The First International Crisis of Artificial Intelligence: Imagining Scenarios and Responses

Shira E. Cohen

With the rise of artificial intelligence (AI), a new but underdeveloped literature has emerged to investigate how AI might impact international conflicts. Just how, this article asks, might the first international AI crises occur, and how might international actors respond to avoid catastrophe? After examining existing literature, this piece argues that the three major applications of AI in military systems that may lead to international crises include: AI-enabled autonomous weapons systems, AI-enabled nuclear weapons systems, and AI-enabled cyber weapons. While it is impossible to pinpoint one scenario or even one type of application that will cause the first international AI crisis, this article draws on the model of the "Doomsday Clock" (maintained since 1947 by the Bulletin of the Atomic Scientists) to help give coherence to possible outcomes and responses. At its core, this article proposes a new model for early warning signs of an impending AI crisis—what it calls, the "AI Doomsday Clock"—which intends to offer a tool to decision-makers that will provide awareness of the early signs of an international crisis in AI, before developing into a potentially disastrous international crisis.

In recent years, there has been a consensus among security experts that technological developments in artificial intelligence (AI) will cause significant adaptations to current warfare.[1] This notion has encouraged many countries to produce their own national AI development strategies, expressing their desires to enter the race for global leadership in the field of AI.

What is known as the "AI arms race" is led by three countries, each with different strategic approaches. First, China, which has set a goal for itself to lead the world in AI by 2030, pursues an aggressive approach that focuses on developing AI that contributes to strategic decision-making. In practice, China is seeking to develop human-level AI that will portray the fluctuating circumstances of a battlefield, advise commanders in their decision-making, and make other tactical contributions. Second, the United States—which in 2019, via executive order by President Donald Trump, created the "American AI Initiative"—follows a more conservative approach that seeks to produce computers that would assist in decision-making, but not make them on their own. This strategy, rather than creating fully autonomous systems, emphasizes "human-machine collaboration", meaning machines that provide data and analysis to human operators who then take action. Last of the three global leaders in the "AI arms race" is Russia, whose president, Vladimir Putin, in 2017 declared, "Whoever becomes the leader in this sphere will become the ruler of the world." For its part, Russia aims to produce better military hardware that relies on AI but leaves decisions about deployment entirely in the hands of human operators. An example of this is the Russian project of self-guided missiles that are able to change course mid-flight in order to evade missile defenses.[2]

As with any arms race once it begins, a process emerges among competitors in which their relentless competition drives an almost constant state of instability.[3] This phenomenon, seen in the current AI arms race, relates to new

1 James Johnson, "Artificial Intelligence & Future Warfare: Implications for International Security," *Defense & Security Analysis* 35:2 (2019): pp. 147-169.

2 Adrian Pecotic, "Whoever Predicts the Future Will Win the AI Arms Race," *Foreign Policy*, March 5, 2019.

3 Lewis Richardson, *Arms and Insecurity: A Mathematical Study of the Causes and Origins of War and Statistics of Deadly Quarrels* (Pittsburgh: Boxwood Press, 1960).

Shira Cohen graduated with a B.A. from the Honors Track in Security, Strategy, and Decision-Making (Highest Honors) and is a postgraduate student in government, diplomacy, and strategy at the Interdisciplinary Center (IDC) Herzliya. She is a research assistant in the field of decision-making, strategy, religion, and emerging technologies. Previously, she was a research assistant at the Institute for the Study of Intelligence Methodology, the Israel National Cyber Directorate, and also the Strategic Planning Department at IDF Planning Directorate. Shira may be reached at Shiraco30@gmail.com

AI-enable technologies in autonomous weapons, nuclear weapons, and cyber weapons; a new technology with the potential to redefine the status quo in military operations is at the center of the competition, potentially leading to unintended consequences at the strategic level, to include crises.[4]

Therefore, this article will seek to examine the following question: How might the first international AI crises occur, and how might international actors respond to avoid catastrophe? Since current AI applications are considered "enabling technology," meaning that AI has had widespread applications similar to the development of electricity or the internal combustion engine, there are many possibilities that might lead to an AI crisis. After examining existing literature, this piece argues that the three major applications of AI in military systems that may lead to international crises include: AI-enabled autonomous weapons systems, AI-enabled nuclear weapons systems, and AI-enabled cyber weapons. While it is impossible to pinpoint one scenario or even one type of application that will cause the first international AI crisis, this article proposes a new model for early warning signs of an impending AI crisis, what it calls the "AI Doomsday Clock." This model intends to offer a tool to decision-makers that will provide awareness of the early signs of an international crisis in AI, before entering a potentially disastrous international crisis. It should be noted that this article will focus on a strategic international crisis in AI, at the military level only and not at the civil level.

A Review of Definitions: Artificial Intelligence, Crises, and Military Technology

Military AI Definition and Applications

In current literature involving international crises and AI, there does not exist a universal definition of the term "AI," as it refers to a wide range of applications and technological functions for improving the performance of automated or autonomous systems.[5] This paper discusses AI in general terms, and therefore, the definition of the Defense Science Board's Summer Study on Autonomy in 2016 for "AI" is adequate. AI is "the ability of a computer system to perform tasks that normally require human intelligence".[6]

With such a broad definition, it is unsurprising that so many different types of AI technology exist. All current AI systems fall into the Narrow AI category, also known as Weak AI, which refers to algorithms that address specific problem sets like image recognition. In contrast, General AI, also known as Strong AI, refers to systems capable of human-level intelligence across a broad range of tasks. The most common approach to Narrow AI is Machine Learning, which involves statistical algorithms that replicate human cognitive tasks by deriving their own procedures through analysis of large training datasets.[7] Another major type is Deep Learning, which is a form of machine learning that uses models of human neural networks in order to produce predictions that can be applied to new information in a hierarchical process.[8] Finally, another key concept in AI is Natural Language Processing, which is a computational process that allows machines to both understand and produce written and spoken language.[9]

With such a variety of potential uses, in the military field, the utilization of AI technology systems can be seen in all dimensions of combat: land, air, sea, intelligence, and command, as well as in diverse military-security fields, such as nuclear, autonomous weapons, and information operations.[10] Research regarding international crises in AI addresses three major applications of AI in military systems that may, under certain circumstances, lead to crises. These include Lethal Autonomous Weapon Systems (LAWS), which use sensor suites and computer algorithms

4 Johnson, "Artificial Intelligence & Future Warfare."

5 James Johnson, "The AI-Cyber Nexus: Implications for Military Escalation Deterrence, and Strategic Stability," *Journal of Cyber Policy* 4:3 (2019): pp. 442-460.

6 Forrest E. Morgan, Benjamin Boudreaux, Andrew J. Lohn, Mark Ashby, Christian Curriden, Kelly Klima, and Derek Grossman, *Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World* (Santa Monica, CA: RAND Corporation, 2020).

7 "Artificial Intelligence and National Security," Congressional Research Service (CRS), 2020.

8 Juergen Schmidhuber, "Deep Learning in Neural Networks: An Overview," *Neural Networks* 61 (2015): pp. 85-117.

9 Prakash M. Nadkarni, Lucila Ohno-Machado, and Wendy W. Chapman, "Natural Language Processing: An Introduction," *Journal of the American Medical Informatics Association* 18:5 (2011): pp. 544-551.

10 Guilong Yan, "The Impact of Artificial Intelligence on Hybrid Warfare," *Small Wars & Insurgencies* 31:4 (2020): pp. 898-917.

to independently identify a target and employ an onboard weapon system;[11] AI-enabled nuclear weapons systems, primarily early warning, intelligence, surveillance, reconnaissance (ISR), and command and control systems; and AI-enabled cyber weapons, which manipulate or destroy code in an adversary system, such as ISR systems or a deployed kinetic-weapons system. Therefore, this paper will focus on these three major applications.

International Crises and Military Technology

Due to limited research literature in the field of AI and international crisis, in order to imagine sources of responses to the first international crisis in AI, one must first examine the influence of military technology on international crises.

The literature regarding the development and proliferation of military technologies shows us that the assimilation of technological innovations has had significant implications for world order, strategic stability, the likelihood of war, and the formation of international crises.[12] The primary ways in which these technologies shape and influence global power relations are through both military and economic means.[13]

The impact of quickly evolving military technologies, as well as the risk that they pose in the form of destabilization and crisis, has led to international efforts to control the distribution, production, development, or deployment of certain military technologies.[14] Two primary examples are nuclear weapons and cyberspace technologies. Both the Cuban Missile Crisis, which led to a drastic change in international relational dynamics,[15] and the 1983 crisis following the Able Archer exercise and the Soviet Union's misconception about its purpose are relevant to the nuclear context.[16] In the cyber context, the Stuxnet attack, which disrupted centrifuge enrichment activities at the uranium enrichment facility in Natanz, Iran, can be classified as an international crisis that drastically changed the way in which cyberspace was viewed.[17] Another example is the cyber attacks on the computer networks of the U.S. Democratic National Committee in 2015 and 2016, which represented the expanding phenomenon of foreign cyber interference in the election.[18]

In recent years, researchers have compared nuclear and cyber technologies in international relations with AI technologies, including the likelihood of their impact on international crises. Payne[19] argues that while each technology have different and distinct characteristics, both nuclear and cyber technologies have the potential to challenge strategic stability, which can lead to an international crisis given the proper preconditions. It should be noted that in cyberspace, there have been recent claims that cyber activities are unlikely to cause escalation as originally expected.[20]

Placing military AI technologies in the same category of emerging military technology as cyber and nuclear technologies may help us understand their true impact on international relations and potential crises.

---

11 "Artificial Intelligence and National Security."

12 Michael C. Horowitz, *The Diffusion of Military Power: Causes and Consequences for International Politics* (Princeton, NJ: Princeton University Press, 2010).

13 Michael C. Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power," *Texas National Security Review* 1:3 (2018): pp. 36-57.

14 Matthijs M. Mass, "How Viable is International Arms Control for Military Artificial Intelligence? Three Lessons from Nuclear Weapons," *Contemporary Security Policy* 40:3 (2019): pp. 258-311.

15 James Johnson, "Artificial Intelligence in Nuclear Warfare: A Perfect Storm of Instability?" *Washington Quarterly* 43:2 (2020): pp. 197-211.

16 Ryan Kiggins, "Big Data, Artificial Intelligence, and Autonomous Policy Decision Making: A Crisis in International Relations Theory?" in Ryan Kiggins ed., *The Political Economy of Robots* (London: Palgrave Macmillan, 2017), pp. 211-234.

17 Louk Fasen, Bianca Torossian, Elliot Mayhew, and Carlo Zensus, "Conflict in Cyberspace: Parsing the Threats and the State of International Order in Cyberspace," *Strategic Monitor* (2019).

18 Michael Buratowski, "The DNC Server Breach: Who Did it and What does it Mean?" *Network Security* 10 (2016): pp. 5-7.

19 Kenneth Payne, "Artificial Intelligence: A Revolution in Strategic Affairs?" *Survival* 60:5 (2018): pp. 7-32.

20 Ben Buchanan, *The Hacker and the State: Cyber Attacks and the New Normal of Geopolitics* (Cambridge, MA: Harvard University Press, 2020); Brandon Valeriano, Benjamin Jensen, and Ryan C. Maness, *Cyber Strategy: The Evolving Character of Power and Coercion* (New York: Oxford University Press, 2018).

International Crises and AI

There are two features of AI technologies that lead to the undermining of strategic stability and crises: the arms races in AI and the characteristics of AI technology.

In the AI realm, there are two arms races taking place simultaneously: the "complex race" and the "narrow race." The "complex race" focuses on developing AI technology that will assist in decision-making processes, and it is currently led by China and the United States. Conversely, the "narrow race" is similar to a traditional arms race, focusing on achieving the best military functions in AI, and it is led by China, the United States, and Russia. These two arms races take place in unison, as countries participating in the complex race still must invest in less advanced AI functions in order to compete in the narrow race. This is due to the fact that a country with more powerful military AI functions, such as drones and submarines with autonomous capabilities, will have a significant advantage on a future battlefield.[21] It should be noted that the preconditions for an international crisis would likely develop more quickly in the narrow race, as some researchers claim that achieving operational AI maturity in the complex race will be possible in 2045 at the earliest and 2070 at the latest.[22]

Along with the AI arms race, one can also study the particular characteristics of AI technology that may directly affect the potential for international crises. First, AI systems act and respond quickly, in a way that may lead to errors or unwanted actions. This rapid response of the systems can make it difficult for operators to address the error before it leads to a crisis.[23] Additionally, merging AI technologies with advanced weaponry may allow opponents to increase their pace of combat. This rapid increase in pace can make it increasingly difficult for commanders to contain or even end events.[24] Second, AI technologies are highly dependent on the reliability and pertinence of the data; if the data is inaccurate, then the system may make non-optimal decisions that could lead to devastating consequences.[25] Third, AI systems still depend on their encoded assumptions, by that of human engineers who risk sowing their biases into the systems that they program. These biases can then lead to errors in AI systems and thus cause instability and even crisis.[26] Fourth, in AI systems that include connectivity between multiple components, the behavior of the AI creates interactions in ways that are not directly visible, not linear, and sometimes not immediately understandable to the operator. This complexity makes it difficult for operators to understand or anticipate system behavior, especially in new environments or in unexpected scenarios. Additionally, this situation may reduce the operator's ability to detect or isolate an error. In times of tension or instability, the detection of an error may be critical, and failure to do this may eventually lead to a true crisis.[27]

Finally, in the field of AI, the future of warfare has much to do with the investments of and costs to the private sector. The gap between the military and private sector in this arena is only growing and may lead to countries' reliance on military systems produced by the private sector, which notably has different interests than those of the state. This situation could lead to the acquisition of inferior systems of weaponry by the states, in addition to systems that do not fully conform to the military interests of the states.[28] Alternatively, in a state with a non-cooperative private sector, it could also cause domestic tension or the state to fall behind in an arms race.

---

21 Pecotic.

22 Mark Graves, "Shared Moral and Spiritual Development Among Human Persons and Artificially Intelligence Agents," *Theology and Science* 15:3 (2017): pp. 333-351.

23 Mass.

24 James Johnson, "Artificial Intelligence, Drone Swarming, and Escalation Risks in Future Warfare," *RUSI Journal* 165:2 (2020): pp. 26-36.

25 Asaf Tzachor, Jess Whittlestone, Lalitha Sundaram, and Sean O. Eigeartaigh, "Artificial Intelligence in a Crisis Needs Ethics with Urgency," *Nature Machine Intelligence* 2 (2020): pp. 365-366.

26 Johnson, "Artificial Intelligence, Drone Swarming, and Escalation Risks in Future Warfare."

27 Mass.

28 M. L. Cummings, Heather M. Roff, Kenneth Cukier, Jacob Parakilas, and Hannah Bryce, "Artificial Intelligence and International Affairs: Disruption Anticipated," *Chatham House Report*, 2018.

A Schema for Types of International Crises in AI

The vague definition of AI technologies, the varied applicability of AI in defense systems, and AI's complex characteristics make it difficult to predict which type of AI technology applications will lead to the first international crisis.

In order to offer some coherence to just how and why the first international AI crisis might emerge, this article lays out the following three scenarios.

First Application – Lethal Autonomous Weapons Systems

First Scenario – Claims of an Accident: An international crisis may result from an accident that has occurred, or that the other party believes has occurred, due to the use of Lethal Autonomous Weapon Systems (LAWS). A possible hypothetical scenario in this context, proposed by Leys,[29] is based on an incident in 2015 whereby Turkey shot down a Russian military plane. However, in the hypothetical, it is a Russian anti-aircraft system located near the Syrian-Turkish border that shot down a Turkish military plane. The Russian military claimed that it suspected that there may have been a malfunction in its autonomous monitoring system. In such a situation, there is no way to determine the truth of Russia's claims, as Turkey demanded U.S. military action against Russia, under the North Atlantic Treaty. This example supports the notion that AI technology adds yet another layer to battlefield uncertainty, similar to the concept put forward by Carl von Clausewitz of the "fog of war". This situation, on the one hand, allows leaders a wider scope of plausible deniability for unsavory state behavior. That is because in some cases, LAWS are designed to operate without a human decision-maker, and therefore states are able to take actions that would be otherwise unavailable to them, under the guise of a machine malfunction. On the other hand, even if both parties recognize that the situation is due to machine malfunction, the malfunction results may be severe and therefore generate pressures that can lead to a crisis.

Second Scenario – Deterrence: Another potential issue stems from the will to deter the opponent that can lead countries to use LAWS before the technology has reached sufficient operational maturity. As long as adversaries fear that the capability may exist, they can be deterred. However, deploying such systems can lead to systems or human errors, and increased sensitivity to disruption and sabotage, as well as a failure of deterrence that may lead to a crisis.[30] For example, the growing competition between China and the United States in the South China Sea may lead either or both to use LAWS, such as submarines before it is technologically mature in order to deter the other from taking action. This premature use can lead to unwanted errors either in the technology or in the decision-making process, eventually leading to a crisis.

Third Scenario – Hyper War: An international crisis may result from "hyper war," meaning a state of conflict in which human decision-making is almost non-existent, and therefore, responses are immediate and potentially destructive.[31] A relevant example, which is borrowed from the financial realm but can also contribute to the field of LAWS, is the 2010 stock market flash crash. According to the U.S. Securities and Exchange Commission, the rapid crash was due to the use of autonomous financial trading algorithms, which led to a trillion-dollar market decline in just a few minutes. In the military field, high-speed LAWS may push the pace of combat to a point at which human commanders might lose control of the process. This challenge of keeping the human in the decision loop is one of the most significant challenges that militaries face regarding the implementation of AI. Additionally, unlike the financial market, there is no overarching authority in international relations that can enforce the AI-based mechanisms if they do fail. Therefore, a similar event in the military field, characterized by an arms race and instability, may quickly lead to an international crisis.[32]

---

29 Nathan Leys, "Autonomous Weapons Systems and International Crisis," *Strategic Studies Quarterly* 12:1 (2018): pp. 48-73.

30 Edward Geist and Andrew J. Lohn, *"How Might Artificial Intelligence Affect the Risks of Nuclear War?"* (Santa Monica, CA: Rand Corporation, 2018).

31 Liran Antebi and Inbar Dolinko, "Artificial Intelligence and Policy: A Review at the Outset of 2020," *Strategic Assessment* 23:1 (2020).

32 Johnson, "Artificial Intelligence in Nuclear Warfare;" Mass.

Fourth Scenario – Withdrawal of Human Military Forces: Replacement of currently deployed human military forces with LAWS may lead to a number of unstable scenarios. First, withdrawing forces can cause a country's allies to question its commitment to the alliance as a result of reduced presence and protection. As Leys[33] noted, although small forces of U.S. troops stationed on allied soil cannot repel a mass invasion, if they are killed, the potential political costs for a U.S. leader who chooses not to respond all but guarantee a military response, restoring U.S. credibility in the event of a military response. It should be noted that in response to the replacement, the ally may work to strengthen its military in order to defend itself, leading to regional tensions. Another scenario resulting from the substitution of human force would be an increase in the use of military force, due to reduced fear of injuries or deaths of soldiers. This scenario would likely increase aggression, therefore increasing tensions between the parties and leading to crises.[34] Finally, this replacement may lead to increased operations below the escalation threshold, such as the use of drones. Such operations may exacerbate ambiguity and incite violent means, thus leading to a crisis.[35]

Second Application – Nuclear Warfare and Artificial Intelligence

First Scenario – Disrupting the Balance of Nuclear Power: All of the leading countries in the AI arms race are nuclear states. Both strategic competition and the possibility of machines to accomplish many tasks that once required human effort or were considered impossible to change shifted the balance of power. That is because it might portend new capabilities that could spur arms races or increase the likelihood of nuclear force being utilized.[36] Additionally, many countries possess nuclear capabilities, which creates an increasing number of pathways by which violence can be escalated, as each country naturally chooses different responses to emerging technologies in the digital age due to inherent cultural differences. An example of the destabilization between nuclear states occurred in 2016 when China captured a U.S. underwater drone, claiming that it posed a danger to Chinese naval navigation. This incident ended with China returning the naval AI drone, after several days of diplomatic controversy. Although this case did not lead to a serious crisis between the nuclear-armed countries, the incident demonstrates the possible risk of escalation caused by the ambiguity surrounding the deployment of new technology.[37]

Second Scenario – Combining AI Systems with Nuclear Systems: This type of combination may drastically alter AI's current capabilities through the assimilation of new nuclear technologies. For instance, recent reports suggest that China is considering the assimilation of AI capabilities into nuclear-powered submarines, and possibly even into nuclear-armed ones. In theory, these capabilities could enable China to receive and process large amounts of signal data, while also supporting a wide range of naval operations and providing the nation with advantages on the naval battlefield. That being said, these new nuclear technological capabilities may disrupt the delicate balance that exists today.[38] It should be noted that military organizations have to overcome a number of hurdles to use the new technology effectively, and therefore, it is important to put into perspective the impact of that new technology on military effectiveness.

Third Scenario – The Use of Autonomous Weapons to Protect Strategic Nuclear Assets: Countries with nuclear weapons can use autonomous weapon systems to protect strategic nuclear assets, by using them to control launch facilities and command and control systems, in ways that affect nuclear deterrence and security. This will become a new and increasingly enticing asymmetric option to undermine an adversary's military readiness, deterrence, and resolve, and thus alter the current balance of power that includes certain equality in the nuclear capabilities of the various sides. For example, a country with nuclear capabilities could use AI drone swarms, which are considered low-risk and low-cost AI-augmented autonomous weapons with ambiguous rules of engagement, to protect its strategic

---

33 Leys.

34 Ibid.

35 Johnson, "Artificial Intelligence & Future Warfare."

36 Geist and Lohn.

37 Johnson, "Artificial Intelligence in Nuclear Warfare."

38 Stephen Chen, "China's Plan to Use Artificial Intelligence to Boots the Thinking Skills of Nuclear Submarine Commanders," *South China Morning Post*, February 4, 2018.

nuclear assets.[39]

Third Application - Cyber Attacks and Artificial Intelligence

First Scenario – Increasing the Potential Scope for Cyber Attacks: Although future attacks will probably be similar in shape to today's attacks, the difference lies with two aspects: increasing the amount of "hackable things," which will allow for hackers to break into new systems and infrastructures; and increasing the number and volume of attacks, in a way that will make the response more complex.[40] When these aspects are combined with cybersecurity risks that are unique to AI systems, the consequences of a successful attack can be severe. For instance, the U.S. Defense Department[41] mentions the "adversarial machine learning" within which AI systems can be harmed by the "contamination" of data, which includes changing algorithms or making other important adaptations that can completely change a system's purpose.[42] Attacks of this kind in greater quantities and on new infrastructures such as electricity, water, and hospitals may increase the possibility of a crisis.[43] It should be noted that scholars have argued recently that cyberspace has not proven as escalatory as early theorists anticipated.[44]

Second Scenario – Increasing the Speed of Attack: Another scenario stems from the fact that the new generation of improved cyber capabilities in the field of AI will increase the risk of accidental escalation due to the increased speed of the warfare itself. Cyber AI reinforces the effects of current, advanced capabilities, thus increasing speed of combat and reducing the decision-making time frame. Additionally, the high speed of AI cyber tools can allow for an attacker, with relatively limited skills, to exploit a narrow window of opportunity and hack his opponent's cyber defenses.[45]

In conclusion, in all three applications described, there are many possible scenarios for international AI crises. As broadly presented as they are, all of these scenarios are likely to occur, and to the best of this author's knowledge, none of them can be refuted with certainty. Therefore, due to the inability to pinpoint one scenario, the following research will offer a model for raising awareness to early warning signs of an impending crisis.

A Model for Early Warning Signs of an International Crisis in AI

While the above section presented ways in which the first international AI crises might emerge, the article now proposes a model intended to help mitigate the impacts of the crisis once it begins. The goal of the model proposed is to offer decision-makers a means that will provide awareness of the early signs of an international crisis in AI before it occurs. Given the successful identification of an impending crisis, decision-makers would be able to act accordingly in order to reduce tensions and ultimately prevent the crisis.

The Original "Doomsday Clock"

The conceptualization of the model is based on the "Doomsday Clock," which has been maintained since 1947 by the Bulletin of the Atomic Scientists (BAS). The clock attempts to estimate the world's proximity to a global catastrophe that could ultimately endanger the future of humanity. Originally, the clock sought to measure the degree of proximity to a nuclear war in the context of the arms race between the United States and the Soviet Union. The estimation is presented on a clock that measures "minutes to midnight," with "midnight" representing the point at which a nuclear war will break out. It should be noted that the longest distance from midnight was 17 minutes in 1991, after the end of the Cold War. On January 23, 2020, the clock was moved to the closest position so far to midnight, at 100 seconds before midnight. The movement of the dial is influenced by relevant international events,

39 Johnson, "Artificial Intelligence in Nuclear Warfare."

40 Antebi and Dolinko.

41 "The National Artificial Intelligence Research and Development Strategic Plan," U.S. Department of Defense, October 2016.

42 Johnson, "Artificial Intelligence, Drone Swarming, and Escalation Risks in Future Warfare."

43 Johnson, "Artificial Intelligence & Future Warfare."

44 Buchanan; Valeriano, Jensen, and Maness.

45 Johnson, "The AI-Cyber Nexus."

and their influence is determined by experts in the field.[46] The clock has become a universally recognized indicator of the world's vulnerability to catastrophe[47] and has been mentioned extensively in mainstream media, including CNN, CBS News, *The Washington Post*, *The New York Times*, and ABC News. This exposure has led to an increase in public awareness to the danger posed by nuclear weapons, thereby influencing decision-makers' awareness as well. It should be noted that since 2007, the BAS has included climate change considerations in the calculations of the doomsday clock.[48]

A Proposal for an "AI Doomsday Clock"

This article explores the idea behind an AI version of the "Doomsday Clock" that would bring awareness to an impending AI crisis. It should be noted that according to the BAS,[49] AI will eventually mature into a transformative national security technology, on par with nuclear weaponry, which explains this paper's propensity to use the nuclear clock in the AI field. This particular clock will consist of three separate dials, with each dial representing one of the three main applications of AI in the military realm. Each of the dials will operate independently because each dial can individually cause an international crisis. Similar to the original clock, whenever an event occurs that brings us closer to or further from a crisis, the dials will move accordingly; midnight will symbolize the point at which an international crisis in AI will occur. It should be noted that since the future of AI is still vague, and because it is impossible to understand all of the arenas in which AI will be integrated, the model allows us to insert additional dials that will correspond to those new areas. Furthermore, regarding the developing applications of AI, the next section will examine the possibilities of linking one moving dial to another; an application in cyberspace may affect, for example, the nuclear.

Regarding the factors that affect the movement of the clock dials, there are "macro factors," which affect the movement of all three dials, and "micro factors," which affect each of the dials separately. All factors are based on the characteristics of AI.

Macro Factors Influencing an "AI Doomsday Clock":

- **International Events** – International events have a definitive effect on the movement of all three dials. For example, signing an international treaty concerning AI will increase the dials' distance from midnight, while an international military incident that increases tensions will bring the dials closer to midnight.
- **A New Player in the Arms Race** – A new country that acquires military AI capabilities, conducts experiments, and assimilates the systems may disrupt the current balance of world power, thus bringing the clock dials closer to midnight.
- **Actions Below the Escalation Threshold** – Actions and events below the threshold of escalation, which are subjective to and determined by each country, may shorten response times, causing decisions that could lead to a crisis. For example, the repeated use of drones for spying or targeted hits may increase tensions, thus bringing the clock dials closer to midnight.

Micro Factors Influencing an "AI Doomsday Clock":

- **Accelerated Development** – Significant advances in one of the three technological applications of AI may undermine stability and thus bring the dials closer to midnight. That is, as part of the AI arms race, only one of the three dials will move in accordance with the type of technology that has been improved.
- **Placing AI Systems in a New and/or Unstable Environment** – Placing AI technologies in a new or unstable environment may exacerbate existing risk, due to new options for increasing the pace of combat. Thus, depending on the technology installed, the corresponding dial will be moved closer to midnight.

---

46 "Frequently Asked Questions," Bulletin of the Atomic Scientists (BAS), (n.d.), accessed at https://thebulletin.org/

47 John Mecklin, "It is Now Two Minutes to Midnight," *Bulletin of the Atomic Scientists*, January 25, 2018.

48 "Frequently Asked Questions," BAS.

49 Ibid.

- **Expanding or Reducing Collaborations** – Institutionalization or the termination of cooperation between countries, states, or the private sector regarding R&D and the acquisition of systems may move the dials. More specifically, cooperation may disrupt the strategic balance of power, causing the relevant dial to move closer to midnight. Similarly, cooperation that promotes stability will move the dials increasingly apart.

In accordance with the BAS "Doomsday Clock," this paper proposes that the new clock will be proprietary to a public body, which will be responsible for moving the clock dials and thus raising awareness of a potential crisis. A public organization would presumably use publicly available information, and therefore, decision-makers may be less inclined to rely upon it for making policy. However, making the information accessible to the general public may raise public awareness and thus create public pressure to reduce the tension. Another advantage of publishing the information by a public body is that in the AI arena, beside the political decision-makers there are many other players, such as the private sector. Therefore, it is important that the information provided by the clock raises awareness for the decision-makers in these sectors as well.

Conclusion

Given the fact that there exists a wide range of AI military applications, there are also many possible scenarios for an AI international crisis. This article has sought to delineate what the three most probable sources of the first AI crises might be—lethal autonomous weapons systems; nuclear; and cyber—and has proposed one means by which to offer early warnings to decision-makers to stave off an impending AI catastrophe, in the form of the "AI Doomsday Clock."

The model's descriptive nature raises the awareness of an impending crisis and therefore makes it possible for decision-makers to take steps to de-escalate tension. The importance of bringing awareness to the very real possibility of an international crisis in AI lies in the fact that AI is an issue that is not a top priority for most governments, due to the relatively new and innovative nature of the field. However, fast-growing interest in AI among countries around the world reflects the importance of the field and thus the need to bring its potential for creating international crises to the forefront of conversation.

In the future, the complex arms race, which focuses on developing AI technology to assist in decision-making processes, will reach greater maturity. At that point, the success or failure of this model will be determined based on how well and how quickly it manages to alert the world before catastrophic AI events transpire. If the model succeeds, a separate, similar model will be formed in order to detect an impending crisis in the complex arms race specifically. Likewise, given this new model's possible failure, adaptations will need to be made in order to address the true needs of this arms race.