2018/12/07/canada-and-france-create-new-international-panel-on-ai/#d2c5c3d2ef22

75. Smith, B. (2017, May 15). The need for a Digital Geneva Convention. Retrieved from https://blogs.microsoft.com/on-the-issues/2017/02/14/need-digital-geneva-convention/

76. Taddeo, M. (2017). Deterrence by Norms to Stop Interstate Cyber Attacks. Retrieved from https://link.springer.com/content/pdf/10.1007/s11023-017-9446-1.pdf

77. Tang, S. (2009). The security dilemma: A conceptual analysis. Security Studies, 18(3), 587-623. doi:10.1080/09636410903133050

78. THE NEW DOGS OF WAR : THE FUTURE OF WEAPONIZED ARTIFICIAL INTELLIGENCE. (2017). Retrieved from http://threatcasting.com/wp-content/uploads/2017/09/ThreatcastingWest2017.pdf

79. The Weaponization of AI and How It Exacerbates the Security Dilemma Between States [Personal interview]. (2019, March). Ambassador Aly Erfan is a Professor of Practice and Program Direction at the Department of Global Affairs and public Policy at the American University in Cairo

80. The Weaponization of AI and How It Exacerbates the Security Dilemma Between States [Personal interview]. (2019, March). Prof. Dalal Mahmoud Al-Sayed is a Professor of Political Science at the Faculty of Economic and Political at Cairo University and Nasser Military Academy

81. The Weaponization of AI and How It Exacerbates the Security Dilemma Between States [Personal interview]. (2019, March). Mona Soliman is a Ph.D candidate at the Faculty of Economic and Political Science at Cairo University and an Associate Researcher at International Politics Journal (Al-Siyasa Al-Dawleya)

82. The Weaponization of AI and How It Exacerbates the Security Dilemma Between States [Personal interview]. (2019, April). Dr/General Mahmoud Khalaf is an Advisor at Nasser Military Academy.

83. The Weaponization of AI and How It Exacerbates the Security Dilemma Between States [Personal interview]. (2019, April). Ambassador Karim Haggag is a Professor of practice at the American University in Cairo

84. The Weaponization of AI and How It Exacerbates the Security Dilemma Between States [Phone interview]. (2019, April). Dr. Waleed Rashad is an assistant professor at the National Center for Social and Criminological Research.

85. Tweedie, M. (2017). 3 Types of AI: Narrow, General, and Super AI. Retrieved from https://codebots.com/ai-powered-bots/the-3-types-of-ai-is-the-third-even-possible

86. Tweedie, M. (2018, June 21). 6 Technologies Behind AI. Retrieved from https://codebots.com/ai-powered-bots/6-technologies-behind-ai

87. Unal, B., & Lewis, P. (2018, December 07). Cybersecurity of Nuclear Weapons Systems: Threats, Vulnerabilities and Consequences. Retrieved from https://www.chathamhouse.org/publication/cybersecurity-nuclear-weapons-systems-threats-vulnerabilities-and-consequences

88. Waltz, K(2009). Nuclear Myths and Political Realities. In The Use of Force(7th ed., pp. 116-132). Rowman and Littlefield

89. Wasser, B., Connable, B., Atler, A., & Sladden, J. (2018, June 07). Comprehensive Deterrence Forum. Retrieved from https://www.rand.org/pubs/conf_proceedings/CF345.html

90. Why we urgently need a Digital Geneva Convention. (2017). Retrieved from https://www.weforum.org/agenda/2017/12/why-we-urgently-need-a-digital-geneva-convention/

91. Wirkuttis, N., & Klein, H. (2017). Artificial Intelligence in Cyber Security. Retrieved from http://www.inss.org.il/publication/artificial-intelligence-cybersecurity/

Appendix:

Annex (1):

The elements of MAD include:

1. *Scale of Destruction:* It basically focuses on the idea of "sparing" rather than damage limitation, (Fairbanks, 2004). It considers number of causalities and degrees of collateral damage and bloodshed. With the increase of inaccuracy in weapon-targeting, the possibility of collateral damage increases, (Fairbanks, 2004). More importantly is the pace of devastation and its extremity, (Jervis, 2009), as well as the speed that causes devastation and damages to occur,

(Schelling, 2008).

2. *Proportionality of Punishment:* As Thomas Schelling and Bernard Brodie pointed out, it is all about reciprocal killing or "mutual kill", (Jervis, 2009). The US Department of Defense coined this phenomenon the "return evil for evil", (Schelling, 2008, p.7). It is also known as deterrence by punishment which measures the extent of punishment and how it will inflict pain upon the attacker.
3. *The Demonstrative Aspect:* It is the "*power to hurt*", a sort of coercive diplomacy by which the defender uses credible threats of inflicting damage and ultimatums, with the aim of influencing the offender's motives., (Schelling, 2008). It is a way of dissuading the offender from carrying out an attack.
4. *Motives and Interests:* The heart of MAD is the psychological factor that contributes to its success. It is the case where leaders are overwhelmed by mutual fear of errors, intentions and conflict of interests, (Jervis, 2009).
5. *Pace of Advancement in Military/Nonmilitary Technology:* Modern technologies favor defense due to their great lethality and mobility, as opposed to infantry technologies and cavalry warfare which favored offense over defense, given that the current technologies are not neutralized by the innovation of novel and more advanced technologies (Van Evera, 2013).
6. *Parity/disparity:* It investigates how the level of parity/disparity in technological advancement and weapon procurement could influence a state's decision and prove the existence of a security dilemma since such an advancement emboldens the strength of a state vis-à-vis its rival, (Jervis, 2009).
7. *Uncertainty:* Uncertainty could arise over rivals' intentions on whether they are malicious or security-seeking, (Tang, 2009) since some weapons are defensive in nature but can be offensively used, (Jervis, 2009).
8. *Lack of Communication:* Uncertainty over intentions reflects the lack of communication between rivals since deterrence requires transparency, as opposed to offense which requires secrecy over power, force, etc., (Van Evera, 2013).
9. *Possible Implications in case of Intentional/Unintentional Error*: The margin of intentional versus unintentional error can be reflected in the case of the

Cuban Missile Crisis. Therefore, it is urgent to raise the question about the effects of error on the expected utility of AI.

10. *Indivisibility of Control:* The core of this idea is the unity of command and control over weapons to make any MAD-like scheme effective, (Fairbanks, 2004).
11. *Wartime Operation*: A group of theorists argued that intensity of war is based on (1) interests at stakes; the more interests at issue, the higher intensity of war and (2) the ability to punish in return for escalatory acts, (Van Evera, 2013).
12. *Second-strike Capability Vs. First-strike Capability:* It is the rational calculus of a first-strike based on the opponent's ability to carry out a second-strike. It nullifies the advantage of a first-strike since there is reciprocal fear of spiral attacks and the first-mover advantage seems dangerous given that it can spur a vicious circle of attack, (Van Evera, 2013). Thomas Schelling, however, argued that the first-strike capability assesses benefits associated with using weapons through preemptive strikes, (Van Evera, 2013). Its advantages include: the feasibility of gaining surprise without detection, the shift in the balance of power, and the dominance of offense when the attacker can defend itself and conquer its rivals and, finally, the extent of political punishment, (Van Evera, 2013).

Annex (2):

Detailed Description of Variables:

Independent Variables:

a. **Human-over-the-loop:** Human supervises the loop, though delegating tasks to machines as it is the case in Air Drones. [1]

[1] De Spiegeleire, S., Maas, M., & Swejis, T. (n.d.). ARTIFICIAL INTELLIGENCE AND THE FUTURE OF DEFENSE. Retrieved from https://www.hcss.nl/sites/default/files/files/reports/Artificial Intelligence and the Future of Defense.pdf

b. **Human-out-of-the-loop:** Human has no control over machines since machines have the power to decide and act. [2]

Dependent Variables:

1) In the era of digital warfare, the degree of military digitalization varies from one state to the other, thus, the degree of vulnerability varies as well. There are three **degrees of dependence on technology**:

I. *Digitally-Independent States:* A military does not have large networks for command and control and its conventional weapons do not require digital technology. Thus, a state is not vulnerable to cyberattacks, (Schneider, 2016). [3]

II. *Digitally-Enabled States:* A state uses technology for the sake of enhancing its network-centered military operations. Such a state utilizes datalinks to convey off/circumvent targeting information. It relies on digitally-enabled applications for cyber intelligence, so as to raise situational awareness. The state's military prefers analogue or hard copy processes. Iran is a perfect example of such a state, (Schneider, 2016).

III. *Digitally-Dependent States*: A state that is highly dependent on technology and its command and control systems are limitless over the horizons and its military has data fusion centers. It implements network-centered operations with the use of datalinks and virtual computing. Virtual computing is highly effective for off-boarding intelligence and for ensuring the optimization of decision-making. More importantly, the state's conventional operations heavily rely on technology, (Schneider, 2016).

2) Sparing: The term "Sparing" is usually associated with MAD. The term "sparing" implies that mutual vulnerability does exist. City-sparing and cyber-sparing were coined by theorists and experts when both nuclear and cyber

[2] Russell, S. J., & Norvig, P. (2010). Artificial Intelligence: A Modern Approach. Retrieved from https://www.cin.ufpe.br/~tfl2/artificial-intelligence-modern-approach.9780131038059.25368.pdf

[3] Schneider, J. (2016). Digitally-Enabled Warfare. Retrieved from https://www.cnas.org/publications/reports/digitally-enabled-warfare-the-capability-vulnerability-paradox

MAD loomed over in the IR literature. Now, with the current inclination to develop AI applications for military purposes, the phenomenon could be referred to as "machine-sparing". The term "machine-sparing" portrays how countries, cities, individuals, cyberspace, and machines are equally subjected to mutual threats or attacks. The scale of destruction exceeds human control if machines have been mandated to act. "Machine-sparing" indicates that militarized AI applications (used without any control or regulation) could destroy the land and thereon.

3) **Latent Violence:** MAD becomes successful when latent violence is used by the defender. The core of nuclear latent deterrence is the deterrence by punishment, (Fuhrmann, 2018)[4] whereas deterrence by denial is the principal element of cyber MAD. But for AI MAD, it is still unclear whether deterrence by punishment, denial or entanglement would be a workable strategy.

4) **Expected Utility, (Slayton, 2017) and Cost-benefit Analysis:** States, as per IR theorists, act rationally and state leaders do not rush to war unless the consequences are cost effective and the interests are vital for state survival. AI applications usually have implications on the governmental decision-making process, this is reflected in; policies, objectives, interests, values and calculations with the increasing tendency to use them across sectors. Furthermore, AI applications shape a state's geographical position, political values and foreign policy. AI applications also promote a state's economic progress, thereby affecting the calculus of war. Thus, the nature of the utilization of nonmilitary capabilities will likely change due to AI applications. The cost-benefit analysis may include:

   I. **Costs of Offense/Defense, (Slayton, 2017)[5]:**

   The cost of military innovation in today's world is crucial for making accurate calculations and developing well-defined strategies and plans. Military technology, like other types of technology and business

[4] Fuhrmann, M. (2018). The Logic of Latent Nuclear Deterrence. Retrieved from http://www.iserp.columbia.edu/sites/default/files/Deterrence without Bombs 2018-0129.pdf

[5] Slayton, R. (2017, February 18). What Is the Cyber Offense-Defense Balance?: Conceptions, Causes, and Assessment. Retrieved from https://muse.jhu.edu/article/648308/pdf

organizations, have both direct and indirect costs, thereby shaping the military strategy.

a) **Direct Cost:** Direct costs usually include the costs of software development and regular updates of software; hardware production; designing effective security systems in both virtual and real realms; weapon production; coding; algorithms and swarms.

b) **Indirect Cost:** Indirect costs are comprised of the allocation of spaces and laboratories; research and development (R&D); the provision of infrastructure; the wages and salaries of software, coding and algorithm developers, as well as the costs of training on coding and algorithms for military staff and personnel.

II. **Comparison, (Handel, 1991)[6]:** Every state investigates the degree of advancement in its military equipment such as its defense system and software (which is not enabled by AI). It compares the size of its military forces and arsenals. It also determines the amount of data possessed and retrieved through surveillance operations. In today's warfare and the information age, each state evaluates its capacities in terms of intelligence operations and espionage. Such evaluation and assessment definitely helps every state to recognize its comparative advantage/strengths and its weaknesses, as opposed to other states/adversaries.

III. **Calculus of War, (Handel, 1991):** Every state should be compelled to cross-examine:

(i) the chances of victory, and how AI applications increase or reduce the chances of victory in the case of considering an offensive AI strike;

(ii) the risks of disrupting AI applications and other similar cyber capabilities in the case of considering a defensive AI counterstrike and in the case of having an amateur security system.

[6] Handel, M. I. (1991). San Tzu and Clausewitz: The Art of War and On War Compared. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a239084.pdf

More importantly, similar to nuclear weapons, the scale of destruction and the number of causalities should be estimated since offensive AI applications could make war much more destructive.

a. **<u>Duration and Scale of Operations:</u>** The duration of war is usually considered by policy-makers since duration defines the scale of operation, (Handel, 1991). This has dramatically changed with the emergence of AI, given that it is faster than the human pace.

b. **<u>Perception of Threats, (Handel, 1991):</u>** The security dilemma is typically exacerbated when a state assumes its interest(s) is/are at stake. The weaponization of AI will redefine threats at all levels; policy-makers and security experts will perceive threats differently since the war battle has been transferred to cyberspace and has shifted from being a war between military personnel to a war between machines and AI-enabled systems. The perception of threats will be based on the degree of dependence on/independence from technology. Yet, the degree of dependence on technology and cyber capabilities is critical in perceiving threats; the implications of using conventional capabilities, either disjointedly or alongside AI capabilities, should be considered. Policy-makers will define threats triggered by the development of AI capabilities as either positive or negative.

c. **<u>Balance of Power:</u>** The inconvenience from the shift in the balance of power comes first, since any shift in the balance of power basically means putting a state's interests at risk and having a influence on a state's decisions and abilities, (Horowitz, 2018)[7]. AI, similar to other capabilities, will alter the balance of power in favor of the superior, as President Putin implied, the top AI application developer

[7] Horowitz, M. C. (2018). Artificial Intelligence, International Competition, and the Balance of Power. Retrieved from https://tnsr.org/2018/05/artificial-intelligence-international-competition-and-the-balance-of-power/

will ultimately hold the most power. Nevertheless, AI should be measured as a variable of power vis-à-vis other sorts of power (i.e. economic, political or geopolitical, etc.).

5) **Estimation of Military and Nonmilitary Capabilities (Highly Advanced/Amateur):** The estimation of non-AI military capabilities in terms of quantity and degree of advancement, should be taken into consideration in order to measure the effect of either highly advanced or amateur AI capabilities on other them. The rationale behind this is to question the significance of other capabilities with the possession of advanced AI.

6) **Means and Levels of Communication (Weak/Strong/Absent/Interconnected Networks):** Since the security dilemma is often tightened as a result of the lack of communication and uncertainty over intentions, levels of communication should be hypothesized as follows:

   (i) weak or strong, if humans have a role,

   (ii) absent or interconnected networks, if humans are absent and out-of-the-loop.

The purpose is comparing levels of communication between states with the presence or the absence of the human aspect.

7) **Estimation of Quantities and Level of Advancement (Equal/Unequal) in AI Applications:** Disparity in nuclear capabilities increased the security dilemma between the two superpowers during the Cold War era. As is the case with nuclear weapons, the disparity in the number of possessed AI applications and the level of advancement in AI software will exacerbate the security dilemma. It is suggested to measure the parity/disparity in AI capabilities as either equal or unequal, so as to help states in their calculations.

8) **Intentions (Malicious/Security-seeking):** Intentions are the cornerstone of MAD and the security dilemma as they create uncertainty. Intentions could either be malicious or security-seeking when the ruling elite has a say in the military decision-making process.

9) **Calculations (Right/Mistaken):** Unlike intentions, machines or software cannot be judged on their intentions but they can be judged on the correctness of their calculation.

10) **Scope of Human Participation in the Decision-making Process (Limited/Unlimited):** Delegating the military's decision-making process to

machines (i.e. giving the machines absolute authority) is still highly unlikely, though the declining role of human beings in military decision-making is worthy of consideration with the emergence of AI.

11) **Degree of Control Over Machines (Absent/Active):** AI will not only undermine role of humans in the decision-making process, but will also make their role almost absent during the course of war. Thus, the degree of human control over machines and software must be measured as either absent or active.

12) **Margin of Error (Human Vs. Machine):** Both human and machine errors are highly possible and highly destructive. Error should be measured as either more common when the human is out of the loop/over the loop.

13) **Command and Control (Reliable/Unreliable), (Slayton, 2017)[8]:** With the development of AI applications for military purposes, the absolute authority, which was once only given to the military's command and control system, has become sharable and divisible with software and machinery. With the adoption of AI, the command and control system is unreliable, given that AI-enabled machines, which could be mandated to make decisions, could be disrupted. Therefore, AI command and control could either be reliable or unreliable based on the degree of human control.

14) **Attribution and Accountability:** Comparable to cyber capabilities, attribution and legitimacy are problematic not only because the difficulty of identifying and proving the identity of the attacker but also the impossibility of rebuking and penalizing a machine. It is also difficult to define accountability of machines in absolute terms, according to international lawyers, who are alarmed by the lack of efficacy and applicability of the Law of Armed Conflicts in the AI realm. It is suggested to use the cyber attribution indicators which include: technical, political and clandestine indicators, (Somara, 2019)[9]. The technical indicators recess IP addresses and makes log

[8] Slayton, R. (2017). What Is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment. *International Security*,*41*(3), 72-109. doi:10.1162/isec_a_00267

[9] (2019). Retrieved from https://www.rand.org/multimedia/video/2019/01/14/accountability-in-cyberspace-the-problem-of-attribution.html?utm_source=WhatCountsEmail&utm_medium=RAND Policy Currents AEM: Email Address NOT LIKE DOTMIL&utm_campaign=AEM:631600804

file analysis, through text-strings, timestamps, C2 infrastructure, malware samples and credentials, (Davis et al, 2017)[10] whereas the political indicators assess the diplomatic knowledge about political motivation and political operatives. Concerning the clandestine indicators or "All-Source Intelligence", they examine classified data obtained by signals-intelligence, human intelligence and open-source intelligence, (Davis et al, 2017) coupled with political insights, (Somara, 2019)[11]. Signals-intelligence (SIGNT) is produced by collecting data from information technology systems, while Human intelligence (HUMINT) is produced by obtaining data from humans, (Davis et al, 2017). For all-source intelligence (OSINT) is produced by using open sources such as the internet to collect and process information, (Davis et al, 2017).

15) **Counterattacks/Counterforce,** (Lieber et al, 2017) **(Probable/Improbable):** There is an endless debate over the rationale of launching a preemptive or a preventive strike amid a high probability of a retaliatory strike. As it was the case with the nuclear weapons, AI could make a second-strike/counterattack probable whether humans are over or out-of-the-loop.

Annex (3):

Brief History of AI:

The small Dartmouth Project, which took place in 1956, marked the birth of Artificial Intelligence, (De Spiegeleire et al, 2017). Since then, AI, as a field of study, had evolved across six main phases. The first phase or the "First AI Spring" (1956-1975), marked the development of neural networks in its primitive forms, is considered as

[10] Davis, J. et al. (2017, June 02). Could Stateless Attribution Promote International Cyber Accountability? Retrieved from https://www.rand.org/pubs/research_reports/RR2081.html

[11] Ibid

the early golden age of AI since AI researchers succeeded in developing tools and prototypes systems capable of performing a limited range of tasks, such as algebra and games, as if they are carried out by humans, (De Spiegeleire et al, 2017). At the peak of the Cold War whereby the grandiose bulk of funds had been allocated to the military sphere, the AI research had slipped into its first winter (1974-1980) and speedy progress had been decelerated. In fact, the Cold War was not the sole reason that contributed to the slippery of AI into its first winter but also the discovery of ample possibilities for developing and underpinning AI algorithms in a manner that could deal with real-world problems, thereby sparking disagreements among AI researchers, (De Spiegeleire et al, 2017). In 1980s, AI research had witnessed its second spring with the advent of expert systems which were actually a group of rule-based programs with limited tasks ranged from answering questions or solving problems, and with massive funds provided by governments for promoting AI research and the establishment of numerous AI companies, (De Spiegeleire et al, 2017). In spite of noticeable sales which reached up to 2 billion by 1988, many AI companies collapsed and AI research had entered its age of darkness for many reasons which included: (i) the development of desktop PCs by Apple and IBM and (ii) the limited utility of expert systems, (De Spiegeleire et al, 2017). Meanwhile, AI programs which were of military significance such as the autonomous battle tank program raised considerable funding, (De Spiegeleire et al, 2017). In an effort to reinvigorate AI research, AI researchers had disregarded their long-term goal of developing human-level AI applications and directed their focus to fragmented subfields by developing applications that solve specific problems, (De Spiegeleire et al, 2017). Due to the increasing utility of AI in logistics, satellite monitoring, spacecraft, traffic management, medical diagnostics and the military, funding had soared up in the mid-2000s, (De Spiegeleire et al, 2017). Tremendous financial contributions from Apple, Facebook, Amazon, Baidu, IBM and Microsoft have furthered AI research since these corporates use AI for developing business models and profit maximization, (De Spiegeleire et al, 2017). In response, AI has reached a tipping point with the proven predictive accuracy of algorithms, the increasing computing power, the Internet of Things and Big Data and cloud infrastructures, (De Spiegeleire et al, 2017).

Annex (3):

The Elements of the Proposed AI MAD Structure:

Mutually Assured Manipulation (MAM)

1. *Scale of Destruction:* The scale of destruction could exceed the destructive potentials of nukes and conventional capabilities since they could be manipulated or disrupted.
2. *Proportionality:* Proportionality could entail proportionality of manipulation, so as to increase uncertainty. Manipulation could be the umbrella of other sources of deterrence.
3. *The Demonstrative Aspect (Latent Violence):* States could employ deterrence by punishment, denial, disruption or manipulation.
4. *The Psychological Factor (Motives and Interests):* Threat of manipulation and the fear of uncertainty would definitely dissuade states from launching a first-strike.
5. *Pace of Advancement in Military/Nonmilitary Technology:* The ever-increasing uncertainty over the adversary's AI capabilities coupled with the high potential of neutralizing a state's defense and C2 systems makes deterrence operative.
6. *Parity/disparity:* Disparity in AI could be reflected in the degree of advancement in AI military applications, while the number of applications would not be of great concern.
7. *Intentional/Unintentional Error*: Errors either triggered by machines or humans could occur because of data manipulation and miscalculations.
8. *Second-strike Capability:* Massive retaliatory attacks are highly probable in AI deterrence.