

**REPORT DOCUMENTATION PAGE**

*Form Approved  
OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.  
**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YYYY) 06/04/2021	2. REPORT TYPE FINAL	3. DATES COVERED (From - To)
---	-------------------------	------------------------------

4. TITLE AND SUBTITLE An Ethically Assisted Future: A Forecasting of Artificial Intelligence Powered Ethical Assistants	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)  CDR Aaron A Robinson, USN  Advisor: Dr. Jason Kelly	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Ethics and Emerging Military Technology (EEMT) Program U.S. Naval War College 686 Cushing Road Newport, RI 02841-1207	8. PERFORMING ORGANIZATION REPORT NUMBER
--	--

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Dr. Tom Creely, Director EEMT Program US Naval War College 686 Cushing Road Newport, RI 02841-1207	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT  
Distribution Statement A: Approved for public release; Distribution is unlimited. Reference: DOD Directive 5230.24

13. SUPPLEMENTARY NOTES  
A paper submitted to the Faculty of the U.S. Naval War College in partial satisfaction of the requirements of the Ethics and Emerging Military Technology Graduate Certificate Program. The contents of this paper reflect my personal views.

14. ABSTRACT  
Artificial Intelligence (AI) can prove a powerful partner for humans in selecting a national military strategy that aligns with nations' internal ethical values and maintains transparent narratives for allies and potential adversaries alike. Leveraging machine learning, crowd-sourced expert opinions, and other development techniques, Ethical Assistants can be created to conform to specific ethical constructs. An ideal Ethical Assistant is aligned to the ethical values of the state or organization developing the Assistant and possesses an equally valid ethical construct of adversaries, allies, and neutral actors affected by the strategies and decisions being weighed.

15. SUBJECT TERMS  
Ethical Assistant, Ethics, EEMT, Ethics in Emerging Technology, Trust in AI

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Director EEMT Program
UNCLAS	UNCLAS	UNCLAS		62	19b. TELEPHONE NUMBER (Include area code)

An Ethically Assisted Future:  
A Forecasting of Artificial Intelligence Powered Ethical Assistants

By:  
Aaron A Robinson  
CDR USN

A paper submitted to the Faculty of the Naval War College in partial satisfaction of the requirements of the Ethics and Emerging Military Technologies Graduate Certificate Program. The contents of this paper reflect my own personal views and are not necessarily endorsed by the Naval War College or the Department of the Navy

## Table of Contents

Abstract	3
Acknowledgements	4
USS Pride of America Incident, 30 June 2031	5
The Purpose of Strategy	10
Strategists Specifically Need <i>Ethical Assistants</i> ?	11
Speed of Decisions and Actions Meets Big Data	13
Government Structure Defining Values	15
The Ability of AI to Facilitate Strategic Genius	18
What is an Ethical Assistant?	20
Military Operations with an Ethical Assistant	23
Artificial Intelligence's Relationship with Ethical Assistants	24
How Different Aspects of AI Design Methodology Play a Role in Developing an Ethical Assistant	25
Other Actors Must be Accounted For	29
Creating an Ethical Assistant Requires a Team of Experts	30
Gaining Trust and Validating Results	31

Programs Assisting Ethical Decision-Making Ability Today	35
Lessons from Other AI decision Aids and Programs	37
Requirement for Multiple Design Methodologies	39
Different Lenses and Structures	41
Using the Iraq Invasion of 2003 as an Example	45
An Ethical Assistant Continues to Provide Support After Initial Decision Points	48
Ethical Assistant’s Inherent Resistance to Gamesmanship	51
Ethical Assistants can be Applicable in all Fields	51
Knowing Ourselves	53
Conclusion	54
Bibliography	55

## Abstract

Artificial Intelligence (AI) can prove a powerful partner for humans in selecting a national military strategy that aligns with nations' internal ethical values and maintains transparent narratives for allies and potential adversaries alike. Leveraging machine learning, crowd-sourced expert opinions, and other development techniques, Ethical Assistants can be created to conform to specific ethical constructs. An ideal Ethical Assistant is aligned to the ethical values of the state or organization developing the Assistant *and* possesses an equally valid ethical construct of adversaries, allies, and neutral actors affected by the strategies and decisions being weighed. AI-powered Ethical Assistants can increase the sensitivity to and value for human lives before selecting a strategy. Ethical Assistants would consider the potential loss of life from military actions that support the strategy proactively versus a reactionary analysis during the execution of the strategy. Specific military operations can also be analyzed to ensure alignment with the overall strategy and values. It will flag inconsistencies between professed values and the effects achieved by Courses of Actions (COAs) in the near, mid, and long term before operations commence. It will lead to debates on the relative merits of competing ethical values with specificity to make the potential impacts felt vice a debate of theoretical ideals. Additionally, the Assistant can continually assess if the selected course of action conforms to the ethical worldview and provides warning flags when missions begin to deviate from prescribed objectives and intents. As advances in AI proceed, an Assistant can recommend changes and alternative courses of action which better conform to the described ethical views. The role of AI, and technology in general, is immense in creating AI Ethical Assistants. Still, the power and efficacy ultimately reside in the self-actualization of human actors evaluating their own ethical systems and values, the ethical system and values of potential adversaries and allies, and the myriad biases that plague each side.

## **Acknowledgements**

I would like to thank Dr. Jason Kelly for graciously sharing his expertise and valuable insights. He was instrumental in aiding me in organizing my thoughts and structuring my paper. I owe him a debt of gratitude for helping me to see this project through. I would also like to thank Dr. Tom Creely for his guidance and mentorship throughout the entire Ethics and Emerging Military Technologies (EEMT) graduate certificate program journey. Additionally I would like to thank Dr. Tim Schultz for inspiring me to apply for the EEMT program and his continued support and instruction throughout my time at the War College. I greatly appreciate the inspirational teachings from Dr. Michael Dennis, Dr. Tim Demy, and Dr. Tom Gibbons that helped focus my thoughts on the applicability of ethics throughout the year. I would also like to extend my grateful appreciation to Ms. Isabel Lopes, Mr. Eamon McKeever, and Ms. Karina Barao for assisting with research and editing. Furthermore, I want to thank Dr. Joan Johnson-Freese, Dr. Andrew Wilson, CAPT Timothy Urban, and LCOL Kenneth Sanders for increasing my ability to see where ethics and strategy intertwine.

## **USS Pride of America Incident, 30 June 2031**

Consider an imagining of the role technology can play in accelerating a strategy and execution mismatch through the lens of a fictional incident in the South China Sea region.

Selected excerpts of the findings of the US Senate Investigation into the USS. Pride of America incident on 30 June 2031 and the subsequent expulsion of all US diplomats from the nations in the greater South China Sea region.

Background: In response to People's Liberation Army Navy (PLAN) fleet maneuvers in preparation for the 110<sup>th</sup> anniversary of the Chinese Communist Party (1 July 2031), US Indo-Pacific Command dispatched an additional Carrier Strike Group to US Seventh Fleet to operate alongside the forward deployed Carrier Task Force Seventy in mid-June. An additional F-22 squadron and two F-35 squadrons arrived on Guam on June 27<sup>th</sup> to augment an ongoing exercise that already included a significant increase in bomber and attack aircraft operating out of Andersen Air Force Base on Guam. Standard operational deployments and activities continued.

Excerpts from Senate Hearing on 9 July 2031:

Admiral White, Commander Indo-Pacific Command: [in response to a statement from the Senator from Missouri on negligent oversight] It is inaccurate to say that the Flag and General Officers and other senior leadership on my staff and under my command were not tightly focused and observant of the conduct of the forces under my command. In addition to standard reports, I held additional daily video teleconferences with all Component Commanders under my command to discuss the developing situation and rising tensions. During the week before the incident, I had additional meetings with

Pacific Fleet, Seventh Fleet, Carrier Task Force 70 and 71, and others closest to the Chinese fleet.

Rear Admiral Green, Commander Carrier Task Force 70: [In response to a question from the Senator from Texas on his understood mission and purpose] Carrier Task Force 70 and 71 were charged with 1) deterring the PLAN from striking Taiwan and 2) to demonstrate the US resolve to support our allies around the South China Sea in the face of any Chinese aggression.

[in response to a question regarding the interactions between the PLAN and US fleets] Our Officers and Sailors conducted themselves in an exemplary manner during multiple 'tests' by the PLAN of our defense-in-depth procedures. Analysis of the three near encounters by our surface vessels reveals that our ships adhered to the Rules of the Road and clearly communicated and captured the encounters, which greatly enhanced the power of our narrative outlining the aggressive and dangerous conduct of the PLA(N). In a similar fashion, our air intercepts were also outstanding. We can arrange for more detailed reviews of these encounters in a closed and classified hearing.

The Senator from Maine: This is the tenth hour of our hearings, and not once has anyone raised the central issue. Who was monitoring the USS. Pride of America? This single destroyer off the Vietnamese and Chinese Coast was 1,000 miles away from the carriers and air activity my friends on this committee are so eager to learn about. Which Admiral or General was following the activity in this region? How did a single ship with a crew of 250 personnel end up sinking 35 unarmed fishing vessels and killing, at last count, 400 people, with another 500 missing? These numbers are laughably low, as we will all learn to our sorrow. Within a day, Vietnam, Cambodia, Malaysia, Indonesia, Singapore, and the

Philippines expelled our Ambassadors or Heads of Missions. Two weeks later, we are still receiving notifications of additional expulsions. ASEAN expelled our monitors as did the Trans Pacific Partnership. Who on this panel was following the reports and actions of the Pride of America? Where was the scrutiny over the conduct in those highly contested and controversial waters?

Admiral Grey, Commander Pacific Fleet: Senator, our watch floor constantly monitors the activities of every ship at sea. The activities of the largest fleet China had put to sea naturally took most of our attention, but we were still observing and advising the Pride of America. Ultimately the Captain felt that the swarming of the numerous small craft around his ship that night presented a clear and imminent danger to his crew and he took actions to ensure the safety of the ship and crew.

Senator from Maine: Admirals, to say I am disappointed in the conduct of you and your fellow Flag and General Officers is a gross understatement. Your negligence in prioritizing the ultimate aims of this great nation has handed the Chinese Communist Party its greatest victory since its founding. Our mission is to preserve the security of our nation and defend our interests. You have irretrievably failed on both accounts. The US has lost any influence within the South China Sea for decades to come. Our values are now seen around the world as hypocritical and self-serving. Innocent fishermen trying to earn a living and feed their families are now dead. These were our friends and allies. We murdered them.

Key findings from unclassified report released 14 August 2031: On 30 June 2031, the USS. Pride of America was conducting freedom of navigation operations in waters claimed by both Vietnam and China. At approximately 0230 local time a group of approximately 60

fishing vessels, at least ten of which had extensive communications antennae, left a nearby island heading straight towards the Pride of America. As the flotilla neared the Pride of America the vessels turned on spot and flood lights, which blinded the bridge crew and lookouts.<sup>1</sup> Without the presence of any PLAN ships, the Pride of America could not believe that these ships were acting without the knowledge of the Chinese military. Minutes later, the data feed on the Pride of America's surface, subsurface, and air drones went offline.<sup>2</sup> Assuming a sophisticated irregular and multi-domain attack was in progress, the Captain authorized the activation of the automated defense systems. Within 15 minutes, the surviving vessels had dispersed, and the Captain ordered the system returned to standby. Damage assessment is 37 fishing vessels sunk, 12 fishing vessels significantly damaged, and 753 individuals killed.

Snapshot from 2035: A UN-sanctioned investigation largely corroborated the U.S investigation's findings. A separate and secret investigation by the Chinese Communist Party has been widely reported in open sources as uncovering that the PLAN Commander of the Southern Naval Region was accepting bribes from clandestine fishing operations to supplement his family income. There was no link reported to the CCP or PLA leadership as a whole. The Commander, his brother, and two sons have not been seen since their arrest while trying to cross into Vietnam on 2 July 2031. Numerous PLAN officers have

---

<sup>1</sup> Lights are commonly used for night fishing of smaller fish and squid. No data indicates the angle or direction of the lighting systems. The investigators were unable to determine if the rapid general increase in light level or specific spot lighting of the bridge were the cause of the momentarily blinding of the bridge crew aboard the USS. Pride of America.

<sup>2</sup> An error in the updating of the encryption cycling programming for the surveillance drones resulted in a loss of communications for approximately 30 minutes before the crew rectified the mismatch. Four subsurface and two surface drones were entangled in fishing nets and lines, but continued to send data back to the USS. Pride of America.

been moved to other regions, but no evidence if these moves are linked to the suspected pay to fish scheme run by the previous Commander.

In 2034, ASEAN allowed a US observer to return to the group. In 2035, Vietnam, Indonesia, and Malaysia again exchanged ambassadors with the US. The Philippines, which lost the most lives during the Pride of America incident, has continued a closer alignment with the CCP. At this time, China, with a proxy Philippines, has almost complete control over the islands and natural resources within the South China Sea. The US, and its military specifically, continue to try to learn what went wrong and why.

How did the US military's values and its decision-making process lead to that fateful morning of 30 June?  
Where in the quest for military superiority did the overarching objective to defend American values and the American way of life disappear and become replaced by maintaining military superiority above all else?

## **The Purpose of Strategy**

The ultimate object of strategy, which may lead to warfare, must be the security of a society's way of life and the society's values. The growth in AI capabilities provides an opportunity to create an assistant that can advise an individual on how her strategic decisions align with her professed ethics based on a range of possible outcomes weighted for probability. For common language in this paper, such a program will be termed an Ethical Assistant. An Ethical Assistant will come to be a vital partner in a world consistently changing and advancing through human creativity aided by leaps in computer processing power.

The harnessing of the same processing power driving change is necessary to assist in selecting national military strategies employing the fruits of and navigating the impacts of these changes. Military power must be viewed and exercised as more than just a counterbalance to an adversary's actions and postures. Scrutiny by leaders and decision-makers must focus on the potential encounters with the most significant risk or reward to achieve the desired strategic ends. Abandonment of core principles and values changes society and nation more effectively than a pure military defeat. The national conscience must be preserved and not subverted into accepting means or ways that betray the people's character. It is easy to imagine a scenario where a miscalculation of the environment or the improper prioritization of competing objectives leads to an outcome similar to the fictional vignette above. Fortunately, a society defines its values, and while allies and adversaries can influence the environment, a nation cannot be forced to abandon its values or conscience. It must willingly choose to abandon its values.

The process of creating and partnering with Ethical Assistants will also reveal our human nature. With an initial self-assessment at the start of the design phase, confronting new biases and contradictions as they arise during the design process will refine self-acknowledgment. The very act and

art of creation of the Ethical Assistant reveals personal and societal values. The benefits of the Ethical Assistant manifest in two intermeshed and co-equal efforts. One is to make the best decisions to align ethics, actions, and aims. The other is to better understand and appreciate the nature and importance of our values. Through creating the Ethical Assistant and then partnering afterward, the design process exposes the true hierarchy of values. These simultaneous outcomes constantly complement one another, further strengthening the power of narrative and actions to achieve our desired aims.

### **Strategists Specifically Need *Ethical Assistants*?**

Every individual confronts the fundamental question: ‘what is my purpose?’ Descartes famously asserted “cogito, ergo sum,” but that only establishes the uniqueness of each individual in the grand journey of life. Aristotle asserted humans are political creatures, that is, social creatures, and as such, humans are drawn to one another and exist in context to one another. Plato’s Allegory of the Cave describes the complexity of the environment and reveals the medium through which humans actually relate to one another: language. Humans describe the world and their desires from their viewpoint, seeking to influence other humans to share in their views. Put another way, individuals unique in thought and perception share stories with each other to further their own vision for the future. The motivator for each individual story is the individual’s ethics. The actions the individual takes to push their story is ethics in practice. In short, an individual’s life is the story of her ethics. Her identity is structured by the projection of her ethical views.<sup>3</sup>

As men and women come together to form larger and larger coalitions of people, the overlap of the common narrative is narrowed to a degree. At the level of a nation-state, the core identity shared by

---

<sup>3</sup> Shannon Valor makes the argument that virtue ethics best captures such a view in the context of today’s technological advances. See Shannon Valor, *Technology and the Virtues* (New York: Oxford University Press, 2016).

the people is the quality that makes them a nation. It is why European drawn borders in the Middle East and Africa fail to unify the inhabitants into a state. The possession of a shared story is the bond that unites a nation, not geographic proximity or skin color, or any single attribute. No other attribute is as central to the characterization of the nation as its collective story. The defense of the nation's identity and its way of life, the security of the state's values, and the practice of its ethics should therefore be the foremost concern in any strategy taken by the nation.

For the United States today, maintaining the moral and ethical high ground is the essential attribute in winning the struggle to shape the international narrative towards American values of freedom and equality. These American ideals bind the citizens together across all strata of society. A commitment to freedom of religion and speech and thought draws the greatest pool of intellectual talent to the United States. It offsets the smaller relative population compared to China or India or Western Europe or any more populous and growing region. These values at the core of the United States' representative democracy also counteract the efficiency advantage of authoritarian and autocratic states, such as China or Russia, in organizing national resources. Not only are the common ethical values of the United States the core of the identity of the state, but they are also the state's greatest advantage in any strategy.

During a war, the greatest long-term risk derives from weakening the ethical identity of the state. The manifestation of the state that existed at the moment the state chose to go to war for a cause will cease to exist if the state abandons the principles that led it to fight in the first place. In addition to altering the very character of the state, it risks losing any allies tied by shared history or values. Which partners would stand with the United States if it began to replicate China? Where would the world's entrepreneurs and inventors go if the legal and economic system no longer appeared free and equitable? Great care is required then in crafting the strategy and operations so that during the war, the ways and means of victory do not become a self-inflicted mortal wound on the victor. Every age

presents unique challenges to overcome the difficulty of balancing passion, reason, and chance in alignment with the desired objectives. However, the nature of the change in the 21<sup>st</sup> century places a particular risk on ethical identity. By partnering with these advances, such as with an Ethical Assistant, strategists can navigate the changing technological environment without giving up humanities' ethical autonomy.

### **Speed of Decisions and Actions Meets Big Data**

Access to resources and information being equal, the speed of quality decision-making is a determining factor in victory or defeat during the conflict. However, organic mental processing speed alone is too slow to maintain pace with the volume of data and the speed of processors going forward.<sup>4</sup> AI will assist in sifting through data to make decisions, and those recommendations must provide ethical context to defend the state's ethical identity. The United States Navy has long held to the doctrine of Mission Command to delegate authority down to the lowest level possible to accelerate and weaponize the decision cycle time for the Navy. The United States Air Force's John Boyd famously described the 'OODA loop' - observe, orient, decide, and act - which leverages agility in decision making to defeat adversaries. America is not alone in identifying the criticality of the pace of thought. "Chinese military doctrine now holds that the side which can make and execute battlefield decisions most quickly will gain the decisive advantage in any conflict."<sup>5</sup> These human-based methods are still relevant and provide

---

<sup>4</sup> The human brain possesses very powerful organic processing power, but is not structured to analyze mass amounts of data in a decision tree format that define speeds of computer processors. For approximate speed calculations and the limitations of comparing the human brain to computer processing power see Kris Sharma, "How Does the Human Brain Compare to a Computer?" *Crucial: Technology*, updated 20 April, 2020, <https://www.crucial.com/blog/technology/how-does-the-human-brain-compare-to-a-computer> and "New Measure of Human Brain Processing Speed," *MIT Technology Review: Biotechnology*, updated 28 August 2009, <https://www.technologyreview.com/2009/08/25/210267/new-measure-of-human-brain-processing-speed/>.

<sup>5</sup> Michele A. Flournoy, "How to Prevent a War in Asia: The Erosion of American Deterrence Raises the Risk of Chinese Miscalculation," *Foreign Affairs*, June 18, 2020, <https://www.foreignaffairs.com/articles/united->

American service members with an advantage over their counterparts today. These decisions need to factor in the ethical implications, as illuminated in the vignette. Still, the incredible increases in communications technology have eroded the comparative advantage of speed for the United States.

In contrast to a doctrine relying on the lowest level capable of taking action, centralized command and control centers can nearly instantaneously direct actions and do so with a pool of experienced and informed personnel. Rather than individuals trying to determine the Commander's intent behind the orders or plan, the Commander herself can direct deviations from her initial orders to confront a changing environment. While this raises many questions on how to develop future Commanders with the requisite decision-making experience to assume the role, it does remove the extraneous expenditure of military resources towards ends not central to the objective. The competition between centralized command and control, long a staple of communist and autocratic regimes, and decentralized command and delegated authority, a trend among western democracies, would rely on training and experience were it not for a larger disruption to decision-making capabilities: data.

By many factors over, a significantly larger amount of data is available for processing when weighing actions and decisions today than even a decade ago. The sheer volume of data and the diversity of the data sets can slow down the rate of decision-making. Still, they may offer the possibility of a higher quality decision and more effective action. However, to keep pace with the higher speed of contemporary warfare, which covers cyber, space, and other new domains alongside more traditional

---

states/2020-06-18/how-prevent-war-asia. Additional sources on Chinese doctrine include Anthony H. Cordesman "China's New 2019 Defense White Paper: An Open Strategic Challenge to the United States, But One Which Does Not Have to Lead to Conflict," Center for Strategic International Studies July 24, 2109, <https://www.csis.org/analysis/chinas-new-2019-defense-white-paper>; Steve Sacks, "China's Military has a Hidden Weakness," *The Diplomat*, April 20, 2021, <https://thediplomat.com/2021/04/chinas-military-has-a-hidden-weakness/>; and the Chinese 2019 Defense White Paper. See also CSIS Chinese Power Project, particularly the Security section: <https://www.csis.org/programs/china-power-project/security>.

environments, individuals cannot pause or hesitate before every action to ponder the implications to the national interest and the nation's ethical identity. The immediacy required for some of these decisions creates a requirement for investment and reliance upon automated systems to augment human operators to create and maintain temporal advances in decision making.

How AI will augment human operators in warfare is currently being debated. The ethical constraints and level of control have not been worked out, but a few trends appear to be forming. Western liberal societies lean towards retaining humans in the loop to a greater extent than autocratic regimes such as China and Russia.<sup>6</sup> This trend between more open societies and more authoritarian regimes is explainable on two levels: the political structure of a state and current relative power positions.

### **Government Structure Defining Values**

A nation's government structure reveals how states view the relationship of the governing to the governed. Oligarchic and authoritarian regimes, such as China and Russia, place a higher value on top-down structures and less value on democratic principles. The larger concern is control and maintaining a minimum level of satisfaction that prevents revolt. Democracies rule by the will of the citizens and find less existential risk in devolving power. The citizens of democracies also retain greater ability to remove or transfer power peacefully to a new administration when the current governing administration fails to adhere to the citizens' perceived state identity. The fall-out from crossing a society's ethical Rubicon might be harsher in an authoritarian state. The repressed society's response

---

<sup>6</sup> For commentary on current doctrine and trends see Elsa Kaina, "China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems," *Lawfare*, April 17, 2018, <https://www.lawfareblog.com/chinas-strategic-ambiguity-and-shifting-approach-lethal-autonomous-weapons-systems> and Austin Wyatt, "Charting Great Power Progress Toward a Lethal Autonomous Weapon System Demonstration Point," *Defence Studies*, 20, no. 1 (2020), 1-20, DOI: 10.1080/14702436.2019.1698956.

may lead to the overthrow of the system of government itself. Still, there is minimal increase in effects until that point, allowing authoritarian states to ignore the impacts of their ethics choices. Democratic administrations may experience increased friction with every decision out of step with even a small portion of its citizens' values.

The second factor pushing China and Russia specifically towards greater autonomy in applying AI-powered unmanned weapon systems is the relative power positions vis-a-vis the United States. While parity in nuclear weaponry can be discussed at length, for the sake of this discussion we will accept the gross assessment that China, Russia, and the United States all retain the ability to cause damage to each other to such extent to be deemed existential to all three.<sup>7</sup> In cyber there is also some level of parity, whether from constraint or capability this paper will not investigate.<sup>8</sup> In economic instruments and conventional military power, the United States retains significant advantages on its own. When supported by Western Europe, Japan, and Australia, which can all be safely assumed to support the United States in a significant global armed conflict, China or Russia find themselves at an even greater disadvantage. To a nation at such a disadvantage, the negative risks of significant nonforecasted results from using autonomous weapon systems or the execution of large preplanned campaigns are less than for the greater power for at least two reasons. First, the existential threat arrives sooner for the weaker power. Secondly, the greater power's strength, in this case, relies on allies with shared values and vision for the world order. A failure by the greater power to live up to a shared ethical standard, in this instance, the United States, may decrease the willingness of its allies to support

---

<sup>7</sup> For an entry point on nuclear deterrence today see Andrew F Krepinevich, Jr., "The Eroding Balance of Terror: The Decline of Deterrence," *Foreign Affairs* 98, Iss. 1 (January/February 2019): 62-69.

<sup>8</sup> For a starting reference see Paul J. Springer, *Encyclopedia of Cyber Warfare* (Santa Barbara, CA: ABC-CLIO, 2017). For a deeper look at China see Nigel Inkster, "Military Cyber Capabilities," *Adelphi series* 55, no. 456 (2015): 83–108, <https://doi-org.usnwc.idm.oclc.org/10.1080/19445571.2015.1181444>. On Russia see Jennifer Pomeroy, Nathan Swartz, Logan Suntzenich, and Kevin Winz, "Russia's Search for Stability: Cyber Capabilities and Military Buildup," *Current Politics and Economics of Russia, Eastern and Central Europe* 33, no. 1 (2018): 101-126, <https://login.usnwc.idm.oclc.org/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fscholarly-journals%2Frussias-search-stability-cyber-capabilities%2Fdocview%2F2189947882%2Fse-2%3Faccountid%3D322>.

the efforts. On the other hand, the allies do not expect the lesser power, Russia or China, to share the same values. The lesser power acting in such a manner does not alter the greater power's or its allies' perception of the lesser power. The risk is more significant for the side more reliant on alliances based on shared values.

The nature of the United States' power also contributes to the divide in risk level for the greater and lesser powers. While the United States is not reliant on the derived power increase from its international allies, the bump is significant. It changes the nature of conflicts from the mobilization of the entire nation into a relatively small increase in cost to the government as partner nations share the burden of engaging in armed conflict around the globe. Negative impacts of autonomous weapons systems may result in outsized costs to the ability for the United States to retain support from its allies. One neutral ship sunk by accident might be acceptable if the appropriate level of apology accompanied by a serious review of the technology is undertaken. A routine acceptance of third-party or friendly losses due to autonomous weapons systems may lead to the withdrawal of overt allied support as the vignette at the start of this paper supposes. It would probably also create significant political pressure from protestors within the United States which in the mid to long term would impact the ability for the government to prosecute the war.<sup>9</sup> In short, Clausewitz's maxim on relative efforts of positive versus negative aims continues to apply to the impacts of autonomous weapon systems where the risk of unintended effects of autonomous weapons is less for the powers seeking to disrupt the status quo than for those powers seeking to defend the current balance of power.<sup>10</sup>

---

<sup>9</sup> These value driven impacts tend to be greater at the beginning of a conflict. As military operations continue, societies appear to become less concerned with an erosion of ethical standards. This paper will delve further into this topic in subsequent sections.

<sup>10</sup> Carl von Clausewitz, *On War*, trans. and ed. Michael Howard and Peter Paret (Princeton: Princeton University Press, 1976), Book I, Chapter 1, Sections 16 & 17. Because the negative aim is the stronger side, the side with the positive aim is weaker and has less to lose in trying disruptive strategies or tactics.

Autonomous weapon systems are just one example of the high-paced nature of war. Militaries continue to use pre-planned responses that direct actions in the face of quickly evolving adversary or environmental changes. The data driving these actions may or may not include all relevant information, but the decision space at that time does not allow for a deeper analysis. For such decisions, like the use of automated weapons systems and pre-planned responses, the full range of impacts need to be assessed well before the units or systems are placed in the environment where they might be employed.

The United States and all liberal democracies must confront the inequalities in using AI in war. The trends in the application of AI, particularly with autonomous weapon systems, at first glance, gives the advantage to authoritarian regimes which place a lower value on conforming to a society's ethical identity. The decision to counter an adversary's autonomous weapon system or AI driven operations of manned assets with faster or 'better' versions of the same type of systems presents grave threats to ethical and cultural identity of a state. The manner of applying a technology, which extends to the creation of the technology, reflects the values of the society that both creates and uses it. The calculus accepted on the battlefield for the value of life, innocent bystander through the willing participant, ultimately reflects the values a society holds for its citizens. Confronting and overcoming the problem set presented by the increasing pace and lethality of warfare needs an outlook that spans the entirety of a nation's values and cuts to the tactical peculiarities of a given conflict.

### **The Ability of AI to Facilitate Strategic Genius**

Clausewitz termed the quality of genius able to discern and act upon the range of variables in war coup d'oeil.<sup>11</sup> Already rare in humans before the exponential growth in data and pace, the

---

<sup>11</sup> Clausewitz, *On War*, Book 1, Chapter 3.

requirements for coup d’oeil are elevated in modern conflicts, which can range from guerilla warfare in the gray zone to open conflict or a hybrid war that complexly weaves all aspects of state power simultaneously. Moreover, a human is not even consciously aware of what patterns she seizes on in these moments of genius in strategic and operational operations. Entire books explore the decisions Napoleon, Nimitz, Patton, Wellington, Washington, and others have made in relatively short periods of time.<sup>12</sup> All brilliant strategists at the strategic and operational level share a deep understanding of historical strategies, the operating environment of the conflict, and the likely effects of the operations on the adversary. The challenge for today’s strategist stems from the rapidity of successive and/or simultaneous operations. The time to reassess the impacts from a completed operation has and will continue to decrease towards nothing. The inertia of any operation will build rapidly and changing a strategy will require heavy costs to counter that weight. AI provides a means for today’s strategists to gain a greater appreciation for the range of likely effects on the adversary in an environment where the chance to alter operations or the strategy is fleeting.

An Ethical Assistant is more than just an aid to achieving greater situational awareness. Because an Ethical Assistant is purposefully designed to focus on the alignment of a strategy with the values of a state, an Ethical Assistant inherently protects the long-term identity and, therefore, existence of a state. Whether engaged in peer versus peer conflict or what may appear lopsided power competitions, a combination of seemingly minor decisions can place the strategic ends at risk. The pace of a conflict leans on programmed responses, regardless of whether executed by humans or machines. Employing autonomous weapons and AI to augment military forces does not remove the principle of command by negation, but there is less time to facilitate the counterorder. The vision behind the responses must

---

<sup>12</sup> A few examples include: Alessandro Barbero, *The Battle: a New History of Waterloo* (New York: Walker & Co., 2005); Craig L. Symonds, *The Battle of Midway: Battle of Midway* (New York: Oxford University Press, 2011); Dominic Lieven, *Russia Against Napoleon: the True Story of the Campaigns of War and Peace* (New York: Viking, 2010); and Edward G. Lengel, *General George Washington: a Military Life* (New York: Random House, 2005).

predict the range of operations, particularly where the probabilities of adverse outcomes overlap to the greatest extent. The strategist can highlight areas where the greatest threats spiral out, and the Commander can focus on these areas. Through the use of an Ethical Assistant in assessing the impacts of a strategy, a strategist will have the ability to see the range of effects and understand how to generate clearer and stronger guidance for mission command, bound and support subordinate Commanders appropriately, identify those areas which present the most risk to achieving the strategic end state, and protect the ethical framework of the state. An Ethical Assistant will be required to winnow through the endless possibilities in the higher paced and still high stakes environment of the future.

### **What is an Ethical Assistant?**

It is necessary to describe an Ethical Assistant and what it is not before delving into a more robust discussion of the requirement for an Ethical Assistant in strategic decisions today and in the future. An Ethical Assistant can be thought of as a counselor solely focused on aligning the proposed course of action, the desired end state, and, most importantly, the ethics. The range of possible adversary responses weighted towards the expected responses, and then the same again for both allies and neutral parties, inform the Ethical Assistant's evaluations. The assistant confronts the complexities of Clausewitz's triangles and Sun Tzu's environmental calculation.<sup>13</sup> From these inputs, the Ethical Assistant provides a range of effects and implications and not a single future. It is not a crystal ball

---

<sup>13</sup> Clausewitz and Sun Tzu are often held up as a western vs eastern approach to the theory of war. Clausewitz's trinity consisted of passion, reason, and chance which interact with the fundamental elements of the people, government, and military. Sun Tzu focused more on the effect of the environment and understanding context to shape the war. To go beyond this gross simplification, see Clausewitz's *On War* and Sun Tzu's *The Art of War*.

foretelling the future but an incredibly complex analytical tool that offers probable and possible outcomes and highlights the inconsistencies which might arise with one's own ethical identity.

The most vital input to the Ethical Assistant after creation is the ethical values of the decision-makers.<sup>14</sup> Defining the ethical structure of her own society sounds like it may be the easiest aspect of the required inputs, but forces her to honestly take a critical look at what she values and why. Another decision-maker from her society may define the fundamental values slightly differently, leading to a different assessment. Boundaries must be established, including an honest assessment of what measures and impacts she deems acceptable to achieve her objectives and end state. These can range from acceptable collateral damage during strikes to possible economic impacts or famine in third-party nations. Critically assessing the ethical framework will not guarantee that a strategy never results in an undesirable or unethical outcome, but demonstrates where the potential to stray from her values is the greatest.

Within the context of a range of possible outcomes, the strategist can identify successful endgames or exit plans which further the political end state and defend the ethical identity of her nation. As an example, the strategist might assess that the United States' identity and influence arise from its position as an anti-imperial and democratic force defending individuals' rights to form nation-states that govern with the consent of the people. Her assessment places a priority to guard against objectives that overwhelming trend towards requiring the United States into playing an imperialist role. Another strategist may determine that the United States identity arises from its role as the champion of individual opportunity within a free market. His calculations may allow for more ambitious objectives or even more limited objectives. The Ethical Assistant uses the range of tools available through AI to

---

<sup>14</sup> Creating an AI is inherently value laden. Distinct and critical ethical decision points exist during an AI's design and creation AND the use of the AI after creation. While the design must anticipate the ethical application of the AI, it does not remove the burden on the individual or group employing or partnering with the AI.

expose probable paths where contradictions exist between the ways, means, or end-state and nation's values. As such, a strategist can identify other, more limited objectives that do not risk the nation's ethical identity. The strategist can then use the range of predicted outcomes to find the culminating point of military operations based on an ethical perspective instead of just political or military lenses.

The role of the Ethical Assistant as an advisor or counselor and not the decision-maker is paramount. As the sciences and technologies empower AI, an Ethical Assistant may acquire the necessary complexity to recommend alternative courses of action or adjusted objectives that align with the desired end state and described ethical framework. However, humans must make the choices. There may be an upper limit to the application of AI recommended options. An AI-powered Ethical Assistant inherently creates non-human solutions for interhuman relationships, some of which may not be achievable due to the constraints of human emotions. The Ethical Assistant is entirely rational within the confines of its design. The Ethical Assistant consistently weighs values of outcomes and choices against the ethical values at the start of the strategy, even while considering the implications of human emotions in strategic contexts. While individual humans may strive to be completely rational when weighing strategic decisions, humans exist in social and emotional contexts. They cannot be relied upon to make purely rational decisions. The effects of the military operations will alter the hierarchy or even the nature of their own values. As a side gains greater advantage, they may see a path for more significant gains than initially chosen. In another instance, suffering higher losses than anticipated may raise the demands for revenge or willingness to use means deemed unethical before the losses. The Ethical Assistant may struggle to predict such choices. The decisions inherent to war, weighing the value of life, death, suffering, and happiness, affect humans, and humans must be the actors. An Ethical Assistant will aid the strategist in sifting through an ever-increasing amount of data and possible outcomes with an inherent ethical approach. Ultimately, the strategist must practice ethics and not abdicate the choices to a program, no matter how beautifully and ethically the program is created.

## **Military Operations with an Ethical Assistant**

Evaluating courses of action is the primary means through which staffs present options to the Commanders to choose. All too easily, these courses of action can become detached from the broader end state and narrowly focus on the short-term objectives. In an ideal strategy, every objective at the tactical, operational, and strategic level is nested inside one another to minimize any wasted effort. The Ethical Assistant allows a Commander and staff to identify the strategy that most efficiently meets the Commander's vision and the grand strategy's end state. It cannot replace the Commander or any other decision maker. The practice of ethics is a human realm and humans are choosing how to act. The Ethical Assistant can illuminate the points of increased risk, allowing the Commander to focus her attention on the most impactful decision points.

The Ethical Assistant's analysis enables the Commander and staff to craft more relevant standing orders and rules of engagement for routine operations outside declared campaigns or operations. For example, through partnering with an Ethical Assistant, it might be possible to shape the nature of the interactions with the PLAN in the South China Sea to more efficiently facilitate the US's vision. The design of the Ethical Assistant for this purpose requires an in-depth understanding of China's viewpoint and surrounding nations and the US's overall vision for the region. When constructing orders for campaigns or operations, the standing orders and rules of engagement can be more narrowly focused on specific responses that impact the objectives but without risking any higher strategic objectives. An Ethical Assistant could identify an opportunity to increase pressure on the CCP if aggressive PLAN maneuvering was presented in a historical context or against international norms. Documenting all interactions with the PLAN is a good start, as the US Navy currently does. These interactions, though, need to be routinely disseminated, not just to counter CCP narrative retroactively.

Towards the end of effectively making ethical decisions during operations, an Ethical Assistant enables operational Commanders to train all subordinate Commanders and personnel generally in the potential decisions that are likely to occur during specific missions through campaign levels.<sup>15</sup> The National Security Council on AI is already looking to change how military personnel are trained on ethics and AI, and here the Ethical Assistant can be a part of the solution.<sup>16</sup> By focusing the Ethical Assistant on a specific level of the conflict or role of the unit or individual, individuals will gain experience in confronting realistic ethical dilemmas. Whether in the present-day gray zone state with China or Russia or preparing for an upcoming mission against the Islamic State, Commanders have to rely on the capability of subordinate personnel to make near-instantaneous decisions that have strategic impacts. Ethical Assistants allow Commanders to train their personnel, test their plans, and craft more effective orders to retain and increase critical advantages in the decision-making cycle while defending and promoting national values.

### **Artificial Intelligence's Relationship with Ethical Assistants**

Though reliant on AI advances to develop an Ethical Assistant, the field of AI itself presents a great risk to the "soul" of the United States, but also great potential. The applications of an Ethical Assistant presuppose limits and bounds on the use of AI within an ethical framework. Still, other AI applications cannot be assumed to have the same. Ethics cannot be an area where human actors cede

---

<sup>15</sup> For continual training to be effective, though, personnel require a strong foundation in ethics education to build upon through their careers. The skill necessary to analyze and interpret data through an ethical lens cannot be assumed to be gained during on-the-job training. In addition to foundational education, and then continued training with the Ethical Assistant, personnel need continuing ethics education.

<sup>16</sup> See *Final Report*, National Security Commission on Artificial Intelligence, March 2021, <https://reports.nscai.gov/final-report/table-of-contents/>. *Appendix C: Key Considerations for the Responsible Development and Fielding of Artificial Intelligence* of the *Final Report*, as well as *First and Second Quarter Recommendations Memos* provide examples of NSCAI's vision in practice.

control, no matter how fast the processing power. Still, the human mind requires assistance to sift through the data and evaluate the impacts of choices. AI will help fill that role.

Towards that end, the federal government must establish itself as a key stakeholder in developing AI to shepherd the advances in line with the nation's values. The United States is establishing itself as a leading developer of AI applications. Corporations that develop AI technologies and applications are strategic players, regardless of how they view their roles. AI applications' ethical structures are a strategic means in great power competition. The government should aggressively lobby and engage with AI developers to counter any potential inadvertent risks to US values and then embed US values into developing the technology and applications. Currently, the US government almost entirely depends on the private sector to develop AI that reflects US values for military operations. Europe's methodology in developing technologies within the context of its values stands as an alternative to the US's current hands-off approach to technological advances' ethical and strategic implications. The United States needs to embrace the strategic importance of AI development's risk and rewards.<sup>17</sup>

### **How Different Aspects of AI Design Methodology Play a Role in Developing an Ethical Assistant**

Many other authors have written extensively on the development of AI and the creation of artificial moral agents and other constructs focusing on applying ethics specifically.<sup>18</sup> I rely upon the

---

<sup>17</sup> While the NSCAI Final Report addresses how the US government should embrace AI, chapter three provides the current vision on AI incorporation in the military. The Blueprints for Action identify forums to strategically push the US vision for AI. See *Final Report*, National Security Commission on Artificial Intelligence, March 2021, <https://reports.nscai.gov/final-report/table-of-contents/>.

<sup>18</sup> Examples include: Wendell Wallach and Colin Allen, *Moral Machines, Teaching Robots Right from Wrong* (Oxford University Press, 2010); Erik Brynjolfsson and Andrew McAfee *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* (New York: W.W. Norton Company, 2016); and *Machine, Platform, Crowd: Harnessing our Digital Future* (New York: W.W. Norton Company, 2018); and David E. Nye, *Technology Matters: Questions to Live With*, (Cambridge, MA: MIT Press, 2007); J. Danaher, "Toward an Ethics of AI Assistants:

works of these experts and the accompanying research to discuss the impact of combining current AI applications and design methodology to create an Ethical Assistant as envisioned in this paper. AI applications today to assist in ethical training, real-time decision-making, or other fields provide important examples and lessons for creating Ethical Assistants.

The Ethical Assistant will evaluate the ethical impacts by applying a bottom-up analysis of cause and effects to the strategist's top-down designed ethical structure. Machine learning algorithms with attendant Testing and Evaluation (T & E) oversight will function alongside traditional programming methods to analyze ethical frameworks and action/counteraction cycles. Integrating selective crowd-sourcing can close the gaps between the aspirational ethical norms of society and the evidenced ethical norms revealed through machine learning. The addition of the crowd-sourced data will further enhance the effectiveness of the Ethical Assistant while also supporting explainability, equitability, and reliability. Combining a variety of approaches to analyze the effects and the alignment of the impacts with the selected ethical framework will also lead to a more holistic analysis of strategies and operations.

Machine learning is fallible to the extent of its initial programming, but its emotionless calculations produce a neutral response to its programmed questions and data set. T & E provides critical checks to ensure alignment with the designer's intent. T & E methods begin at the start of the design process to eliminate bias. Selected experts in their respective fields of military operations, ecological and environmental sciences, political and social sciences, and ethics will shape the nature of

---

an Initial Framework," *Philosophy of Technology* vol 31 (2018); Thomas Arnold and Matthias Scheutz, "The "Big Red Button" is too Late: An Alternative Model for the Ethical Evaluation of AI Systems," *Ethics and Information Technology* 20, (2018): 59–69, <https://doi.org/10.1007/s10676-018-9447-7>; Daniel Kasenberg, "Quasi-Dilemmas for Artificial Moral Agents," *ICRES 2018: International Conference on Robot Ethics and Standards*, Troy, NY, August 20-21, 2018, <https://doi.org/10.13180/icres.2018.20-21.08.012>; and Nick Bostrom and Eliezer Yudkowsky, "The Ethics of Artificial Intelligence," *Machine Intelligence Research Institute*, accessed July 29, 2019, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.477.9269&rep=rep1&type=pdf>. Additional works can be found in the bibliography. Of particular note, Tuft's Human Robot Interaction Lab webpage provides links for further exploration on the topic: <https://hrilab.tufts.edu/publications/>.

the relationships within the data the Ethical Assistant.<sup>19</sup> As the Ethical Assistant progresses through machine learning, additional raw data sets are provided which will truly test the original theories and begin to unlock the true potential of independent machine learning. The Ethical Assistant, designers, and strategist will uncover unknown biases and overlooked data points as incongruent results between curated and raw data sets highlight critical effects and relationships not currently captured in ethical reviews or during pre- and post-mission analyses. Rather than negatively judging the specific machine learning application, as has been the norm for recent high-profile cases of previously unrecognized bias with facial recognition or chat bots, these discoveries should be embraced as a step forward in eliminating ignorance and inequality.<sup>20</sup> The experts, strategists, and Ethical Assistant will be wiser due to the T & E continuously correcting these biases.

Machine learning can improve collateral damage estimates, already heavily reliant on complicated algorithms. Machine learning's ability to identify trends and relationships within immense amounts of data unlocks the potential to grasp a more nuanced understanding of how military force ripples across society and the natural world. By expanding the breadth and time horizon of the damage

---

<sup>19</sup> This list is not intended to be all encompassing.

<sup>20</sup> Tom Simonite, "The Best Algorithms Struggle to Recognize Black Faces Equally," *Wired*, July 22, 2019, <https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/>; Clare Garvie and Jonathan Frankle, "Facial-Recognition Software Might Have a Racial Bias Problem," *Atlantic*, April 7, 2016, <https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/>; Drew Harwell, "Federal Study Confirms Racial Bias of Many Facial-Recognition Systems, Cast Doubt on Their Expanding Use," *Washington Post*, Dec 19, 2019, <https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/>; Kade Crockford, "How is Face Recognition Surveillance Technology Racist?," ACLU, June 16, 2020, <https://www.aclu.org/news/privacy-technology/how-is-face-recognition-surveillance-technology-racist/>; Alex Najibi, "Racial Discrimination in Face Recognition Technology," *Science in the News*, October 24, 2020, <https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/>; Amy Kraft, "Microsoft Shuts Down AI Chatbot After it Turned into a Nazi," CBS News, updated March 25, 2016, <https://www.cbsnews.com/news/microsoft-shuts-down-ai-chatbot-after-it-turned-into-racist-nazi/>; Sarah Perez, "Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism," Tech Crunch, March 24, 2016, <https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/>; and Alyssa Newcomb, "Microsoft's Chat Bot 'Tay' Gets a Time-Out After Rude Comments," abc News, March 25, 2016, <https://abcnews.go.com/Technology/microsofts-chat-bot-tay-time-rude-comments/story?id=37926075>.

assessments, the strategist can begin to comprehend connections that are nearly invisible today. On a longer time frame, operations that appear successful in the short term might actually be counter-productive.

Considering a potential flare up with Iran, the Ethical Assistant will need to analyze the wider impacts on the Iranian populace, US standing in the region with other nations, and the tactical likelihood of losses when determining the way ahead. Allowing food, fuel, or other sundries to continue to move may win over the populace in a longer-term vision while cutting off all trade might secure a faster military solution. The strategist must ask the right questions in the right manner. Still, machine learning provides the cleanest opportunity for individuals to minimize emotional bias and allow humans to choose between competing values.

Machine learning will also contribute to the crafting of the ethics-focused algorithms at the Ethical Assistant's core. It may help to view the Ethical Assistant's programming as two distinct parallel components. One component focuses on maximizing all AI and machine learning potential to provide a reliable range of outcomes from a set strategy and operations. Simultaneously, the other half of the Ethical Assistant is constantly testing and evaluating those outcomes for 1) achieving the stated objectives and 2) for alignment with the ethical values of the state.

In an environment where every decision and action impact all subsequent actions and decisions, one of the Ethical Assistant's strengths is identifying the decision points with the widest impact. The discoveries from machine learning on the cyclical effects of actions and the subsequent counteractions when using military force will improve those same algorithms and provide a better depiction of the ethical decision points in the same cycles. Ethicists can focus their attention on these nodes and drill down into the nature of how to analyze the competing values. The practitioners of the different ethical schools will each possess their own unique methodology, and the resulting recommendations and

mitigation methods will be equally unique. Machine learning will speed the process of developing more accurate algorithms and models to assess the alignment of military operations with a nation's desired ethical values.

### **Other Actors Must be Accounted For**

To accurately present the critical ethical decision points within a range of likely outcomes during a conflict, the Ethical Assistant will need to possess the ability to predict the range of likely adversary responses. The designers need to avoid searching for the perfect simulation of reality instead of guiding inherently imperfect human actions. The same combination of methodologies used to describe the impact of actions and decisions on the environment and populations will need to be employed to describe the motivators for an adversary's actions and the impact of those actions. The complexity grows as the Ethical Assistant includes neutral and allied parties. The greater the ability to parse out separate groups with distinctive cultural, ethical, and national identities, the even more complex the calculations and successive iterations become.

The possible combinations are infinite, and for this reason machine learning is the only viable option available currently that can attempt to explore these interactions. Monte Carlo Tree Search algorithms represent one way in which the more likely branches of actions can be described and analyzed without mapping the entirety of the infinite scope of possibilities. With the survival of nations and people's lives at stake, blind trust in mapping winning strategies or rational decisions cannot suffice. The most likely courses of action will need to be checked against the possible courses of action. The possible actions that cross a certain risk threshold will receive more extensive analysis. The Ethical Assistant will require multiple methodologies for establishing cause-and-effect relationships to avoid overlooking critical connections and preserving the uniqueness of individual humans.

Imagining an Ethical Assistant designed to advise a strategist in mainland southeast Asia, separate algorithms can be combined which acknowledge the unique cultures and values of Vietnam, Laos, Cambodia, Thailand, and Myanmar. The effects on the people versus the governments are distinct in each state. The values of the governments and their perspective view of their legitimacy and their vulnerabilities differ as well. These differences will create unique responses from each state to any US actions. As improvements in computational speeds and design methods continue, greater fidelity in the range of likely effects will be possible to account for subsets or factions in the states or across states. While designers and strategists can leverage AI advances, they must vigilantly avoid projecting US values and decision-making onto other actors.

### **Creating an Ethical Assistant Requires a Team of Experts**

Creating an Ethical Assistant requires a team of experts in various fields and how to relate their fields to other fields. An Ethical Assistant can be designed in numerous ways, in the same vein that nuclear submarines or spacecraft can be designed differently. While exceedingly complicated holistically, each part within the larger vision is an engineering marvel on its own. The growing awareness of humanity's interdependence with the natural world will increase the factors for considerations. The already daunting enterprise to describe collateral damage assessments alongside the interpretation of fleet positioning and troop movements while judging the impacts of an information operations campaign online will only grow. Experts on each of these operations' intent and possible effects will need to work with experts in other areas. This dream team of intellectuals and practitioners will also have to mesh with an ethicist to describe these actions in a manner that the Ethical Assistant's ethical engine can evaluate to determine when to highlight decision points and flag actions as incongruent with the larger objectives. The larger the team involved in crafting, testing, and challenging

the Ethical Assistant, the higher the likelihood that the values of the state will be preserved, even knowing the Ethical Assistant will be imperfect.

The selection of the design team members provides an additional opportunity to reinforce the values the Ethical Assistant is protecting. Transparency is a critical aspect of team construction that amplifies or detracts from the target values of the Ethical Assistant. The team members share their biases and values in their advice and assessments. An Ethical Assistant designed to emphasize equality and democratic principles but designed by a team that lacks diversity or operates in a strict hierarchical structure would create skepticism of its ability to support the stated values. The representation on the team is an explicit data point to judge the commitment to the Ethical Assistant's values and a vital first step in demonstrating the trustworthiness of the Ethical Assistant.

The role of AI, and technology in general, is immense in creating Ethical Assistants. Still, the power and efficacy ultimately reside in the self-awareness of human actors of their ethical systems and values, the ethical system and values of potential adversaries and allies, and the myriad biases that plague each side.

### **Gaining Trust and Validating Results**

An Ethical Assistant will need to gain the trust of an ever-increasing pool of individuals. First, each subcomponent's designers will need to have faith that their part of the Ethical Assistant's programming behaves according to their intent and design. Next, each component will need to interact appropriately with the other components. Suppose the Ethical Assistant is assessing removing a power plant from an adversary's grid. It must 1) correlate the various means of removing the power plant from the grid to the cost in direct lives and collateral damage from the strike method, 2) short through long term impacts of the loss of power capacity, 3) foresee what options the local population will take to

mitigate the impacts, 4) and so on. The Ethical Assistant must provide a full reckoning of the impact of removing the power plant before it is capable of checking the nesting of the action within the larger context of the strategy for alignment with the desired ethical values. Finally, the strategist who commissioned the Ethical Assistant must trust that it captures her specifications fully. She must have confidence that she can alter the Ethical Assistant to align it better when improved data is found and refine the scope of the flagged ethical decision points. Gaining the trust of the strategist is only the first great test in the creation of an Ethical Assistant. The strategist must gain the confidence of the Combatant or Operational Commanders and convince them of the appropriateness of the strategist's course of action, which relies on trust in the Ethical Assistant. Similarly, the strategist and Ethical Assistant of the chosen plan may need to convince other strategists and decision-makers at a higher level.

The use of historical examples to train an Ethical Assistant also provides the means to gain the trust of the strategist and others. As part of developing and designing the various aspects of the Ethical Assistant, developers can hold competitions focusing on explaining certain decisions during previous wars or entire conflict periods. Opening these competitions up to any entries will create opportunities to leverage crowdsourcing to find novel approaches that dedicated development labs may have overlooked but can leverage after discovery. Inviting scrutiny of ethical judgments of historical events also expands the lens on how to evaluate decisions and the following effects. As the design and execution of successive competitions improve with the results of the previous events, the insights and knowledge gained will increase. Whether from new techniques identified by the entries or the design team itself addressing weaknesses identified by the entries, the designers can improve the Ethical Assistant to account for different cultures, technology levels, and historical vantage points. The first iterations of the Ethical Assistant's programming may be too tailored to explicitly account for a set conflict or period. Still, they will rapidly improve and expand with testing of more extended periods. If the

Ethical Assistant can reliably direct the strategist to the critical decision points during similar conflicts, the strategist and the senior Commanders can have confidence in the Ethical Assistant. The wider the range of conflicts that an Ethical Assistant can explain, the greater the strength of its ethical engine in drawing decision-makers to the critical ethical junctures.

Underpinning the trust in the quality of the Ethical Assistant's analyses is the Ethical Assistant's adherence to the Department of Defense's Ethics Principles and the department's Responsible AI tenets.<sup>21</sup> The five AI Ethics Principles are built upon the Department of Defense's strong ethical framework.<sup>22</sup> I will highlight a few specific properties of the Ethical Assistant related to the principles.

**Responsible.** Ethical Assistants should not have the ability to direct actions.

The first principle establishes the importance of human leadership in directing all systems, including Ethical Assistants and other AI, down to simple tools. It holds humans accountable for the actions of or stemming from AI applications. Further, an Ethical Assistant can highlight the decision areas to determine if actions are responsible. By training or focusing decision space on the actions which may lead to unethical or competing ethical values, the Ethical Assistant allows the strategist to evaluate if an application of force is appropriate. With the growing applications of AI and the fielding of autonomous weapons systems by other actors, Commanders in the field or at sea will not have the time to deliberatively respond if they have not already considered the ramifications of their defensive or offensive postures. The human must choose to proceed with a plan knowing the risk that an undesirable effect could still arise.

---

<sup>21</sup> *Implementing Responsible Artificial Intelligence in the Department of Defense*, Deputy Secretary of Defense, 26 May 2021.

<sup>22</sup> *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense, Primary Document*, Defense Innovation Board, October 2019. <https://innovation.defense.gov/ai>.

**Equitable.** Ethical Assistants need to be transparent on the relative weight given to one group of human lives compared to any other group of humans.

Unrecognized biases have profound impacts on the social and economic realities of life today. In the application of military force, unrecognized bias could lead to greater death or suffering of overlooked subsets of populations. By pairing with an Ethical Assistant, strategists are better prepared to find the biases and confront the underlying reason for the biases.

**Traceable.** AI experts must be capable of auditing the methodologies, data sources, and design procedures to include T & E.

To establish trust in these analyses, all aspects of the Ethical Assistant must be auditable. T & E programs need to be transparent in how they modify the algorithms. How the Ethical Assistant came to a decision should be understandable to a technical professional and generally explainable to a layperson.

**Reliable.** Ethical Assistants need to deliver quality analysis.

In addition to consistently turning out quality results, Ethical Assistants will undergo testing and assurance to guarantee the program's safety, security, and robustness. The program must also be capable of securing sensitive or classified data.

**Governable.** Ethical Assistants need to be designed and engineered to support human actors.

At the heart of every ethical debate sits a choice facing an individual human. To meet this aim, the strategist must always possess the ability to control and adapt the Ethical Assistant. Ethical Assistants should also include an ability to detect and highlight concerns on misalignments of values, learned or directly input, or a shift outside programable acceptable deviations. The burden falls on a human to interpret the information provided and determine the nature of the corruption or

misalignment. Perhaps the Ethical Assistant is only exposing fallacies within the strategist's priorities or bias in data sets. Still, it could also highlight intentional efforts by third parties to hack and exploit the AI and its American Values.

Ultimately, Ethical Assistants must reflect the values of American society.

### **Programs Assisting Ethical Decision-Making Ability Today**

Ethical decision-making tools are already employed by organizations today. Some of these programs focus on educating and training individuals to make value-based decisions with no right or wrong answer supported by statistical analysis, including in the medical community. The medical community uses AI-based programs to present health care providers challenging ethical decision points to foster thoughtful reflection on how the individual would handle similar situations. From Carnegie Mellon University, Bruce McClaren designed SIROCCO to bridge the gap with a casuistry approach to ethics that prepared medical professionals to make timely ethical decisions. He contends that this training requirement comes from a gap between the abstract, open-textured nature of ethical principles and concrete facts of a given situation.<sup>23</sup>

This medical training application is distinct from an AI analysis of best medical practices regarding prescribing drugs or surgery decisions. The Defense Innovation Board's pursuit of sharing its substantial pathology resources to advance AI research into the accurate identification of medical conditions is an analytical use.<sup>24</sup> IBM's development of Watson to identify and recommend cancer

---

<sup>23</sup> Wendell Wallach, *Moral Machines: Teaching Robots from Right and Wrong* (Oxford: Oxford University Press, 2011), 131-91; Bruce McLaren, "Computational Models of Ethical Reasoning: Challenges, Initial Steps, and Future Directions," *IEEE Intelligent Systems* 21, no. 4 (July-August 2006): 29-37.

<sup>24</sup> Lander, Eric. *Defense Innovation Board Joint Pathology Center (JPC) Repository Enhancement*. March 2020. [https://innovation.defense.gov/Portals/63/documents/Meeting%20Documents/March%202020/DIB\\_JPC\\_Recommendations\\_Final.pdf](https://innovation.defense.gov/Portals/63/documents/Meeting%20Documents/March%202020/DIB_JPC_Recommendations_Final.pdf)

treatment is another example.<sup>25</sup> How a doctor should apply the findings to a specific patient whose ethnicity, age, or other condition is not represented in the sample population is an ethical dilemma that arises from the manner of the analysis. The decision on how to proceed in this situation has long term effects on all the following decision points regarding surgery, medication, and long-term care physically, mentally, and emotionally. Training programs for ethical decisions in the medical field relied on previous case studies with in-depth information.

These two AI applications, training individuals for decision making and analyzing data, will often go hand in hand, as in my vision for an Ethical Assistant. By focusing an individual's training on those decisions, when a similar moment arises, the provider will recognize the importance of the decisions and be better prepared to address them. Similarly, a strategist can use an Ethical Assistant to highlight to combatant and operational Commanders and other senior decision-makers the critical ethical decision points on which the strategy's success relies.

AI is used to direct humanitarian assistance aid decisions through near and real-time data compiled from publicly available social media, economic indicators, and satellite imagery, combined with dedicated ISR monitoring and other specific assets. Humanitarian crisis responders do not blindly follow AI advice but incorporate the modeling into the broader decision-making process. There may be points where the AI advice can be allowed to execute autonomously, such as how much water is needed at specific sites or recognizing when local expenditure accounts need funding to complete the human selected options. In general, however, the selection of the course of action requires human input. The consequences affect the ability of some humans to survive at the risk of a decreased likelihood of others. These life and death decisions must be weighed and accepted by other humans. If all decisions are turned over to a programmed response, even an intelligently designed AI, the actors have

---

<sup>25</sup> IBM's Watson did not succeed as hoped, but demonstrates an important step in pursuing AI assistance in the medical field.

cheapened the value of humanity. They have washed their hands of any empathy for the suffering of the individuals. AI can assist individuals in preparing to make ethical decisions and weigh options during the decision-making. Ethics, though, is a human endeavor concerned with turning the values of humans into action.

### Lessons from Other AI decision Aids and Programs

The success of AI projects in numerous fields, from gaming to design, demonstrates the realistic potential for building an Ethical Assistant. Deep Blue's success against Garry Kasparov came after years of deliberate research and tailoring to beat Kasparov's style. AlphaGo's success represented a leap forward in machine learning, both from the possible combinations of moves and universal methodology against all Go players. AI gaming has moved from the perfectly defined environments of Chess and Go to partial and non-deterministic contests.<sup>26</sup> The move demonstrates the ability for non-deterministic value-laden judgments on the part of AI programs. Outside of gaming, AI is also successfully creating original works in less-defined structured domains of poetry, literature, and many news service blurbs.<sup>27</sup> AI has also successfully written musical compositions indistinguishable from human-only composed

---

<sup>26</sup> Sebastian Risi and Mike Preuss, "From Chess and Atari to StarCraft and Beyond: How Game AI is Driving the World of AI," *Künstl Intell* vol 34 (2020): 7–17, <https://doi.org/10.1007/s13218-020-00647-w>; Marc Ouellette and Steven Conway, "A Feel for the Game: AI, Computer Games and Perceiving Perception," *Eludamos. Journal for Computer Game Culture* 10, no. 1 (2019):9-25; Reid McIlroy-Young, Siddhartha Sen, Jon Kleinberg, and Ashton Anderson. "Aligning Superhuman AI with Human Behavior: Chess as a Model System." *KDD '20: Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, August 2020: 1677-1687. <https://doi.org/10.1145/3394486.3403219>.

<sup>27</sup> A selected sampling: GPT-3, "A robot wrote this entire article. Are you scared yet, human?," *the Guardian*, September 8, 2020, <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>; Stephen Marche, "The Computers Are Getting Better At Writing," *The New Yorker*, April 30, 2021, <https://www.newyorker.com/culture/cultural-comment/the-computers-are-getting-better-at-writing>; and Jey Han Lay, Trevor Cohn, Timothy Baldwin and Adam Hammond, "This AI Poet Mastered Rhythm, Rhyme, and Natural Language to Write Like Shakespeare," *IEEE Spectrum*, April 30, 2020, <https://spectrum.ieee.org/artificial-intelligence/machine-learning/this-ai-poet-mastered-rhythm-rhyme-and-natural-language-to-write-like-shakespeare>;

pieces.<sup>28</sup> AI is also presenting potential changes in the field of Architecture. The filling in and assistance to a human-crafted design is an evolutionary area that could potentially disrupt architectural firms' current paradigm, either by forcing further consolidation or freeing up accessibility to architectural services to a wider client base.<sup>29</sup> The revolutionary disruption of fully AI created visions for designs presents particular unique challenges to the architectural field. These applications and their successes provide great promise for the wider application of AI to new areas and the potential to create new research and learning methodologies.

The successes also provide a cautionary reminder to the limits of reliance on AI. The further humans step back from the design process, the less 'human' the resultant becomes. The methodology used by AI is always non-human, and it's because of this nature, human-machine pairing and machine learning provide insights that humans cannot grasp on their own. The non-human nature also drives the imperative of maintaining human oversight at critical points.<sup>30</sup> AI-designed architecture is otherworldly at times because the programming does not grasp all the contextual relationships that humans inherently know but fail to account for in describing the limits.<sup>31</sup> AI is not a superior human design but a distinct

---

<sup>28</sup> See Bartu Kaleagasi, "A New AI Can Write Music as Well as a Human Composer," *Futurism*, March 9, 2017, <https://futurism.com/a-new-ai-can-write-music-as-well-as-a-human-composer>; and also Omar Lopez-Rincon, Oleg Starostenko, and Gerardo Ayala-San Martín, "Algorithmic music composition based on artificial intelligence: A survey." *2018 International Conference on Electronics, Communications and Computers (CONIELECOMP)* (2018): 187-193. DOI: 10.1109/CONIELECOMP.2018.8327197; and Artemis Moroni, Jonata Manzolli, Fernando Von Zuben and Ricardo Gudwin, "Vox Populi: An Interactive Evolutionary System for Algorithmic Music Composition," *Leonardo Music Journal*, Vol. 10 (2000): 49–54, accessed On 03 May 2021, <http://direct.mit.edu/lmj/article-pdf/doi/10.1162/096112100570602/1673952/096112100570602.pdf>. Note that all three identify a vision for AI to work with humans in composing music, even as AI is capable of composing music on its own.

<sup>29</sup> Kathleen O'Donnell, "Embracing artificial intelligence in architecture," *The American Institute of Architects*, Accessed May 2, 2021, <https://www.aia.org/articles/178511-embracing-artificial-intelligence-in-archit:46>; Stanislas Chaillou, "AI & Architecture," *Medium*, February 26, 2019, <https://medium.com/built-horizons/ai-architecture-4c1ec34a42b8>; and Dr. Kai Oberste-Ufer, "7 Ways Artificial Intelligence is Revolutionizing Architecture," *Dormakaba*, February 9, 2019, <https://blog.dormakaba.com/7-ways-artificial-intelligence-is-revolutionizing-architecture/>.

<sup>30</sup> In the previous articles on AI and architecture, note the desire for human partnering to temper the effects of AI design to meld with humans.

<sup>31</sup> See Chaillou's closing comments: Reiner-roth, Shane, "Would artificial intelligence like to say something about architecture?," *Archinect News*, April 26, 2019, <https://archinect.com/news/article/150133371/would-artificial-intelligence-like-to-say-something-about-architecture>.

approach to human concerns from a non-human constrained methodology. AI's decisions on achieving an outcome conform to deliberately placed bounds in the initial programming and reinforced through its learning sets and T & E. Humans must oversee the critical decision points to maintain our humanity and our values in the output.

### **Requirement for Multiple Design Methodologies**

The array of Ethical Assistants will require multiple design methodologies reflective of the world's plurality of ethical approaches and values. These distinct approaches must still share some common traits to ensure the maximum benefit from focused analysis on values and ethical lenses in weighing military actions. Sun Tzu and Clausewitz each identify the imperative to know the values and ideological narratives at the center of any conflict. Clausewitz's assertion that war is politics by other means implies Sun Tzu's admonition to know yourself and your enemy. The realm of politics is the competition for shared values and narratives. You cannot win over supporters if you do not understand the product you are selling. In addition to ethical lenses, biases and other disciplinary perspectives shape how a group acts ethically. The three main International Relations theories are commonly employed lenses to view actions on the world stage, but are distinct from ethical frameworks.<sup>32</sup> The designing of and partnering with an Ethical Assistant requires understanding the interplay between ethics and other disciplinary approaches to understanding interactions, such as International Relations. Ethical Assistants be uniquely designed for different parties, taking into account the perspectives of the strategist and her state and their interpretation of other actors' perspectives.

---

<sup>32</sup> I am referring to Realism, Liberalism, and Constructivism. For an overview of these and other theories of IR see Robert Jackson and Georg Sorensen, *Introduction to International Relations: Theories and Approaches, Fifth Edition* (Oxford: Oxford University Press, 2013).

Strategy does not simply originate fully formed at the highest levels of government and flow down through the organization, but will instead compete within organizations, rising and flowing down in a constant revision which refines and ideally improves the strategy. Strategy is not static but constantly adapts to new facts that can constrain or expand the aims and means. Factions (Departments, Services, etc.) are incentivized to ensure strategies incorporate their worldviews. Acknowledging the competitive nature, it seems inevitable that multiple Ethical Assistant models could compete to generate the 'correct' assessment of a strategy within a larger organization. Too many Ethical Assistants, like too many independent strategies, would bog down the refinement and improvement of the idea.

Before expanding the resources to create a new separate and distinct Ethical Assistant, the designers should examine the differences in lenses and ethical framework from currently fielded and trusted Ethical Assistants. Suppose the ethical differences between two parties comes down to a difference within the hierarchy of values or stated objectives, but the ethical structure and perception of other actors are shared. In that case, it may not make sense to design an entirely new Ethical Assistant. The same Ethical Assistant design can be used for each, with slight tweaks accounting for different weights in values. On the other hand, if one organization priorities the ethics of care and its counterpart applies a utilitarian approach, even if aligned on objectives and perceptions of all other parties, the Ethical Assistants would most likely require fundamental design differences necessitating two distinct Ethical Assistants.

There is a separate vital requirement for competing Ethical Assistant architectures to ensure commonly shared biases do not escape unnoticed within an Ethical Assistant. Even with a robust T & E program, errors can occur and be missed due to how data is presented to the Ethical Assistant. The variations in results between two Ethical Assistants could represent many distinct scenarios. In one outcome, different fundamental programming on what is critical concerning the strategy or the values

of the state can lead to substantive debate on the aim of the strategy or operation. Another analysis of the comparative results could highlight unrecognized bias programmed into a system. Two or three distinctly designed Ethical Assistants within a larger organization such as the Department of Defense, should provide a sanity check for the range of potential effects of proposed strategies and create enough unique insights to foster more innovative discussions over the priorities in a strategy.

The differences between the Ethical Assistant's analyses will increase the awareness of the strengths and weaknesses of the plan and the knowledge of her ethical values. The enhanced understanding provides the strategist and the organization the ability to focus resources to mitigate the weaknesses, leverage strengths, and shape training and development. The increased awareness of how the underlying values drive the strategy increases the decision space for the national-level decision-makers. They and the nation are less likely to be caught off guard or deceiving themselves on the values behind their actions.

### **Different Lenses and Structures**

Today some strategists, civilian and military, substitute international relations viewpoints for ethical structures and values. Realism, Liberalism, and Constructivism are ways to view the interactions between sovereign governments, but these are not the same as ethical frameworks. Some overlap with certain ethical philosophies better than others, but international relations theories are models for comparing state interactions, not applying ethical values. An individual focused on Care Ethics may struggle to synthesize their ethics with Realism as easily as an Egoist. Regardless, Liberalism does not connote a single way of weighing the variables on whether to conduct missile strikes or send in a raid or even whether to go to war in the first place. Both Ethics and international relations are lenses through which to assess the environment around us. IR evaluates the rules that govern interactions between nations while ethics address the rightness of our actions against our standard for ourselves.

It would be easier to build in multiple IR lenses than ethical frameworks. Ethical frameworks revolve around internal judgments, which build and carry forward to interpret the next judgment point. International relations theories order events and describe the environment. There are some inherent judgment points if an individual has an aversion to Realism or Liberalism, but the goal is to explain relationships, not assign value. It is the equation of value that determines ethical merit.

When designing the Ethical Assistant, we try to minimize all our biases and create an idealized lens and assign the values appropriately. If the strategist believes that Realism is the only explanation for international relations, the Ethical Assistant can view the data and analysis in that light. An Ethical Assistant could even be programmed to view Realism or Liberalism on a sliding scale or Realism with specific cutouts that state where a Realist view specifically makes an apparently Liberalist support for UN support necessary. These lenses on analyzing data and strategy would emphasize or deemphasize specific requirements or effects. A Constructivist may find herself tweaking her Ethical Assistant to enhance its awareness of cultural overlap and narratives as the strategic objectives change. The interplay between IR theory and ethical lens generally focuses on the scope of the data to include and manner of assigning values.<sup>33</sup>

An IR lens takes the data in, and the Ethical Framework characterizes the data and provides the structure the data fills in. The weight of the value shapes the potential relationships between the data points. Flavors of consequentialism could be played with if the Ethical Assistant was programmed in such a manner. Creating a single Ethical Assistant that switches between Utilitarianism and Virtue Ethics would require revaluing such significant portions of the dataset as to essentially be two separate Ethical Assistants that share some common design features. This indicates that there can be portions of Ethical

---

<sup>33</sup> Shannon Valor (*Technology and the Virtues*) makes a strong case that only virtue ethics allows humans to navigate today's technological environment. The same shifts in computing power and data analysis that drives the changes in the environment may yet enable other methods that allow others to argue that other ethical frameworks remain viable.

Assistants that can be commonly designed but would also present exposure to shared biases if not compared against additional Ethical Assistants with differently designed segments. In the future, technology and clever human and AI designers can solve these issues and embed numerous ethical structures and IR lenses to explore potential military strategies between adversaries. Still, each significant shift in ethical system would require a redefining of the relationship between the data.

Just as the design of an Ethical Assistant can be designed for different ethical lenses, it can account for the differing values between US, Iranian, or Chinese ethical constructs. A strategist focused on Just War Theory to explain US decisions on when to use force could incorporate interpretations from an Islamic perspective alongside US and western understandings of Just War Theory, particularly *jus in Bello*. Shared historical and religious lineage contributes to the relatively close interpretation of the theory and could aid in terminating or de-escalating conflicts. This overlap could shape information operations or strike targets during any operations in the Middle East or other predominantly Muslim regions. Understanding the significant overlap in this ethical outlook regarding war conduct doubles the imperative to conform as exhibiting unethical behavior will decrease support at home and anger populations in the region.

Putting similar considerations into light regarding China requires much more nuance and awareness of internal Chinese actors. The West has separated *jus ad bellum* and *jus in bello*, but Chinese theorists and the Chinese Communist Party have not embraced the division.<sup>34</sup> The recent buildup of Confucian traditions could be interpreted as reinforcing some shared aspects of *jus in bello*. However, the support also needs to be considered as merely cover for establishing inroads amongst western societies. Knowing the minds of the CCP is different than understanding the opinions of the

---

<sup>34</sup> Nadine Godehardt, "The Chinese Meaning of Just War and Its Impact on the Foreign Policy of the People's Republic of China," *GIGA Working Paper* 88, (Sep 2008), <https://www.giga-hamburg.de/en/publications/11577630-chinese-meaning-just-impact-foreign-policy-people-republic-china/>. This paper offers analysis of multiple researchers providing a good starting point to delve further into the topic.

Chinese people. The distinction between the ruling hegemony and the general populace depends on the specific strategy and its intended audience. The baseline assumption has generally been that the CCP and/or the Chinese people and the US are all evaluating a potential conflict through a similar Just War framework. It is more probable that the frameworks of each nation are more complex than a single ethical structure and the two complex frameworks within each nation are very distinct from one another.

The actual possible combinations that could be involved are sharply greater than any relationship to the total number of people involved. Each Ethical Assistant is designed by humans designing what they believe about other humans and what they believe about themselves. The strategists and designers must focus on understanding the range of risk and reward in adherence to the ethical framework. Programming the parameters to trigger the highlighting of these ethical decision points is critical. Does the Ethical Assistant present the entire range of calculated possible outcomes or cut off at a certain deviation point? Using multiple Ethical Assistants of varying design to assess a strategy via different lenses and ethical frameworks, the strategist can shed light from different angles to highlight the most critical decision points within the strategy and its subsequent execution. There is a sweet spot that balances acknowledging significant differences in ethical structures and values and perspectives, generating enough algorithms to ensure biases or blind spots are exposed, and efficiently using the resources available. The practitioners working on and with the Ethical Assistant will need constant reminding of these limitations. The Ethical Assistant is not a clairvoyant; a probability is not a foregone conclusion.

## Using the Iraq invasion of 2003 as an example

Consider how Ethical Assistants could have aided the US' case for the invasion of Iraq in 2003. The US argued that the Iraqi regime's continued subterfuge around its weapons of mass destruction program was against the international community's values and principles, mainly as Al Qaeda could use the weapons to strike at Western targets. The behavior necessitated the overthrow of the Iraqi regime.<sup>35</sup> What would an Ethical Assistant designed by the US have brought to such an argument? How could other actors have used their own Ethical Assistants to alter the narrative of the US plan? How would an Ethical Assistant adjust operations?

First off, a US Ethical Assistant would have had to evaluate the aim of regime change in light of strengthening the international liberal order and preventing greater risk to the western world. The Ethical Assistant designed to value the international structure of laws and treaties would compare the situation to historical precedents, both action and inaction, to generate an array of likely effects stemming from various responses to similar situations. The Ethical Assistant would have highlighted the actions against the Iraq regime as lacking absolute ethical continuity as the US and the UN had not interfered during previous uses and development of chemical weapons. The Iran-Iraq war provides one such example. The same comparison would apply every time a nation ignores international agreements or unilaterally withdraws from them, to include the US. In short, the Ethical Assistant would highlight that the values the US was really demonstrating to the world was concern 1) for the risk to itself and its western allies from Al Qaeda receiving weapons and technology from Saddam Hussein's regime and 2) the continued normal operations of the western dominated economic order vis a vis oil production and control in the Middle East. Acknowledging these factors as the primary motivators does not decrease the US perspective to go to war. Instead, it should clarify whether President Bush wanted to convey

---

<sup>35</sup> Esther Pan, "IRAQ: Justifying the War," *Council on Foreign Relations*, February 2, 2005, <https://www.cfr.org/background/iraq-justifying-war>.

when he was also promoting a greater focus on democratization and human rights versus stability. If nothing else, the myriad of books that were published on the centrality of oil and the hypocrisy of western values following the invasion of Iraq could have focused on a substantive normative debate over US and western values instead of the coverup campaigns and exposés.

In addition to assessing the actual conduct of the war and revealing its cost in human, economic, and environmental terms, the Ethical Assistant would have also analyzed the rebuilding and end state goals. Having information from the war with Iraq in 1991 and the information from the Iran-Iraq war, the Ethical Assistant would have access to essential data on the military and non-military effects of operations in the theater during combat. The study of post regime change would also reflect lessons gleaned from both imposing governments by outside powers and revolutions and occupations. An Ethical Assistant could have highlighted the danger to barring all Baathists from the government, from both the ethical double standard of assumed guilt and what such a ruling would drive individuals to do to support their families. Moreover, the Ethical Assistant would focus on the critical decision zones vital to establishing the nature of the newly created institutions: immediately following the regime's collapse or the handover from military to civilian control of the country. These decision zones are apparent in hindsight. If these areas were highlighted to a broader audience beforehand, perhaps there would be different outcomes. Instead, decision-makers in Iraq became bogged down with details that appear vital in the moment, but faded in importance over time.

Many prominent voices did voice their concerns on the invasion and the post-invasion period. These warnings were often ignored as party politics or preemptive responsibility deferment if the outcome did not turn out as hoped. The Ethical Assistant can overcome these types of attacks on its assessment because the Ethical Assistant is the partner of the strategist. The Ethical Assistant does not have any agenda other than supporting the strategist to assess the strategy better and improve its execution by identifying the decision areas that portend the greatest risks or rewards. Beyond the

strategist's perception of the Ethical Assistant, other parties have reason to also view the Ethical Assistant as an impartial analyst.

The Ethical Assistant could have also provided the US government with a sense of how other actors would respond. What would Russia do in the UN Security Council? How would the Democratic Party's decisions impact popular support inside the US? There are not enough resources and time to describe and program every actor and so the designers and the strategist will have to make choices on to what degree to factor in other perspectives. They will need to be transparent on these decisions as they determine the scope of the ethical impact analysis. Within the Middle East, the US would focus on Israel, Saudi Arabia, Egypt, and Iran as key power brokers. Would they have captured the role the Iraqi Kurds played or other groups?

During the occupation of Iraq, the treatment of women and families provided critical openings for opposition forces to recruit new members and gather resources. The US's tactically successful raids were also regenerating the forces the US was targeting. Though many became disillusioned with a "hearts and minds" outlook, the intent was to control the narrative. Many of the US' tactical actions provided ammunition for the opposition forces to push their narrative. The Ethical Assistant would have zeroed in on these competing objectives and the competing values in the conflict.

Other nations would have access to their own Ethical Assistants and the technologies powering them. If the US had been confronted with a potential range of casualties and economic and environmental damage, the US would be in a position to debate the merit in trading the loss of life and the suffering for a more peaceful and hopefully virtuous world. Such a debate may not have changed the decisions of other nations to support or oppose the US-led invasion of Iraq. Still, US leaders, directed to consider possible decision points of particularly impactful effects, could have avoided some of the pitfalls that occurred after the invasion began. Furthermore, if planning councils routinely confront

'unbiased' assessments of their ethical rationale behind their strategies, perhaps the councils will assess their ethical structures more closely.

### **An Ethical Assistant Continues to Provide Support After Initial Decision Points**

Attendant to the shared human condition is failure and imperfection. Humans make mistakes. Each complex decision will overlook some facet or lend too much credit to specific data. The result will not necessarily play out as desired or foreseen. Human systems are complex. As stated before, Ethical Assistants will not predict the future, but focus the strategist's attention to the most critical decision points within a likely range of outcomes. Individuals will have to evaluate the environment to make the decision they believe will achieve their aims aligned with their ethical values. The decision will not be perfect but contribute to creating a more favorable outcome.

The initial ethical decision points are just one stage in a ceaseless cycle of ethical decision-making. The Ethical Assistant provides essential support to the strategist and decision-makers after the first decision. The same data sets which the Ethical Assistant monitored to learn and analyze the decision also allow the Ethical Assistant to monitor the effects of the decision. Some of these data sets may even be real or near-real time. The strategist has already provided the Ethical Assistant evaluation criteria for the overall aims and values. These criteria were the initial inputs for the Ethical Assistant's assessments to support the strategist. With the Ethical Assistant's partnership, the strategist can continue monitoring, evaluating, and acting to maintain alignment with the ultimate aims and within the identified ethical framework.

The ability to evaluate, either iteratively or continually, is critical because ethics is an effects-based field. Monitoring the impact of decisions enables clearer insight into the nature of the ethical dilemmas in decisions. Overlooked impact areas or outcomes can be included for future analysis. The

intent behind an action and decision is not enough to justify an action, unless the knowledge of the action's wider implications is understood. Ethical decisions can still lead to undesired consequences, but the ethical framework must account for the ethical rationale behind the decision. A deontological rule is only ethically supportable if the rule is applied correctly. Adhering to a specific virtue can only be prioritized if the intent of the virtue applies to the situation. The strategist can prioritize honesty but not if transparency in one facet of a strategy covers up the larger intent. Ethical values must be consistent with the effects.

One means of supporting the intended ethical outcomes of a strategy is to advertise the ethical aims. Many actors share significant stakes in furthering foundational US values such as religious freedom, self-determination, gender equality, and a host of others bundled under human and personal rights. Often the impetus stems from supporting their democratic form of governments which rely on at least some of these values. Some Non-governmental Organizations also champion tightly focused niches in the broad umbrella of human rights. International organizations such as the United Nations or World Bank, alongside third-party nations, also have vested stakes in promoting values that also align with the US. The Ethical Assistant's focus on the ethical aims alongside other aims in achieving the desired end state provides a focal point which parties can coalesce around constructively. While other aspects of the strategy may be contentious between various parties, the US can harness some support and resources from parties where the values overlap. Doing so will continue to further the US's meta-narrative and defend its prominent position in the world.

Strategic decisions can also include presumed future decisions or actions that will take place to balance out a plan and ensure the achievement of an aim or objective. Suppose the ethical value of an effort relies on future actions or responses. In that case, the ethical merit of the overall strategy rings hollow if those actions are not taken. The nation will lose credibility and soft power. Through real or near real-time analysis of the effects of a strategy, deliberative surveys and discussions, and harnessing

the support of external actors, an Ethical Assistant and strategist can keep focus on and more efficiently achieve the desired aims and objectives. The nation will not lose sight of critical aspects or fail to commit the resources at the necessary future point. The larger strategic vision will remain in place.

Additionally, the Ethical Assistant's design supports the incorporation of new data, enabling the strategist and decision-makers to modify actions and decisions to maintain alignment with the values and aims of the US. The security of American values remains at the center of all strategy discussions. The strategy may not win universal support but sends a crystal-clear message to all potential allies and adversaries on the benefit of siding with the US in furthering democratic principles and human rights.

Imagine that the ancient Greek polis had been able to partner with Ethical Assistants. Would Pericles' words and strategic vision have remained more visible and central to Athens after his death? Thucydides' depiction of Pericles' net assessment entering the Peloponnesian War highlighted the dangers Athenians would later willingly choose to pursue. It is possible that Pericles' vision for Athens' victory was not in line with the Athenian assembly that remained after his death. The answer could only have been found in the expression of the individual values of the citizens of Athens, the reasons why they voted and acted. The Melian Dialogue and the decisions on the Sicilian campaign would have a different character if debated within the Athenian assembly. What were the real values that drove Athens to fight? It is also possible that the values changed during the war or that the war itself altered Athenian values. Was it a conscious decision? The Athenian assembly is a unique thought experiment as within its membership, democracy ruled. The assembly could have debated these questions and values explicitly before taking action. Today we look back on the consequences of actions to derive value-laden decisions and strategies, which even Thucydides' records cannot fully reveal.

**Ethical Assistant's Inherent Resistance to Gamesmanship:** Would an Ethical Assistant become simply another political tool to manage the narrative instead of a steering aid to guide a value-driven narrative?

In the highly publicized polarity of today's political environment, there can be a natural inclination to dismiss the resiliency of an Ethical Assistant's focus on the genuine values driving a strategy. Indeed, nothing prevents a strategist from creating artificial values or masking the "real/hidden" objective behind the stated purposes in the strategy. However, the overall transparency of the inputs into the Ethical Assistant subsumes the hidden motives and values. The Ethical Assistant and the strategy as designed become the debated vision and either selected or rejected. A strategist and the decision-makers who support the Ethical Assistant and the strategy become accountable for the transparent values and objectives expressed in the strategy.

The ulterior motives or values might be fascinating from a psychoanalytical viewpoint, but the attainment of these 'real/hidden' objectives is still visible and captured in the overall analysis. The 'real/hidden' objectives would still align with the overall ethical structure and values if they survived the Ethical Assistants review. The most probable value-laden decision points are still determined based on the Ethical Assistant and strategy's transparent and explicit values and objectives. In revealing the aims, values, and ethical structures, the strategists bind themselves to the visions laid out by the design of the Ethical Assistant and the nature of their strategies. The explicitly expressed values and agreed-upon aims burn away the influence of a strategist's disingenuous values or ulterior motives. The strategist's supposed fiction becomes the new reality from which they cannot escape.

### **Ethical Assistants can be Applicable in all Fields**

The application of military force is only one of many powers that states wield in international affairs. International economic policies carry as much force to shape lives over time, and can be perhaps

more insidious as the effects unfold over time. The relationship between cause and effect can be murky. The intent may slowly shift over time without a full comprehension of the relationship to the starting point. In any area where complex processes affect large swaths of people uniquely, an Ethical Assistant can help decision-makers craft and weigh strategies to achieve the desired aims.

How would an Ethical Assistant have helped the Clinton Administration and Congress with implementing the North American Free Trade Agreement? The administration freely admitted that globalization and free trade with Canada and Mexico specifically would alter the landscape of the US economy.<sup>36</sup> The vast increase in gross domestic product justified the costs associated with the changes. In recognition of the sacrifices imposed on some American workers and their families as industries were rearranged and current jobs lost, the government would invest in job training and new industries in the regions most impacted. By investing in these areas, the nation would be showing its connection and concern to all Americans. The rhetoric sounded good, but the follow-through was severely lacking.

Twenty-five years later, the US is dealing with the fallout from failing to follow through on promises to American citizens regarding the shared distribution of economic gains purchased with the dignity of specific American industries. How might a running scorecard for NAFTA have changed any actions or decisions by Congress or the Presidential administrations? Might the exposure of ultimate aims and intermediate ways and means of congressional policies by analysis in partnership with Ethical Assistants create more accountable elected officials? Would the communities of the former factories

---

<sup>36</sup> See Robert E. Scott, "The High Price of 'Free' Trade: NAFTA's Failure has Cost the United States Jobs Across the Nation," *Economic Policy Institute*, November 17, 2003, [https://www.epi.org/publication/briefingpapers\\_bp147/](https://www.epi.org/publication/briefingpapers_bp147/); Robert E. Scott, "Investors Cleaning Up While Workers Suffer," *Economic Policy Institute*, June 23, 2003, [https://www.epi.org/publication/webfeatures\\_viewpoints\\_nafta\\_jobs/](https://www.epi.org/publication/webfeatures_viewpoints_nafta_jobs/); Andrew Chatzky, James McBride, and Mohammed Aly Sergie, "NAFTA and the USMCA: Weighing the Impact of North American Trade," *Council on Foreign Relations*, Updated July 1, 2020, [https://www.cfr.org/backgroundunder/naftas-economic-impact#:~:text=Summary%20NAFTA%20was%20a%20landmark%20trade%20deal%20between,States%20for%20contributing%20to%20job%20losses%20and%20outsourcing.](https://www.cfr.org/backgroundunder/naftas-economic-impact#:~:text=Summary%20NAFTA%20was%20a%20landmark%20trade%20deal%20between,States%20for%20contributing%20to%20job%20losses%20and%20outsourcing.;); and Gwen Ifill, "The 1992 Campaign: The Democrats; With Reservations, Clinton Endorses Free-Trade Pact," *New York Times*, October 5, 1992, <https://www.nytimes.com/1992/10/05/us/1992-campaign-democrats-with-reservations-clinton-endorses-free-trade-pact.html>.

have had greater leverage if their needs had been captured and then tracked by an Ethical Assistant? Would constituents around the country enjoying increased living standards have been as content if they saw the visible link to the suffering of their fellow Americans? Again, the proof would only be in actions that these individuals could have made in the past. It will never be known.

### **Knowing ourselves**

The entirety of constructing an Ethical Assistant is a massive project with real difficulties and complexities. Compiling the data. Writing the code. Gathering experts and transforming their knowledge to an accessible form for the Ethical Assistant. No less arduous a task and with equal risk to any other attendant component, the strategist must delve into and fully understand her own ethics and values.

The critical component is the knowledge revealed about our human nature in the process of creating and partnering with an Ethical Assistant. Self-knowledge is the essential starting point. The refining of the self-assessment is the most significant benefit. It may seem odd to expend so much time and effort to create a supremely complex AI program whose most significant benefit is a mirror into human thought processes and ethical calculations. The very act and art of creating the Ethical Assistant is the revealing (invoking Heidegger) of values. As individuals and collectively as a society, women and men must confront the changing relationships with technology from an external view to the messier and more entangled role technology is taking inside individual social decisions. Biases and superficial arguments will be worn away. Through creating the Ethical Assistant and then partnering afterward, the true hierarchy of values is exposed. The self-knowledge gained can then be employed to defend, strengthen, and align actions. The Ethical Assistant supports the strategist by bringing forward the most

complex and critical decision points, allowing the strategist to expend decision space where it is most impactful: weighing our own competing values.

## **Conclusion**

Ethical Assistants provide a means for humans to take educated control over decisions involving data beyond the ability of any human to comprehend fully. The central question can again focus on what we stand for and how our actions represent those values. Humans bear responsibility for the direction of our development, not technology, no matter how alluring the prospect.

Nations face the same dilemma of aligning values and actions. An Ethical Assistant provides a more straightforward depiction of the ethical choices facing nations and will contribute to coalescing societal debates around the state's values. No application and no single human can guarantee that every decision of a state meets an ethical standard, let alone the standards of a multitude of people. By shifting the conversation towards the alignment of actions and values, an Ethical Assistant will further progress towards a more perfect union and enhance the security of the state in the process.

## Bibliography

- Anderson, Michael and Susan Leigh Anderson. *Machine Ethics*. Cambridge: Cambridge Press, 2010.
- Arnold, Thomas and Matthias Scheutz. "The "Big Red Button" is Too Late: An Alternative Model for the Ethical Evaluation of AI systems." *Ethics and Information Technology* 20 (2018): 59–69. <https://doi.org/10.1007/s10676-018-9447-7>.
- Bauer, William A. and Veljeko Dubjevic. "AI Assistants and the Paradox of Internal Automaticity." *Neuroethics* 13 (2020): 303–310. <https://doi.org/10.1007/s12152-019-09423-6>.
- Barbero, Alessandro. *The Battle: a New History of Waterloo*. New York: Walker & Co., 2005.
- Barbour, Ian. *Ethics in an Age of Technology*. HarperCollins, 1993.
- Bode, Ingvild and Hendrik Huelss. "Autonomous Weapons Systems and Changing Norms in International Relations." *Review of International Studies* 44, no. 3 (2018): 393–413. doi:10.1017/S0260210517000614.
- Bousquet, Antoine. "Chaoplexic Warfare or the Future of Military Organization." *International Affairs* 84, no. 5 (2008): 915-929.
- Bostrum, Nick, Yudkowsky, Eliezer. "The Ethics of AI." Machine Intelligence Research Institute. Accessed July 29, 2019. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.477.9269&rep=rep1&type=pdf>.
- Brose, Christian. "The New Revolution in Military Affairs: War's Sci-Fi Future." *Foreign Affairs*, May/June 2019, 122-34.
- Brynjolfsson, Erik and Andrew McAfee. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W.W. Norton Company, 2016.
- Daniel Byman. "Why Drones Work: The Case for Washington's Weapon of Choice." *Foreign Affairs* 92, no. 4 (Jul, 2013): 32-43.
- Chaillou, Stanislas. "AI & Architecture." *Medium*, February 26, 2019. <https://medium.com/built-horizons/ai-architecture-4c1ec34a42b8>.
- Chapa, Joseph O. "Moral Philosophy as a Force Protection Measure." *The Strategy Bridge*, July 19, 2019: 1-8.
- Chatzky, Andrew, and James McBride, and Mohammed Aly Sergie. "NAFTA and the USMCA: Weighing the Impact of North American Trade." *Council on Foreign Relations*, Updated July 1, 2020. <https://www.cfr.org/backgrounder/naftas-economic-impact#:~:text=Summary%20NAFTA%20was%20a%20landmark%20trade%20deal%20between,States%20for%20contributing%20to%20job%20losses%20and%20outsourcing>.
- Cordesman, Anthony H. "China's New 2019 Defense White Paper: An Open Strategic Challenge to the United States, But One Which Does Not Have to Lead to Conflict See also CSIS Chinese Power Project," *Center for Strategic International Studies*, July 24, 2109. <https://www.csis.org/analysis/chinas-new-2019-defense-white-paper>.

- Cordesman, Anthony H. with Grace Hwang. "US Competition with China and Russia: The Crisis-driven Need to Change US Strategy." *Center for Strategic International Studies*, January 6, 2021. <https://www.csis.org/analysis/biden-transition-and-us-competition-china-and-russia-crisis-driven-need-change-us-strategy>
- Cukier, Kenneth. "Ready for Robots? How to Think About the Future of AI." *Foreign Affairs*, July/August 2019, 192-98.
- Crockford, Kade. "How is Face Recognition Surveillance Technology Racist?" ACLU, June 16, 2020. <https://www.aclu.org/news/privacy-technology/how-is-face-recognition-surveillance-technology-racist/>.
- Cronin, Audrey Kurth. "Why Drones Fail: When Tactics Drive Strategy." *Foreign Affairs* 92, no. 4 (Jul, 2013): 44-54.
- Danaher, J. "Toward an Ethics of AI Assistants: an Initial Framework." *Philos. Technol.* vol 31(2018): 629–653. <https://doi.org/10.1007/s13347-018-0317-3>.
- Danzig, Richard. "Technology Roulette: Managing Loss of Control as Many Militaries Pursue Technological Superiority." *Center for a New American Security, Technology & National Security (June 2018)*.
- Defense Innovation Board. *AI Principles: Recommendations on the Ethical Use of AI by the Department of Defense, Primary Document*. Defense Innovation Board, October 2019. <https://innovation.defense.gov/ai>.
- Defense Innovation Board. *AI Principles: Recommendations on the Ethical Use of AI by the Department of Defense, Supporting Document*. Defense Innovation Board, October 2019. <https://innovation.defense.gov/ai>.
- DeFranco J, DiEuliis D, Bremseth LR, Snow JJ, Giordano J. "Emerging Technologies for Disruptive Effects in Non-Kinetic Engagements." *HDIAC Journal*, Vol 6, no. 2 (2019): 49-55. <https://www.hdiac.org/wp-content/uploads/emerging-technologies.pdf>.
- Deputy Secretary of Defense. *Implementing Responsible Artificial Intelligence in the Department of Defense*, Deputy Secretary of Defense, Washington, D.C, 26 May 2021.
- Ethical Application of AI Framework*. American Council for Technology-Industry Advisory Council (ACT-IAC), October 8, 2020. <https://www.actiac.org/act-iac-white-paper-ethical-application-ai-framework>.
- Final Report*. National Security Commission on AI, March 2021. <https://reports.nscai.gov/final-report/table-of-contents/>.
- Flournoy, Michele A. "How to Prevent a War in Asia: The Erosion of American Deterrence Raises the Risk of Chinese Miscalculation." *Foreign Affairs*, June 18, 2020. <https://www.foreignaffairs.com/articles/united-states/2020-06-18/how-prevent-war-asia>.

- Garvie, Clare and Jonathan Frankle. "Facial-Recognition Software Might Have a Racial Bias Problem." *Atlantic*, April 7, 2016. <https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/>.
- GPT-3. "A Robot Wrote This Entire Article. Are You Scared Yet, Human?" *Guardian*, September 8, 2020. <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>.
- Godehardt, Nadine. "The Chinese Meaning of Just War and Its Impact on the Foreign Policy of the People's Republic of China." *GIGA Working Paper* 88, (Sep 2008). <https://www.giga-hamburg.de/en/publications/11577630-chinese-meaning-just-impact-foreign-policy-people-republic-china/>.
- Han Lau, Jey and Trevor Cohn and Timothy Baldwin and Adam Hammond. "This AI Poet Mastered Rhythm, Rhyme, and Natural Language to Write Like Shakespeare." *IEEE Spectrum*, April 30, 2020. <https://spectrum.ieee.org/artificial-intelligence/machine-learning/this-ai-poet-mastered-rhythm-rhyme-and-natural-language-to-write-like-shakespeare>.
- Harris, Melissa. "DIB's Recommendations for Digitization Can Help Tackle Diseases Like Coronavirus." *GovernmentCIO*, March 6, 2020. <https://governmentciomedia.com/dibs-recommendations-digitization-can-help-tackle-diseases-coronavirus>.
- Harwell, Drew. "Federal Study Confirms Racial Bias of Many Facial-Recognition Systems, Cast Doubt on Their Expanding Use." *Washington Post*, Dec 19, 2019. <https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/>
- Hinman, Lawrence M. *Ethics: A Pluralistic Approach to Moral Theory, 5th Ed.* Wadsworth Cengage Learning.
- Hughes, Thomas P. *Human-Built World: How to Think About Technology and Culture.* University of Chicago Press, 2004.
- Ifill, Gwen. "The 1992 Campaign: The Democrats; With Reservations, Clinton Endorses Free-Trade Pact." *New York Times*, October 5, 1992. <https://www.nytimes.com/1992/10/05/us/1992-campaign-democrats-with-reservations-clinton-endorses-free-trade-pact.html>.
- Inkster, Nigel. "Military Cyber Capabilities." *Adelphi series* 55 no. 456 (2015): 83–108. <https://doi-org.usnwc.idm.oclc.org/10.1080/19445571.2015.1181444>.
- Interim Report.* National Security Commission on AI, November 2019. [https://www.nscail.gov/wp-content/uploads/2021/01/NSCAI-Interim-Report-for-Congress\\_201911.pdf](https://www.nscail.gov/wp-content/uploads/2021/01/NSCAI-Interim-Report-for-Congress_201911.pdf).
- Jackson, Robert and Georg Sorensen. *Introduction to International Relations: Theories and Approaches, Fifth Edition.* Oxford: Oxford University Press, 2013.
- Jarod, Russell. "Making AI Intelligible." Boston University. Accessed July 29, 2019. <http://www.bu.edu/articles/2018/making-artificial-intelligence-intelligible/>.
- Jarrah, Mohammed Hossein. "Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making." *Business Horizons* 61, iss 4 (July-August 2018):577-586.

- Johnson, James S. "Artificial Intelligence: A Threat to Strategic Stability." *Strategic Studies Quarterly*, Spring 2020: 16-39.
- Kaina, Elsa. "China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems." *Lawfare*, April 17, 2018. <https://www.lawfareblog.com/chinas-strategic-ambiguity-and-shifting-approach-lethal-autonomous-weapons-systems>.
- Kaleagasi, Bartu. "A New AI Can Write Music as Well as a Human Composer." *Futurism*, March 9, 2017. <https://futurism.com/a-new-ai-can-write-music-as-well-as-a-human-composer>.
- Kaplan, David M. *Readings in the Philosophy of Technology*. Lanham, Md: Rowman & Littlefield Publishers, 2004.
- Kasenberg, Daniel, Vasanth Sarathy, Thomas Arnold, Matthias Scheutz and Tom Williams. "Quasi-Dilemmas for Artificial Moral Agents." *ICRES 2018: International Conference on Robot Ethics and Standards*, Troy, NY, August 20-21, 2018. <https://doi.org/10.13180/icres.2018.20-21.08.012>.
- Kelsay, John. *Arguing the Just War in Islam*. Cambridge, Massachusetts: Harvard University Press, 2007.
- Kraft, Amy. "Microsoft Shuts Down AI Chatbot After It Turned into a Nazi." CBS News, updated March 25, 2016. <https://www.cbsnews.com/news/microsoft-shuts-down-ai-chatbot-after-it-turned-into-racist-nazi/>.
- Kranzberg, Melvin. "Technology and History: 'Kranzberg's Laws.'" *Technology and Culture*, Vol 27, No 3 (July 1986): 544-560.
- Krepinevich, Andrew F, Jr. "The Eroding Balance of Terror: The Decline of Deterrence." *Foreign Affairs* 98, Iss. 1 (January/February 2019): 62-69.
- Kuhn, Thomas S. *The Structure of Scientific Revolutions, 4<sup>th</sup> Edition*. Chicago, University of Chicago Press, 2012.
- Lander, Eric. *Defense Innovation Board Joint Pathology Center (JPC) Repository Enhancement*. March 2020. [https://innovation.defense.gov/Portals/63/documents/Meeting%20Documents/March%205,%202020/DIB\\_JPC\\_Recommendations\\_Final.pdf](https://innovation.defense.gov/Portals/63/documents/Meeting%20Documents/March%205,%202020/DIB_JPC_Recommendations_Final.pdf)
- Lavenda, David. "Artificial Intelligence Vs. Intelligence Augmentation." *Network World*, August 5, 2016. <https://www.networkworld.com/article/3104909/artificial-intelligence-vs-intelligence-augmentation.html>.
- Legenhausen, Hajj Muhammad. *Islam and Just War Theory*. Qom, Iran: The Imam Khomeini Education and Research Institute, February 7, 2008.
- Lengel, Edward G. *General George Washington: a Military Life* First edition. New York: Random House, 2005.
- Lieven, Dominic. *Russia Against Napoleon: the True Story of the Campaigns of War and Peace*. First American edition. New York: Viking, 2010.

- Lopez-Rincon, Omar and Oleg Starostenko and Gerardo Ayala-San Martín. "Algorithmic Music Composition Based on AI: A Survey." *2018 International Conference on Electronics, Communications and Computers (CONIELECOMP)* (2018): 187-193. DOI: 10.1109/CONIELECOMP.2018.8327197.
- Lucas, George R., Jr. "State Sponsored Hacktivism and the Advent of the "Soft War." *Ethics and Armed Forces*, Issue 2014 vol. 2: 16-23.
- Maedche, A., Legner, C., Benlian, A. et al. "AI-Based Digital Assistants." *Bus Inf Sys Eng* 61 (2019): 535–544. <https://doi.org/10.1007/s12599-019-00600-8>.
- Manikonda, Lydia, Aditya Deotale, and Subbarao Kambhampati. "What's up with Privacy?: User Preferences and Privacy Concerns in Intelligent Personal Assistants." *arXiv*, Cornell University, submitted November 20, 2017. arXiv:1711.07543v1.
- Marche, Stephen. "The Computers Are Getting Better At Writing." *New Yorker*, April 30, 2021. <https://www.newyorker.com/culture/cultural-comment/the-computers-are-getting-better-at-writing>.
- McAfee, Andrew and Erik Brynjolfsson. *Machine, Platform, Crowd: Harnessing Our Digital Future*. New York: W.W. Norton Company, 2018.
- McIlroy-Young, Reid and Siddhartha Sen and Jon Kleinberg and Ashton Anderson. "Aligning Superhuman AI with Human Behavior: Chess as a Model System." *KDD '20: Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, August 2020: 1677-1687. <https://doi.org/10.1145/3394486.3403219>.
- McLaren, Bruce M. "Computational Models of Ethical Reasoning: Challenges, Initial Steps, and Future Directions." *IEEE Intelligent Systems* 21, no. 4 (July-August 2006): 29-37.
- McLaren, Bruce. "Extensionally Defining Principles and Cases in Ethics: an AI Model," *AI Journal*, Volume 150, (November 2003): 145-181.
- Mengshu, Zhan. "A Brief Overview of Alexander Wendt's Constructivism." *E-International Relations*, May 19, 2020. <https://www.e-ir.info/2020/05/19/a-brief-overview-of-alexander-wendts-constructivism/>.
- MIT Technology Review. "New Measure of Human Brain Processing Speed." *MIT Technology Review: Biotechnology*. Updated 28 August 2009. <https://www.technologyreview.com/2009/08/25/210267/new-measure-of-human-brain-processing-speed/>.
- Moroni, Artemis and Jonata Manzolli and Fernando Von Zuben and Ricardo Gudwin. "Vox Populi: An Interactive Evolutionary System for Algorithmic Music Composition." *Leonardo Music Journal*, Vol. 10 (2000): 49–54. Accessed On 03 May 2021. <http://direct.mit.edu/lmj/article-pdf/doi/10.1162/096112100570602/1673952/096112100570602.pdf>.
- Najibi, Alex. "Racial Discrimination in Face Recognition Technology." *Science in the News*, October 24, 2020. <https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/>.

- “New Measure of Human Brain Processing Speed.” *MIT Technology Review: Biotechnology*. Updated 28 August 2009. <https://www.technologyreview.com/2009/08/25/210267/new-measure-of-human-brain-processing-speed/>.
- Newcomb, Alyssa. “Microsoft’s Chat Bot ‘Tay’ Gets a Time-Out After Rude Comments.” ABC News, March 25, 2016. <https://abcnews.go.com/Technology/microsofts-chat-bot-tay-time-rude-comments/story?id=37926075>.
- Nye, David E. *Technology Matters: Questions to Live With*. MIT Press, 2007.
- Oberste-Ufer, Kai, PhD. “7 Ways AI is Revolutionizing Architecture.” *Dormakaba*, February 9, 2019. <https://blog.dormakaba.com/7-ways-artificial-intelligence-is-revolutionizing-architecture/>.
- O’Donnell, Kathleen M. “Embracing AI in architecture.” *The American Institute of Architects*. Accessed May 2, 2021. <https://www.aia.org/articles/178511-embracing-artificial-intelligence-in-archit:46>.
- Oesterheld, Casper. “Machine Ethics and Preference Utilitarianism.” Last modified May 25 ,2015. <https://reducing-suffering.org/machine-ethics-and-preference-utilitarianism/>.
- Orend, Brian. “Kant’s Just War Theory.” *Journal of the History of Philosophy* 37, no 2. (April 1999): 323-353.
- Ouellette, Marc and Steven Conway. “A Feel for the Game: AI, Computer Games and Perceiving Perception.” *Eludamos. Journal for Computer Game Culture* 10, no. 1 (2019):9-25.
- Pan, Esther. “IRAQ: Justifying the War,” *Council on Foreign Relations*, February 2, 2005, <https://www.cfr.org/background/iraq-justifying-war>.
- Parker, Russell Worth. “Terrorism and Just War: A Comparative Analysis of Western and Islamic Precepts.” *The Strategy Bridge*, October 22, 2017. <https://thestrategybridge.org/the-bridge/2017/10/22/terrorism-and-just-war-a-comparative-analysis-of-western-and-islamic-precepts>
- Perez, Sarah. “Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism.” *Tech Crunch*, March 24, 2016. <https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/>.
- Pomeroy, Jennifer, Nathan Swartz, Logan Suntzenich, and Kevin Winz. "Russia’s Search for Stability: Cyber Capabilities and Military Buildup." *Current Politics and Economics of Russia, Eastern and Central Europe* 33, no. 1 (2018): 101-126. <https://login.usnwc.idm.oclc.org/login?url=https%3A%2F%2Fwww.proquest.com%2Fscholarly-journals%2Frussias-search-stability-cyber-capabilities%2Fdocview%2F2189947882%2Fse-2%3Faccountid%3D322>.
- Postman, Neil. *Technopoly: The Surrender of Culture to Technology*. Vintage Press, 1992.
- Powers, Thomas. “Prospects for a Kantian Machine.” *IEEE Intelligent Systems, IEEE*. 21, no. 4 (July-August 2006): 46- 51. 10.1109/MIS.2006.77.

- PriceWaterhouseCoopers. "GlobalAI Study: Exploiting the AI Revolution." Accessed September 15, 2019. <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligencestudy.html>.
- Risi, S., Preuss, M. "From Chess and Atari to StarCraft and Beyond: How Game AI is Driving the World of AI." *Künstl Intell* vol 34 (2020): 7–17. <https://doi.org/10.1007/s13218-020-00647-w>
- Rossi, Francesca. "AI Ethics." PowerPoint presentation. October 14, 2020.
- Sacks, Steve. "China's Military has a Hidden Weakness." *The Diplomat*, April 20, 2021. <https://thediplomat.com/2021/04/chinas-military-has-a-hidden-weakness/>.
- Sandvik, Kristin Bergtora, and Kjersti Lohne. "The Rise of the Humanitarian Drone: Giving Content to an Emerging Concept." *Millennium Journal of International Studies*, (June 27 2014). <https://doi.org/10.1177/0305829814529470>.
- Schulzke, Marcus. "Autonomous Weapons and Distributed Responsibility." *Philosophy & Technology* 26, no. 2 (June 2013): 203-219. doi:http://dx.doi.org.usnwc.idm.oclc.org/10.1007/s13347-012-0089-0. <https://login.usnwc.idm.oclc.org/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fscholarly-journals%2Fautonomous-weapons-distributed-responsibility%2Fdocview%2F1354336648%2Fse-2%3Faccountid%3D322>.
- Scott, Robert E. "The High Price of 'Free' Trade: NAFTA's Failure has Cost the United States Jobs Across the Nation." *Economic Policy Insitute*, November 17, 2003. [https://www.epi.org/publication/briefingpapers\\_bp147/](https://www.epi.org/publication/briefingpapers_bp147/).
- Scott, Robert E. "Investors Cleaning Up While Workers Suffer." *Economic Policy Institute*, June 23, 2003. [https://www.epi.org/publication/webfeatures\\_viewpoints\\_nafta\\_jobs/](https://www.epi.org/publication/webfeatures_viewpoints_nafta_jobs/).
- Sharma, Kris. "How does the human brain compare to a computer?" *Crucial: Technology*. Updated 20 April, 2020. <https://www.crucial.com/blog/technology/how-does-the-human-brain-compare-to-a-computer>.
- Simonite, Tom. "The Best Algorithms Struggle to Recognize Black Faces Equally." *Wired*, July 22, 2019. <https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/>.
- Soltes, Eugene. "Why It's So Hard to Train Someone to Make an Ethical Decision." *Harvard Business Review*, January 11, 2017. <https://hbr.org/2017/01/why-its-so-hard-to-train-someone-to-make-an-ethical-decision>.
- Sparrow, Robert and George Lucas. "When Robots Rule the Waves?" *Naval War College Review*: Vol. 69: No. 4 , Article 7 (2016).
- Springer, Paul J. *Encyclopedia of Cyber Warfare*. Santa Barbara, CA: ABC-CLIO, 2017.
- The State Council Information Office of the People's Republic of China. "China's National Defense in the New Era." Foreign Languages Press Co. Ltd: Beijing, China: 2019. <http://www.xinhuanet.com/english/download/whitepaperonnationaldefenseinnewera.doc>.
- Symonds, Craig L. *The Battle of Midway: Battle of Midway*. New York: Oxford University Press USA - OSO, 2011.

- Valor, Shannon. *Technology and the Virtues*. New York: Oxford University Press, 2016.
- Wallach, Wendell and Colin Allen. *Moral Machines, Teaching Robots Right from Wrong*. Oxford University Press: 2010.
- Webb, Whitney. "New Pentagon-Google Partnership Suggests AI Will Soon Be Used To Diagnose Covid-19." *Unlimited Hangout*, September 17, 2020.  
<https://unlimitedhangout.com/2020/09/reports/new-pentagon-google-partnership-suggests-ai-will-soon-be-used-to-diagnose-covid-19/>.
- Wendt, Alexander. *Social Theory of International Politics*. Cambridge, Cambridge University Press: 1999.
- Winner, Langdon. *Autonomous Technology: Technics-out-of-control as a Theme in Political Thought*. MIT Press, 1977.
- Wyatt, Austin. "Charting Great Power Progress Toward a Lethal Autonomous Weapon System Demonstration Point." *Defence Studies*, 20:1, 1-20. DOI: 10.1080/14702436.2019.1698956.
- Yao, Mariya, Marlene Jia, and Adelyn Zhou. *Applied AI: A Handbook for Business Leaders*. TOPBOTS Inc., 2018.