The Potential Impact of Artificial Intelligence on Preventive Diplomacy from a Balance-of-Threat Perspective

The Potential Impact of Artificial Intelligence on Preventive Diplomacy

from a Balance-of-Threat Perspective

Daniel F. Oriesek

A Thesis in the Field of International Relations

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

May 2022

Abstract

When the Peace of Westphalia was signed in the cities of Osnabrück and Münster in 1648, it not only ended the Thirty Years' War but also, and more importantly, laid the foundation for the contemporary concept of territorial sovereignty and created a system of independent nation-states. These states mutually agreed not to interfere in each other's domestic affairs while maintaining order by checking each other's ambitions through the creation of a general equilibrium of power.

An essential component in enabling such an equilibrium was the development of a method of influencing the decisions and behavior of the governments and peoples of nation-states so that issues of concern could be resolved by means of dialogue, negotiation, and other measures short of war or violence. This development represented the birth of modern diplomacy.

Over the ensuing centuries, the concepts of equilibrium and balance of power were further developed until Kenneth Waltz integrated and structured these ideas as part of his conception of structural realism. According to this theory, an anarchic self-help system and shifts in the relative distribution of capabilities mean that balances of power recurrently form in the international system and drive international relations (Waltz, 1979).

A few years later, Stephen M. Walt expanded this idea into what he referred to as *balance of threat.* Walt suggests that states balance not only against actual power but also against real or perceived threats. From Walt's point of view, balance-of-power theory is not incorrect but rather incomplete. Walt acknowledges that power is a critical factor in the level of threat posed by a state but also suggests that a threat must also include other elements, such as geographic proximity, offensive capabilities, and perceived intentions (Walt, 1985, 1987).

As technical developments, particularly in the form of digitalization and artificial intelligence (AI), transform the way in which people work, and as computer processing

power is approaching—if it has not already exceeded—its human equivalent, we need to consider how such factors will change the way diplomacy is conducted in the future.

Because the fields of both AI and diplomacy are relatively broad, this thesis focuses on the field of preventive diplomacy and, within this field, the actions that are directly linked to monitoring and maintaining a balance of threat. With regards to AI, this thesis adopts the system of categorization put forward by Nick Bostrom and focuses on the first category, artificial narrow intelligence (ANI), while also speculating as to what may lie ahead when we reach the second category, artificial general intelligence (AGI), also known as human-level AI (Bostrom, 2014).

This thesis considers the following question: "How might AI change the possibilities and conduct of preventive diplomacy in terms of managing balances of threat for more stability in the world?" Based on the individual assessments of a diverse pool of experts in the fields of diplomacy, security policy, foreign affairs, and AI, this thesis concludes that ANI will make significant contributions to the field of preventive diplomacy from the perspective of better managing balances of threat.

In sum, the experts agree that the potential benefits of AI outweigh the potential risks. The majority of experts believe there are four areas where ANI can make contributions: 1) *monitoring*, where ANI can contribute to providing a more complete and timely picture of potential changes in the balance of threat and could increasingly be used to develop a system of early indicators to warn of ongoing changes in the balance; 2) *validation*, where ANI can contribute to validating and improving the quality of information regarding potential changes in the balance of threat, thus increasing the trustworthiness of such information; 3) *anticipation*, where ANI can contribute to anticipating the reactions of players in a balance-of-threat system and thus could provide valuable decision-making support; and 4) *solution finding*, where ANI can contribute to detecting relevant players and their relations to one another in a given context as well as assist in detecting seemingly unrelated issues that could expand room for negotiations and solution finding.

Experts were divided as to whether ANI could make contributions in other areas, largely due to doubts that current-level ANI is sufficiently advanced for such tasks. The experts are also divided over whether ANI alone could improve the overall effectiveness of preventive diplomacy, as they fear that political forces and individual interests might work against such efforts.

The experts do not seem particularly concerned about the potential of AI to replace diplomats or security professionals within the next decade, but they are concerned that AI is potentially biased and could be manipulated. The experts see the greatest danger with AI as its potential to replace human decision making or the possibility that decision makers may become overly reliant on results generated by AI.

Therefore, the technical possibilities afforded by current ANI should be used to obtain better validated information more quickly, to better anticipate possible (re)actions by stakeholders, and to identify previously unknown paths to finding solutions. Relying on ANI for these purposes could significantly contribute to identifying potential problems in the balance of threat early on so that the balance can be restored diplomatically before arms races, unhappy alliances, or even preemptive wars ensue. Simultaneously, however, we must remain critical of ANI's proposals and at least conduct a plausibility check before incorporating its recommendations into our decisions.

Dedication

To all my brothers and sisters in the armed forces, no matter where you are in this world—thank you for your service and sacrifice and for upholding the balance of threat needed to keep us all safe.

Acknowledgments

I am very grateful to all the people who contributed to the completion of this research.

My gratitude goes to Doug Bond, my research advisor, who helped me structure my initial ideas, found me a great thesis director, and was always there for me when I needed him.

I would also like to thank Bruce Schneier, my thesis director, who, with his concise but razor-sharp suggestions, helped me stay grounded and focused on what is feasible today without getting lost in wishful thinking and science fiction, as is so easy to do when talking about AI.

Thanks go also to the Geneva Center for Security Policy and Colonel (GS) Christian Bühlmann for giving me the opportunity to present and test my initial thoughts regarding this work in an online forum with academics and practitioners from around the globe. Thank you also for assisting me in reaching out to potential survey participants.

Thanks to my colleagues at work, my fellow students at Harvard, and members of the diplomatic, security, foreign policy, and academic community for participating in the survey or bouncing thoughts and ideas about this project with me.

Last but not least, my thanks go to my family, whose encouragement, love, and support during difficult times gave me the strength to turn all the negativity into the positive energy necessary to complete this fascinating journey at Harvard.

Table of Contents

List of Tables

List of Figures

Glossary of Abbreviations

AI	artificial intelligence
AGI	artificial general intelligence
ANI	artificial narrow intelligence
ASI	artificial superintelligence
CIA	Central Intelligence Agency
COMINT	communications intelligence
cps	calculations per second
ELINT	electronic intelligence
EXCOMM	Executive Committee of the National Security Council
GEOINT	geospatial intelligence
HUMINT	human intelligence
IC	integrated circuit
ICBM	intercontinental ballistic missile
IMINT	imagery intelligence
MASINT	measurement and signatures intelligence
MOFSET	metal-oxide-semiconductor field-effect transistor
MRI	magnetic resonance imagery
OSINT	open source intelligence
PHOTINT	photo intelligence
PI	photointerpreter
RBM	results-based management
RSI	recursive self-improvement
SIGINT	signals intelligence
TELINT	telemetry intelligence
UN	United Nations

UNSC	United Nations Security Council
USAF	United States Air Force
WBE	whole brain emulation

Chapter I

Introduction and Overview

The overarching goal of this thesis is to investigate the potential impact that recent and upcoming developments in artificial intelligence (AI) will likely have on the field of preventive diplomacy from a balance-of-threat perspective. To achieve this objective, my research began with a review of underlying concepts in the literature, an investigation into the current state of AI technology, followed by the development of an initial set of hypotheses. Based on these hypotheses, specific questions were formulated and programmed into an online survey, which experts in the fields of diplomacy, security policy, foreign affairs, and AI were invited to complete. Through follow-up interviews, particular aspects of individual responses were further investigated before these collective expert insights were consolidated in this thesis.

Starting Point: Balance of Power/Balance of Threat

Today's world order still follows the basic logic of balance of power. Over time, it appears that the balance-of-power mechanism has moved away from having a random, unstable, and adversarial nature and taken the form of a spontaneous dynamic formed between states seeking to oppose the most powerful actor in the system; moving instead toward a more sustainable, conscious approach in which balance is the result of the common goal of the states involved in maintaining it.

While the theory of *offensive realism* advocates seeking power and influence to achieve security through domination and hegemony, *defensive realism* argues that the anarchical structure of the international system encourages states to maintain moderate and reserved policies to attain security (Lobell, 2010). Based as it is on a conscious and collective approach to achieving a balance of power, defensive realism should be more common within the international system of states.

Balance-of-power theory falls short because it focuses on power as defined by military capability. In doing so, it omits other critical components of power, such as geographic proximity, offensive capabilities, and perceived intentions. This shortfall is corrected by applying *balance-of-threat theory*, which not only expands upon the elements of power but also separates the notion of power from the notion of threat (Walt 1985).

A Necessary Lubricant: Diplomacy

Conflicts originate and escalate on a continuum between peace and war. Diplomacy, and particularly preventive diplomacy, is an important mechanism by which, through dialogue, negotiation, and other measures short of violence, to resolve issues before they escalate into open conflict or war (Encyclopedia Britannica, n.d.).

Preventive diplomacy can be applied in bilateral settings as well as, more importantly, in multilateral settings, where international organizations such as the United Nations (UN) can play key roles as mediators and facilitators. This form of diplomacy has proven quite effective in preventing conventional interstate conflicts, but has been less successful with respect to civil conflicts.

The UN, through the United Nations University, developed a framework to measure *ex post* the success of preventive diplomacy. Thus far it has identified six success factors on the basis of several case studies in which the framework has been applied. However, while the framework serves well for assessing historical situations and the degree of success of preventive diplomacy, it has not yet provided practical applications in terms of foresight (e.g., allowing the prevention or de-escalation of potential crises before they happen).

A Potential Game Changer: Artificial Intelligence

The field of AI has made significant advances since it was introduced over 60 years ago. Originally, AI was applied to a small set of use cases, such as winning at a chess game, and was for the most part based on mathematical modeling and rule-based automation. With improvements in machine learning—particularly since the development of deep learning (Lecun, et al., 2015; Singh, 2017) and the capacity to process vast amounts of data (also referred to as "big data")—the applications of AI have become relevant to various other fields, and AI is now capable of mimicking or facilitating, if not supplanting, human interaction. In particular areas, advanced AI algorithms can already perform tasks at far higher speeds, with greater reliability, and at much lower cost than humans.

Kathleen Walch (2019) attempted to categorize the many applications of AI by assigning different AI use cases to seven patterns. The patterns mentioned by her can be summarized as follows:

1) Hyperpersonalization (AI developed for treating each customer as an individual);

2) Autonomous systems (mechanical systems controlled by AI with the goal of reducing the need for manual labor);
3) Predictive analytics and decision support (using machine learning and other cognitive approaches to AI to understand how data on past or existing behaviors can facilitate the prediction of future outcomes or help humans make decisions about future outcomes based on these patterns);
4) Conversational/human interactions (AI that allows machines to communicate as humans would);
5) Identifying patterns and anomalies (AI that is particularly good at detecting patterns or identifying anomalies and outliers);
6) Recognition systems (AI that is particularly good at finding, classifying, and identifying images, video, audio, or objects);
7) Goal-driven systems (AI that is good at learning the rules of games or other forms of interaction and then learning how to excel at them).

According to Walch, these seven patterns can be applied individually or in various combinations depending on the specific solution to which AI is being applied. Some examples of solutions are:

- *Chatbots* that facilitate the process of data collection through standardized and guided conversations (voice, text) with human subjects. In doing so, they help bridge the human-machine interface.
- *Image recognition systems* that are capable of recognizing objects or faces on the basis of other datasets, which can range from datasets of similar fashion styles to mug shots in police databases. Another important application here is detection of

change, such as recognizing changes in aerial or other imagery over periods of time. This is highly useful for quickly detecting relevant changes that might indicate the massing of troops in particular areas or the creation of new installations (such as nuclear launch sites) or changes along a roadside hinting at the presence of buried improvised explosive devices.

- *Improved communication systems* that can target content and channels and match recipients in need of particular information with smaller quantities of more relevant information. Such systems contribute to a reduction in information overload, helping decision makers to focus, improve their productivity, and engage in better-informed decision making. Language translators—particularly when coupled with voice recognition technology—can elevate intercultural communication to a new level, representing another example of the use of AI in this area. Simple examples include spam filters and ad blockers.
- *Screening support systems* that are useful in areas such as credit scoring, pre-selection of candidates for job interviews, security clearance screening, and processing college applications. Such systems leverage technology to free human experts from mundane and time-consuming routine work so they can focus on more value-adding parts of their jobs, such as conducting candidate interviews. Other examples can be seen in intelligence-gathering applications, where algorithms screen all available information on the internet and are to some degree increasingly able not only to find relevant articles based on key words but also to understand the essence of the content of these articles in multiple languages. This is a task that can now be highly automated, with systems carrying out—almost in

real-time—work that previously required armies of analysts and language experts working for days. As a consequence, analysts can now focus on relevant discoveries and insights and the meanings thereof rather than on first assembling the necessary information.

- *Decision support systems* used in healthcare, where AI can provide clinical decision support and help doctors to more quickly obtain reliable and consistent second opinions. The objective is not to replace the human decision maker but rather to increase the decision-maker's confidence and reduce stress and anxiety when a decision needs to be made. A case in point, with solutions once again incorporating image recognition and detection technology: the screening of X-rays or magnetic resonance imagery (MRI) for anomalies such as cancer.
- *Cyber or physical security systems* that use AI to detect vulnerabilities or anomalous user behavior in corporate systems or within restricted perimeters. While attackers may still infiltrate systems and perimeters, these systems quickly detect malicious and unusual activities so defensive measures can be taken before attackers cause any major damage.
- *Logistics and supply chain systems* that incorporate AI include autonomous trucks, taxis, cars, drones, robotic picking systems, and intelligent storage systems, which allow logistic fulfillment to be carried out seven days a week and thus significantly shorten delivery times.
- *Anticipating the behavior* of all kinds of individuals and organizations based on past behavior and with reference to the behavior of similar individuals or organizations. This is already conducted in areas such as parole decisions, risk

assessment, and sports betting. This application has significant potential but is also subject to major potential pitfalls, such as building bias into an algorithm by having it learn from biased datasets.

While all these applications are still considered to fall into the category of artificial narrow intelligence (ANI), technology has now reached a point where supercomputers can match the pure processing power of the human brain. While teaching a machine to make intelligent, human-like decisions is a significant challenge, there are at least three promising approaches that could lead to the development of systems that possess human-like intelligence (i.e., artificial general intelligence (AGI)) within the next 20 years.

A Solid Basis: The Concepts of Balance of Power, Balance of Threat, and Diplomacy

While the basis for this research, namely balance-of-power theory, balance-of-threat theory, and diplomacy, are quite well covered in the literature, research on the digitalization of diplomacy or the application of AI in a diplomatic context is still nascent. While research on the former topic focuses mostly on new realities for the conduct of diplomacy and diplomats in light of the prevalence of social media and the ever-more-direct communication and power of influence groups, research on the latter fields is focused on diplomatic efforts necessary to regulate the use of AI, with a special focus on ethics and autonomous weapon systems.

Significant technological advances in AI now promise to provide, if not complete solutions, at the very least support for dealing with increasingly complex problems. This is particularly the case with respect to the ability to access and connect vast amounts of data, recognize patterns, ascertain the connections between seemingly unrelated issues, and better understand sometimes irrational decisions on the basis of learnings from the past, all of which has potential for strengthening the impact of preventive diplomacy and maintaining relative stability in the international system.

Structure of This Thesis

Chapter II summarizes the historical events that are commonly referred to as the Cuban Missile Crisis. I chose this example because, in my view, it sets the contextual stage for this research by providing a tangible illustration of how a perceived, possibly imaginary, threat can set in motion balance-of-power mechanisms. It also exemplifies the value of preventive diplomacy, which in this case was able to resolve what might otherwise have been a fatal nuclear exchange.

Chapter III presents a literature review, and goes into more detail concerning the concepts that form the basis against which the application of AI is assessed. This review includes the concepts of balance of power, balance of threat, diplomacy, and preventive diplomacy, with a particular focus on the role of the UN and attempts to measure the effectiveness of preventive diplomacy.

Chapter IV describes the research methodology and lays out the thesis hypotheses, which serve as the basis for the questions included in the expert survey.

Chapter V presents an analysis of the survey results, summarizes the findings, and reflects on the limitations of this research.

Finally, Chapter VI provides the conclusion, makes some recommendations for future work on the topic and offers some final thoughts.

I am hopeful that this thesis will encourage researchers and policymakers alike to work toward adopting this technology and take advantage of the opportunities it affords while remaining aware of the potential dangers it presents.

Chapter II

Context: The Cuban Missile Crisis

Let us never negotiate out of fear. But let us never fear to negotiate.
—John F. Kennedy

Between 1958 and 1960, in the midst of the Cold War and with John F. Kennedy running for president of the United States, the term “missile gap” was coined and became a key issue in Kennedy’s election campaign. The term captured the perceived superiority of the number and power of the USSR’s nuclear missiles in comparison to those of the US.

Although it has become clear over time that such a gap never existed, the perception of its existence emerged based on excessive figures provided by the United States Air Force (USAF) and the Central Intelligence Agency (CIA), as well as exaggerated estimates made by the Gaither Committee (Renshon, 2009). This committee, named after its chairman, H. Rowan Gaither, was tasked in 1957 by President Dwight Eisenhower with assessing the USSR’s strength, and creating a strategy for the US to strengthen its defensive systems and better prepare for a potential nuclear attack (Snead, 1997).

In 1961, the USSR had only four intercontinental ballistic missiles (ICBMs) of the R-7 *Semyorka* type, a derivative of the missile that launched the *Sputnik* satellite into space on October 4, 1957. By October 1962, the actual number of Soviet ICBMs could have been around a dozen, but some US intelligence sources at the time estimated the number to be as high as 75 (Correll, 2005).

Although the possibility of the USSR overpowering the US in terms of ballistic missiles was not realistic, the perception of this threat alone sufficed to motivate President Eisenhower in April 1958 to instruct the Department of Defense to plan the deployment of three squadrons (45 missiles) of Jupiter (SM-78/PGM-19) medium-range ballistic missiles to France. After French President Charles de Gaulle refused to allow the US to transport and station the missiles on French territory, the missiles were instead deployed to Italy in 1959 (two squadrons and 30 missiles) and Turkey (one squadron and 15 missiles, with one more squadron planned for future deployment (US Army, 2004)).

The 45 Jupiter missiles deployed to Italy and Turkey were perceived by the USSR as indicating a US forward strike capability (the missiles had an effective range of 2400 km/1491 mi), meaning the missiles stationed in Italy could hit the Soviet Union, its satellite states, and armed forces. The missiles based in Turkey were capable of striking Moscow and other major Russian cities directly (Missile Defense Project, 2018).

In response to the perceived threat posed by the Jupiter missiles to the USSR and in light of the failed 1961 invasion of Cuba staged by exiled Cubans and the CIA in the Bay of Pigs, Soviet First Secretary Nikita Khrushchev agreed to Cuban Prime Minister Fidel Castro‘s request to station nuclear missiles on the island to deter a future invasion. In the summer of 1962, the secret construction of a number of missile launch facilities began.

The US detected the Soviet/Cuban construction efforts on October 14, 1962, following an over-flight of Cuba by a U-2 reconnaissance plane (Chayes, 1974). The US subsequently perceived the missiles stationed in Cuba as indicating a forward strike capability on the part of the USSR and as posing a direct threat because those missiles allowed the USSR to effectively target most of the continental US.

What followed has come to be referred to as the Cuban Missile Crisis, also known as the October Crisis of 1962. The event lasted for a little over a month: from October 16, when President Kennedy called the first meeting of the Executive Committee of the National Security Council (EXCOMM), until the formal removal of the sea blockade around Cuba on November 20. The most critical period consisted of the 13 days from October 16–28, when after a series of US actions signaling resolve, a promise from the US not to invade Cuba, and secret negotiations over the removal of the Jupiter missiles from Turkey, the Soviets finally agreed to withdraw the missiles from Cuba (Stilwell, 2018). This confrontation in the fall of 1962 is often considered to be the closest the Cold War ever came to escalating into full-scale nuclear war (Scott & Gerald, 2015).

The example of the Cuban Missile Crisis illustrates the conceptual point of departure of this research, namely the interconnection between the concepts of balance of power, balance of threat, and diplomacy (as a tool to prevent further escalation into a full-blown war). While the doctrine of mutually assured destruction between two opposing global blocks served well during the Cold War to ensure relative global stability, the situation has fundamentally changed and continues to evolve. With the end of the Cold War, the global balance of power shifted from a bi-polar (US–Soviet Union) order to a unipolar one (US) and subsequently to an increasingly multipolar (US, China, Russia, and other regional powers) order of great powers (Baron, 2013; Brooks & Wohlforth, 2008).

Within such developments, not only does the club of nuclear powers grow more extensive (nine countries currently have nuclear weapons: China, France, India, Israel, North Korea, Pakistan, Russia, UK, US), but new technologies such as quantum computers and hypersonic weapons have the potential to alter existing balances of power and balances of threat. While hypersonic weapons may not have the potential to alter the strategic balance of power, they can cause serious uncertainties at the theater level, such as by challenging the presence of US aircraft carriers in the South China Sea or the reaction times of defendants in general (Kunertova, 2021). The threatening behavior, armaments, and/or open nuclear aspirations of some states can cause other states to perceive greater threats, which can cause belligerent states to react (e.g., by arming themselves, exiting arms reduction treaties, or seeking nuclear weapons) and thus affect the global balance of threat.

To maintain relative stability in the international system of nation states and to mitigate the risk of escalation into open conflict, preventive diplomacy, whether executed bilaterally or within the context of multilateral organizations such as the UN, will likely increase in importance and may well benefit from advancements in the field of AI. However, before a framework can be assembled in which the problem can be studied in more detail, the literature on the concepts of balance of power, balance-of-threat, diplomacy, preventive diplomacy, and AI needs to be reviewed.

Chapter III

Literature Review

There is abundant literature on the concept of balance of power. As such, several definitions of this concept have been formulated.

The Geopolitical Backbone: Balance of Power

In their 1966 collection of essays titled “Diplomatic Investigations: Essays in the Theory of International Politics,” Herbert Butterfield, Martin Wight, and H. Bull (1966) identified nine different ways in which the balance of power concept has been used. All these definitions can be classified on the basis of the principle of intentionality as outlined by Federico Chabod (in Woolf, 2002) and Morgenthau (1972).

Following this principle, two conceptions of balance of power emerge. The first can be characterized as adversarial (Little, 2007), and assumes that an equilibrium of power is the outcome of a spontaneous dynamic within a system of states. It is the threat of hegemony that promotes the formation of leagues and alliances created with the objective of opposing the most powerful actor in the system. The resulting balance is the natural but unstable result of the constant struggle of all participants for survival.

In contrast, the second conception considers a balance of power to be the result of a conscious and common goal of the states involved in the international system. Under this approach, denominated associational great powers seek to create a stable mechanism that will ensure international security through cooperation and, if necessary, war. This approach also implies that each participant in the international system is constantly reflecting on the power dynamics in the system and that any actor involved in tipping the scales does so while theoretically reflecting upon the potential outcome and reaction of the great powers (Little, 2007).

In 1648, two peace treaties were signed in the Westphalian cities of Osnabrück and Münster to end the Thirty Years' War. These treaties, today collectively referred to as the Treaty of Westphalia (originally entitled the "Peace Treaty Between the Holy Roman Emperor and the King of France and Their Respective Allies"), laid the foundation for the contemporary concept of territorial sovereignty and created a system of independent nation-states. These nation-states mutually agreed not to interfere in each other's domestic affairs and to maintain order by checking each other's ambitions through establishing a general equilibrium of power. Based on these characteristics, the Treaty of Westphalia could be defined as a conscious and plural attempt to realize a European balance of power (Morgenthau, 1972).

However, this view was not shared by all. Among others, Osiander (1994) argues that there was no systematic thinking, as laid out in the second conception of balance of power, during the negotiations leading to the Treaty of Westphalia. While the Treaty provided some stability in the short term for war-ravaged Europe, it did not fundamentally change the medieval system. The conservative approach of the great powers of Europe led to the preservation of the status quo and existing structures rather than the emergence of a new shared international order. Post-Westphalian Europe consisted of a series of nation-states but was not yet a society of states. While nation-states recognized each other as autonomous entities and shared some common rules, the contemporary procedures applied within the group of nation-states were still set out to restore order inside the Holy Roman Empire, not to establish a wider international community (Osiander, 1994).

The War of Devolution (1667–1668) and the rise of Louis XIV of France with his aggressive foreign policy, provoked a reaction from the dominant European powers and led to alliances established in order to avoid continental hegemony. The League of Augsburg subsequently united almost all the leading European powers against France in the Nine Years' War (1688–1698).

However, the ultimate crisis for the system of nation-states under the Westphalian order came in the form of the Spanish Succession, which unless otherwise resolved, could have resulted in one monarch ruling the kingdoms of both France and Spain. While Louis XIV had initially agreed to divide the Spanish territories between France and the Holy Roman Empire, he subsequently changed his mind and attempted to incorporate the Spanish throne into his existing kingdom. This was unacceptable to the other nation-states in the system, and once again war was declared. The election of Charles VI of Habsburg as Holy Roman Emperor in 1711 became a catalyst for peace negotiations, as the major powers realized that the internationalization of succession questions going forward was absolutely necessary to avoid uncontrollable power shifts owing to succession schemes (Luard, 1992).

Unlike the Treaty of Westphalia in 1648, the Treaties of Utrecht (also commonly known as the Peace of Utrecht), which were negotiated and signed between April 1713 and February 1715, emphasized the aspect of cooperation and explicitly included the concept of maintaining a balance of power as a supreme, general interest for the common good. As a consequence, the European nation-states agreed to a more even distribution of power. France renounced the Spanish throne, and all parties accepted several changes of territories. The treaties ended the bipolar nature of the European system of nation-states up to this point and replaced it with a set of more equal actors under the supervision of Great Britain, the real winner of the Spanish Succession (Clark, 2007; Luard, 1992; Osiander, 1994).

The Peace of Utrecht laid the foundation for a phase of relative tranquility and, at least until the outbreak of the French Revolution in 1789, wars ceased to be a means through which to achieve supremacy, becoming instead an instrument by which to enforce security if necessary. The new society of states would now be governed by great powers. This fluid and dynamic society of states evolved in the "golden age" of the concept of balance of power (Morgenthau, 1972).

Other scholars have adopted the idea of balance of power being a conscious and common goal of the states involved in the international system (e.g., Earle, 1943; Hume, 1752; Rucellai, 1733; Von Gentz, 1806) over the centuries following the Peace of Westphalia. However, it was Waltz who most profoundly affected international relations theory by integrating and structuring the idea of balance of power as part of his conception of *structural realism*, also referred to as neorealism (Waltz, 1979). According to Waltz, power (i.e., military capability) is the most important factor in international relations, and the nature of the international structure is defined by the ordering principle of anarchy (i.e., the lack of a formal central authority) as well as by the distribution of capabilities measured by the number of great powers within the system (Waltz, 1979).

The structural realist theorists derived from Waltz's structural realism can be divided into two opposing camps with competing assumptions and policy prescriptions: those advocating for offensive realism (Brooks, 1997; Frankel, 1996; James, 2002) and those advocating for defensive realism (Lynn-Jones, 1995; Miller, 1996; Nexon, 2009; Schweller, 2003; Snyder, 2013; Taliaferro, 2000; Walt, 2002).

Offensive realists seek power and influence to achieve security through domination and hegemony. Defensive realists believe that within an anarchical international system, states have an incentive to maintain defensive policies to ensure their own security. According to defensive realists, aggressive expansion as advocated for by their offensive counterparts upsets the balance of power and reduces the willingness of states to conform to the balance-of-power theory. Defensive realists do not deny the reality of interstate conflict or the temptation for state expansion, but they do believe that these incentives are the exception rather than the rule. Defensive realism points toward "structural modifiers," such as the security dilemma, geography, and elite beliefs and —perceptions, to explain the outbreaks of conflict (Lobell, 2010).

For the purpose of this thesis research, it is assumed that the majority of nation-states in the international community act according to defensive realism. They have a sustained underlying interest in maintaining and, if necessary, restoring an equilibrium of power in the international community—even if this means taking collective offensive measures against nation-states that work against such common interest.

The Cause of Reaction: Balance of Threat

Only a few years after the appearance of Waltz's work on structural realism, Stephen Walt expanded on the idea of the balance of power in his 1985 article "Alliance

Formation and the Balance of Power" and subsequently in his 1987 book *The Origins of Alliances*. In what Walt calls a *balance of threat*, he suggests that nation-states balance themselves not only against actual power but also against real or perceived threats. From Walt's point of view, the balance-of-power theory is not wrong but, rather, incomplete. Walt acknowledges Waltz's view that power (defined as military capability) is a critical factor in the level of threat posed by a state. However, Walt suggests that threat must also include other elements, such as geographic proximity, offensive capabilities, and perceived intentions.

As the example of the Cuban Missile Crisis shows, in a confrontation between great powers, relative power based on actual numbers and capabilities certainly matters. But what is more important is the perceived threat, which originates from assumed (and not always accurate) numbers and the perceived capabilities (e.g., the capability to directly strike major cities with nuclear warheads or the assumed effect of hypersonic weapons) of the potential opponent. Balance-of-threat theory thus modified balance-of-power theory, particularly structural realism/neorealism, by separating power from threat. Realist analyses prior to this newly introduced perspective had assumed that greater power (i.e., an augmentation of military capabilities) on the part of an opponent would go along with an increase in offensive intentions.

Walt argues that there is insufficient empirical evidence to allow for such a general statement, and that balance-of-threat theory offers a better explanation. Balance-of-threat theory would explain examples of nation-states that are rising in terms of power but do not display offensive intentions, resulting in other nation-states not perceiving them as threats and consequently not feeling the need to balance against them. The US during the Cold war represents a case in point. Although the US was more powerful than the USSR, contrary to the balance-of-power theory, more states chose to ally with the US than with the USSR because the US displayed intentions that were less aggressive than those of the USSR. Even after the USSR collapsed and the US became the undefeated unipolar superpower, it remained formally allied with NATO, Japan, South Korea, and several other countries. It seemed likely that were the US to withdraw its forces, this would trigger requests for a continued US presence. Furthermore, during this period, counterbalancing coalitions, as predicted by the balance-of-power theory, did not materialize (Walt, 2009).

This research makes use of balance-of-threat theory because it seems to have a higher predictive validity. Also, the concept of diplomacy ties in nicely as a means of preventing misunderstandings and de-escalating perceived threats before they turn into real threats and conflict.

Diplomacy: Lubricant of the International System

Diplomacy is defined in the *Encyclopedia Britannica* as the established method of influencing the decisions and behaviors of foreign governments and peoples through dialogue, negotiation, and other measures short of war or violence. Modern diplomatic practices have their roots in the establishment of the European system of nation-states following the Thirty Year's War. Important milestones in the further development of modern diplomacy were the Treaty of Westphalia, which established the first balance-of-power system based on nation-states; the Peace of Utrecht, which established a more conscious international society with great powers assuring the balance of power; and the Congress of Vienna in 1815, which established a system of diplomatic rank and protocol.

Thus, the basis for diplomatic conduct and rights was established in mid-17th-century Europe. After these diplomatic practices and rights were adopted throughout the rest of the world, they were formalized by the 1961 Vienna Convention on Diplomatic Relations, which is still valid today (Black, 2010).

As history has shown, there is a close interconnection between peace, diplomacy, and war. Prussian general and military theorist Carl von Clausewitz (1832) noted: "War is an act of violence intended to compel our opponent to fulfil our will" (in Paret & Howard, 2008, p. 75). He also stated: "War is not an independent phenomenon, but the continuation of politics by different means" (in Paret & Howard, 2008, p. 7). Both of von Clausewitz's statements imply that there is a smooth transition between non-kinetic and kinetic forms of conflict resolution and vice versa. Therefore, I believe that one of the most important roles—if not the most important role—of diplomacy is the resolution of issues and disputes as a means of avoiding war.

More recently, research and articles have addressed what is called "digital diplomacy," or "e-diplomacy," or "cyber diplomacy" (DiploFoundation, n.d.). The sources focus on what digitization means in terms of interest-group involvement, and the challenges for government representatives with regard to communication through social media and other digital platforms.

From a diplomatic perspective, AI is addressed primarily as an issue that calls for a structure of governance. This issue seems particularly urgent when regulating the development and deployment of autonomous weapons systems (DiploFoundation, 2019). However, that topic is outside the scope of this thesis. The following sections consider the roots of classical diplomacy.

Preventive Diplomacy: Avoiding Open Conflict

While preventive diplomacy can be broadly understood as the use of diplomatic means to resolve disputes between two nation-states before they escalate into war, this bilateral perspective does not go far enough. With the founding of the League of Nations in 1920, preventive diplomacy became a consideration for multilateral organizations as well as nation-states. With the horrors of World War I still present in the minds of participating leaders, the stated goal of the League of Nations was to prevent future wars through collective security and disarmament and to settle international disputes through negotiation and arbitration (League of Nations, 1919). The effectiveness of the League of Nations was certainly flawed given that it did not prevent World War II, but its existence ultimately led to the founding of its successor organization, the United Nations, in 1945.

Preventive Diplomacy: Is the United Nations Effective?

Whether the UN has proved successful is widely debated. Shortly after its founding, Leland Goodrich observed in 1947:

> The United Nations for what it quite properly is, a revised League, no doubt improved in some respects, possibly weaker in others, but nonetheless a League, a voluntary association of nations, carrying on largely in the League tradition and by the League methods. (p. 21)

Figure 1 shows that, notwithstanding the founding of the UN, the overall number of conflicts in the world has increased.

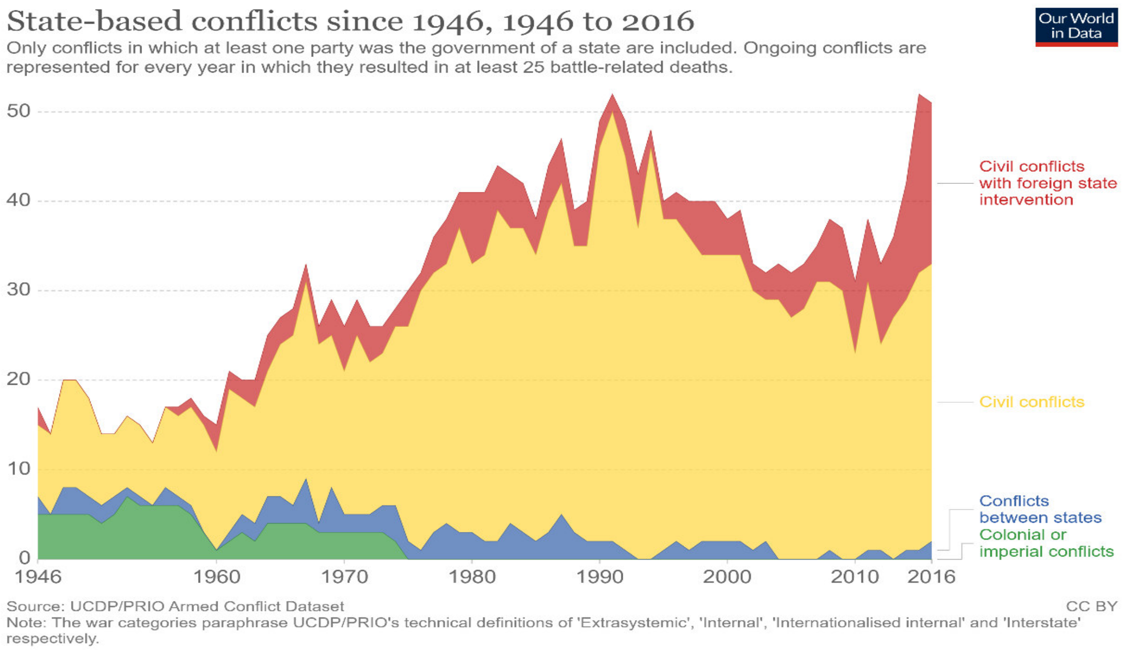

Figure 1. State-Based Conflicts, 1946 to 2016.

Source: UCOP/PRIO Armed Conflict Dataset.

Inspection of Figure 1 shows that the number of civil conflicts (yellow), and the number of civil conflicts involving the intervention of a foreign state (red), have risen sharply, while other state-based conflicts, such as colonial or imperial conflicts (green), have disappeared. In addition, the number of conflicts between nation-states has also declined (blue). An optimistic assessment would indicate that the UN does play a useful role in the world and that it does effectively prevent outright conflict between nation-states—the original purpose of the UN. This achievement seems even more remarkable in light of the fundamental shifts in global power structures that have occurred since 1945.

However, critics fault the UN for failing to address the challenges of rising civil conflicts with or without foreign state involvement. They argue that the UN needs fundamental reforms, citing some of the UN's most criticized failures:

- 1948: Israeli occupation of Palestine, ongoing
- 1948 Kashmir dispute, still unresolved
- 1978-1979 genocide in Cambodia
- 1991 civil war in Somalia, ongoing
- 1994 genocide in Rwanda
- 1995 Srebrenica massacre
- 2003 Darfur conflict in Sudan, ongoing
- 2003-2011 invasion of Iraq
- 2011 Syrian civil war, ongoing
- 2013 humanitarian crisis in South Sudan, ongoing
- 2014 civil war in Yemen, ongoing
- 2017 Rohingya crisis in Myanmar, ongoing.

What are some reasons for the ineffectiveness of the UN when dealing with civil conflict or regional disputes? One concerns the way in which the UN Security Council (UNSC) was established and governed. Its five permanent members (China, France, Russia, UK, US) all have the power to veto any proposed action (Carswell, 2013). This causes turmoil when dealing with topics that run counter to the interests of any of the permanent members. A recent example is Russia, which continually vetoes any plan to assist Syria in peacefully ending its ongoing conflict because such peace would run counter to Russia's interests in the region and could impede its close ties with the Syrian regime led by Bashir Assad.

The UN also struggles because its resources are limited. The UNSC is hesitant to become entangled in complex conflict situations, acknowledging that it simply cannot become involved in every conflict around the world. This is in part because the UNSC can block any proposal that runs counter to the interests of any of the permanent members. But it is also because UNSC members in general are hesitant to take action in unclear conflict situations. For example, in the case of the 1994 Rwandan genocide, the UNSC refused to send in more peacekeepers, and did not allow UN troops in the country to engage with the perpetrators unless doing so was necessary for the troops' own self-defense.

Because nation-states interact in an anarchic environment in which the UN is not a "world government" with supreme executive powers, and because nation-states have agreed not to interfere in each other's internal affairs, the actions of the UN are, by design, limited. This means that although the UN is undergoing a number of reforms that may make it more effective and quicker to act also militarily, its global influence and ability to act will always be limited unless profound changes in the governance of the UN, particularly that of the UNSC, take place (Heinbecker & Goff, 2005).

Preventive Diplomacy: The Origin of Crimes Against Peace

The US recognizes 195 countries in the world, of which 193 are part of the UN (except Vatican City and Palestine). The UN Charter prohibits member states from attacking other UN member states. This should theoretically suffice to end all wars between nation-states, but unfortunately reality does not always accord with this charter.

As the world is imperfect, the concept of a "crime against peace," reflected in the Nuremberg Trials that followed World War II, comes into play in such situations. This concept legitimizes collective action against states that violate peace (i.e., under this premise, it is acceptable for state actors to interfere in the internal affairs of a nation-state and thus annul the common agreement of noninterference dating back to 1648). A crime against peace takes place when a nation-state involves itself in "starting or waging a war against the territorial integrity, political independence or sovereignty of a state, or in violation of international treaties or agreements." At Nuremberg, this was held to be the crime that makes all war crimes possible (UN War Crimes Commission, 1946).

This definition raises a number of questions. For example, how would a civil conflict within a country's frontiers qualify as a "crime against peace" when the territorial integrity, political independence, or sovereignty of another nation-state is not affected? What if the respective state has never signed the international agreement that is allegedly being violated? How can a party to an internal civil conflict (e.g., an armed group) violate international treaties or agreements if it is not even authorized to sign any documents on behalf of the state in which it operates?

Despite its institutional shortcomings and often unclear legal basis for dealing with internal civil conflict, the UN can and should play a key role in preventing, containing, and hopefully assisting in ending different types of conflicts in the world. Even if it cannot and should not become involved in every conflict, the UN can take a leadership role within the international community and provide both established processes for resolution and an indispensable platform for diplomatic efforts in the form of dialogue, negotiations, conferences, arbitration, and mediation, as well as for the codification of international law.

Preventive Diplomacy: UN Definitions

Six months after taking office in January 1992, UN Secretary-General Boutros Boutros-Ghali issued a report titled, "An Agenda for Peace: Preventive Diplomacy, Peacemaking and Peacekeeping." The report drew upon ideas and proposals conveyed to the Secretary-General by governments, regional agencies, nongovernmental organization, institutions, and individuals from many countries. The report acknowledged that sources of conflict can be complex and run deep, and it identifies broad measures intended to address the sources of conflict. Such measures include enhancing respect for human rights and fundamental freedoms, engaging in sustainable economic and social development, and curtailing the existence and use of massively destructive weapons.

The report provides definitions for preventive diplomacy, peacemaking, and peacekeeping, and states that from the UN perspective, the three are integrally related:

1. *Preventive diplomacy* refers to action taken to avoid disputes between parties escalating into conflicts, and to limit the spread of conflicts when they do occur.
2. *Peacemaking* refers to action taken to bring hostile parties to agreement through such peaceful means as those foreseen in Chapter VI of the Charter of the United Nations. Chapter VI deals with the peaceful settlement of disputes, and requires that countries involved in disputes that could lead to war first attempt to seek solutions through peaceful methods such as "negotiation, enquiry, mediation, conciliation, arbitration, judicial settlement, resort to regional agencies or arrangements, or other peaceful means of their own choice" (United Nations, 1945).
3. *Peacekeeping* refers to the deployment of a UN presence in the field, hitherto with the consent of all the parties concerned and normally involving UN military and/or police personnel and, frequently, civilians. Peacekeeping is a technique that expands the possibilities for both the prevention of conflict and the making of

peace (Boutros-Ghali, 1992). Chapter VII of the United Nations Charter sets out the UN Security Council's powers with respect to peacekeeping. It allows the Council to "determine the existence of any threat to the peace, breach of the peace, or act of aggression" and to take military and nonmilitary action to "restore international peace and security" (United Nations, 1945). Preventive measures include conflict early warning, fact-finding, confidence-building measures, early deployment, humanitarian assistance, and the creation of demilitarized zones (Boutros-Ghali, 1992).

Preventive Diplomacy: How Should Its Success Be Measured?

How can one prove that diplomacy contributed to the prevention of an open conflict if said conflict never broke out? Measuring the effect of preventive diplomacy is a daunting task that more closely resembles art than science. The difficulty begins with developing a common definition and a framework. Based on the framework, the highly complex and variable situations that constitute pre-conflict, conflict, and post-conflict situations in very specific contexts can be made comparable and perhaps even quantifiable so as to provide the basis for an *ex post* assessment of different field cases.

The "Assessment Framework for UN Preventive Diplomacy," developed by the United Nations University Centre for Policy Research, represents one such attempt (United Nations University, 2018a). It breaks with the common UN logic of results-based management (RBM), and instead encompasses the idea that certain inputs (i.e., resources) combined with a set of activities (i.e., negotiations, field activities, training) lead to outputs (i.e., better trained security sector actors) that over time will lead to an outcome (i.e., lower sectarian violence) and ultimately an impact, such as improved stability in a country or a region (United Nations Joint Inspection Unit, 2017).

The reason for the break is the underlying RBM assumption that the UN can dedicate resources to conducting activities, which will cause a result, which will in turn contribute to a desired change (the "impact"). This notion underlies most UN evaluations and is the basis on which the organization (and aid and development agencies worldwide) typically assess progress and allocate funds. The problem with this approach is that it assumes a causality that cannot be generalized in highly complex conflict situations; what is required is a more nuanced assessment of how different measures of preventive diplomacy, in which combination and application along the timeline, led or did not lead to successful de-escalations of conflict.

To this end, the team at the United Nations University (2018a) has developed a six-step assessment framework that can be applied by practitioners as elements of the overall framework:

1. An evidence-based context analysis describing the key factors driving and inhibiting escalation, the interests of those who control the situation, the overall difficulty of resolving the crisis, and the outcome;
2. A causal analysis that identifies the major factors that influenced the decision of the conflict actors to engage in or refrain from violence;
3. A counterfactual argument describing what would have happened had there been no outside intervention by preventive diplomatic actors, including those deployed by the UN;

4. An analysis of the extent to which the decision(s) of the conflict actors can be attributed to the UN's intervention, weighed against other interventions and relevant factors;
5. An analysis of what enabled and/or inhibited the UN's ability to contribute to preventing escalation, including issues related to mandate, strategy, resources, coordination with other actors, and external factors. This analysis incorporates a distinction between those issues that are external to the UN and those that are within the UN's control;
6. An assessment of the extent to which the UN intervention was linked to sustainable peace, looking both at whether the intervention itself was inclusive and at what longer-term capacities for conflict prevention were left in place.

A team led by Laurie Nathan of the University of Notre Dame applied the preventive diplomacy assessment framework described above to a number of crises and situations. Among them were the 2008–2010 crisis in Guinea, the situation in Lebanon during the Syrian crisis (2011–2017), the Malawi crisis 2011–2012, the 2015 Nigerian elections, the Independence Referendum in Southern Sudan (2010–2011), and the popular uprising in Yemen from April to November 2011 (United Nations University, 2018b).

The project intended to answer the following three questions:

1) Was the preventive diplomacy successful or unsuccessful in each case?

2) Why was the UN's preventive diplomacy successful or unsuccessful?

3) How sustainable or unsustainable were the successful outcomes?

In seeking answers to these questions, the team followed the logic of preventive diplomacy, which defines success as “a shift from volatile and escalatory conflict dynamic to a dynamic of containment and de-escalation.” The team considered three actors:

1) the conflict parties (i.e., those with the power to decide whether to escalate to large-scale violence in a given setting);

2) the preventive diplomacy intervener (the actor attempting to steer the conflict parties’ decision in a nonviolent direction);

3) other actors with influence over the conflict parties.

The assumption was made that the conflict parties were at all times the primary decision makers (United Nations University, 2018b).

From the analysis of the case studies investigated by the team, six success factors emerged, which are cited below.

1. The conflict parties had not yet decided to resort to large-scale violence. This created the potential for successful diplomatic interventions.
2. The parties consented to preventive diplomacy by the UN. Where consent was not forthcoming at the outset, it had to be won by the UN. Alternatively, the UN at times deferred to a regional organization that took the lead.
3. There was a high level of international and regional cooperation and unity. The main dynamics were that the UNSC was united; key international and regional actors supported UN leadership on preventive diplomacy; and/or UN preventive diplomacy was undertaken in partnership or coordination with other international actors.

4. International leverage was used effectively. This was especially true of soft leverage, which included the Secretary-General's involvement through an envoy; a unified stance on the part of the international community; and the deployment of UN resources and technical expertise to support prevention efforts. The cases did not reveal a clear pattern regarding coercive forms of leverage.
5. The UN envoy had the right set of attributes and skills. These often included deep knowledge of the conflict and the parties, a regional or cultural affinity with the parties, and skills in terms of communication and persuasion.
6. There was good internal UN coordination and cooperation. The UN Country Team and the UN regional offices are crucial partners in preventive diplomacy efforts by envoys.

Success factors 1, 2, and 5 are generally applicable to preventive diplomacy efforts in both multilateral and bilateral settings; success factors 3, 4, and 6 come to bear only in settings where the UN or another multilateral organization (e.g., the Organization for Security and Co-Operation in Europe, African Union) and the international community decide to act and then do so effectively.

With respect to sustainability, the authors mention the operational nature of preventive diplomacy, meaning that it serves to avoid escalation of conflict at a given moment but does not *per se* eliminate the structural causes of a conflict. In successfully de-escalating momentary tensions, preventive diplomacy can support the creation of the political space required to attend to the requisite structural reforms (Nathan, et al., 2018). The preventive diplomacy framework developed and the case studies investigated by the United Nations University Centre for Policy provide a structured approach for assessing and obtaining valuable ex post insights into what factors have a positive impact on assuring preventive policy success in different situations.

For the purpose of this research, I build upon the logic of this framework since it has the potential to provide a suitable structure and basis for deep learning about preventive diplomacy based on historical events and context. This form of machine learning is a prerequisite for developing useful AI in this subject area and hopefully an *ex ante* capacity for early conflict prevention.

The Potential Game Changer: Artificial Intelligence

The term "artificial intelligence" (AI) was coined by John McCarthy (Dartmouth College), Marvin Minsky (Harvard University), Nathaniel Rochester (IBM), and Claude Shannon (Bell Labs) in a proposal for a conference in 1956 at Dartmouth College called the Dartmouth Summer Research Project on Artificial Intelligence (McCarthy et. al., 2006). Significant advancements have been made since then. In 1958, McCarthy developed the programming language LISP (short for "List Processing"), a language that greatly eased programming attempts to model human thought at the time. Around the same time, the integrated circuit (IC) was invented and, with the advent of the metal-oxide-semiconductor field-effect transistor (MOFSET), the foundation was laid for the development of high-density IC chips.

Based on these developments and technological progress, in 1965 Gordon Moore, co-founder of Fairchild Semiconductors and CEO and co-founder of Intel, projected a doubling of the number of transistors in dense ICs approximately every two years (Moore, 2006). His projection was found to be so accurate that it is commonly known today as Moore’s Law. In Table 1 below, I highlight some important milestones in computer history that laid the groundwork for the development of AI.

Table 1. Milestones in Computer History.

DATE	MILESTONE
1969	ARPAnet was first switched on. It was the first large-scale, general-purpose computer network connecting different kinds of computers together.
1971	Intel introduced the first microprocessor. In the same year, networked email and the “@” character were introduced on the ARPAnet.
1973	Packet Radio Network (PRNET) was introduced. This laid the foundation for what would later become modern mobile networks.
1976	Cray-1 supercomputer and the Apple-1 were introduced, and, in 1985, the U.S. National Science Foundation Network (NSFNET) was created. The creation of this network was a major factor in helping internet protocols (TCP/IP) win out over rival protocols, such as OSI, SNA, and DECNET.
1986	Daniel Hillis of Thinking Machines Corporation moved AI a step forward when he developed the controversial concept of massive parallelism in the Connection Machine CM-1. The machine was able to complete several billion operations per second. The machine’s system of connections and switches let processors broadcast information and requests for help to other processors in a simulation of brain-like associative recall. Using this system, the machine could work faster than any other computer at the time on a problem, which could be parceled out among the many processors.
1989	David Levy became the first master chess player to be defeated by a computer.
1990	At CERN in Switzerland, English programmer and physicist Tim Berners-Lee laid the foundation for what would later become the “World Wide Web.” Only one year later, after the National Science Foundation changed its policy, the Internet became a publicly accessible network with no commercial restrictions.
1997	IBM’s Deep Blue chess computer finally defeated the current world chess champion, Garry Kasparov, after several previous attempts had failed (Anderson, 2017).
2002	iRobot launched its Roomba autonomous robotic vacuum cleaner, which uses a cleaning algorithm that allows it to clean a room while detecting and avoiding obstacles.
2005	Stanford Racing Team’s autonomous vehicle “Stanley” drove 175 miles autonomously on a desert course.
2010	China’s Tianhe supercomputers became operational, with a peak speed of one petaflop (one thousand trillion calculations per second).

DATE	MILESTONE
2011	Siri, the digital voice assistant, was introduced as a feature of Apple's iPhone 4S.
2013	Edward Snowden exposed the NSA surveillance program PRISM, under which the NSA had collected data with the assistance of companies such as Microsoft, Facebook, and Google.
2015	Bill Gates joined Elon Musk and Stephen Hawking in expressing their anxieties regarding AI, which sparked a debate on the topic. While proponents of AI foresee unprecedented human achievement, Gates and others expressed the view that while intelligent machines may benefit humankind in the short term, a future might come to pass in which more advanced super-intelligent machines could pose a grave threat to human existence (Computer History Museum, n.d.).

Source: thesis author

Since then, major progress has been made in ANI, and AI-based engines and assistants have become commonplace in consumer-focused industries, being used in cars, mobile phones, spam filters, web interfaces, streaming portals, online translators, voice recognition systems, and navigation systems, to name just a few.

In 2016, Google's Alpha Go beat Lee Sedol, the winner of 18 world Go titles, after playing a move that had never before been conceived of by a human player (DeepMind, n.d.); in the same year, Singapore launched the first self-driving taxi service (Watts, 2016).

Besides consumer-focused applications of ANI, sophisticated ANI systems are also widely used in other sectors and industries. These include the military, manufacturing, and finance, where algorithmic high-frequency AI traders now account for more than half of equity share trades on U.S. markets (Bostrom, 2014). Furthermore, AI can be found in technology developed for expert users, such as systems and devices that help doctors make diagnoses. Most famous was IBM's *Watson*, which commanded enough factual information and understood speech sufficiently well to soundly beat the most accomplished Jeopardy champions in 2011 (Gabbat, 2011).

How can ANI be developed into AGI? According to Tim Urban and others (Kurzweil, 2005; Urban, 2005b; Vinge, 1993), there are two major steps. The first is increasing computational power; the second is making machines intelligent.

Key Factor 1: Increasing Computational Power

If an AI system is to approach the level of sophistication where it could be as intelligent as the human brain, it first must have the raw computing capacity equal to a brain. To this end, the processing capacity needs to be increased; this capacity can be measured by the total calculations per second (cps) the system can execute.

To estimate the number of cps a human brain can execute, one can attempt to determine the maximum cps of each structure in the brain and then add the resulting figures together. Ray Kurzweil used a shortcut for this method by taking experts' professional estimate for the cps of one structure of the brain and that structure's weight compared to that of the whole brain. He then multiplied the cps proportionally to obtain an estimate for the total cps of the human brain. He repeated the procedure by taking estimates from different experts of different regions of the brain, and the total was consistently in the region of approximately 10^{16}, or 10 quadrillion (=10,000,000,000,000,000) cps (Kurzweil, 2005).

At the end of 2013, China's *Tianhe-2* supercomputer outperformed that number by a factor of almost 2.5, clocking in at about 34 quadrillion cps. However, *Tianhe-2* needs 720 m^2 of space and 24 megawatts of power, at a total setup cost of $390 million.

Japan's *Fugaku* supercomputer is, as of January 2021, the world's fastest supercomputer, exceeding processing power of the human brain by a factor of a little over 40, clocking in at 442 quadrillion cps and a total program cost of around $1 billion. The *Fugaku* still occupies the space of a good-sized sports stadium and uses roughly 28.3 megawatts of power (Fulton, 2020). In comparison, the human brain runs on approximately 20 watts of power (Urban 2015).

Kurzweil suggests that one think about the state of computers by considering how many cps one can buy for $1,000. When that number reaches the 10-quadrillion cps mark (which he predicted back in 2005 would be roughly now), this would mean that AGI could become a part of daily life (Kurzweil, 2005). The possibility of achieving the hardware requirements for AGI proved controversial when Kurzweil's book *The Age of Spiritual Machines* was published (Kurzweil, 1999), but his 2005 prediction has now become a fairly mainstream view among informed observers. The current controversy instead revolves around the algorithms necessary to make a computer intelligent.

Key Factor 2: Making Machines Intelligent

There is no single correct or generally approved strategy by which to make computers intelligent. According to Urban, there are three prevalent strategies:
1) plagiarize the brain, 2) attempt to make evolution do what it did before but for us this time, and 3) make this challenge the computer's problem, not ours (Urban, 2005b).

Plagiarizing the brain. In her 2010 article in *Wired,* Priya Canapati quotes Ray Kurzweil from his 2005 book:

> "The singularity is near: when humans transcend biology." [Reverse-engineering the human brain] would be the first step toward creating

> machines that are more powerful than the human brain. These supercomputers could be networked into a cloud computing architecture to amplify their processing capabilities. Meanwhile, algorithms that power them could get more intelligent. Together these could create the ultimate machine that can help us handle the challenges of the future.

While critics argue that reverse-engineering the human brain is excessively complicated, Kurzweil argues that neuroscientists, computer engineers, and psychologists have been working to simulate the human brain so they can ultimately create a computing architecture based on how the mind works. Kurzweil further observes: "The key to reverse-engineering the human brain lies in decoding and simulating the cerebral cortex – the seat of cognition" (Canapati, 2010).

In her article Canapati goes on to explain that the human cerebral cortex contains about 22 billion neurons and 220 trillion synapses. To be capable of running a software simulation of the human brain, a machine with a computational capacity of at least 36.8 petaflops and a memory capacity of 3.2 petabytes would be needed (Canapati, 2010). Ten years later, technological developments have surpassed the required computational capacity presented by Canapati. For example, *Fugaku*, running on double precision [64 bit], is capable of 488 petaflops and has a total memory of over 5 petabytes.

However, according to Kurzweil, "The objective is not necessarily to build a grand simulation—the real objective is to understand the principle of operation of the brain." Kurzweil estimates that about a million lines of code could be enough to simulate the human brain, a view shared by Terrence Sejnowski, a professor and the head of the Computational Neurobiology Laboratory at the Salk Institute for Biological Studies (Canapati, 2010). Kurzweil goes on: "Even a perfect simulation of the human brain or cortex will not do anything unless it is infused with knowledge and trained" (Canapati, 2010).

One example of computer architecture simulating the functioning of the brain is the artificial neural network. Such a network starts out as a network of transistors (technical neurons), which are connected to each other with inputs and outputs. Initially, the network does not "know" anything. It must subsequently learn, which it does by attempting to perform a task, such as recognizing a set of pictures. Initially, its neural firings and subsequent guesses at recognizing the pictures will be completely random. However, once the network receives feedback that it has recognized a picture correctly, the transistor connection pathways that led to the success are strengthened and when it does not recognize the picture correctly, those corresponding transistor connection pathways weakened. After many such trial and feedback loops, the network will over time develop efficient neural pathways, resulting in the machine becoming optimized for the particular task. The human brain learns in a similar but more sophisticated way.

The extreme form of plagiarizing the brain is called whole brain emulation (WBE), which refers to the theoretical possibility of modeling the function of the brain using a one-to-one model (Sandberg & Bostrom, 2008). One approach could be to slice a real human brain into thin layers, scan each one, use software to assemble an accurately reconstructed 3-D model, and then implement the model on a powerful computer. The result would be a computer that is technically capable of everything the brain is capable of. However, it would still need to learn and gather information (Urban, 2015). If engineering reaches an extreme level of sophistication, engineers may be able to emulate a real brain with such accuracy that the brain's full personality and memory will be intact once the brain architecture has been uploaded to a computer. This way, the brain of a recently deceased person could be "re-awakened" on a computer, making for a robust form of human-level AGI (Urban, 2015).

So how far are researchers from possibly achieving WBE? In "Whole Brain Emulation: A Roadmap," an article published in 2008, Anders Sandberg and Nick Bostrom postulate that should current technological trends continue, this event may occur by the middle of the century (Woronko, 2019). In 2014, researchers were able to emulate the brain of a millimeter-long flatworm and upload the resulting software into a Lego robot's body (OpenWorm, n.d.). The flatworm brain consists of a total of 302 neurons, while the human brain contains 100 billion. While this gap could indicate that there is still much work to be done, exponential progress might get us there faster than we think (Urban, 2015).

Artificial Evolutionary Approach. Building a computer as powerful as the human brain is possible. Whether WBE can be achieved remains to be seen, but it appears to be at least theoretically possible. Should WBE prove too complex to achieve, however, another option would be to emulate evolution instead. This could theoretically be accomplished by means of a genetic algorithm, a search heuristic inspired by Charles Darwin's theory of natural evolution. Such an algorithm reflects the process of natural selection, whereby the fittest individuals are selected for reproduction in order to produce the offspring that constitute the next generation (Mallawaarachchi, 2017).

Urban explains the working of genetic algorithms as follows: The algorithm works with a performance and evaluation process at its center, which is executed repeatedly. The algorithm emulates the conditions under which biological creatures “perform” by living life and are “evaluated” by whether they manage to reproduce or not. In practice, this amounts to a group of computers attempting to complete tasks, with the most successful computers being bred with each other by having half of each of their code bases merged to produce a new computer. The less successful computers are eliminated over time. Over many iterations, this selection process would produce increasingly better computers (Urban, 2015).

A major challenge associated with the genetic algorithm approach is the creation of an automated “evaluation” and “breeding” cycle that allows the evolutionary process to run on its own. Another challenge is compressing a process that took nature billions of years into the timeframe of a few decades (Urban, 2015).

In contrast, those who attempt to implement an artificial evolutionary model have some major advantages over natural evolution. First, evolution has no foresight and works randomly, producing more unhelpful mutations than helpful ones. In an artificial environment, by contrast, the process can be controlled in such a way that it is only driven by beneficial developments (Urban, 2015). Second, natural evolution does not aim for anything in particular beyond survival. As such, some environments may even select against higher intelligence (e.g., on the basis of a larger or more complex brain using too much energy). In an artificial setting, however, the evolutionary process could be directed specifically towards increasing intelligence (Urban, 2015). Third, to select for intelligence, natural evolution had to drive innovation in other areas to enable the development of intelligence in the first place, such as by changing the way in which cells produce energy. By removing such additional burdens (e.g., by using electricity), it could be possible to significantly accelerate the evolutionary process. Nevertheless, it is still not entirely clear whether improving and accelerating the evolutionary process will suffice to make this a viable strategy (Urban, 2015).

Turn the Problem Over to the Computer. A third possible strategy is for scientists to build a computer possessed of two major skills: conducting research on AI, and coding relevant changes to itself. Such a system would not only learn but also constantly improve its own architecture. The system's main job would be to determine how to make itself more intelligent. This concept is called recursive self-improvement (RSI). Systems that operate according to this principle create new software iteratively, and the newly created software generates a more intelligent system using the current system. Applied over many iterations, this process should eventually lead to AGI before moving on to artificial superintelligence (Prithvinath Reddy, 2020; Urban, 2015). There is some debate as to how soon AI will reach human-level general intelligence. According to a survey of hundreds of scientists, the common view is that this will more likely than not be achieved by 2040 (Bostrom, 2014).

In conclusion, there is sufficient literature on balance of power, balance of threat, and diplomacy to lay a solid foundation for this research. If one thinks of the world as a system of nation-states that seek to consciously balance themselves against current and emerging threats as a means of assuring security, then understanding the dynamic of how threats emerge and are perceived is key.

In their 2014 paper "The Construction of Threats: The Iran Nuclear Crisis as a Textbook Example for Jack Mezirow's Transformative Learning Theory," Bock and Eschenbacher offer some important insights. They present a simple model that demonstrates the influence of experience and expectations on human perceptions, and on this basis conclude: “Threats are not given but socially constructed against the background of the experiences states made [sic].” As an example, they use the case of Iran and how it is perceived by others.

In the following sections, I use the model shown in Figure 2 as a framework upon which to base my approach to answering the research question.

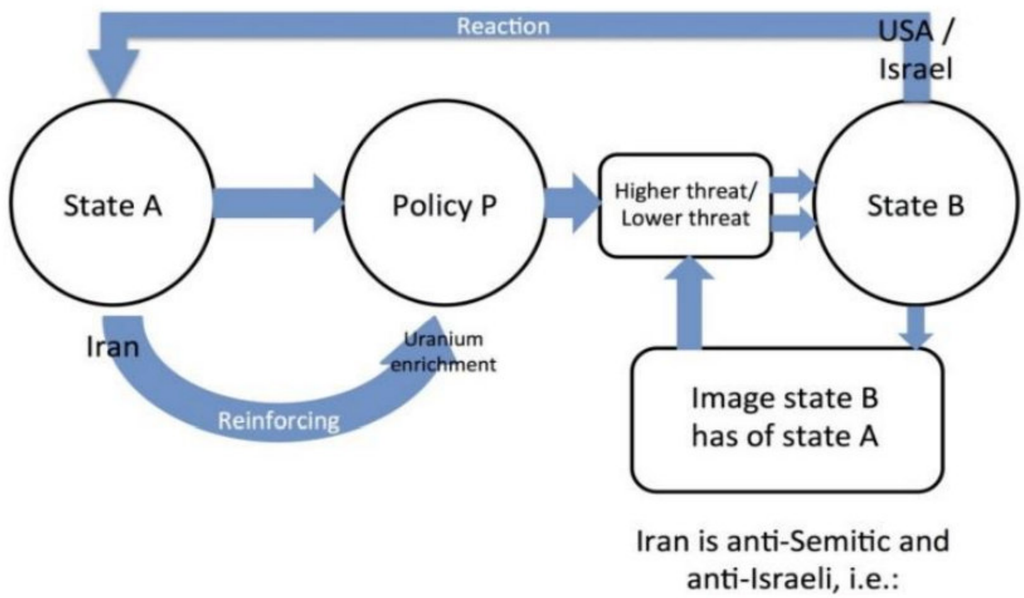

Figure 2. The Original Bock and Eschenbacher Threat Perception Model
Applied to Iran, the US, and Israel.

Source: Bock & Eschenbacher, 2014, p. 77.

Chapter IV

Research Methodology

The research conducted for this study consists of a combination of desk research (using secondary sources) and fieldwork (using primary sources) in the form of structured interviews with experts on AI as well as experts on diplomacy, security policy, and foreign policy. My initial goal was to interview approximately 10 experts from each field for a total of about 40 interviews. I invited participants to complete an online survey. The selection of experts invited to participate was based on their practical experience in the field and their possession of experience from the Cold War as well as in more recent settings; participants were also selected with an eye to including a range of nationalities and exposure to different geographical parts of the world.

After completion of the online survey, participants could indicate whether they would be available for a follow-up discussion of their answers. Based on this voluntary indication of availability, I conducted a limited number of follow-up interviews, during which I further focused on certain comments made by the participants and was able to obtain a better understanding of how they think ANI would work in practice.

The reader should bear in mind, however, that the opinions expressed by the experts represent only a contemporary picture of a subset of experts based on the current state of knowledge and developments in the fields of AI and diplomacy, respectively. During the time it took to write this thesis, new developments of relevance to the topic took place and continue to evolve. One such development is the recent standoff between

Russia and NATO in Ukraine; also what is currently unfolding with protests against the government in Kazakhstan—both events taking place after the interviews. Therefore the results of this research should be not taken as absolute but rather as an indication where these experts think we are and where developments could go in the future.

After presenting results for each of the five areas where ANI could potentially make contributions, namely 1) monitoring, 2) validation, 3) anticipation, 4) solution finding, and 5) collective learning, I have elaborated on what the experts stated that anticipated specific applications of ANI could look like based on exemplary scenarios. Where doing so would be meaningful, I referenced what I heard from the experts vis-à-vis the 1962 Cuban Missile Crisis (refer back to Chapter II), and then described where ANI might have made a difference.

My survey was based on the model depicted in Figure 3, which is a modification of the Bock and Eschenbacher threat perception model (refer back to Figure 2).

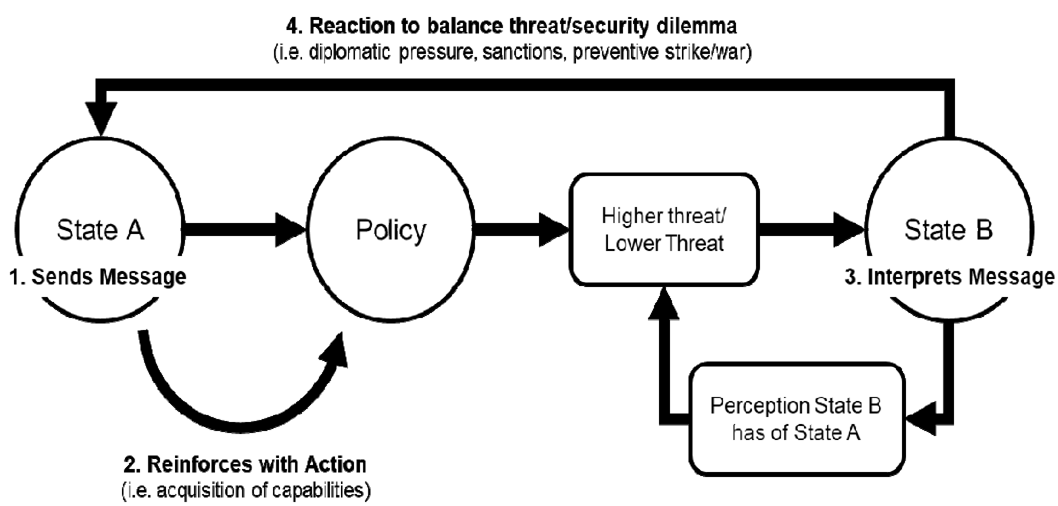

Figure 3. Balance-of-Threat System: Origin and Balancing of Interstate Threats.

Source: adapted from original Bock & Eschenbacher (2014).

Actual or perceived threats originate from State A making statements and possibly reinforcing those statements with actions (i.e., the acquisition of military capabilities), which will in sum determine State A's stance in terms of foreign security policy. State B observes the rhetoric and behavior of State A and constantly interprets its observations in order to determine whether the words and actions of State A result in it constituting a greater, equal, or lesser threat to its own security interest (i.e., peace and, ultimately, survival).

Depending on the perceived threat level, which is a function of the perceived intentions of State A to actually do what it is threatening to do, the capability of State A to do what it is threatening to do, and a general assessment of the trustworthiness of State A, State B will react to the perceived threat. Such reactions may involve attempting to re-balance the threat (i.e., security dilemma) by taking a stronger posture and/or augmenting State B's own military capabilities, entering alliances with other states, seeking the support of the international community, or, in the worst-case scenario, launching a preventive war.

Before continuing, there are three key assumptions at the core of this research. These are:

- Assumption 1: Balance-of-threat theory is valid and provides a way to create stability in the international system.
- Assumption 2: Preventive diplomacy can contribute to managing and maintaining balances of threat to avoid open conflict.
- Assumption 3: AI cannot fully replace human diplomatic agents, at least not in the near future (approximately 20 years).

Based on these assumptions and the model shown earlier in Figure 3, five areas for potential contributions by AI were identified (see Figure 4 below). Each area is described below in more detail, and initial hypotheses are presented.

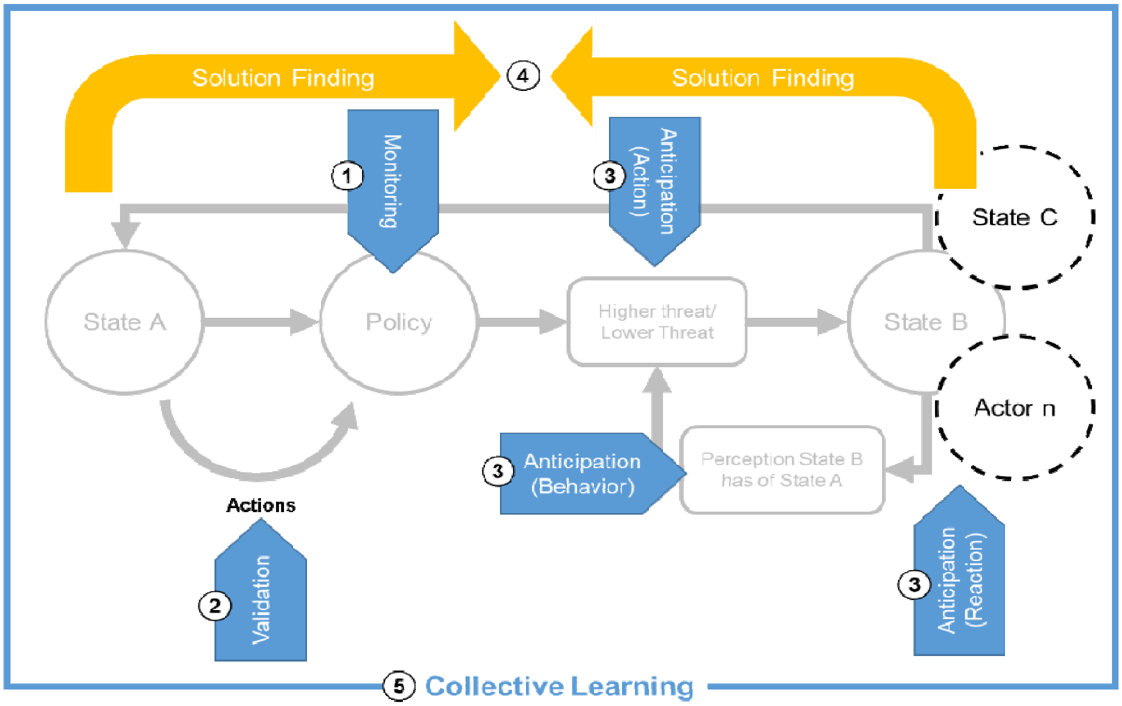

Figure 4. Five Potential Areas Where AI Could Contribute
in Terms of Balancing Threats.

Source: adapted by thesis author from Bock & Eschenbacher, 2014.

Monitoring

Monitoring refers to all activities related to watching the messages, actions, and policy changes of other states and actors, as well as other environmental factors. Monitoring detects actual or impending shifts in the overall balance of threat and pinpoints the sources of such changes. Systems that use AI can rapidly process large amounts of data from a broad range of sources and further detect connections between individual data points, allowing them to detect patterns.

Initial Hypotheses Regarding Monitoring

H1: AI can provide a more complete, timely, and reliable (cross-referenced) situational picture.

H2: AI can be used to help create an effective early warning system for changes in balances of threat.

Validation

Validation refers to all activities related to assessing the validity of information. Validation can be conducted on information in general but is particularly relevant for cross-referencing messaging (i.e., in the media) and actual actions. An example would be a comparison of reports of an alleged ceasefire violation in the media and what actually happened on the ground. This activity is extremely important in establishing the reliability of information, which, in times of information warfare, fake news, and other attempts to manipulate the public, has become ever more valuable.

Initial Hypothesis Regarding Validation

H1: AI can assist in the discernment of false information from correct information and thus contribute to better decision making and overall trust in information and sources.

Anticipation

Anticipation refers to the ability to foresee events. Within the balance-of-threat system, this could be in the area of anticipated *action* (Baarslag, et al., 2015) by State A (the cause of the change in the system), likely *behavior* of State A (i.e., in reaction to actions of State B or other actors), and *reactions* of other states and actors in the system as a response to the actions and reactions of States A and B. Combining the three types of anticipation in a wargaming setup allows for exploration of future and likely actions and reactions in a particular balance-of-threat system, thus revealing likely paths of future development before any specific action is taken (Dorn, Webb, & Pâquet, 2020; Oriesek & Schwarz, 2021, 2008; Perla, 1990).

Initial Hypotheses Regarding Anticipation

H1: AI can facilitate the anticipation of the specific likely action of Actor A in a given context.

H2: AI can facilitate the anticipation of the general likely behavior of Actor A in a given context or in reaction to specific actions of Actor B and other participants in the system.

H3: AI can facilitate the anticipation of the likely reactions of other participants in the system to specific actions-reactions of Actors A and B.

Solution Finding

Solution finding refers to the action of finding enough common ground among participants in the balance-of-threat system to allow the system to return to a stable state and conflict to be avoided. This requires a thorough understanding of who the relevant participants are, their individual needs and interests, and how issues of concern affect each of the participants (De Jonge & Sierra, 2017; Dorn, Webb, & Pâquet, 2020). Another important aspect is networked thinking and the ability to identify and potentially link cooperation on one issue to interactions on a second issue, a process called "linkage" (Frieden, et al., 2019).

Initial Hypotheses Regarding Solution Finding

H1: AI can facilitate the identification of the relevant participants in a balance-of-threat system and how they relate to one another (system analysis).

H2: AI can facilitate the anticipation of the likely reactions of the relevant participants in relation to a specific issue or set of issues.

H3: AI can facilitate the detection of seemingly unrelated issues that, in combination with the disputed issue, offer opportunities for win-win situations and conflict resolution (linkage).

Collective Learning

Collective learning refers to the systematic and structured capture, dissemination, and implementation of lessons learned within the balance-of-threat system with the intention of making the system better (i.e., more stable and peaceful). The Assessment Framework for UN Preventive Diplomacy could be a starting point for a structure that would enable such collective learning.

Initial Hypotheses Regarding Collective Learning

H1: AI can help to assess, consolidate, and evaluate relevant information as a means of accomplishing the goal of effectively determining lessons to be learned.

H2: AI can improve the overall effectiveness of preventive diplomacy (e.g., by employing a genetic algorithm or functioning as an RSI system with the objective of maximizing the effectiveness of preventive diplomacy).

H3: AI can support the dissemination of critical information and provide decision support for the most relevant actors in a given context.

H4: AI will challenge existing information power structures by leveling the playing field, shifting the definition of competitive advantage away from the asymmetric possession of proprietary information toward having the capacity for superior analysis and interpretation of publicly available information.

Survey Questionnaire and Interviews

Based on the balance-of-threat system and the initial hypotheses, a questionnaire was created (see Appendix 1). After receiving the necessary clearances from Harvard University's Institutional Review Board, the survey was programmed and uploaded. A closed group of participants was invited via an email containing a link to the survey; upon consenting to participate in this study, participants were granted access to the questions, which on average took about 30 minutes to complete. Toward the end of the questionnaire, the participants were asked whether they would be willing to participate in a follow-up interview.

I contacted a subset of participants upon their consenting to participate and providing contact information for the purpose of gaining a more in-depth understanding of their answers and general views on the topic. The follow-up interviews took about 45 minutes on average. No direct quotes are used, nor will the names of participants be disclosed. However, some of the statements made by participants are included, albeit in a generic, generalized, and anonymous way. The online survey was conducted between June 1 and August 31, 2021. The follow-up interviews were conducted between October 1 and December 12, 2021.

Chapter V

Analysis of Results

After a general introduction to the composition of the sample, this chapter provides an analysis of the responses given by the experts. The presentation and analysis cover five areas (Monitoring, Validation, Anticipation, Solution Finding, and Collective Learning), where AI could potentially make a contribution, followed by a section on risks and benefits. The chapter concludes by noting possible limitations of this research.

Composition of the Sample

A total of 31 experts participated in the online survey. Of these, 19 were Swiss, six were American, two were German, two were Swedish, one was Spanish, and one was Senegalese. Of these experts, almost one-third have 20+ years and almost three-quarters have 10+ years of professional experience in their respective fields (Figure 5).

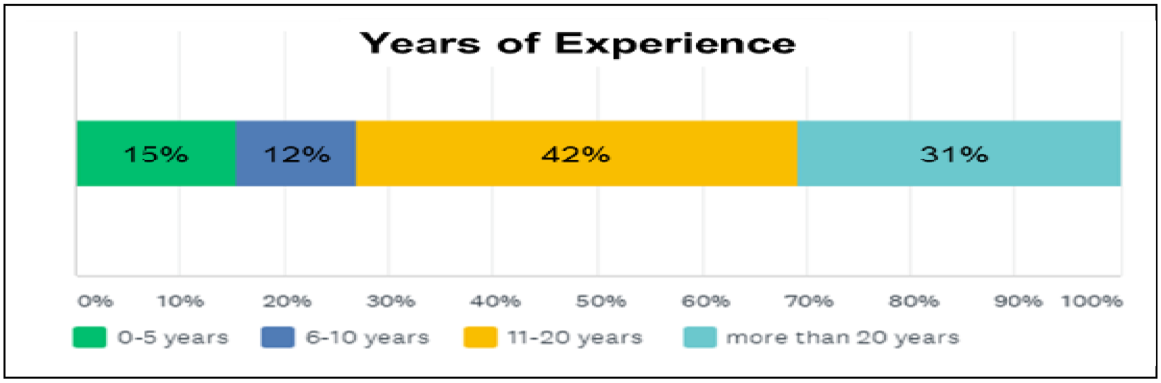

Figure 5. Years of Experience of Survey Participants.
Note: (% values rounded)

Source: thesis author

About two-thirds of the survey participants have worked in or covered the European continent; the remainder have covered or worked in Africa, Asia, or the Americas.

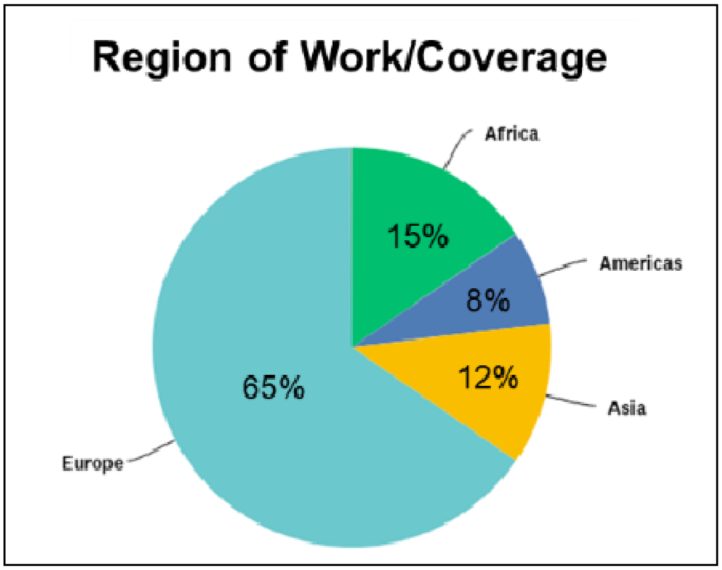

Figure 6. Region of Work/Coverage of Survey Participants.
(Note: % values rounded)

Source: thesis author

About half of the participants classify themselves as experts in security policy; these include defense attachés, researchers, and high-level civilian policy experts. About one-fifth classify themselves as experts on AI; these include researchers as well as AI practitioners working primarily in the defense context. Approximately one-fifth classify themselves as experts on foreign policy; these include experts on foreign policy in the capitals as well as practitioners who work in the field, dealing with development programs or other foreign policy initiatives abroad. The remainder are high-level diplomats at the level of ambassador who have practical experience from multiple postings abroad.

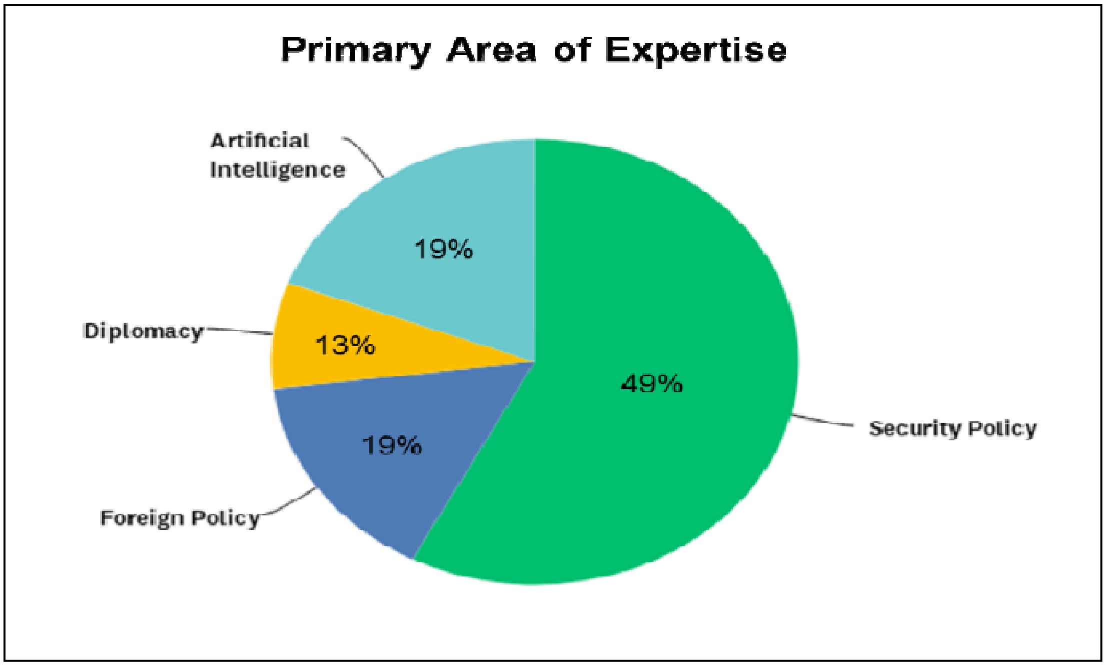

Figure 7. Primary Area of Expertise of Survey Participants
(% values rounded)

Source: thesis author

Findings: Composition of the Sample

With a total of 31 respondents who have significant work experience in environments across the globe and a good mix of working experts, researchers, and AI practitioners, the sample is large and sufficiently diverse to provide relevant insights into the research topic.

Monitoring

There were three questions with respect to monitoring relative to whether the experts believed that ANI could contribute to obtaining a (1) more complete *and* (2) timely picture of changes in the balance of threat, and (3) whether ANI could thus be used to set up an effective early warning system for detecting changes in balances of threat. These questions were based on knowledge that ANI today is particularly good at recognizing changes, such as changes in imagery, rhetoric, or behavior (6th pattern: Recognition), as well as detecting anomalies in a broader context (5th pattern: Identifying patterns and anomalies). These applications of ANI could provide timely warnings should anomalies hint that the balance of threat could potentially change and thus provide early indications that diplomatic efforts should be undertaken. The majority of the experts agreed that ANI could make a significant contribution to obtaining a more complete (see Figure 8) and a more timely picture of potential changes (see Figure 9) in the balance of threat.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	8%	0%	67%	25%

Figure 8. Expert Assessment:
Getting a More *Complete* Picture of Changes in the Balance of Threat.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	4%	9%	58%	29%

Figure 9. Expert Assessment:
Getting a More *Timely* Picture of Changes in the Balance of Threat.

In contrast, the more skeptical experts raised concerns as to how messages or rhetoric could be comprehensively captured by ANI. These concerns related to communication, which consists not only of what is said but also what is not said (e.g., body language). Here, the experience and ability of diplomats and other high-level negotiators to "read the room" in its entirety is a skill that, at least for the time being, cannot easily be replaced by ANI.

The majority of experts believed that ANI could be used to "set up" or at least "contribute to" an early warning system to detect changes in the balance of threat. Technical questions remain regarding whether ANI could capture enough of the necessary input variables to fully capture what constitutes the balance of threat, but there is a broad consensus that a functioning early warning system would be very beneficial and potent (see Figure 10).

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	17%	4%	58%	21%

Figure 10. Expert Assessment:
Usefulness of ANI to Set Up an Early Warning System.

When asked about other applications of ANI with respect to monitoring balances of threat, the experts mentioned several interesting points. For one, they largely agreed that ANI can contribute to obtaining a more detailed "threat picture," which would result from the ability to include, manage, and cross-reference a wider array of sources than would otherwise be possible, or that would only be possible with immense human resources (i.e., analysts screening vast quantities of source material from multiple channels in multiple languages). Second, they see the potential of using these data to develop “impending scenarios” (i.e., what is likely to happen if no diplomatic intervention takes place), which would be particularly useful for establishing a fact base and platform when entering negotiations. One respondent raised the idea of what he called “digital ambassadors,” or algorithms that Nation A and Nation B would run over the available fact base and explore options for courses of action before they enter actual negotiations. Note that the ultimate negotiation remains with the human agents—a theme that will emerge again in the section on risk.

When asked how the advent of AGI would change their assessment, the experts agreed that the overall quality of decision support with respect to precision, speed, networking of factors, and consequent predictability would improve. The opinions differed with respect to whether AGI would be reliable and sufficiently ethical to replace human agents altogether.

Findings: Monitoring

The experts agreed that ANI could make significant contributions toward monitoring balances of threat. This is because ANI could contribute to a more complete and timely picture of potential changes in the balance of threat. In addition, it could be used to set up or at least support some type of early warning system to detect changes in the balance of threat. The major contributions of ANI would stem from including, managing, and cross-referencing a wider array of sources than would otherwise be possible or that would be possible only through reliance on immense human resources.

As an example, during the Cuban Missile Crisis, it took about 27 hours from the time Maj. Richard Heyser began crossing Cuba in his U2 reconnaissance plane in the early morning of October 14 until the photointerpreters (PIs) at the National Photographic Interpretation Center received copies of the images; it was another six hours until they were certain they had identified large surface-to-surface missiles capable of carrying nuclear warheads. It took another five hours of quality checks passing through various levels of the hierarchy until the information reached National Security Advisor McGeorge Bundy at around 21:00 on October 15 (GlobalSecurity.org, 2011).

During the Cuban Missile Crisis, aerial images taken by the U2 aircraft (Figure 11) were the only source of information available in those moments, and it took a relatively long time to process—approximately 38 hours until the information was ready and available to decision makers. Today, that same type of information could be obtained within minutes by applying ANI.

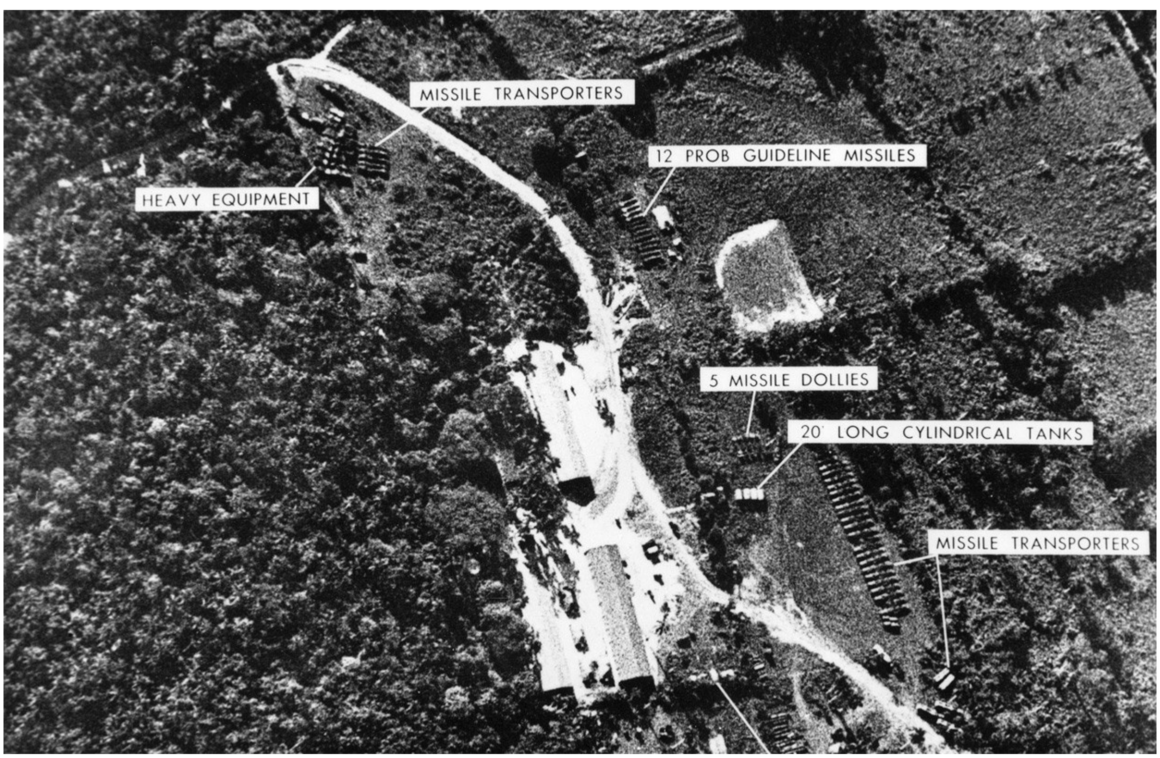

Figure 11. Ballistic Missile Base in Cuba, 1962.

Source: National Geographic.org.

Given that sensor and transmission technologies have developed significantly since the 1960s, they need to be excluded from a comparison of how the same task would be performed today. Thus the focus of this comparison is on the six hours needed by the PIs in 1962 to interpret the aerial images and identify the missiles, and the potential of AI to include and cross-reference multiple sources.

The process of interpreting aerial images has changed significantly over the years. Today, taking an aerial image from a plane, satellite, or drone and determining what it depicts or what has changed in comparison to an earlier image requires minutes, preferably seconds. Without going into technical details, one can imagine a multi-step process involving pixel comparison, zooming in on areas that have potentially changed, and applying approaches such as light detection and ranging to identify what has actually changed. Such changes could involve alterations to a building or a change in its height, which can be detected despite changes in vegetation (Du, et al., 2016). If such an approach for detecting objects and changes were linked to a database containing the shapes, dimensions, and material details of military objects, shapes, reflections, and other details in an image, it could be compared within seconds against entries in the database; ANI could use such a database to calculate the probability of a match.

In the case of the missiles on Cuba, the ANI approach would certainly include comparing a recently transmitted, high-resolution satellite image of the suspected area with images of that same area taken at an earlier point in time. If an object were clearly located, the ANI would show its location on the image and state with, for example, a 95% probability that the image depicts the components of an yet unassembled R-12 missile.

Cross-referencing sensor information. If a match is not immediately clear (i.e., the probability of [something] is below 80%), the ANI would do what PIs usually do if an image is inconclusive: it would attempt to double-check its initial finding by comparing what it thinks is an R-12 missile with other images from the same area taken by other sensors (planes, drones, other satellites) or possibly in other spectrums (i.e., forward-looking infrared). By cross-referencing the information from these different sensors, ANI would adjust the probability of a match based on whether it finds different identifying characteristics for an R-12, such as heat signatures or shapes taken from a different angle, or leave the probability as it was if no additional information can be obtained.

Thus ANI could make a significant contribution by not only improving the speed of the initial matching but also by automating the "double-checking" procedures to arrive at better probability estimates. There would be no guarantee that an object could always be identified with 100% certainty, but if image features revealed an object and this object was present in a database, the likelihood of ANI identifying such an object would be high.

Final decision making. As a consequence of the application of ANI, the National Security Advisor of today would be given information regarding specific missiles on Cuba more quickly than his 1962 counterpart. In addition, that information would have been "double-checked" and thus would be as reliable as possible. The Advisor may still be confronted with a degree of uncertainty (i.e., 5% uncertainty if the probability of a match is determined to be 95%), but his counterpart in 1962 would have had to deal with the same issue. The ultimate responsibility for deciding what to do with this information would still lie with human decision makers.

Toward an early-warning system. The ANI-based approach described above with regard to the example of missiles on Cuba could be applied to any global area of interest. While permanent 24/7 aerial coverage of the entire planet is currently not possible, it would be absolutely possible, through prioritization and automation, to cover designated areas of interest in the world at regular intervals. Therefore, such areas of interest could be screened for changes, and unusual developments (e.g., massing of military equipment in a particular area) would be detected quickly.

Implementing such a system in areas of interest for regional or global balances of threat would constitute the nucleus of an effective early warning system. Detecting changes early and determining what these changes are would be helpful in identifying developments that could upset regional or global balances of threat (i.e., the massing of 120,000 Russian troops near the Ukrainian border, as seen in Figure 12).

Figure 12. Russian Army Camp Bordering Ukraine, 2021.
Source: www.maxar.com/images

Such a system would provide decision makers with valuable information and allow them to engage in preventive diplomacy before tensions escalate into armed conflict, thus contributing to greater stability in a balance-of-threat system.

Validation

The question of validation sought to determine whether experts believed that ANI could contribute to improving the reliability of information used to detect changes in the balance of threat. This question is based on the knowledge that ANI today is good at detecting anomalies in the greater context (5th pattern: Identifying patterns and anomalies). The application of ANI would enable a wider inclusion of sources due to its ability to validate information from different sources while screening for potentially fake news and information operations (e.g., attempts to influence what people believe) of various degrees of sophistication.

The majority of experts agreed that ANI could make a significant contribution to validating and thus improving the quality of information regarding potential changes in the balance of threat (Figure 13). They particularly pointed to the speed with which a wide array of sources can be harvested and compared as representing a major win in terms of time used and the quality of the resultant database.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	4%	14%	59%	23%

Figure 13. Expert Assessment: Improving the Reliability of Information.

Skeptical experts raised concerns that such applications would only work if enough information were available, which is not always the case, especially in remote conflicts. This could lead to an incomplete picture (i.e., few or no statements from news or other communication channels), while other information (i.e., from satellite imagery or use of the electromagnetic spectrum) would possibly not suffice to piece together a picture of the situation on the ground that would be adequate enough to provide a solid basis for negotiations, let alone decision making.

When asked about other applications of ANI to validate information in the context of balances of threat, the experts did not offer any new insights beyond noting that ANI can be used for the fusion of many different and independent sources/sensors and thus be used to validate facts by cross-referencing the information from one source with that from one or more other sources/sensors.

An interesting point raised was the notion of "trust," which is the basis for any type of validation. Here, the participants noted that validated results based on ANI should be distributed using a trust-based system (i.e., by using blockchain technology). Such a system should ensure that people have greater trust in the correctness and the integrity of the information validated using ANI. When asked about how the advent of AGI would change their assessment, the experts seemed to agree that this would also improve the overall quality of the validation process, and that if AGI could actually understand the concept of trust or autonomously create trust, then this could further contribute to information verification and the quality of the end product, namely validated information.

Findings: Validation

The experts agreed that ANI could make significant contributions to validating and improving the quality of information regarding potential changes in the balance of threat in a given region of the world. Such contributions would rely on ANI improving the speed and array of sources that could be tapped into in order to validate information by cross-referencing it with information from other sources, thus increasing the trustworthiness of such information.

As in the example of applying ANI to monitor and ultimately detect missiles on Cuba, ANI can speed up the processing of information (i.e., satellite images), and then cross-reference this information with other sensors (i.e., images from planes, drones) to obtain more reliable results. In the Cuba example, while the boost in confidence came from cross-referencing images from different sensors (satellites, planes and drones), this was merely an application in the discipline of imagery intelligence (IMINT).

For the purpose of more general "validation" of information, ANI can be applied to a much broader set of sources and in a number of disciplines of intelligence collection. The intelligence community refers to five major disciplines of intelligence collection (Lowenthal & Clark, 2016):

- HUMINT: Human intelligence is the collection of information from human sources.
- SIGINT: Signals intelligence refers to electronic transmissions that can be collected by ships, planes, ground sites, or satellites. One important subcategory of SIGINT is COMINT, or communications intelligence, which refers to the interception of communications between two parties.

- IMINT: Imagery intelligence is sometimes also referred to as PHOTINT or photo intelligence. GEOINT, or geospatial intelligence, involves the analysis and visual representation of security-related activities on the earth. It is produced through the integration of imagery with geospatial information.
- MASINT: Measurement and signatures intelligence is a little-known collection discipline that is concerned with weapons capabilities and industrial activities. MASINT includes the advanced processing and use of data gathered from overhead and airborne IMINT and SIGINT collection systems. TELINT, or telemetry intelligence, is sometimes used to indicate data relayed by weapons during tests, while ELINT, or electronic intelligence, can indicate electronic emissions picked up from modern weapons and tracking systems. Both TELINT and ELINT can be types of SIGINT and contribute to MASINT.
- OSINT: Open-source intelligence refers to a broad array of information and sources that are generally available, including information obtained from the media (newspapers, radio, television, etc.), professional and academic records (papers, conferences, professional associations, etc.), and public data (government reports, demographics, hearings, speeches, etc.).

In 1962 in Cuba, the evidence regarding the presence of Soviet missiles was quite clear once the construction of sites had begun, and the main components of a missile launch site became visible on aerial photographs. However, overall developments and the situation on the ground were not detected until at a very late stage in the process. One could even argue that had the construction of the missile site not been detected by the U2 and the PIs, there would have been a high probability that the Cubans (and the Soviets) would have obtained a forward strike capability without the US noticing for some time. *Cross-referencing intelligence collection disciplines.* Had modern ANI been applied to integrate multiple sources from the different intelligence collection disciplines, the 1962 Cuban missile scenario would have unfolded differently. First, OSINT, COMINT, and HUMINT sources would have indicated that the Soviets and the Cubans had intensified their exchanges after Fidel Castro met with then Vice President Nixon during his visit to Washington in 1959 (Nixon, 1969). From OSINT, possibly HUMINT (if the US had trustworthy agents on the ground in Moscow and Havana), and COMINT, the US would have learned that the Jupiter missiles deployed to Turkey represented a major concern to the Soviets and that Fidel Castro, especially after the failed 1961 invasion of Cuba at the Bay of Pigs, was deeply concerned about the US attacking him again. By putting the pieces together from OSINT (newspapers, radio appearances, television broadcasts, public speeches) and COMINT (phone, cable, letter intercepts), ANI could have identified multiple indications that Soviet leader Khrushchev and Castro were conspiring to develop a solution to both their problems.

Meanwhile, SIGINT, HUMINT, OSINT, and IMINT obtained from within the Soviet Union would have provided evidence of the existence of the R-12 and R-14 missiles, and MASINT from flight tests would have provided data concerning the range of the R-12 and R-14 missiles (1,292 miles and 2,500 miles, respectively). With this information alone, ANI could have determined locations around the globe where the deployment of such missiles would pose a direct threat to the continental US, which the

CIA determined for the ExComm 1962 immediately after the CIA had obtained knowledge of the presence of the missiles in Cuba (Figure 14).

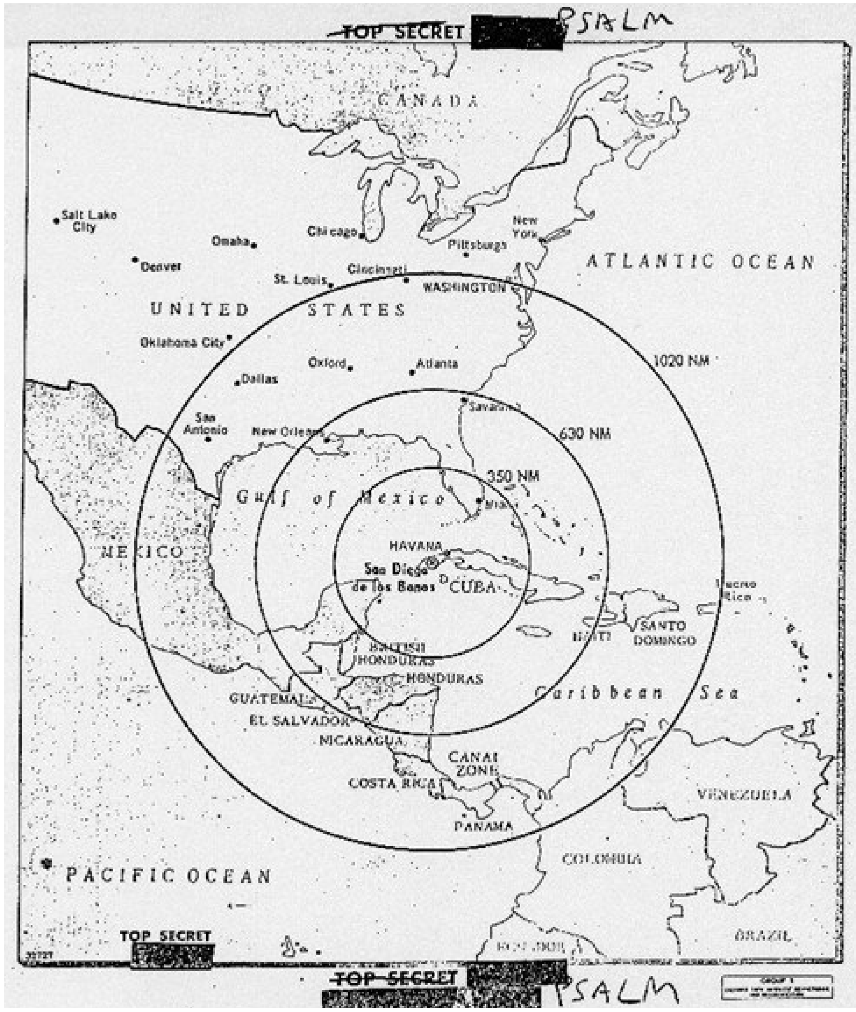

Figure 14. CIA Briefing Map for ExComm Meeting, October 16, 1962.

Source: https://nsarchive2.gwu.edu/nsa/cuba_mis_cri/17.jpg

Today, MASINT information would be fed into a database where it would be updated on a regular basis with other pieces of information, such as images taken by defense attachés during parades or statistics found in different reports or mentioned in technical publications or other OSINT sources. The same database would also contain images of and performance data on Russian ships, meaning that when a large group of ships of different classes set sail for Cuba, this would not have escaped the IMINT, especially when the ships began to enter the Caribbean Sea, an area of interest for the US that is under almost permanent surveillance.

As the Cubans and Soviets slowly began to move from their initial intent to solve the problems, to actually building missile sites on Cuba, ANI would have tracked at least some of the developments and cross-referenced, for example, the destination of the ships approaching Cuba with both Cubans and Soviet rhetoric. ANI also would have double-checked this information with statements from HUMINT reports that a significant number of military and engineering personnel had left the Soviet Union via ships destined for Cuba as well as random newspaper clippings discussing Soviet ships with strange cargo and personnel entering or exiting ports, refueling, and so forth on their way to Cuba. Assembling all these pieces would eventually lead ANI to determine with high probability that Soviet missiles and personnel were in transit to Cuba. This insight would be based on predictive algorithms similar to those that are today widely used by companies such as Apple or Google to, for example, provide users with traffic information roughly around the time they leave work based on user geo- and behavior data collected over time.

Because the conclusion drawn by the ANI in the above scenario would be based on multiple, unrelated sources hinting at the same development, the reliability and hence the validity of the information would be rated as "high." The validation of such information would almost certainly have led decision makers to pay attention to whether the predicted developments would actually occur. Upon being convinced that the predicted events were actually taking place, decision makers could have immediately begun engaging in diplomatic action intended to deescalate the situation before they were actually confronted with the presence of missiles in Cuba. Thus, the naval blockade of Cuba in 1962 could most probably have been avoided. ANI would have contributed to a faster recognition of a problem that could potentially have turned into a crisis and would have provided decision makers with additional time to find solutions short of kinetic actions or war.

Battling fake news. Aggregating sources for the purpose of cross-referencing information is a valuable tool with which to counter the ever-increasing amount of intentional or unintentional misinformation. Because many newspapers and other news channels rely on the same sources (i.e., news agencies) for a good part of their content, this fact is deliberately exploited by both state and non-state actors. There are many examples of instances in which actors have attempted to spread so-called "fake news" by sharing false information via different social media channels, hoping that established news channels and (as the content appears with increasing frequency throughout the media landscape) the public will ultimately believe the veracity of the fake content.

So-called "deep fakes" are the most brazen form of fake news. It is now possible to create videos in which politicians' appearances are completely faked and they appear to utter statements they never actually made in reality. Figure 15 schematically depicts how a source with a green screen (left) is modified by merging it with an image of President Obama against a background. The source and the Obama image are then merged to give the impression that President Obama moved and spoke just as the individual in the source did.

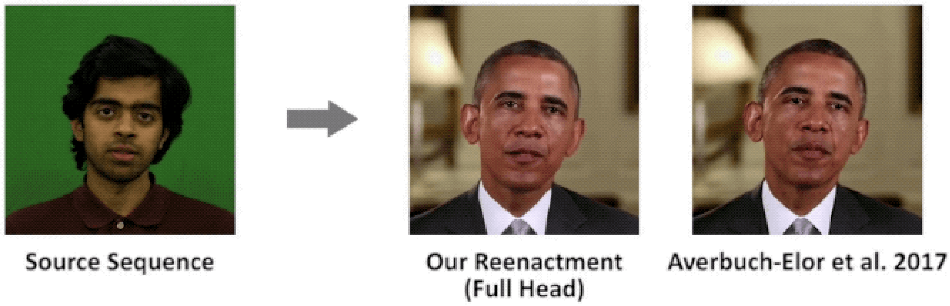

Figure 15. Deep Faking a Politician.

Source: https://www-hwzdigital-ch.translate.goog/anwendungen-mit-artificial-intelligence/?_x_tr_sl=de&_x_tr_tl=en&_x_tr_hl=en&_x_tr_pto=sc (translated page, animated)

Today, ANI is being applied to check connections between images, text, video, and audio, identify irregularities, and unmask deep fakes. In the animated version of Figure 15, for example, the algorithm would have immediately detected that as the fake Obama speaks, the shape of the lamp in the background changes as well. This is a fairly reliable indication that the content has been tampered with. In the context of this research, the focus is not on technical solutions for detecting fake videos but rather on applying an approach similar to that used for validating intelligence for the purpose of filtering out or at least identifying questionable, inconsistent, or incorrect information.

To apply such an approach to a media headline showing an image of Russian troops amassing close to the Ukrainian border (see Figure 16), ANI would check whether the image had been deep faked or used in another context (e.g., by scanning all available OSINT image databases). It would also check whether the same statement appears in other news channels and, more importantly, cross-reference the alleged information in the headline and/or article with other independent sources. It would attempt to confirm the location by cross-referencing the image with databases of satellite images and HUMINT reports from people on the ground. It would simultaneously access IMINT and MASINT for images and technical data and attempt to confirm the type of tanks, trucks, artillery pieces, and so forth. Should the equipment not be among that found in Russian arsenals, there is a high probability that the image is fake.

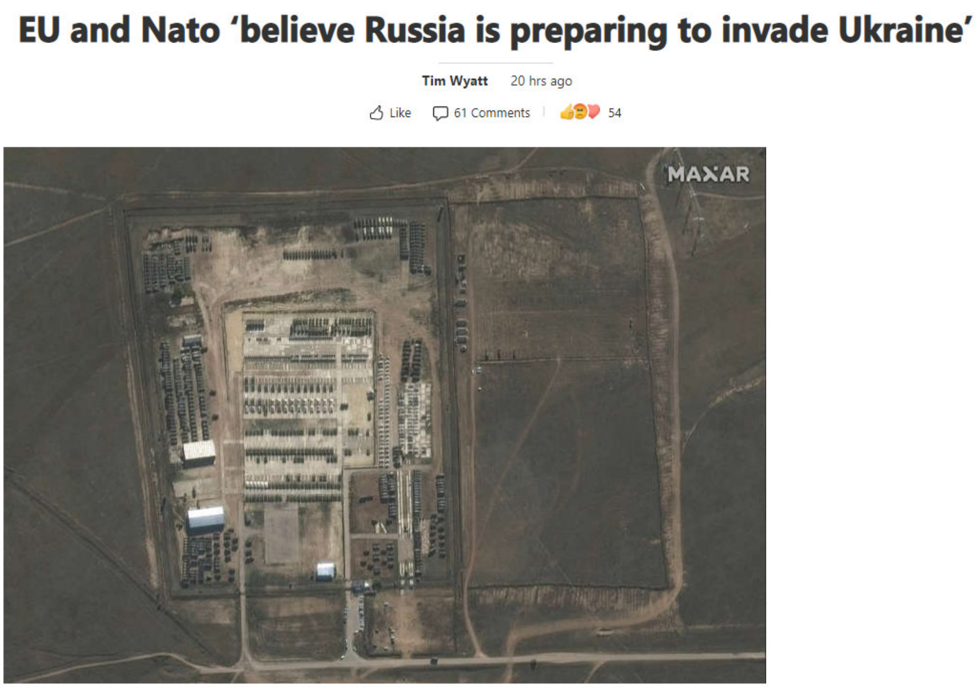

Figure 16. December 6, 2021: Headline with Image of
Russian Military Deployment in Crimea, Close to the Border with Ukraine.

Source: satellite image, Maxar via Reuters, 2021

By accessing COMINT information, ANI, even if it cannot access the content of actual conversations due to encryption, could at least map out the locations of the busiest nodes in the communication networks and thus pinpoint the location(s) of leadership elements. By assembling and validating relevant information in this way and performing the necessary checks, ANI would come to the conclusion that Russia is indeed amassing troops and preparing for a potential invasion of Ukraine. Even when provided with such reliable facts, decision makers will still have to decide what to do next, but they will be able to do so with much greater confidence than without these facts.

Anticipation

The first question with respect to anticipation relates to whether the experts believed that ANI could contribute to an understanding and anticipation of the likely reaction of a state or non-state actor in a given situational context. This question is based on the knowledge that ANI today can be applied to understanding how past or existing behaviors help predict future outcomes or help humans make decisions about future outcomes (3rd pattern: Predictive analytics and decision support). The application of ANI would provide advice on how certain actors will most likely behave given data from their past behavior combined with the parameters of a given situational context, such as environmental changes or economic factors. The majority of experts agreed that ANI could contribute to anticipation of likely reactions and would thus provide valuable decision support.

The skeptical experts raised concerns that reactions and decisions are tied to human psychology, which makes them more difficult to anticipate, as they may also include unpredictable and irrational behaviors that are challenging to predict due to a lack of patterns from the past. Other experts argued that because ANI is based on past patterns and does not understand the concept of intent combined with capability per se, it could therefore not detect potential actions based on such observations.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	18%	14%	41%	27%

Figure 17. Expert Assessment: Understand and Anticipate Likely Behavior in Situational Context.

The second question with respect to anticipation aimed to determine whether the experts believed that ANI could contribute to understanding and anticipating the likely behavior of a state or non-state actor in reaction to the specific actions of other actors and participants in the system. The underlying objective was to expand upon the quasi actor-centric approach of the first question (which looked at actors' likely reactions to a change in the environment) to account for a more dynamic environment, with other actors taking actions and other stakeholders in the system reacting, thus incorporating an added layer of complexity. Once again, the majority of experts agreed that ANI could make a contribution with respect to anticipating likely reactions and thus would provide valuable decision support. The more skeptical experts raised concerns similar to those they raised in response to the first question, while some expressed the belief that the task is simply too complex for ANI.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	23%	9%	45%	23%

Figure 18. Expert Assessment: Understand and Anticipate Likely Behavior in Relation to Actions of Other Actors and Stakeholders in the System.

Based on the first two questions, the third question went a step further by seeking to determine whether the experts believed that ANI could be used to set up a dynamic and comprehensive behavioral model of all major actors in a system. In this case, most of the experts were more cautious. While none thought it to be impossible, and about half agreed to varying degrees that it was possible, approximately one-third doubted that it could be done. The consensus among skeptics seemed to be that such a task would be too complex for ANI to handle. This was their shared opinion mostly because such a model would not be static but rather one that is constantly changing as the actors, and policies of the actors, change over given periods of time. One skeptic questioned the quantity of resources that would go into developing such a model and whether such efforts would make sense as a use of resources in comparison to simply sending human experts to carry out the assessment.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	32%	18%	41%	9%

Figure 19. Expert Assessment: Could ANI be Used to Set Up a Dynamic Comprehensive Model?

When asked about other applications of ANI in the area of anticipation, the idea of digital “national matches” hosted by the UN was brought up. As with the idea of digital ambassadors, nations would digitalize their policies based on agreed-upon rules and meet on neutral systems to play out potential disputes with each other before engaging in actual negotiations. The idea would be that all participants could explore “what-if” scenarios and determine what would likely happen should other participants react in a certain way. Armed with this in-depth understanding, negotiations could be more targeted toward the real issues at hand.

Another point raised by the experts was that ANI could provide support in detecting all relevant actors, even those previously thought to be of no importance in a given conflict situation. Once identified, the resulting information would help human decision makers account for all relevant players when making their decisions.

When asked about how the advent of AGI would change their assessment, the experts seemed to agree that this would provide the necessary prerequisite (i.e., the capability for handling enormous complexity) for building dynamic behavioral models that account for all relevant actors. Such an advance would also enable improvements to the human–machine interface that would make it more natural to interact with such an AGI (i.e., making interaction with the machine more like speaking or interacting with a real person).

Note: The sections referring to anticipation and solution finding partly overlap with respect to building models for anticipating likely behavior of actors in a balance of power system. Therefore the analysis of findings for both sections is presented in the next section on solution finding.

Solution Finding

The first question with respect to solution finding pertained to whether the experts believed that ANI could contribute to system analysis (a method for identifying the relevant participants in a balance of threat system and the kind of relations these participants have with regard to each other). This question is based on the knowledge that ANI today is very good at sorting through large amounts of data from different sources and recognizing relevant variables, such as the relevant participants in a system (6th pattern: Recognition). In parallel with identifying the relevant variables, ANI also efficiently identifies how these variables relate to one another, enabling it determine, for example, whether participants in a system have a positive or negative relation to one another and how strong said relation is (5th pattern: Identifying patterns and anomalies).

The majority of experts agreed that ANI today could make contributions to system analysis, while the more skeptical experts doubted that ANI could effectively capture the psychological aspects of relationships between actors. The skeptics also suggested that the detection algorithm would be biased either by inherent bias in the historical data used or by the biases of its developers.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
5%	10%	14%	38%	33%

Figure 20. Expert Assessment: Could ANI Contribute to System Analysis?

The second question with respect to solution finding aimed to determine whether the experts believed that ANI could contribute to anticipating the likely reaction of the relevant participants in relation to a specific issue or set of issues. This question is closely linked to the second question in the anticipation section, differing only in that it is asked from the specific perspective of an existing system analysis (i.e., the participants and their relations to one another have been identified in a previous step).

About two-thirds of the experts agreed that ANI today could contribute to anticipating the likely reactions of participants to a specific issue or set of issues in a given system. The more skeptical experts expressed doubt that ANI could effectively capture the human factor, namely hidden intentions, which cannot be captured by the system if such intentions are not spoken or otherwise explicitly communicated.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	14%	19%	43%	24%

Figure 21. Expert Assessment: Could ANI Anticipate Reactions to Issues in a Given System?

The third question with respect to solution finding attempted to determine whether the experts believed that ANI could help detect seemingly unrelated issues that, in combination with the disputed issue, offer opportunities for win-win situations and conflict resolution. This question relates to the concept of linkage.

In the context of the Cuban Missile Crisis, linkage referred to linking the immediate dispute over the presence of Soviet missiles in Cuba with the issue of previously deployed U.S. Jupiter missiles in Turkey. This opened up an avenue for the US to resolve the immediate crisis in Cuba by making concessions regarding an issue that had been troubling the Soviets for some time. From the U.S. point of view, giving up the Jupiter missiles in Turkey was a concession worth making given the immediate threat to the US posed by the stationing of Soviet missiles in Cuba. The attractiveness of this course of action was strengthened by the fact that the Jupiter missiles were at the end of their life cycle and thus represented a far smaller loss for the US than the perceived win their removal represented from the Soviet perspective.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	19%	14%	48%	19%

Figure 22. Expert Assessment: Could ANI Contribute to Finding Solutions?

About two-thirds of the experts agreed that at its present level of sophistication, ANI could contribute to identifying seemingly unrelated issues that could lead to linkage and thus create win-win situations and expand the room for solution finding. The more skeptical experts doubted that current ANI has the capability to handle the complexity of identifying seemingly unrelated issues, as this would require an in-depth understanding of the content given that each conflict situation is unique, meaning that historical data may be difficult to come by. They did, however, seem to agree that as ANI improved toward full AGI, such a solution could become feasible.

When asked about other applications of ANI to solution finding, the experts mentioned topics such as using ANI to create a knowledge base of historical events. Such a base could be drawn on to compare past conflicts with current issues and gain a better understanding of the mindset of one's counterparts on the opposing side.

One very interesting proposal concerned the combination of ANI technology with "smart contracts." The idea is that the conflicting parties should define a common "end state" for conflict resolution and, as the negotiation and solution finding progresses, the parties could see how their progress in particular areas of the negotiation would contribute to or detract from reaching the end state. In other words, the solution would be a dynamic decision support system that would make the consequences of elements in the negotiations visible to all parties involved. The expert suggesting this idea also mentioned that at the outset of engaging in such a smart contract negotiation, common ground should be established (i.e., by the parties mutually agreeing to accept human rights laws or similar legal frameworks) before the negotiation begins.

When asked about how the advent of AGI would change their assessment, the experts seemed to agree that AGI would make the solution finding by means of AI more feasible, as AGI would be better able to not only use learning based on past experiences but also understand content and context better and thus make suggestions (e.g., via digital ambassadors).

Findings: Anticipation and Solution Finding

Besides agreeing that ANI could be helpful with respect to anticipation, the experts also concurred that contemporary ANI could make contributions to system analysis (i.e., detecting relevant players and their relations, including the strength of the relations between relevant players in a given context). They also believed that ANI could assist in anticipating the likely reactions of individual participants in such a system. However, they doubted the ability of current-level ANI to create a dynamic, comprehensive behavioral model of all the players in a balance-of-threat system. Such doubts are mainly due to the technology's limitations in that its workings are based on pattern recognition on the basis of data from the past and that ANI lacks the ability to understand the concept of intentions and capabilities, which are sometime hidden. The experts believed that ANI could contribute to identifying seemingly unrelated issues that could lead to linkage (e.g. bring more issues to the table), create win-win situations, and expand the space for solution finding.

While the experts acknowledged that contemporary ANI is not yet capable of predicting all the dynamic and likely actions and reactions of actors in a system, there are feasible applications of contemporary ANI that can at least contribute to a better understanding of individual actors and their likely reactions to specific events or developments. Figure 23 presents a simplified form of the results of a so-called system analysis.

Once the most relevant actors (A, B, C, D, and E) have been identified, the ANI would then screen the others. The same approach could be applied in other languages, either by screening in each language or first translating a foreign language into English (using applications such as www.deepl.com) and then using the algorithm for screening in English.

Based on the results of this type of screening, the arrows in Figure 23 would be either red (= negative) or green (= positive) and thick (= strong relationship) or thin (= weak relationship). Such a system analysis would provide a basic understanding of the relevance of particular actors in a given context and how they relate to one another.

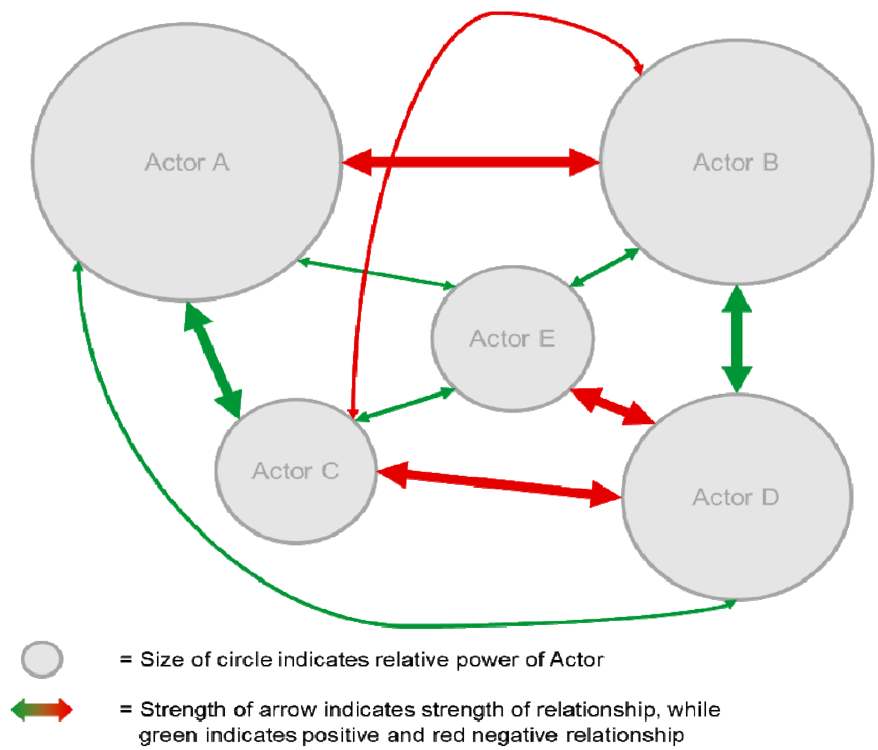

Figure 23. Example of System Analysis.

Source: thesis author

Figure 23 displays two main actors (A and B), who are in conflict with each other while each is allied with a primary ally (C for A and D for B). In this context, E is caught in the middle with good but weak relationships toward A, C, and B and a very poor relationship with D. Should E's relationship with D deteriorate further and D possibly start to mobilize troops or acquire new capabilities for a possible attack on E, then E might opt to balance this threat (security dilemma) by either acquiring capabilities itself, mobilizing troops (if it has), or aligning itself more closely with potential allies A and C; however, it is less likely to do so with B due to E's strong relationship with D.

This example illustrates how visualizing and updating such a system analysis in quasi-real time would help decision makers better understand why some actors behave in certain ways. With the advent of AGI, the prerequisites for conducting such a system analysis and modeling the likely behaviors of actors would only become more sophisticated, as the process would be based on genuine understanding of content and context as opposed to only understanding fragmented aspects of a situation that could be assembled into a greater whole.

Based on information harvested from OSINT sources, possibly combined with information obtained primarily from HUMINT and COMINT, ANI could certainly and quickly detect the most relevant actors in a given system. To this end, one can imagine defining a set of relevant OSINT sources, then defining the more specific context (e.g., all documents relating to a conflict in Bosnia and Herzegovina) before screening these documents for all mentions of different actors. The initial result would be similar to a so-called "word cloud," in which the most common words in a given set of documents are mentioned, and those that appear with greater frequency are presented in a larger-size font (see Figure 24).

Figure 24. Example of a Word Cloud.

It is interesting to note that as of December 15, 2021, Deputy Secretary of Defense Kathleen Hicks was briefed by US military commanders in the Pacific on a newly developed software tool that calculates what they call "strategic friction," which may arise with China depending on activities conducted by the US or by its allies in the Pacific region (Stone, Lewis, & Perry, 2021). While there is no information available on the details of how the system works, it is a first step in developing a more comprehensive, multi-actor tool to predict the likely behavior of participants in a system.

When provided with a better understanding of the roles, relationships, and needs of the individual actors in a system, ANI could also be used to screen for topics that may be critical to one actor while being of negligible or no importance to other actors. Through identifying such issues through the use of ANI, decision makers could identify opportunities to be brought up at the negotiating table. Had the US appreciated the perceived threat the *Jupiter* missiles in Turkey posed to the Soviets, the US could have used that bargaining chip much earlier in the negotiating process and extracted certain concessions from the Soviets elsewhere in turn for withdrawing the missiles. Had the US deescalated the *Jupiter* threat early on, Khrushchev may have suggested during their initial meeting that Castro abandon the idea of stationing R-12 and R-14 missiles on Cuba. In summary, understanding positions better and deescalating potential threats earlier would contribute to making escalatory measures such as arms races, alliances, or even preventive wars less likely.

Collective Learning

The first question with respect to collective learning sought to determine whether the experts believed that ANI could contribute to assessing, consolidating, and evaluating relevant information toward the goal of establishing effective lessons learned. This question is based on the knowledge that ANI is good at taking information from the past and analyzing it with a particular objective (5th pattern: Identifying patterns and anomalies; 7th pattern: Goal-driven systems). In the context of this research, the objective would be to identify variables and behavior that would effectively contribute to conflict resolution in similar situations in the future.

The majority of experts agreed that current ANI could make such contributions, but a significant portion—almost a quarter of participants—disagreed. The skeptics doubted the ability of current ANI to effectively identify or create real lessons learned, given that it uses a quantitative approach and is not able to understand the content and context necessary to carry out such a task. Another argument put forward by skeptics is that interests and threats are not stable and continuously change. So even if ANI could identify the current balance-of-threat system and draw lessons learned from this system, there is no guarantee that the lessons will still be usable in the future when important variables have changed.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	24%	14%	43%	19%

Figure 25. Expert Assessment: Could ANI Contribute to Creating Lessons Learned?

The second question with respect to collective learning attempted to determine whether the experts believed that ANI could improve the overall effectiveness of preventive diplomacy. While the majority of experts did believe ANI could make such a contribution, it was striking that almost 30% of respondents did not have a clear opinion on this matter. The remainder were skeptical, mainly because they doubted that ANI, based on quantitative approaches, is capable of performing such a task.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	9%	29%	38%	24%

Figure 26. Expert Assessment: Could ANI Improve the
Effectiveness of Preventive Diplomacy?

The third question with respect to collective learning aimed to determine whether the experts believe that ANI could support the dissemination of critical information and decision support to the most relevant actors in a given context.

This question was based on the assumption that if relevant actors in a conflict were provided with relevant information in a timely manner (i.e., scenarios concerning what would likely happen were they to take a particular course of action), this would encourage them to abstain from otherwise harmful courses of action.

While the majority of experts agree that ANI could play a role here, the skeptics have concerns about the political component of such an approach. Who would define the criteria for the scenarios to be developed? Would concerned players accept the scenarios as feasible? How would the scenarios reflect the national interests of the parties involved if all or part of their interests, agendas, or intentions are not known to the others? Others believe that this task is simply too complicated for current-level ANI.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	29%	5%	47%	19%

Figure 27. Expert Assessment: Could ANI Support the Dissemination of Critical Information?

The last question with respect to collective learning attempted to determine whether the experts believed that ANI could challenge existing information power structures. This question was based on the assumption that if the results from ANI technologies (i.e., superior analysis of publicly available information such as media products, satellite imagery, etc.) were widely available (i.e., available to all states in the international system), the competitive advantages of certain nations created by their access to proprietary information (through agents, special satellites, electronic reconnaissance, etc.) would lose value, resulting in the playing field becoming more level.

While the majority seemed to agree, almost one-third were unsure. The skeptics doubted that this would be a feasible scenario, although their objections were less due to technical limitations and more due to political ones. They also expressed the perception that the superior analysis of publicly available information would serve a complementary role rather than a primary one; they doubted that key players would dismantle their intelligence services, which allow them to obtain information intentionally hidden from actors in the international system—information that could hardly be uncovered by ANI.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	14%	29%	48%	9%

Figure 28. Expert Assessment: Could ANI Challenge Current Information Power Structures?

When asked about other applications of ANI with regard to collective learning, the experts mentioned "swarm intelligence," a concept from biology referring to the collective behavior of a group of animals, especially social insects such as ants, bees, and termites, in which individual creatures each follow very basic rules. With respect to AI, this is an approach to problem solving using algorithms based on the self-organized collective behavior of social insects—in this case different people in the nation-state system.

Another idea put forward was to use ANI to analyze all the learning curves of preventive diplomacy from the perspective of all actors involved, thus collectivizing their knowledge and accounting for cultural differences—a factor often not sufficiently considered when negotiating common solutions. Furthermore, according to one expert, ANI could be used to monitor agreed conflict-resolution measures and their effectiveness.

Such knowledge could be applied to future situations as well, which would close the feedback loop between agreed measures and their ultimate effectiveness and make this knowledge available for future conflict resolution.

When asked how the advent of AGI would change their assessments, the experts seemed to agree that AGI would be capable of the analysis of content and context and thus would also possess an improved ability to detect hidden agendas and intentions. AGI could also potentially detect previously unrecognized linkages to other factors currently not included in the analysis and ultimately provide a more significant contribution to the detection of real lessons learned that could be applied to future challenges.

Findings: Collective Learning

The experts were divided over the question of whether ANI could make contributions to assessing, consolidating, and evaluating relevant information with the goal of determining effective lessons learned. This dichotomy among the experts' opinions stemmed from doubts that current-level ANI possesses the ability to identify or create real lessons learned, given that it functions on the basis of a quantitative approach and is unable to understand the content or context necessary for such a task. Their understanding is that while ANI could go through documents and datasets post-crises, it could only discover common patterns based on statistical approaches and apply heuristic techniques to identify recommended solutions. It could not understand what it is dealing with, which is a prerequisite for identifying real knowledge gaps and arriving at lessons that should be learned.

Therefore, this research concludes that at this time ANI cannot make any significant contribution in this particular area. This situation may change, however, when

AGI reaches a level such that it can offer new possibilities in terms of screening and understanding information with the goal of identifying lessons learned.

Not all the experts are convinced that ANI could improve the overall effectiveness of preventive diplomacy. This is partly due to the aforementioned limitations of ANI's quantitative approach but also due to political forces that might work against the free dissemination of information to all participants. If such information were withheld, there would be no deterrence on the basis of knowing what would likely happen, at least not for those players from whom such information was withheld. In this context, the experts were neither convinced of nor opposed to the idea that ANI could potentially change existing information power structures.

Risks

In addition to being asked to consider the potential of ANI with respect to preventive diplomacy from a balance-of-threat perspective, the experts were also asked about their perception of the potential risks of the technology and its application. This was done through four questions that asked the experts for their assessment of the following issues: 1) ANI taking away the jobs of diplomats and other security professionals, 2) the inherent bias of algorithms, 3) the potential benefits as opposed to the potential risks, and 4) the potential benefits as opposed to the potential loss of privacy. These four questions were followed by three open questions that allowed the experts to freely reflect on where they saw the greatest dangers associated with and the greatest potential of applying ANI—and later, potentially AGI—in setting preventive diplomacy and maintaining balances of threat/peace.

Regarding the first question, the vast majority of the experts were not concerned currently that ANI could potentially take away the jobs of diplomats or security professionals. This is based on the knowledge that ANI is currently not at the level where it could perform complex tasks that are currently performed by humans and that require combining knowledge and skills in different areas. This may change with the advent of AGI, but for the time being, skills such as "reading the room" and detecting hidden agendas and intentions are simply too complex for ANI to handle. The minority that do believe AI could eventually replace human agents believe this on the basis of anticipating a future in which AGI will be available.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
35%	35%	20%	10%	0%

Figure 29. Expert Assessment: Risk of ANI Replacing Diplomats or Other Security Professionals.

Regarding the second question, the vast majority of experts expressed concerns about ANI being potentially biased for two reasons. On the one hand, their concerns stemmed from the fact that ANI's "intelligence" comes from learning based on data from and about the past. Consequently, depending on the dataset (e.g., delinquency rates among certain ethnicities), an algorithm may learn based on inherent biases in the data (e.g., a justice system that treats certain ethnicities unfavorably) and build these biases into its decision support for the future. On the other hand, people training the algorithm control the datasets that are used for its training, and the possibility cannot be ruled out that developers would build their own biases into systems through this initial dataset selection.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	10%	20%	60%	10%

Figure 30. Expert Assessment: Risk of Bias in ANI.

The majority of experts seem to agree that the potential benefits of AI outweigh the potential risks. In doing so, they acknowledge that this is not only a question of one's belief in technology but also a question with ethical, legal, and political dimensions.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	5%	30%	50%	15%

Figure 31. Expert Assessment: Potential Benefits Outweigh Potential Risks of ANI.

When weighing the potential benefits of AI against the potential loss of privacy, the experts seemed somewhat uncertain. More than half did not offer a position, while another one-third only somewhat agreed that the potential benefits outweighed the potential risks. This is most likely a consequence of the open form of the question, and one could argue that the two are not necessarily linked. Some experts argued that AI could actually contribute to protection of privacy if handled correctly.

Completely Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Completely Agree
0%	5%	55%	35%	5%

Figure 32. Expert Assessment: Potential Benefits Outweigh Potential Loss of Privacy Through ANI.

When asked the open question as to where they see the greatest danger of applying ANI and later AGI in the setting of preventive diplomacy and maintaining balances of threat/peace, the experts' answers were quite clear. The greatest danger they see comes from replacing humans with machines, particularly with respect to possible over-reliance on the results generated by said machines without understanding how the machines arrived at their conclusions and without taking the step of applying critical thinking to such results. This danger may increase as AGI and systems with potentially greater autonomy in decision making (such as autonomous weapons systems or nuclear defense systems) become a possibility. A second possible danger mentioned less frequently was that of machines being fed incorrect information or being manipulated to arrive at specific results.

When asked the open question about where they see the greatest potential in terms of applying ANI and later AGI in setting preventive diplomacy and maintaining balances of threat, the top three areas mentioned by the experts were: (1) improving the range and quality of information upon which decisions are based (i.e., by inclusion of data, interpretation of data, and reduction of complexity to enable diplomats to only focus on the important); (2) making the consequences of potential courses of action transparent to all parties involved, which would also act as a deterrent (by providing a timely recognized picture of the situation; by providing accurate, validated and trustworthy information; and by providing realistic depictions of the scenarios that would be likely to transpire given the stakeholders involved), and (3) contributing to monitoring and keeping actors honest and accountable for their actions.

Lastly, the experts were asked if they thought AGI could eventually replace human agents in the conduct of preventive diplomacy. Here, the experts acknowledged that technically AGI certainly could do so, and that the logic of evolution suggests that such a situation will eventually come to pass. However, they doubted that this would be technically feasible for at least another 10 years. They also had quite strong opinions that even if it was capable of doing so, AGI should never replace human agents. AGI could certainly support decision making, but the final decision should remain with humans who would act as a safety net or plausibility check against the course of action proposed by the machine. The machine might be effective at including and interpreting data but still have serious limitations when it comes to assessing the human aspects of a situation or actors, such as psychology, irrational behavior, moods, or current state of health.

Findings: Risks

The vast majority of experts were not concerned that ANI could potentially eliminate the jobs of diplomats or security professionals—at least not within the next decade. The experts were, however, concerned that ANI can potentially be biased due to bias in the information used for training machines (e.g., algorithms and datasets that are biased toward the particular national interests of one or more actors and thus will produce results in favor of those actors) as well as bias on the part of the developers, who may consciously or unconsciously choose to include or exclude certain sets of data.

The majority of experts seemed to agree that the potential benefits of AI outweigh the potential risks, although they seemed undecided when it came to assessing the benefits of AI against the potential loss of privacy associated with this technology. According to the experts, the greatest danger associated with applying ANI—and in the future, AGI—in setting preventive diplomacy and maintaining balances of threat/peace lies in replacing humans with machines or over-reliance on the results generated by machines without understanding how a given machine arrived at its conclusion. This danger may increase as technology progresses toward AGI. The experts seemed to agree that it is absolutely necessary to take the results generated by machines and apply human critical thinking to those results before making any decision, let alone taking action on the basis of such results. The experts see another danger in the fact that machines could be fed incorrect information or that they could be manipulated by their developers or third parties to arrive at specific results.

Research Limitations

This research did not discuss the technical details of AI beyond the point necessary for understanding its potential applications to preventive diplomacy from a balance-of-threat perspective. Consequently, the hypotheses were based on general knowledge regarding what ANI can currently do (i.e., seven patterns, specific known applications), and AGI was generalized as a form of AI that could basically perform as well as, if not outperform, a human. The research thus does not provide any specific insights for AI developers beyond initial ideas regarding other fields of application for AI technology.

Furthermore, this research did not include testing the feasibility of governance assumptions in some of the questions. Therefore it cannot answer the question of whether a centrally managed (i.e., by the UN) and commonly available (i.e., to all nation-states, and potentially other actors) AI system for the improvement of preventive diplomacy could even be implemented in the first place.

This research also focused only on the application of AI within the context of balance-of-threat systems, and within those on the levers of preventive diplomacy for stabilizing such systems. It did not consider other potential applications of AI within the fields of diplomacy or security policy.

Chapter VI

Conclusion

This research aimed to identify specific areas where AI could potentially contribute to the conduct of preventive diplomacy in terms of managing balances of threat to create more stability in the world. I summarize the results in the following sections.

Summary of Results

Based on a quantitative and qualitative analysis of answers to a structured online survey, with follow-up interviews from a diverse pool of experts in the fields of diplomacy, security policy, foreign affairs, and AI, I conclude that those experts agree that AI will make significant contributions to the field of preventive diplomacy in order to better manage balances of threat.

The results indicate that when asked about the specific possibilities of ANI today, the majority of experts agreed on four areas where ANI can make a contribution (see Figure 33):

1) *Monitoring*, where ANI can contribute to a more complete and timely picture of potential changes in the balance of threat, and where ANI could increasingly be used to develop a system of early indicators to warn of ongoing changes in the balance.

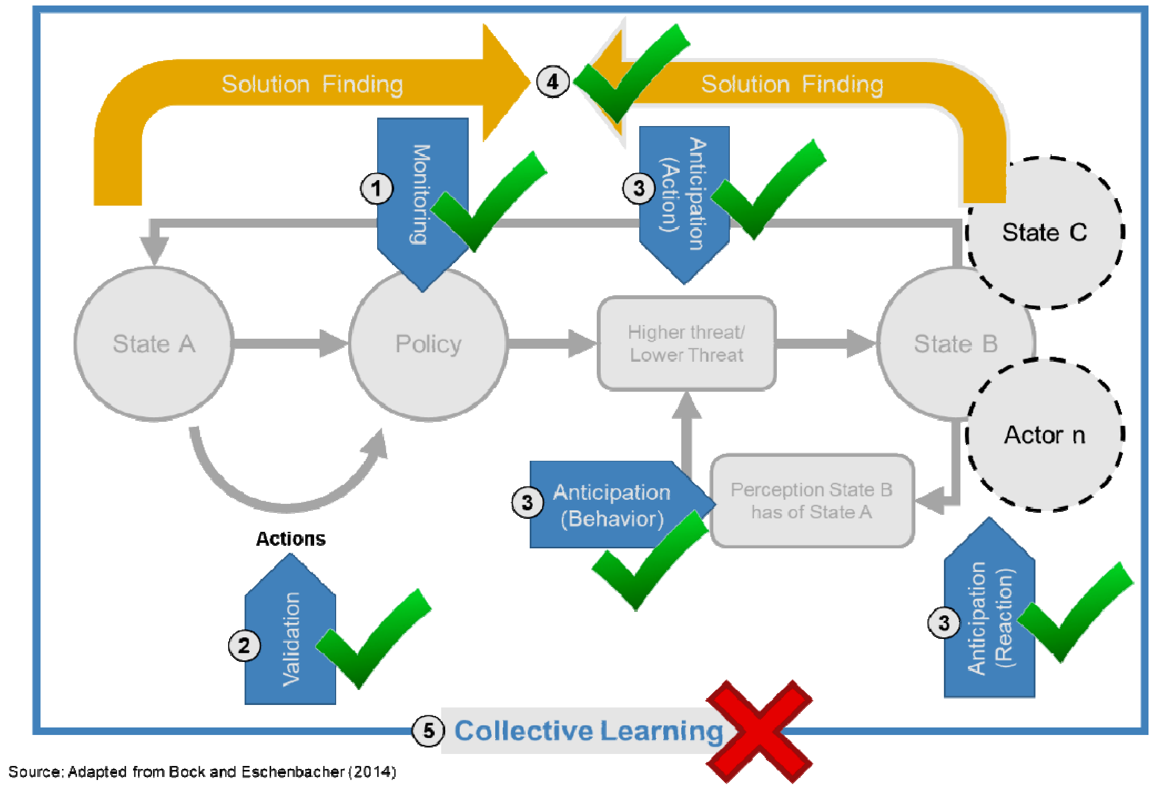

Figure 33. Summary of Results.

2) *Validation*, where ANI can contribute to validating and improving the quality of information as well as screening out "fake news" regarding potential changes in the balance of threat, thus increasing the trustworthiness of such information.

3) *Anticipation*, where ANI can contribute to anticipating the reactions of players in a balance-of-threat system and thus could provide valuable decision support.

4) *Solution finding*, where ANI can contribute to detecting the relevant players and their relations to one another in a given context as well as contribute to detecting seemingly unrelated issues that could be used to expand the room for negotiations and solution finding.

The experts were divided over the question of whether ANI could potentially make a contribution to collective learning, because a significant number of experts doubt that current ANI is sufficiently advanced for such a task. They are also divided over the question whether ANI could improve the overall effectiveness of preventive diplomacy. They fear that besides technical limitations, political forces might work against the free dissemination of information to all participants—a needed prerequisite for learning and deterrence in a setup designed to benefit all participants. In this context, they are neither convinced of nor opposed to the idea that ANI could potentially change existing information power structures.

With respect to risks, the experts were little concerned that AI could potentially take away the jobs of diplomats or security professionals within the next 10 years. The experts were concerned that AI is potentially biased and could be manipulated. In sum, the experts agreed that the potential benefits of AI outweigh the potential risks.

Going forward, the experts predicted that as we approach AGI, the contributions of AI in the four identified areas will get even better; then contributions to the area of collective learning and setting up dynamic and comprehensive behavioral models should become feasible.

In the context of approaching AGI, the experts saw the greatest danger of applying AI in the setting of preventive diplomacy and maintaining balances of threat in replacing humans with machines or overreliance on the results generated by machines without understanding them. Most of the experts therefore caution that it is absolutely necessary to take the results generated by machines and apply human critical thinking to those results before taking any decision or action.

To summarize the assessments of the individual experts, decision makers should use the technical possibilities of ANI today to obtain better validated information more quickly, to better anticipate possible (re)actions of stakeholders, and to identify novel means of arriving at solutions. Using ANI for these approaches should significantly assist decision makers to identify potential problems in the balance of threat early on, so they can restore the balance diplomatically before arms races, unhappy alliances, or even preemptive wars ensue.

Simultaneously, however, decision makers must remain critical of ANI's proposals and at least conduct plausibility checks before incorporating recommendations into decisions. Currently, it is only possible to speculate about what may be possible in the future with AGI. Nevertheless, today, many of the technological prerequisites and approaches needed for eventually reaching AGI already exist. Until the day we reach AGI, however, the human decision maker will remain at the center, and the need for diplomatic solutions, whether executed by humans or machines, will continue to exist for the foreseeable future.

Recommendations for Future Research

The opinions expressed by the experts in the survey and the interviews represent only a contemporary assessment by a given subset of experts based on the current state of knowledge and developments in the fields of AI and diplomacy, respectively. To better understand the validity and implications of these results, and as technology progresses, such expert assessments should be repeated with a broader array of experts covering geographical areas under-represented in the current research.

Based on my followup interviews, some interesting ideas were surfaced, of which at least two should be investigated in further research. First, the idea of the "digital ambassadors," where two or more nation-states could duel virtually before taking any actions in reality. Second, the idea of applying an AI-supported "smart contract" approach for diplomatic negotiations with the idea of conflicting parties commonly defining a "desired end state" and then negotiating and finding avenues for solution, while observing in real time how their decisions in particular areas would contribute to or detract from reaching that end state.

Final Thoughts

Developments in AI are progressing at great speed, and AI applications are spreading into almost all aspects of our lives. It is important, therefore, to look today at potential applications of AI in new fields for tomorrow.

At the same time, the world is undergoing fundamental changes. Starting with shifts in global power structures, we see the global proliferation of technology for good and bad, even as we move toward an ever more networked global society. While this brings many advantages, it has shown us in painful ways how vulnerable and dependent we have become on one another. Likewise, the big problems of our time, such as climate change, pandemics, migration, cyber and/or organized crime no longer stop at national borders, and solving those problems requires a joint effort by all stakeholders.

However, as we are still based on the Westphalian system of nation-states and are governed in the end by an arrangement of national interests in a system of anarchy, maintaining balances of power and threat still form the backbone for the relative stability in the world. It is therefore only a logical conclusion that we should make every possible effort to apply technological advances, such as AI, to reinforce our system of preventive diplomacy to better maintain this backbone of stability. If AI enables decision makers to identify potential conflicts early, then they can seize the opportunity to undertake diplomatic steps early based on more reliable information, with greater sensitivity as to the needs and positions of key stakeholders. Whether machines will or should take over responsibility for such tasks some day in the future, however, is another discussion.

With this thesis, I want to make a contribution to and launch the debate into this little explored, important, and fascinating field while providing initial thoughts and specific ideas that can be further developed going forward.

Appendix

Questionnaire

This research is based on two major assumptions. The first is that overall stability in the world is upheld by a balance-of-threat system. This means that we have an equilibrium of nation-states within the international system that is based on balance of power but in which states decide to upgrade their capabilities (military and otherwise), enter or dissolve alliances or, in extreme cases, wage preventive wars on the basis of the real or perceived threat represented by other states.

The second assumption is that preventive diplomacy can contribute to managing and maintaining balances of threat to avoid open conflict. "Preventive diplomacy" includes any diplomatic action taken to prevent conflicts from becoming violent or to prevent conflicts involving low-level violence from spreading or escalating into large-scale violence. The focus is thus on interstate, not intrastate, conflict. Figure A-1 below illustrates the dynamic of State A posing a real or perceived threat to State B, which triggers a reaction by State B.

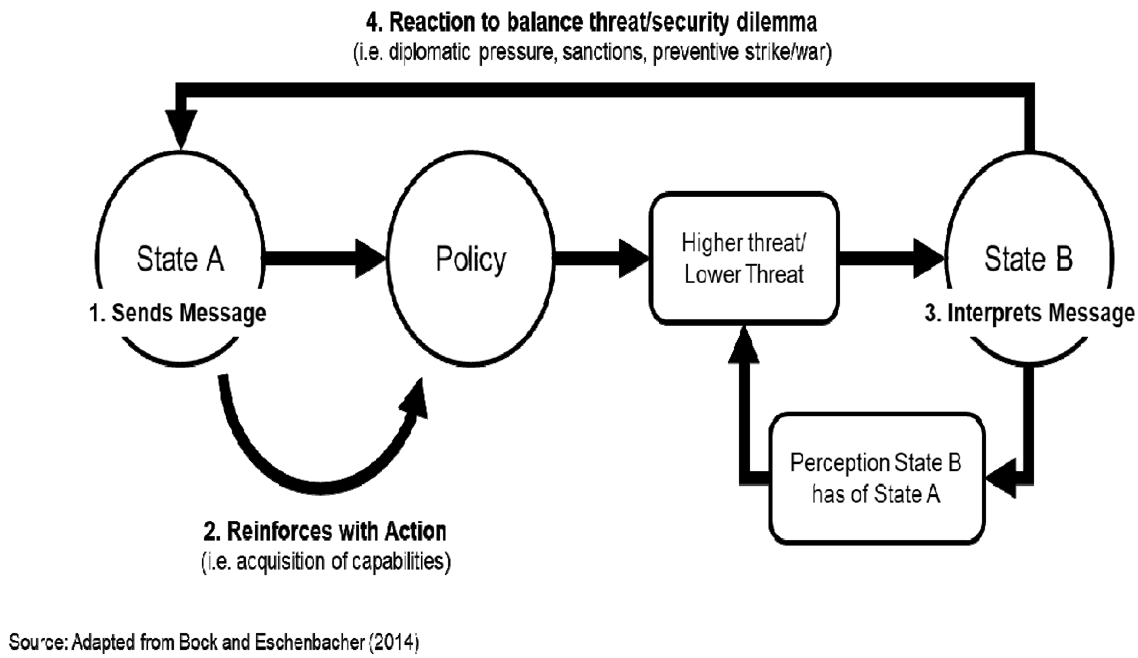

Figure A-1. Balance-of-Threat System: Origin and Balancing of Interstate Threats.

Actual or perceived threats originate when State A makes statements, and possibly reinforces those statements with actions (e.g., through the acquisition of military capabilities), which will in turn determine its foreign policy. State B observes the rhetoric and behavior of State A and continually interprets what it observes to determine whether the words and actions of State A constitute an actual, greater, equal, or lesser threat to its own security interest (i.e., peace and, ultimately, survival).

Depending on the perceived threat, which is a function of the perceived intention by State A to actually do what it is threatening to do, the capability of State A to actually carry out its intentions, and a general assessment of the trustworthiness of State A, State B will react to the perceived threat. Such reactions could consist of attempting to re-balance the threat (i.e., security dilemma) by taking a stronger posture and/or augmenting State B's own military capabilities, entering alliances with other states, seeking the support of the international community, or, in the worst-case scenario, triggering a preventive war.

Artificial general intelligence (AGI) refers to a computer that is, across the board, as intelligent as a human and therefore can support humans or even perform a broad range of interconnected tasks previously exclusively reserved for humans. The next evolutionary step beyond AGI would be ASI, which can be defined as “an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills.” In this context, this research seeks to identify specific areas in which AGI (also referred to as strong AI or human-level AI) can contribute to the practice of preventive diplomacy as a means to maintain a stable balance of threat.

With this background information, below is a digital copy of the survey and questions I provided to each survey participant for his/her consideration.

General Section

Information about Survey Participant
What is your primary area of expertise? (security policy, foreign policy, diplomacy, artificial intelligence)
How many years of experience do you have in the field?
What is your nationality? (dropdown list)
In which regions did you work/which regions did you cover geographically? (Choose the top three in descending order of experience [dropdown list])
How would you characterize your function? (Defense attaché, ambassador or other high-level diplomat, high-level military, high-level civilian government employee, academic/researcher, other [please specify])
What is your sex? (male, female)

Open Entry Questions
Based on your understanding of the point of departure of this research, where do you see the biggest potential for AGI in the context of preventive diplomacy and maintaining balances of threat (please be specific)?
Do you think AGI could eventually replace human agents in the conduct of preventive diplomacy? Why or why not (please be specific)?

Section on Monitoring

Monitoring refers to all activities related to watching the messages, actions, and policy changes of other states and actors, as well as other environmental factors. Monitoring allows for the detection of actual or impending shifts in the overall balance of threat and can enable the pinpointing of the sources of such changes. AGI can rapidly process large amounts of data drawn from a broad range of sources and detect connections between individual data points, allowing it to detect patterns.

<table>
<tr><td rowspan="2">Do you agree or disagree with the following statements?
(scale from 1 = completely disagree, 2 = somewhat disagree, 3 = neither disagree nor agree, 4 = somewhat agree to 5 = completely agree)</td><td colspan="4">Expert Weight[1]</td></tr>
<tr><td>SP</td><td>FP</td><td>DI</td><td>AI</td></tr>
<tr><td>AGI can contribute to obtaining a more complete picture of changes in the balance of threat (i.e., by detecting changes in rhetoric, behavior, acquisition of new capabilities, etc. from a variety of sources).</td><td>2</td><td>2</td><td>1</td><td>2</td></tr>
<tr><td colspan="5">Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?</td></tr>
<tr><td>AGI can contribute to obtaining a timelier picture of changes in the balance of threat.</td><td>2</td><td>2</td><td>2</td><td>2</td></tr>
<tr><td colspan="5">Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?</td></tr>
<tr><td>AGI could be used to set up an effective early warning system for early detection of changes in balances of threat.</td><td>2</td><td>2</td><td>1</td><td>2</td></tr>
<tr><td colspan="5">Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?</td></tr>
<tr><td colspan="5">Do you see other potential applications of AGI with regards to monitoring balances of threat (please be specific)?</td></tr>
</table>

[1] Expert weight can range between 1 (low) and 2 (high) and is attributed according to the relative level of expertise respondents have with respect to a particular question. SP= = Expert on Security Policy, FP= = Expert on Foreign Policy, DI= = Diplomat in the Field, AI= = Expert on Artificial Intelligence.

Section on Validation

Validation refers to all activities related to trying to assess the validity of information. This is important for information in general but is particularly relevant with respect to cross-referencing messaging (i.e., in the media) and actual actions. An example would be comparing reports of an alleged ceasefire violation in the media with what actually happened on the ground. This activity is extremely important for establishing the reliability of information, which, in times of information warfare, fake news, and other attempts to manipulate the public, has become ever more valuable.

Do you agree or disagree with the following statements? (scale from 1 = completely disagree, 2 = somewhat disagree, 3 = neither disagree nor agree, 4 = somewhat agree, to 5 = completely agree)	Expert Weight[2]			
AGI can contribute to improving the reliability of information about detected changes in the balance of threat (i.e., by validating changes from more than one source, thus screening out fake news or detecting information operations).	2	2	2	1
Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?				
Do you see other potential applications of AGI with respect to validating information within the context of balances of threat (please be specific)?				

[2] Expert weight can range between 1 (low) and 2 (high) and is attributed according to the relative level of expertise respondents have with respect to a particular question. SP= = Expert on Security Policy, FP= = Expert on Foreign Policy, DI= = Diplomat in the Field, AI= = Expert on Artificial Intelligence.

Section on Anticipation

Anticipation refers to the ability to foresee events. Within the balance-of-threat system, this could be in the area of the anticipated action by State A (the cause of the change in the system), the likely behavior of State A (i.e., in reaction to something State B or other actors do) and the reactions of other states and actors in the system as a response to the actions and reactions of States A and B. Combining the three types of anticipation in a wargaming setup would allow for the exploration of the likely future actions/reactions in a particular balance-of-threat system, enabling states to see likely paths of future development before taking any particular action.

Do you agree or disagree with the following statements? (scale from 1 = completely disagree, 2 = somewhat disagree, 3 = neither disagree nor agree, 4 = somewhat agree, to 5 = completely agree)	Expert Weight[3]			
	SP	FP	DI	AI
AGI can contribute to understanding and anticipating the likely reaction of a state or non-state actor within a given situational context (e.g., economic crisis, environmental changes).	2	2	1	2
Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?				
AGI can contribute to understanding and anticipating the likely behavior of a state or non-state actor in reaction to a specific action of another actor or other participants in the system (e.g., the reaction to aggressive rhetoric and whether it is backed or condemned by other actors).	2	2	2	2
Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?				
AGI could be used to set up a dynamic comprehensive behavioral model of all major actors in a system.	2	2	1	2
Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?				
Do you see other potential applications of AGI with regard to anticipation within the context of balances of threat (please be specific)?				

[3] Expert weight can range between 1 (low) and 2 (high) and is attributed according to the relative level of expertise respondents have with respect to a particular question. SP= = Expert on Security Policy, FP= = Expert on Foreign Policy, DI= = Diplomat in the Field, AI= = Expert on Artificial Intelligence.

Section on Solution Finding

Solution finding refers to the ability to find enough common ground among the participants in the balance-of-threat system to allow the system to return to a stable state and avoid conflict. This requires a thorough understanding of who the relevant participants are, what their individual needs and interests are, and how they relate to each other.

A common approach to solution finding is so-called system analysis, which provides an overview of the main actors and how close and amicable or hostile their relationships with one another are. Based on this systemic view, the importance of networked thinking and the ability to identify and potentially link cooperation on one issue to interactions on a second issue (linkage) becomes evident.

Do you agree or disagree with the following statements? (scale from 1 = completely disagree, 2 = somewhat disagree, 3 = neither disagree nor agree, 4 = somewhat agree, to 5 = completely agree)	Expert Weight[4]			
	SP	FP	DI	AI
AGI can contribute to system analysis, the identification of the relevant participants in a balance-of-threat system and how they relate to one another (i.e., by conducting screening of all available information and evaluating content to identify the strength and tone of the respective relationships).	2	2	1	2
Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?				
AGI can contribute to correct anticipation of the likely reaction of the relevant participants in relation to a specific issue or set of issues.	2	2	2	2
Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?				
AGI can assist in the detection of seemingly unrelated issues that, in combination with the disputed issue, offer opportunities for win-win situations and conflict resolution (linkage).	2	2	2	1
Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?				
Do you see other potential applications of AGI with regard to solution finding within balance-of-threat systems (please be specific)?				

[4] Expert weight can range between 1 (low) and 2 (high) and is attributed according to the relative level of expertise respondents have with respect to a particular question. SP= = Expert on Security Policy, FP= = Expert on Foreign Policy, DI= = Diplomat in the Field, AI= = Expert on Artificial Intelligence.

Section on Collective Learning

Collective learning" refers to the systematic and structured capture, dissemination, and implementation of lessons learned within the balance-of-threat system with the intention of making the system better (i.e., more stable and peaceful). The Assessment Framework for UN Preventive Diplomacy could serve as a starting point for a structure for systematic collective learning.

Genetic algorithms are search heuristics that are inspired by Charles Darwin's theory of natural evolution. Such algorithms reflect the process of natural selection, wherein the fittest individuals (or, in this case, the best approaches) in each generation are selected for reproduction in order to produce offspring that will be better adapted to the situation or environment.

The concept of a recursive self-improving (RSI) system requires computers with two major skills: conducting research on AI, and coding relevant changes to themselves. Such systems would not only learn but also constantly improve their own architecture. The main job of RSI systems is to determine how to make themselves more intelligent.

Do you agree or disagree with the following statements? (scale from 1 = completely disagree, 2 = somewhat disagree, 3 = neither disagree nor agree, 4 = somewhat agree, to 5 = completely agree)	Expert Weight[5]			
	SP	FP	DI	AI
10) AGI can contribute to assessing, consolidating, and evaluating relevant information towards the goal of identifying effective lessons learned.	2	2	1	2
Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?				
11) AGI can improve the overall effectiveness of preventive diplomacy (i.e., by employing a genetic algorithm or functioning as a recursive self-improving system with the objective of maximizing the effectiveness of preventive diplomacy).	2	2	2	2
Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?				

[5] Expert weight can range between 1 (low) and 2 (high) and is attributed according to the relative level of expertise respondents have with respect to a particular question. SP= = Expert on Security Policy, FP= = Expert on Foreign Policy, DI= = Diplomat in the Field, AI= = Expert on Artificial Intelligence.

Do you agree or disagree with the following statements? (scale from 1 = completely disagree, 2 = somewhat disagree, 3 = neither disagree nor agree, 4 = somewhat agree, to 5 = completely agree)	Expert Weight[5]			
	SP	FP	DI	AI
12) AGI can support the dissemination of critical information and decision support to the most relevant actors in a given context (i.e., by providing them, in a timely manner, with realistic scenarios of what will happen if they behave in a specific way, thus encouraging them to abstain from an intended course of action that may have negative consequences).	2	2	2	1
Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?				
13) AGI will challenge existing information power structures by leveling the playing field, shifting the definition of competitive advantage away from the asymmetric possession of proprietary information towards having the capacity for superior analysis and interpretation of publicly available information, a capacity that will be more broadly available.	2	2	1	2
Follow up if value of previous question is 2 or less: Can you elaborate on why you disagree?				
Do you see other potential applications of AGI with regards to collective learning or preventive diplomacy and the management of balances of threat (please be specific)?				

Closing Section

Thank you very much for participating in this online survey.

Beyond the goal of testing some of the initial hypotheses, I am also very interested in your additional views on the topic, and would like to follow up with you personally. Please also indicate if you are interested in receiving a copy of the final thesis.

Would you be available for a follow-up by telephone or video call? (Yes/no, suggested time slots)
Would you like to receive an electronic copy of the final thesis? (Yes/no, email address)

*** End of Questionnaire ***

References

Anderson, M. R. (2017). Twenty years on from Deep Blue vs. Kasparov: How a chess match started the big data revolution. The Conversation website. https://theconversation.com/twenty-years-on-from-deep-blue-vs-kasparov-how-a-chess-match-started-the-big-data-revolution-76882.

Baarslag, T., Hendrikx, M. J. C., Hindriks, K. V., & Jonker, C. M. (2015). Learning about the opponent in automated bilateral negotiation: A comprehensive survey of opponent modeling techniques. *Autonomous Agents and Multi-agent Systems, 30*(5), 849-898.

Baron, J. (2013). *Great power peace and American primacy: The origins and future of a new international order*. Palgrave Macmillan.

Black, J. (2010). *A history of diplomacy*. Reaktion.

Bock, A., & Eschenbacher, S. (2014). *The social construction of threats. The Iran nuclear crisis as a textbook example for Jack Mezirow's transformative learning theory.* New Prairie Press.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Boutros-Ghali, B. (1992). An agenda for peace. *International Relations, 11*(3), 201-218. https://doi.org/10.1177/004711789201100302.

Brooks, S. G. (1997). Dueling realisms. *International Organization, 51*(3), 445-77.

Brooks, S. G., & Wohlforth, W. C. (2008). *World out of balance: International relations and the challenge of American primacy*. Princeton University Press.

Butterfield, H., Wight, M., & Bull, H. (1966). *Diplomatic investigations: Essays in the theory of international politics*. Allen & Unwin.

Canapati, P. (2010). Reverse-engineering of human brain likely by 2030, expert predicts. *Wired*. https://www.wired.com/2010/08/reverse-engineering-brain-kurzweil/

Carswell, A. J. (2013). Unblocking the UN Security Council. *Journal of Conflict & Security Law, 18*(3), 453-80.

Chayes, A. (1974). *The Cuban missile crisis: International crises and the role of law*. Oxford University Press.

Clark, I. (2007). *Legitimacy in international society*. Oxford University Press.

Computer History Museum. (n.d.). *Timeline of computer history.* http://www.computerhistory.org/timeline

Correll, J. T. (2005, August 1). Airpower and the Cuban missile crisis. *Air Force Magazine.* https://www.airforcemag.com/article/0805u2/

De Jonge, D., & Sierra, C. (2017). D-Brane: A diplomacy playing agent for automated negotiations research. *Applied Intelligence, 47*(1), 158-77.

DeepMind. (n.d.). The Google DeepMind challenge match. https://deepmind.com/alphago-korea

DiploFoundation (2019). Mapping the challenges and opportunities of artificial intelligence for the conduct of diplomacy. https://www.diplomacy.edu/sites/default/files/AI-diplo-report.pdf

DiploFoundation. (n.d.). Digital diplomacy. https://www.diplomacy.edu/e-diplomacy

Dorn, A. W., Webb, S., & Pâquet, S. (2020). From wargaming to peacegaming: Digital simulations with peacekeeper roles needed*. International Peacekeeping*, 27(2), 289-310.

Du, S.; Zhang, Y.; Qin, R.; Yang, Z.; Zou, Z.; Tang, Y.; Fan, C. (2016). Building change detection using old aerial images and new LiDAR data. *Remote Sensing 8* (12): 1030. https://doi.org/10.3390/rs8121030.

Earle, E. M. (1943). Power politics and American world policy. Review of Nicholas J. Spykman's *America's Strategy in World Politics*. *Political Science Quarterly 58*(1), 102-104.

Encyclopedia Britannica. (n.d.). Definition of Diplomacy. https://www.britannica.com/topic/diplomacy

Frankel, B. (1996). Restating the realist case: An introduction. *Security Studies, 5*(3), 9-20.

Frieden, J., Lake, D. A., & Schultz, K. A. (2019). *World politics* (Fourth edition.). Norton.

Fulton, S. (18 Nov 2020). Top 500: Japan's Fugaku still the world's fastest supercomputer. Data Center Knowledge. https://www.datacenterknowledge.com/supercomputers/top500-japan-s-fugaku-still-world-s-fastest-supercomputer

Gabbat, A. (17 Feb 2011). IBM computer Watson wins Jeopardy clash. *Guardian.* https://www.theguardian.com/technology/2011/feb/17/ibm-computer-watson-wins-jeopardy#:~:text=Ken%20Jennings%2C%20the%2074%2Dtime,Terminator%2C%20they%20lost%20each%20time

Global Security (2011). The Cuban missile crisis of 1962: Presenting the photographic evidence abroad. https://www.globalsecurity.org/intell/library/cubakent.htm

Goodrich, L. M. (1947). From League of Nations to United Nations. *International Organization 1*(1), 3-21.

Heinbecker, P., & Goff, P. M. (2005). *Irrelevant or indispensable? The United Nations in the twenty-first century.* Wilfrid Laurier University Press.

Hume, D. (1752). *Political discourses.* A. Kincaid and A. Donaldson.

James, P. (2002). *International relations and scientific progress: Structural realism reconsidered.* Ohio State University Press.

Kunertova, D. (2021). Hypersonic weapons: Fast, furious . . . and futile? *RUSI Newsbrief, 41*(8).

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology.* Viking.

Kurzweil, R. (1999). *The age of spiritual machines: When computers exceed human intelligence.* Viking.

League of Nations. (1919). *The covenant of the League of Nations.* Chapman.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.

Little, R. (2007). *The balance of power in international relations: Metaphors, myths and models.* Cambridge University Press.

Lobell, S. E. (2010). *Structural realism/offensive and defensive realism.* Oxford Research Encyclopedias. https://oxfordre.com/internationalstudies/internationalstudies/view/10.1093/acrefore/9780190846626.001.0001/acrefore-9780190846626-e-304

Lowenthal, M. M., & Clark, R. M. (eds.). (2016). *The five disciplines of intelligence collection*. Thousand Oaks, CA: CQ Press.

Luard, E. (1992). *The balance of power: The system of international relations, 1648-1815.* MacMillan.

Lynn-Jones, S. M. (1995). Offense-defense theory and its critics. *Security Studies, 4*(4), 660-691.

Mallawaarachchi, V. (2017, July 8). Introduction to genetic algorithms—Including example code. Towards Data Science. https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth Summer Research Project on artificial intelligence: August 31, 1955. *AI Magazine, 27*(4), 12.

Miller, B. (1996). Competing realist perspectives on great power crisis behavior. *Security Studies, 5*(3), 309-357.

Missile Defense Project. (15 Jun 2018). "SM-78 Jupiter" Missile Threat. Center for Strategic and International Studies. https://missilethreat.csis.org/missile/jupiter/

Moore, G. E. (2006 [1965]). Cramming more components onto integrated circuits. *Electronics 38* (8). Reprinted in *IEEE Solid-State Circuits Society Newsletter, 11*, no. 3 (April 19, 1965): 33-35.

Morgenthau, H. J. (1972). *Politics among nations: The struggle for power and peace* (5^{th} ed.). Knopf.

Nexon, D. H. (2009). The balance of power in the balance. *World Politics, 61*(2), 330-359.

Nixon, R. (1969). *Six crises*. Doubleday.

OpenWorm. (n.d.). Building the first digital life form. https://www.openworm.org

Oriesek, D., & Schwarz, J. O. (2008). *Business wargaming: Securing corporate value*. Gower.

Oriesek, D., & Schwarz, J. O. (2021). *Winning the uncertainty game.* Taylor and Francis.

Osiander, A. (1994). *The states system of Europe, 1640-1990: Peacemaking and the conditions of international stability*. Oxford University Press.

Perla, P. P. (1990). *The art of wargaming: A guide for professionals and hobbyists.* Naval Institute Press.

Prithvinath Reddy, P. (30 Jan 2020). Artificial superintelligence: A recursive self-improvement model. https://easychair.org/publications/preprint_open/dWLr

Renshon, J. (2009). Assessing capabilities in international politics: Biased overestimation and the case of the imaginary 'missile gap'. *Journal of Strategic Studies, 32*(1), 115-147.

Rucellai, B. (1733). *Bernardi Oricellarii: De bello Italico commentarius*. Londini.

Sandberg, A., & Bostrom, N. (2008). Whole brain emulation: A roadmap. Future of Humanity Institute. https://www.fhi.ox.ac.uk/brain-emulation-roadmap-report.pdf

Schweller, R. L. (2003). The progressiveness of neoclassical realism. (pp. 311-347). In C. Elman & M. Elman (eds.), *Progress in international relations: Appraising the field.* MIT Press.

Scott, L., & Gerald, H. G. (2015). *The Cuban missile crisis: A critical reappraisal.* Taylor & Francis.

Singh, A. (30 Jan 2017). Deep learning will radically change the ways we interact with technology. *Harvard Business Review*. https://hbr.org/2017/01/deep-learning-will-radically-change-the-ways-we-interact-with-technology

Snead, D. L. (1997). Eisenhower and the Gaither Report: The influence of a committee of experts on national security policy in the late 1950s. Ph.D. Dissertation. University of Virginia. https://digitalcommons.liberty.edu/fac_dis/57

Snyder, J. (2013). *Myths of empire.* Cornell University Press.

Stilwell, B. (2 April 2018). The Cuban missile crisis: 13 days that almost ended the world. https://www.wearethemighty.com/articles/tick-tock-cuban-missile-crisis/

Stone, M., Lewis, M., & Perry, M. (16 Dec 2021). U.S. builds new software tool to predict actions that could draw China's ire. https://www.reuters.com/world/china/us-builds-new-software-tool-predict-actions-that-could-draw-chinas-ire-2021-12-15/

Taliaferro, J. W. (2000). Security seeking under anarchy: Defensive realism revisited. *International Security, 25*(3), 128-161.

United Nations. (1945). Charter of the United Nations. https://www.un.org/en/about-us/un-charter

United Nations Joint Inspection Unit. (2017). Results-based management in the United Nations system: High-impact model for results. https://www.unjiu.org/sites/www.unjiu.org/files/jiu_note_2017_1_english_0.pdf

United Nations University. (2018a). Assessment framework for UN preventive diplomacy: An approach for UN mediators and international policymakers. https://i.unu.edu/media/cpr.unu.edu/post/2729/UN-Preventive-Diplomacy-Assessment-Framework.pdf

United Nations University. (2018b). Policy paper and case studies: Capturing UN preventive diplomacy success: How and why does it work? https://collections.unu.edu/eserv/UNU:6572/UN-Preventive-Diplomacy-Policy-Paper-and-Case-StudiesWEB.pdf

United Nations War Crimes Commission. (1946). *The crime against peace.* G. Blanchong et cie.

United States Army. (3 Jun 2004). Chapter 7: Deployment. https://history.redstone.army.mil/miss-jupiter.html

Urban, T. (2015). The AI revolution: The road to superintelligence. https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

Vinge, V. (1993). *The coming technological singularity: How to survive in the post-human era.* San Diego State University. https://edoras.sdsu.edu/~vinge/misc/singularity.html

Von Clausewitz, C., Howard, M. E., & Paret, P. (2008). *On war.* Princeton University Press.

Von Gentz, F. (1806). *Fragments upon the balance of power in Europe. Making of modern law*. M. Peltier.

Walch, K. (17 Sep 2019). The seven patterns of AI. *Forbes*. https://forbes.com/sites/cognitiveworld/2019/09/17/the-seven-patterns-of-ai/?sh=70963ea512d0

Walt, S. (2009). Alliances in a unipolar world. *World Politics 61*(1), 86-120. doi:10.1017/S0043887109000045

Walt, S. M. (2002). The enduring relevance of the realist tradition. (pp. 197-234). In I. Katznelson & H. Milner (eds.), *Political Science*. Norton.

Walt, S. M. (1987). *The origins of alliance.* Cornell University Press.

Walt, S. M. (1985). Alliance formation and the balance of world power. *International Security, 9*(4), 3-43.

Waltz, K. N. (1979). *Theory of international politics.* Addison Wesley.

Watts, J. M. (25 Aug 2016). World's first self-driving taxis hit the road in Singapore. *Wall Street Journal.* https://www.wsj.com/articles/worlds-first-self-driving-taxis-hit-the-road-in-singapore-1472102747

Woolf, S. (2002). Reading Federico Chabod's *Storia dell'idea d'Europa* half a century later. *Journal of Modern Italian Studies 7*(2), 269–292. https://doi.org/10.1080/13545710210138009

Woronko, M. (6 Feb 2019). Is it natural to upload consciousness? *Medium*. https://medium.com/the-philosophers-stone/it-it-natural-to-uplad-consciousness-a0fb4ab1f9d1