

Article

Not peer-reviewed version

---

# Algorithmic Diplomacy and the Geopolitics of Artificial Intelligence: Machine-Driven International Relations in a Data-Oriented Global Landscape

---

[Safran Almakaty](#)\*

Posted Date: 10 March 2026

doi: 10.20944/preprints202603.0683.v1

Keywords: algorithmic diplomacy; artificial intelligence governance; North–South inequality; nuclear deterrence; machine-mediated international relations; and computational power asymmetry



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Algorithmic Diplomacy and the Geopolitics of Artificial Intelligence: Machine-Driven International Relations in a Data-Oriented Global Landscape

Safran Safar Almakaty

Journalism and New Media Department, College of Media and Communication, Imam Mohammad Ibn Saud Islamic University (IMSIU); ssamakaty@imamu.edu.sa

## Abstract

The integration of artificial intelligence (AI) and algorithmic decision-making into international relations represents a significant evolution that current theoretical frameworks have not fully accommodated. Established International Relations (IR) paradigms—realism, liberalism, and constructivism—typically conceptualize technology as a tool for statecraft rather than as an influential factor actively shaping the decision-making environment. This paper asserts that such a conceptual omission is increasingly untenable. Utilizing an integrative analysis spanning IR theory, AI ethics, security studies, and political economy, this work examines three principal aspects of "algorithmic diplomacy": (1) the ways in which algorithmic bias within systems operated by international organizations perpetuates structural inequalities between North and South; (2) the impact of AI incorporation into nuclear command and control systems on the logic of Mutually Assured Destruction (MAD) and crisis management timelines; and (3) the emergence of computational power imbalances in trade and climate negotiations as a novel form of diplomatic influence. The conclusion presents an initial framework for machine-mediated international relations and emphasizes the necessity for normative and governance responses at both national and multilateral levels.

**Keywords:** algorithmic diplomacy; artificial intelligence governance; North–South inequality; nuclear deterrence; machine-mediated international relations; and computational power asymmetry

---

## 1. Introduction

The image of the diplomat—deliberate, contextually sensitive, deploying human judgment accumulated over decades of lived experience—has long anchored the practice and theory of international relations. The twenty-first century, however, has introduced a new actor into the negotiating chamber: one that calculates rather than reflects, optimizes rather than deliberates, and scales rather than adapts. As states, international organizations, and non-state actors increasingly deploy AI-driven systems to process intelligence, model adversary behavior, allocate humanitarian resources, and manage nuclear arsenals, the architecture of international diplomacy is undergoing a silent but profound transformation.

The academic literature has been slow to respond. The dominant traditions in International Relations—Waltzian structural realism, Keohanian liberalism, and Wendtian constructivism—were formulated in an era when technology, however consequential, could reasonably be treated as an exogenous instrument wielded by rational state actors (Waltz, 1979; Keohane & Nye, 2001; Wendt, 1999). Cyber-warfare scholarship has advanced understanding of digital conflict (Deibert, 2020), and information operations research has illuminated how data environments shape political outcomes (Zuboff, 2019). What remains critically underdeveloped, however, is a comprehensive theory of how AI systems themselves—through their embedded logics, training data, and optimization objectives—

reshape the environment within which international decisions are made and norms are negotiated. This paper proposes to describe this emergent domain as "machine-mediated international relations."

The stakes of this conceptual gap are not merely academic. Real world consequences are already materializing. Refugee allocation systems deployed by the United Nations High Commissioner for Refugees (UNHCR) exhibit documented biases that disadvantage applicants from the Global South (Bansak et al., 2018). The United States, Russia, and China are actively integrating AI-assisted decision support into their nuclear command, control, communications, and intelligence (C3I) architectures, raising fears of inadvertent escalation that Cold War deterrence theory never anticipated (Acton, 2018; Geist & Lohn, 2018). Meanwhile, technologically advanced economies deploy sophisticated computational modeling tools in trade and climate negotiations that effectively allow them to anticipate and thereby neutralize the bargaining positions of less-endowed counterparts (Kissinger et al., 2021; Farrell & Newman, 2023).

These developments demand more than incremental adjustments to classical theories—they require a fundamental rethinking of the ontological and normative status of technology in international affairs. When algorithmic systems become embedded in the core processes of diplomacy and security, they do not simply amplify human intent; they introduce new logic, priorities, and vulnerabilities that may diverge from or even undermine the interests of their deployers. The opacity, speed, and scale of AI-driven decision-making complicate traditional mechanisms of accountability and crisis management, raising urgent questions about agency, responsibility, and control in a world where human judgment is increasingly mediated—or displaced—by machines.

This paper proceeds from the premise that the rise of algorithmic diplomacy is not a peripheral development but a central challenge for the theory and practice of international relations in the twenty-first century. By drawing on interdisciplinary scholarships in IR, AI ethics, security studies, and the political economy of data, the analysis aims to move beyond treating technology as a passive instrument of statecraft. Instead, it interrogates how AI systems actively shape the environments in which power is exercised, interests are articulated, and outcomes are negotiated. In doing so, it seeks to illuminate the new forms of leverage, risk, and inequality emerging in a data-driven world and to lay the groundwork for a theory of machine-mediated international relations that is attuned to the realities of algorithmic governance on the global stage.

## 2. Theoretical Framework: From Human Diplomacy to Machine-Mediated International Relations

The practice of diplomacy has traditionally relied on human judgment, contextual awareness, and accumulated experience to navigate the complexities of international relations. However, the emergence of artificial intelligence and algorithmic systems is reshaping these foundations, introducing new actors and logic into the diplomatic sphere. This section explores how classical International Relations (IR) theories—realism, liberalism, and constructivism—have generally regarded technology as a passive instrument of statecraft, failing to account for its potential to actively reshape the environment in which international decisions are made. By tracing the evolution of technological influence from a mere tool to a possible agent, this framework sets the stage for understanding the profound implications of AI-driven systems on global governance, security, and negotiation. The goal is to articulate a paradigm for machine-mediated international relations, addressing the conceptual and normative gaps left by traditional theories and preparing the groundwork for a comprehensive analysis of algorithmic diplomacy in a data-driven world.

### 2.1. *Technology in Classical IR Theory*

Major International Relations theories have not sufficiently addressed technology as an independent structural factor. For realists in the tradition of Waltz (1979), the international system is anarchic, and its structure is defined by the distribution of capabilities among states. Technology

enters this framework as a component of state power—weapons system, an industrial capacity, a surveillance apparatus—but always in service of the sovereign actor who wields it. Mearsheimer's (2001) offensive realism similarly treats technological advantage as a resource that states accumulate in the competitive pursuit of security, rather than as something that can independently reshape the terms of competition. The possibility that technology might itself constitute a strategic environment, operating with a degree of autonomous logic that partially escapes human control, sits uneasily within these frameworks.

Liberals fare somewhat better. Keohane and Nye's (2001) concept of complex interdependence acknowledged that technological connectivity could reshape state interests and create new forms of mutual vulnerability. Their concept of "soft power" (Nye, 2011) incorporated information and communication technologies as instruments of influence. Yet even here, the liberal tradition treats technology as a medium through which human agency operates, not as a co-constitutive force that can restructure the terms on which agency is exercised. The possibility that the logic of an algorithm—its training data, its objective function, its architecture—might systematically privilege certain outcomes and occlude others before any human actor has made a conscious choice falls outside this framework's theoretical purview.

Constructivism, with its emphasis on ideational structures and intersubjective norms, offers perhaps the most promising foundation for understanding machine-mediated IR. Wendt's (1999) argument that "anarchy is what states make of it" implies that the normative frameworks through which states understand their interests are socially constructed and historically contingent. By extension, the introduction of AI systems into diplomatic practice is not merely a technical development but a normative one: it is reshaping what counts as evidence, what counts as an acceptable decision process, and what counts as a legitimate outcome in international negotiations. Constructivism has not, however, systematically engaged with the question of whether—and how—algorithmic systems participate in the social construction of international reality. Expanding the constructivist project to account for machine agents is one of the central theoretical tasks this paper identifies.

## 2.2. *Technology as Agent versus Tool*

The distinction between technology as a passive tool and technology as an active agent has gained traction in Science and Technology Studies (STS) and critical data scholarship. Pasquale (2015) articulates the epistemic dimension of this problem through the concept of the "black box society": when consequential decisions are made by algorithms whose inner workings are opaque to those affected, accountability frameworks collapse. Crawford's (2021) *Atlas of AI* extends this critique to the material and geopolitical dimensions of AI infrastructure, demonstrating that the apparent immateriality of digital systems conceals vast extractive relationships—in rare earth minerals, precarious labor, and behavioral data—that recapitulate colonial patterns of accumulation.

For IR theory, the agent/tool distinction has specific and weighty implications. If AI systems are merely tools, then existing frameworks for state responsibility, diplomatic accountability, and norm compliance remain essentially intact: the state that an AI system deploys is accountable for its behavior, in the same way it is accountable for any weapons system it operates. But if AI systems are agents—if their embedded logic, training data biases, and optimization pressures create systematic behavioral tendencies that partially escape the intentions of their deployers—then new theoretical and normative frameworks become necessary. Russell's (2019) works on the "control problem" in AI, which frames the central challenge as aligning AI objectives with human values rather than simply programming AI to follow instructions, supports the latter interpretation. Bostrom (2014) extends this reasoning further, arguing that sufficiently advanced AI systems may develop instrumental strategies that diverge from human intentions in ways that are difficult to predict or contain—a concern that, while formulated in the context of hypothetical superintelligent systems, has practical relevance for the increasingly autonomous AI-assisted decision architectures being deployed in international governance and security contexts today. The AI system that fails to behave as intended

is not merely malfunctioning; it expresses the priorities embedded in its design environment, which may diverge systematically from the interests of those it nominally serves.

### *2.3. Toward a Theory of Algorithmic Diplomacy*

This paper proposes the concept of "algorithmic diplomacy" to describe the emerging domain of international relations in which AI systems play constitutive—not merely instrumental—roles in shaping diplomatic processes and outcomes. While Bjola and Holmes (2015) have explored how digital technologies are reshaping diplomatic practice—examining how social media, digital communications, and information technologies are transforming the conduct of diplomacy, their framework does not extend to the constitutive role of AI in restructuring the diplomatic environment itself. The concept of algorithmic diplomacy, as developed here, addresses this gap by foregrounding the ways in which AI systems actively shape the conditions under which diplomatic engagement occurs.

This framework rests on three core propositions. First, algorithmic systems embedded in international institutions carry political values and power asymmetries that reflect the conditions of their production, primarily within the technologically advanced economies of the Global North. Second, the speed, scale, and opacity of AI-assisted decision-making are qualitatively transforming the risk landscape of international security, particularly in the nuclear domain. Third, disparities in computational capacity between states constitute a new dimension of international power with profound implications for the legitimacy and fairness of multilateral negotiations. Each proposition is developed in the sections that follow.

## **3. Algorithmic Bias in Global Governance: Reinforcing North–South Inequalities**

The integration of artificial intelligence and algorithmic systems into international governance has brought both transformative opportunities and significant challenges. While these technologies promise greater efficiency and enhanced decision-making across organizations such as the United Nations, World Bank, and humanitarian agencies, they also risk perpetuating—and in some cases amplifying—existing global inequalities. In particular, the structural biases encoded in algorithms often reflect the interests, data environments, and priorities of technologically advanced nations, predominantly in the Global North. As these systems increasingly mediate the allocation of resources, the management of crises, and the determination of development trajectories, their impact on North–South relations become a central concern. This section examines how algorithmic bias operates within the frameworks of global governance, highlighting its role in reinforcing historical patterns of exclusion and disadvantage for populations and states in the Global South. By analyzing the architecture, deployment, and consequences of AI-driven decision-making, we aim to illuminate the mechanisms through which data-driven systems contribute to the reproduction of international power asymmetries.

### *3.1. The Architecture of AI in International Organizations*

International organizations have increasingly adopted AI and algorithmic decision-making tools to manage tasks of unprecedented scale and complexity—from monitoring compliance with sanctions regimes to allocating humanitarian assistance in crises affecting many millions of displaced people. The motivations are genuine: the UNHCR manages the consequences of displacement for over 100 million people worldwide (UNHCR, 2023), a caseload that no human bureaucracy could process with the speed and consistency that the humanitarian imperative demands. The World Food Program employs predictive modeling to anticipate food security crises, and the World Bank uses algorithmic tools to assess creditworthiness and development potential in client states.

The political economy of AI development, however, ensures that systems deployed by these organizations are overwhelmingly designed in, and calibrated to the data environments of, the Global

North. As Couldry and Mejias (2019) demonstrate in their analysis of "data colonialism," the extraction of behavioral data from Global South populations to train AI systems that are developed and owned primarily in the North—while returning those systems in forms that serve the assumptions and interests of Northern developers—constitutes a new variant of the colonial relationship. When international organizations deploy AI systems that systematically misrepresent or discount the circumstances of Southern populations, the result is not neutral inefficiency, but structured inequality encoded in technical infrastructure.

### 3.2. Algorithmic Bias in Humanitarian and Development Contexts

The landmark study by Bansak et al. (2018), published in *Science*, demonstrated both the promise and the peril of algorithmic refugee resettlement. Their machine learning model significantly improved employment outcomes for refugees by optimizing matching between individuals and resettlement locations. However, the analysis also revealed that model performance was substantially stronger for refugee populations from countries with larger historical datasets—predominantly countries with longer migration histories to Western nations—than for newer or smaller refugee populations, which are disproportionately from sub-Saharan Africa. This finding illustrates a fundamental structural problem: AI systems learn from historical data, and historical data in international humanitarian response reflects decades of policies shaped by political and racial hierarchies (Benjamin, 2019; Eubanks, 2018).

Buolamwini and Gebru (2018), in their landmark study of commercial facial recognition systems, demonstrated that classification accuracy varied dramatically across demographic groups, with error rates for darker-skinned women exceeding those for lighter-skinned men by over thirty percentage points—a finding that illustrates how algorithmic bias is not incidental but structural, embedded in the training data and design choices that constitute the system. While their study focused on commercial classification technology, the implications extend directly to the international governance domain: when algorithmic systems trained on skewed datasets are deployed in contexts where the stakes involve refugee protection, development assistance, or sanctions enforcement, the consequences of demographic bias are amplified by the power asymmetries inherent in the international system.

An algorithm trained on thirty years of resettlement decisions will replicate, and in many cases amplify, the patterns encoded in that data—including patterns of preference, exclusion, and neglect. Eubanks (2018) has documented this dynamic in domestic U.S. contexts, demonstrating how automated welfare and public safety systems systematically penalize already-marginalized populations who are simultaneously most dependent on public systems and least capable of contesting algorithmic determinations. The same logic applies at the international level: when the UNHCR's resettlement algorithm under-serves sub-Saharan African refugees because their data profile is underrepresented in training sets, the effect is to compound historical disadvantage with contemporary algorithmic neglect.

The problem extends beyond humanitarian contexts into development finance. International financial institutions employ credit-scoring algorithms and economic vulnerability assessments that systematically disadvantage low-income countries. O'Neil (2016) demonstrates that proxy variables commonly used in such models—including infrastructure quality, governance indicators, and historical borrowing behavior—function as mathematical proxies for poverty itself, creating self-fulfilling cycles of disadvantage. When the World Bank or IMF uses such models to determine aid eligibility or conditionality, the political implications are significant: algorithmic outputs acquire an aura of objectivity that insulates politically consequential decisions from democratic challenge, a phenomenon Pasquale (2015) identifies as one of the defining features of the black box society writ large.

### 3.3. The Data Colonialism Thesis and Its Implications for Global Governance

Couldry and Mejias's (2019) data colonialism thesis provides a macro-structural framework for these dynamics. They argue that the global infrastructure of data extraction—through which behavioral, economic, and social data from populations worldwide is harvested to train AI systems primarily owned by corporations and state agencies in the United States and China—constitutes a structurally new iteration of the colonial relationship. Classical colonialism extracted natural resources from periphery to core; data colonialism extracts informational resources from populations in the Global South while concentrating on the productive benefits of AI at the core.

For international governance, this means that the apparent universalism of AI-assisted decision-making in international organizations is structurally compromised. The algorithms deployed are not neutral tools; they are artifacts of political economies that embed specific distributions of voice, evidence, and value. As Acemoglu and Johnson (2023) argue in their sweeping analysis of technological power and inequality, the direction of AI development reflects choices made by a small number of technologically powerful actors—choices that systematically favor applications profitable in wealthy markets over applications that address the needs of the poor globally. When this dynamic operates within international institutions nominally committed to equity and universality, the tension becomes constitutive rather than incidental.

The normative implications are severe. If international organizations are to uphold their foundational mandates of equity, universality, and human dignity, they must develop governance frameworks requiring algorithmic impact assessments, demanding training data transparency, and creating meaningful mechanisms through which affected communities can contest algorithmic determinations. The UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) represents a meaningful step in this direction, as does the OECD Principles on AI (2019). However, as Floridi et al. (2018) observe, translating broad ethical principles into operational governance mechanisms remains deeply contested—particularly across the North–South divide, where the distributional stakes of that translation are highest.

**Table 1.** AI Applications in International Organizations and Documented Bias Concerns.

Organization	Application Domain	AI/Algorithmic Approach	Key Equity/Bias Concern	Policy Response (as of 2024)
UNHCR	Refugee resettlement matching	Machine learning optimization (employment-outcome prediction)	Underperformance for underrepresented nationalities; historical data gaps for sub-Saharan African and smaller refugee populations	UNHCR Protection Data Strategy (2021); calls for mandatory algorithmic impact assessment
World Food Program (WFP)	Food security early warning (FSIN model)	Satellite data fusion and predictive analytics	Sparse ground-truth data in remote/fragile regions biases predictions toward historically monitored areas	FSIN data-gap mapping initiatives; WFP Innovation Hub AI Ethics Guidelines
World Bank	Development lending and country risk assessment	Algorithmic credit/governance scoring	Infrastructure deficit proxies reinforce colonial-era inequalities; historical underfunding	World Bank Data Governance Framework (2021);

Organization	Application Domain	AI/Algorithmic Approach	Key Equity/Bias Concern	Policy Response (as of 2024)
			creates circular self-fulfilling disadvantage	ongoing Quota and Voice Reviews
UN Security Council	Targeted sanctions screening	Automated financial transactions and entity screening	Disproportionate false-positive flagging of Global South financial institutions and individuals; asymmetric error distribution	ISIL/AI-Qaida Sanctions Ombudsperson mechanism; academic calls for algorithmic audit
IMF	Economic vulnerability and crisis prediction	Large-scale macroeconomic simulation models	Systematic underestimation of crisis risk in developing economies; limited data quality from low-income country submissions skews model calibration	IMF Data Standards Initiative; quota reform; Capacity Development program
IOM	Migration flow prediction and humanitarian logistics	Big data analytics (social media, mobile data harvesting)	Surveillance concerns: data sourced from migrants without meaningful consent; Northern-coded categorizations distort mobility patterns	IOM Data Protection Manual; growing GDPR extraterritorial pressure

Sources: Bansak et al. (2018); Benjamin (2019); Buolamwini & Gebru (2018); Couldry & Mejias (2019); Eubanks (2018); Floridi et al. (2018); O'Neil (2016); OECD (2019); UNESCO (2021); UNHCR (2023).

#### 4. Automated Deterrence: AI, Nuclear Command and Control, and the Transformation of MAD

The intersection of artificial intelligence (AI) and nuclear command and control systems mark a profound shift in the logic and practice of deterrence. Historically, the doctrine of Mutually Assured Destruction (MAD) relied on a delicate balance: the assurance that any nuclear aggression would be met with catastrophic retaliation, fostering a kind of uneasy stability between rival states. This framework was predicated on human judgment, deliberate decision-making, and the capacity for diplomatic intervention during crises. However, the accelerating integration of AI technologies into nuclear architectures introduces new dynamics that challenge the foundational assumptions of MAD. Automated systems promise rapid threat detection and response, but also bring risks associated with speed, opacity, and adversarial vulnerability—factors that may undermine the very stability MAD was designed to protect. In this section, we examine how AI-driven automation transforms the deterrence landscape, exploring both its strategic attractions and its destabilizing consequences for global security.

##### 4.1. The Logic of MAD in the Pre-AI Era

Mutually Assured Destruction (MAD) represented one of the twentieth century's most counterintuitive strategic innovations: the deliberate cultivation of mutual vulnerability as a



guarantee of stability. Brodie (1959) articulated the foundational logic—that a potential aggressor would be deterred from launching a first strike if it could be confident that a devastating retaliatory second strike was guaranteed to follow. This logic relied on key assumptions: that decisions allowed for human review, both sides could interpret signals correctly, and rational actors would opt for de-escalation over nuclear conflict (Sagan & Waltz, 2003).

What the deterrence literature frequently obscures is how precarious this stability was even before the advent of AI. Schlosser (2013), in his comprehensive history of U.S. nuclear command and control, documents numerous episodes in which technical malfunctions, misidentified radar signatures, and communication failures brought the world within minutes of accidental nuclear war. The 1983 Petrov incident—in which a Soviet early warning officer correctly identified a satellite malfunction as a false alarm, choosing to override the system's indication of an incoming American strike—is only the most celebrated of many such near-misses (Lewis et al., 2014). Stability under MAD, it emerges, depended not merely on rational deterrence calculations but on the contextual judgment of fallible human operators working in degraded, ambiguous information environments. It is precisely this human buffer that AI integration now threatens to remove.

#### 4.2. AI Integration in Nuclear Command and Control Architectures

All three of the world's largest nuclear powers are pursuing the integration of AI-assisted systems into their nuclear C3I architectures, though the precise character and extent of this integration remain classified in each case. The strategic motivations are transparent: AI offers faster threat detection, more rapid fusion of sensor data from multiple sources, and the potential to identify adversary launch preparations before they become visible to conventional intelligence platforms (Horowitz, 2018; Kania, 2017). For states confronting the prospect of a decapitating first strike designed to eliminate their command infrastructure before human orders can be issued, delegating authority to faster automated response systems represents an apparently rational adaptation.

The strategic implications, however, are deeply destabilizing. Acton (2018) has argued that the increasing integration of conventional and nuclear command systems—what he calls "entanglement"—creates new pathways to inadvertent escalation. When AI-assisted systems designed for conventional operations run on the same infrastructure as nuclear early-warning systems, a conventional cyber intrusion targeting that infrastructure could be read by automated systems as a precursor to nuclear decapitation, potentially triggering an automated nuclear response before any human decision-maker has been alerted. Geist and Lohn (2018), in their RAND Corporation analysis, identify several specific pathways through which AI increases nuclear risk, including the prospect of pattern-matching algorithms incorrectly classifying a novel crisis scenario as analogous to a historical pattern of pre-launch preparation—a danger that grows as AI systems are increasingly trained on historical datasets that may not generalize to genuinely unprecedented geopolitical configurations.

Cummings (2017), in her Chatham House analysis, further emphasizes that the increasing speed and autonomy of AI-assisted military systems may outpace the ability of human operators to exercise meaningful oversight, particularly in high-stakes environments where the margin for error is vanishingly small. Her findings underscore that the challenge is not merely one of technical reliability but of systemic design: as human roles in the decision chain are compressed or eliminated, the capacity for contextual judgment—the very faculty that averted catastrophe in historical near-miss incidents—is progressively degraded.

Scharre (2018) extends this analysis to the accountability problem inherent in autonomous military systems. When an autonomous system makes a decision that escalates a crisis, there is no clear chain of human responsibility. For nuclear systems, this accountability gap is not merely an ethical problem but a strategic one: the deterrence framework depends fundamentally on the credible communication of intent between adversaries, and when intent is generated or interpreted by opaque algorithmic systems, the communication channel is epistemically corrupted. The adversary cannot know whether a threatening signal represents the considered judgment of political leadership or an algorithmic artifact, a distinction that is foundational to the functioning of deterrence.

### 4.3. New Instabilities: Speed, Opacity, and Adversarial Vulnerability

Three specific characteristics of AI systems generate novel instabilities in the nuclear deterrence framework that have no direct Cold War precedent. The first is speed. Cold War deterrence was supported by backchannel communication mechanisms—most famously the Moscow–Washington hotline—that allowed decision-makers to clarify intent during a crisis, providing a margin of time in which human judgment could override preliminary escalatory impulses. When AI systems process sensor data and generate response recommendations in milliseconds, this window for diplomatic intervention effectively disappears. Altmann and Sauer (2017) describe this as the stability–instability paradox of autonomous systems: the same speed that makes AI attractive for defensive early warning makes it catastrophically dangerous for crisis management, because the system cannot be paused while diplomats confer.

The second destabilizing characteristic is opacity. The "black box" problem that Pasquale (2015) identified in civilian algorithmic governance acquires catastrophic proportions in the nuclear domain. If neither the deploying state's political leadership nor adversary decision-makers can understand the decision logic of an AI early-warning or response system, the result is a form of strategic blindness in which neither party can accurately interpret the other's signals. Kissinger, Schmidt, and Huttenlocher (2021) capture this concern by observing that AI systems can "achieve goals by means that humans do not fully understand," and that this epistemic uncertainty is particularly dangerous in domains where a single misinterpretation can be irreversible. Classical deterrence theorists assumed that adversaries could model each other's decision processes with sufficient accuracy to calculate credible threats; AI systems undermine this assumption at its foundations.

The third destabilizing characteristic is adversarial vulnerability. AI systems, unlike human decision-makers, can be deliberately deceived through the manipulation of their input data. A sophisticated adversary could, in principle, inject false data into an AI early-warning system through cyber intrusion, deceptive sensor signals, or the use of "adversarial examples" carefully engineered inputs designed to fool machine learning classifiers into producing specific erroneous outputs (Brundage et al., 2018). Jervis's (1978) security dilemma, in which defensive preparations are misread as offensive intent, acquires a new and more dangerous dimension when those "readings" are performed by algorithms that can be deliberately manipulated. The prospect of an adversarial AI attack designed to simulate a nuclear launch—thereby triggering an automated counter-launch—represents a nightmare scenario with no Cold War analog and no existing arms-control framework capable of addressing it.

**Table 2.** The Evolution of Nuclear Deterrence in the Age of AI: Comparative Dimensions.

Dimension	Cold War MAD Framework	AI-Assisted Command and Control	Strategic Stability Implication
Decision timeline	Minutes to hours; deliberate human review at each stage	Potentially milliseconds to seconds; AI processing precedes human notification	Crisis management windows eliminated; backchannel diplomacy rendered operationally irrelevant
Human judgment	Central and legally required at all stages of escalation	Reduced to advisory or confirmatory role in fast-response architectures	Loss of contextual reasoning; human override may be technically impossible in high-tempo scenarios

Dimension	Cold War MAD Framework	AI-Assisted Command and Control	Strategic Stability Implication
Signal processing and interpretation	Radar, satellite, and HUMINT, with human interpretation	AI-automated sensor fusion; pattern-matching against historical datasets	Novel crisis configurations may be incorrectly classified; training data biases can generate false alerts
Adversarial vulnerability	Physical destruction or jamming of physical infrastructure	Cyber intrusion, adversarial examples, data poisoning of training sets	Algorithmic systems can be deliberately manipulated to misclassify benign signals as hostile
Accountability structure	Clear chain of political and military command responsibility	Diffuse; deploying state retains legal responsibility, but mechanism of error is opaque and contested	Deterrence communication corrupted when intent is produced by uninterpretable systems
Crisis communication channels	Dedicated backchannel (e.g., Moscow–Washington Direct Communications Link)	AI systems may process diplomatic signals as data inputs to threat models, stripping contextual meaning	Diplomatic signals may be misread or filtered out by algorithmic processing before reaching decision-makers
Conventional–nuclear entanglement risk	Conventional and nuclear C2 systems largely architecturally separate	Increasing integration of conventional AI-assisted C4ISR with nuclear C3 infrastructure	Conventional cyber or EW attacks on shared infrastructure may inadvertently trigger nuclear early-warning responses
Arms control verification	Human-verifiable physical inspections (INF Treaty, START)	AI-generated assessments may be unverifiable even by the deploying state	Compliance verification becomes profoundly more difficult; existing treaty architecture inadequate

Sources: Acton (2018); Altmann & Sauer (2017); Brodie (1959); Brundage et al. (2018); Cummings (2017); Geist & Lohn (2018); Horowitz (2018); Jervis (1978); Kissinger et al. (2021); Lewis et al. (2014); Sagan & Waltz (2003); Scharre (2018); Schlosser (2013).

## 5. The "Black Box" of Negotiation: Computational Power Asymmetries in Multilateral Diplomacy

In the evolving landscape of international diplomacy, the rapid integration of artificial intelligence and advanced computational technologies has fundamentally reshaped the dynamics of multilateral negotiation. While traditional diplomatic processes have relied on personal expertise, strategic information management, and institutional memory, today's negotiations increasingly unfold within environments dominated by data analytics, machine learning, and predictive modeling. These developments have created profound asymmetries in computational power

between states, with technologically advanced actors wielding tools that enable them to aggregate, analyze, and simulate vast quantities of information. As a result, the negotiation process itself has become a "black box" opaque, complex, and often inaccessible to parties lacking comparable digital infrastructure. This section explores how computational power disparities influence diplomatic outcomes, alter bargaining leverage, and challenge the foundational principles of transparency and equity in international negotiations.

### *5.1. AI as a Structuring Force in International Negotiation*

Diplomatic negotiation has always involved information asymmetry. The skilled negotiator conceals her reservation price while probing her counterpart's, and her advantage derives precisely from the opacity of her true position. Fearon (1995), in his rationalist theory of war, argued that armed conflicts arise because states cannot credibly commit to revealing their private information about costs and resolve; the tragic corollary is that information asymmetry makes war possible. The introduction of AI modeling into multilateral negotiation fundamentally alters this dynamic: computational tools capable of aggregating vast datasets, modeling adversary preferences, simulating thousands of negotiation scenarios, and predicting domestic political constraints allow technologically superior parties to effectively "see through" the opacity that weaker counterparts depend upon for their bargaining leverage.

This is not a theoretical conjecture. Advanced economies have deployed sophisticated data analytics and econometric modeling in trade negotiations for decades. What is new in the current era is the application of machine learning to this problem: the capacity to train models on vast archives of prior negotiating positions, parliamentary debates, legislative histories, economic data, and even diplomatic communications—where these are accessible through signals intelligence—in order to construct probabilistic models of an adversary's true reservation price in a given negotiation. Kissinger, Schmidt, and Huttenlocher (2021) identify this capability as representing a qualitative shift in the intelligence environment of diplomacy, one that reduces the strategic privacy on which weaker parties depend and that has no equivalent in traditional diplomatic practice.

### *5.2. Trade Negotiations and the Algorithmic Advantage*

The consequences of computational asymmetry in trade negotiations are visible in the structure and outcomes of recent major agreements. The United States–Mexico–Canada Agreement (USMCA), the negotiations around the Trans-Pacific Partnership framework (TPP/CPTPP), and ongoing deliberations concerning digital trade governance within the World Trade Organization (WTO) all involve parties with radically different analytical capabilities. While the United States and the European Union can deploy teams of economists armed with computable general equilibrium models, satellite-derived economic data, and behavioral predictive analytics, many smaller trading partners—particularly from the Global South—must conduct negotiations with substantially more limited resources.

Lee (2018), in his analysis of the AI rivalry between the United States and China, argues that access to large and diverse datasets constitutes the primary source of competitive AI advantage, enabling the construction of more accurate models of economic and social behavior. In the context of trade negotiations, this means that the party with superior AI modeling capacity can simulate not only the economic effects of proposed tariff schedules but also the domestic political constraints, coalition vulnerabilities, and behavioral dynamics of the opposing delegation. The party being modeled is, in effect, more transparent to its counterpart than to itself—a radical inversion of traditional bargaining dynamics.

Allison's (2017) Thucydidean framing of the U.S.–China rivalry is instructive here: the competition for technological supremacy between these powers is not only a military competition but a contest for the ability to structure the information environment within which international negotiations occur. When China deploys AI-assisted models in Belt and Road Initiative debt negotiations with developing country partners—with access to comprehensive economic and

political data harvested from those partners' own infrastructure—or when the United States uses advanced econometric modeling to predict the positions of WTO dispute settlement panels, the ostensibly level playing field of international law becomes tilted by invisible computational infrastructure. Farrell and Newman (2023), writing on "weaponized interdependence," provide a complementary analytical frame: technologically dominant states leverage their structural position within global information networks not merely for intelligence gathering but as a direct instrument of coercive influence, using the threat of algorithmic exclusion—from financial messaging systems, credit rating algorithms, and supply chain visibility platforms—as a form of economic statecraft.

### 5.3. *Climate Diplomacy and the Computational Divide*

The global climate negotiations conducted under the United Nations Framework Convention on Climate Change (UNFCCC) illustrate the computational power divide in multilateral diplomacy with particular clarity. Climate negotiations involve extraordinary scientific, economic, and political complexity, and parties capable of modeling these trade-offs with greater precision possess a systematic advantage in framing the terms of agreement. The European Union's Copernicus Climate Change Service, the United States' suite of integrated assessment models (IAMs), and China's rapidly expanding national climate modeling capacity stand in sharp contrast to the analytical resources available to Small Island Developing States (SIDS), Least Developed Countries (LDCs), and most African Union member states.

Rolnick et al. (2022) document the expanding applications of machine learning across the climate domain, from satellite-based emissions monitoring and deforestation detection to energy system optimization and high-resolution climate impact projection. What their analysis makes clear is that the most powerful AI applications in climate science—those that could provide accurate, locally specific projections of climate damages, which are critical for negotiating specific loss-and-damage provisions and adaptation finance allocations—require training data and computational infrastructure that is overwhelmingly concentrated in the Global North. This creates a negotiating paradox of profound political significance: the most climate-vulnerable nations depend on the ability to document and project climate damage to their territories, but the AI tools required to do so with the precision that wealthy-country negotiators will recognize as credible are precisely inaccessible to them.

When a Pacific atoll nation seeks specific loss-and-damage compensation at COP30, it does so in a system in which the computational models that quantify its losses are owned and operated by the very parties from whom compensation is sought. The constructivist insight that "intersubjective knowledge" shapes international outcomes (Wendt, 1999) apply with full force: when the computational tools that generate shared understanding of climate risk are controlled by one coalition of parties, the "shared" knowledge that structures negotiations is structurally partial. Zegart (2022), analyzing the intelligence revolution enabled by AI and big data in the security context, argues that computational intelligence superiority—the capacity to process and act on information more comprehensively and rapidly than one's counterpart—represents a new and undertheorized dimension of international power. Applied to the negotiating context, her analysis suggests that the G7's aggregate computational dominance constitutes a structural power asymmetry in all major multilateral forums, one that is largely invisible and therefore exceptionally difficult to contest through conventional diplomatic or legal channels.

**Table 3.** Computational Power Asymmetries in Multilateral Negotiations: Selected Actors and Contexts (2018–2024).

Actor/Party	AI/Data Infrastructure Level	Key Capability	Primary Negotiation Context(s)	Core Equity/Strategic Concern
United States	Very High: USTR econometric infrastructure; NSA signals intelligence; government-academic AI partnerships; private sector collaboration	Predictive trade preference modeling; adversary reservation-price mapping; real-time behavioral analytics of counterpart delegations	WTO dispute resolution; USMCA; IPEF; G7; climate finance; sanctions design	Structural advantage enables anticipation of counterpart's true bottom line; computational capacity used to structure negotiating agenda before talks begin
China	Very High: State-integrated AI ecosystem; comprehensive domestic social and economic data infrastructure; BRI-derived partner-state data; civil-military fusion	Large-scale financial modeling; debtor-state vulnerability analysis; political stability modeling of BRI partners; long-horizon scenario simulation	BRI debt restructuring; WTO; RCEP; SCO frameworks; climate negotiation positioning	Data harvested from BRI partner infrastructure is subsequently used to model those partners' negotiating behavior, constituting systematic informational exploitation
European Union	High: Copernicus satellite system; Eurostat; Joint Research Centre IAMs; ECMWF climate modeling	Strong climate impact assessment; trade-flow simulation; regulatory standard-setting capacity amplified by Brussels Effect	Paris Agreement/COP; EU–Mercosur; WTO digital trade; loss-and-damage finance	EU regulatory standard-setting via AI Act may inadvertently disadvantage non-EU trading partners unable to meet compliance requirements
India	Medium: ISRO satellite capacity; Aadhaar data infrastructure; growing IIT research capacity	Moderate trade modeling; strong pharmaceutical and IP data analytics; limited integrated climate modeling	TRIPS pharmaceutical access; climate finance and adaptation; BRICS economic frameworks; WTO dispute settlement	Data sovereignty assertiveness increasing; significant gaps in multilateral climate and financial modeling constrain independent negotiating position

Actor/Party	AI/Data Infrastructure Level	Key Capability	Primary Negotiation Context(s)	Core Equity/Strategic Concern
African Union / African States	Low to Medium: Fragmented national infrastructure; significant data sovereignty gaps; reliance on external platforms	Largely dependent on World Bank, IMF, and NGO-provided models; emerging regional capacity through AU Data Policy Framework	AfCFTA; UNFCCC loss and damage; WTO TRIPS; debt restructuring (Paris Club/G20)	Near-complete dependence on creditor- and donor-provided models for key negotiations; highest climate vulnerability combined with lowest analytical capacity
Small Island Developing States (SIDS)	Very Low: Minimal national AI/data infrastructure	Dependent on IPCC synthesis reports, NGO advocacy data, and UN system tools; limited capacity for independent quantitative analysis	UNFCCC 1.5°C threshold negotiations; Warsaw International Mechanism; Loss & Damage Fund	Greatest exposure to climate risk; least computational power to independently quantify and assert damages; structurally and asymmetrically disadvantaged in loss-and-damage negotiations
G7 Aggregate	Very High: Collective modeling infrastructure vastly exceeds all developing country coalitions combined	Near-comprehensive coverage of multilateral negotiation domains; capacity for real-time scenario updating and position adjustment during active negotiations	All major multilateral fora	Systemic concentration of computational power within G7 constitutes a structural power asymmetry that reinforces existing hierarchies across the international system

Sources: Allison (2017); Farrell & Newman (2023); G7 (2023); Kissinger et al. (2021); Lee (2018); Rolnick et al. (2022); UNESCO (2021); Wendt (1999); Zegart (2022)

#### 5.4. The Erosion of Transparency and the Challenge to Equity

The cumulative effect of computational power asymmetries is a profound erosion of transparency and equity at the heart of multilateral negotiations. When the informational environment is shaped by actors with the resources to develop and deploy advanced AI systems, the negotiation process itself becomes increasingly opaque to less technologically advanced parties. These actors face the double bind of not only lacking the tools to independently verify the claims and projections presented by their counterparts but also of being systematically disadvantaged in the

construction of negotiating agendas, the framing of key terms, and the setting of technical benchmarks. The black box nature of algorithmic mediation means that decisions and outcomes are often justified by reference to models or simulations that are inaccessible, unexplainable, or unverifiable for those outside the computational elite.

These dynamic risks entrenching a feedback loop in which existing disparities in digital infrastructure and analytical capacity are reinforced by the very processes designed to address global challenges. For example, the allocation of climate adaptation finance or the design of digital trade rules may be disproportionately shaped by the technical priorities and data resources of the most powerful negotiating parties, leaving the concerns and lived realities of less-resourced states marginalized. The result is not only a practical inequity in outcomes but also a legitimacy deficit for international institutions that claim to operate based on sovereign equality and fairness.

### *5.5. Prospects for Redress: Building Computational Equity*

Addressing these challenges requires a concerted effort to democratize access to data, modeling tools, and AI infrastructure. Initiatives such as open-source climate models, shared data repositories, and international partnerships for technical assistance can help reduce the gap between computational haves and have-nots. Multilateral organizations—including the United Nations, the World Bank, and the World Meteorological Organization—have begun to recognize the importance of technical capacity-building as a core component of equitable diplomacy. However, these efforts remain nascent and are often limited by resource constraints, intellectual property barriers, and the strategic interests of dominant actors.

More fundamentally, the challenge is not only technical but also political and normative. The international system must develop new mechanisms for algorithmic transparency, model explainability, and the meaningful participation of all stakeholders in the construction and validation of computational tools used in negotiation. Without such reforms, the risk is that multilateral diplomacy will increasingly operate within a "black box" paradigm—one that privileges the technologically advanced, marginalizes the vulnerable, and undermines the foundational principles of transparency, equity, and sovereign equality in international relations.

## **6. Toward a Framework for Algorithmic Diplomacy**

The accelerating integration of artificial intelligence (AI) systems into the fabric of international relations is transforming both the substance and the structure of global diplomacy. As states and international organizations increasingly rely on algorithmic tools to inform decision-making, negotiate outcomes, and allocate resources, new forms of complexity and asymmetry are emerging. These developments challenge traditional assumptions about institutional design, strategic stability, and bargaining dynamics, demanding a reconceptualization of how diplomacy is conducted in an era of pervasive machine mediation. This section introduces a framework for algorithmic diplomacy, examining how AI systems are not merely instrumental technologies but constitutive actors with the power to shape negotiation environments, redefine interests, and recalibrate the distribution of influence across the international system. By exploring conceptual foundations, governance responses, and normative challenges associated with algorithmic mediation, we seek to illuminate the urgent need for new theoretical and practical approaches to global governance in the age of AI.

### *6.1. Machine-Mediated IR: Conceptual Foundations*

The three preceding sections have documented empirically distinct but theoretically interconnected phenomena: the structural biasing of global governance through AI systems, the destabilization of nuclear deterrence through algorithmic command and control, and the emergence of new forms of bargaining asymmetry through computational power disparities. Each phenomenon is, in isolation, amenable to treatment within existing IR frameworks—as a problem of institutional design, strategic stability, or rationalist bargaining theory. Taken together, however, they suggest the



need for a more integrated theoretical framework—one that takes seriously the constitutive role of AI systems in structuring the environment within which international politics unfolds.

The concept of machine-mediated international relations, as developed in this paper, captures this novelty. By this term, we mean a condition in which AI systems play not merely instrumental but constitutive roles in international politics—determining what information is processed, how interests are formulated, what options appear visible, and whose voices are legible to international institutions. This condition is distinct from earlier forms of technologically mediated diplomacy—the telegraph, telephone, satellite communications—because the mediation is not neutral. Unlike a telephone, which transmits human communication without substantially transforming it, an AI system transforms information according to embedded logics, training data, and optimization objectives that reflect the political economy of its production. The telephone did not choose what to transmit, to whom, or according to what priorities; the algorithm does, and in doing so, it acts.

Several conceptual building blocks of a machine-mediated IR framework can be identified. The first draws on Fearon's (1995) rationalist tradition: algorithmic capability constitutes a form of "private information" in international negotiations—one that creates systematic advantage for its possessors while undermining the transparency assumptions that underlie cooperative international institutions. The second draws on constructivism (Wendt, 1999): AI systems actively participate in the social construction of international "reality." When international organizations use algorithmic models to define what counts as a refugee, a food security emergency, or a climate-vulnerable nation, those definitional choices are politically consequential acts that shape international norms and resource allocation. The algorithm is, in this sense, a norm entrepreneur operating at enormous scale. The third building block draws on critical IR and political economy (Coudry & Mejias, 2019; Crawford, 2021; Acemoglu & Johnson, 2023): the design, ownership, and governance of AI systems must be foregrounded as central questions of international political economy, not merely as technical matters best left to engineers and their national regulatory bodies.

## 6.2. Governance Responses: Progress and Critical Limitations

The international community has not been passive in the face of these challenges. The OECD Principles on AI (2019) established an early multilateral baseline for responsible AI governance, emphasizing transparency, accountability, and the protection of human rights and democratic values. The UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) extended this framework to explicitly address the global development dimensions of AI ethics, recognizing the vulnerabilities of low- and middle-income countries and calling for capacity-building as a component of responsible AI deployment. The EU AI Act (European Parliament & Council of the European Union, 2024), while primarily a domestic regulatory instrument, establishes high-risk AI system standards with significant extraterritorial reach—through the mechanism of the Brussels Effect—that may progressively shape AI deployment in international governance contexts. At the level of summit diplomacy, the Bletchley Declaration (AI Safety Summit, 2023) and the G7 Hiroshima AI Process (G7, 2023) have begun to construct nascent governance architecture, though one that remains dominated by technologically advanced economies and that has yet to produce binding international commitments. These are meaningful developments, but they remain insufficient to the scale and urgency of the challenges documented in this paper. Dafoe's (2018) AI governance research agenda—perhaps the most comprehensive academic framework published to date—identifies three fundamental obstacles to effective multilateral AI governance: coordination problems among states with competing interests, verification challenges that make compliance with any governance framework extremely difficult to confirm, and the difficulty of constructing shared norms across actors with radically different political cultures, economic systems, and security priorities. All three obstacles are visible in the current governance landscape. In the nuclear domain specifically, the absence of any binding international instrument governing the use of AI in nuclear command and control represents a governance vacuum of potentially catastrophic proportions, one that existing

arms control frameworks—designed for a world of human operators and verifiable hardware—are structurally ill-equipped to fill.

Brundage et al. (2018), in their comprehensive analysis of the malicious use of artificial intelligence, identify a range of AI-enabled risks—from automated disinformation campaigns to adversarial manipulation of military systems—that require coordinated international responses. Yet the asymmetric distribution of AI capabilities creates a structural impediment to cooperation: the parties most capable of engineering effective governance frameworks are simultaneously the parties with the greatest material interest in preserving their computational advantages. This dynamic mirrors the broader challenge of great-power coordination in international institutions, but with the additional complication that the power asymmetry in question is largely invisible, technically opaque, and rapidly evolving.

### *6.3. Normative Challenges and the Agenda for Reform*

The normative challenges of algorithmic diplomacy are as fundamental as the technical ones. The legitimacy of international governance rests on some version of the principle of sovereign equality; the idea that states, regardless of their material capabilities, have formally equal standing in international law and institutions. The introduction of AI-assisted decision-making into the core of global governance institutions threatens this principle by creating a new form of structural inequality that is largely invisible, technically complex, and therefore exceptionally difficult to contest through conventional diplomatic or legal channels. When the outcome of a climate negotiation is partially determined by which delegation has access to AI models that can predict the other's bottom line, the formal equality of the negotiating parties is undermined by technological asymmetry that leaves no visible trace in the negotiating record.

What is required is a comprehensive agenda for algorithmic equity in international relations, structured around at least four essential components. The first is mandatory algorithmic impact assessment for AI systems deployed in international governance contexts—analogue to environmental impact assessments—with specific attention to differential effects on Global South populations, and with results that are publicly available for scrutiny by affected parties. The second is substantive and adequately resourced capacity building: genuine transfer of computational resources, training data, technical expertise, and governance capacity to low-income countries, as a precondition for meaningful participation in AI-mediated multilateral negotiations rather than a peripheral add-on. The third is transparency and explainability requirements for AI systems informing consequential international decisions, ensuring that parties whose interests are affected can meaningfully understand and contest algorithmic determinations. The fourth—and perhaps most urgently—is the establishment of a dedicated multilateral instrument for nuclear AI risk reduction, drawing on but substantially extending existing arms control frameworks, to address the specific destabilization risks identified in the preceding section.

Suleyman and Bhaskar (2023), writing about the broader political economy of the AI revolution, argue that the concentration of transformative AI capability in the hands of a small number of technologically dominant states and corporations represents one of the defining power shifts of the twenty-first century. Managing this shift in ways consistent with international stability, global equity, and the rule of law is not merely a technical governance challenge; it is one of the central diplomatic and political imperatives of the coming decades.

### *6.4. Operationalizing Algorithmic Diplomacy: Pathways and Obstacles*

Translating the agenda for algorithmic equity into operational reality requires both institutional innovation and sustained political will.

First, international organizations must develop new mechanisms to monitor, audit, and assess the deployment of AI systems within their own decision-making processes. This involves establishing independent oversight bodies with the authority to evaluate the fairness, accuracy, and social impact of algorithmic tools used in negotiations, resource allocation, and crisis response. Such

bodies should be empowered to recommend corrective actions and ensure compliance with agreed transparency and equity standards.

Second, cross-sectoral partnerships between governments, academia, civil society, and the private sector are essential to pool technical expertise and foster innovation in algorithmic governance. These partnerships can help develop open-source AI tools tailored to the needs of less-resourced states, facilitate knowledge transfer, and promote inclusive participation in the design and validation of computational models that inform international decision-making. The engagement of diverse stakeholders is critical to ensure that algorithmic diplomacy reflects a plurality of perspectives and values, rather than reinforcing the interests of the computational elite.

Third, capacity-building efforts must be embedded within broader frameworks for sustainable development and digital inclusion. This means integrating AI literacy, data governance, and technical infrastructure support into existing international development programs, with dedicated funding streams and measurable outcomes. Special attention should be paid to empowering local communities and marginalized groups to participate meaningfully in algorithmic governance, both at the national and international levels.

Finally, the international community must confront the challenge of norm diffusion and cultural adaptation. The ethical standards and regulatory principles governing algorithmic diplomacy should be developed through inclusive, iterative processes that respect local contexts and traditions. This includes recognizing the diversity of legal frameworks, governance models, and social values that shape how different societies approach AI and digital technologies. Only through genuine dialogue and mutual learning can the legitimacy and effectiveness of algorithmic governance be secured.

### *6.5. The Future of Algorithmic Diplomacy*

As AI systems continue to evolve and permeate the structures of international relations, the stakes for algorithmic diplomacy will only grow. The risk is that, without proactive reform, technological asymmetries will deepen existing divides and exacerbate tensions between states and peoples. However, with visionary leadership and collaborative effort, it is possible to harness the transformative potential of AI for the common good—strengthening transparency, fostering equity, and renewing the legitimacy of global governance institutions. The challenge is formidable, but the opportunity is historic: to build a system of international relations in which machine mediation serves, rather than subverts, the principles of justice, inclusion, and peace.

## **7. Discussion**

The findings presented in the preceding sections converge on a central proposition: that the integration of artificial intelligence into the structures and processes of international relations is not a peripheral technical development but a transformative force that reconfigures the distribution of power, the construction of knowledge, and the calculus of risk across the international system. This discussion section synthesizes the paper's three empirical strands, situates them within broader debates in International Relations theory and AI governance scholarship, addresses limitations of the analysis, and identifies directions for future research.

### *7.1. Synthesis of Findings Across Domains*

The three empirical domains examined in this paper—algorithmic bias in global governance, AI in nuclear command and control, and computational power asymmetries in multilateral negotiations—appear at first glance to occupy distinct analytical spaces, corresponding respectively to institutional design, strategic stability, and bargaining theory. However, the analysis reveals a deeper structural unity. In each domain, the introduction of AI systems transforms the informational and decisional environment in ways that exceed the capacity of existing theoretical frameworks to capture. In humanitarian governance, algorithms trained on historically biased datasets reproduce and amplify patterns of exclusion that classical institutional theory treats as amenable to procedural

reform; but the opacity and technical complexity of algorithmic bias resist the transparency mechanisms on which such reform depends. In nuclear command and control, AI-assisted systems compress decision timelines and introduce adversarial vulnerabilities that classical deterrence theory—premised on deliberate human calculation and credible communication of intent—cannot accommodate. In multilateral negotiations, computational power disparities create a form of structural advantage that operates beneath the threshold of visibility available to conventional diplomatic or legal challenges, undermining the principle of sovereign equality that legitimates the multilateral order.

What unites these phenomena is the constitutive role of AI systems in shaping the environments within which international politics occurs. The concept of machine-mediated international relations, as proposed in this paper, captures this unifying dimension. It signals a condition in which the architecture of algorithmic systems—their training data, optimization objectives, design assumptions, and ownership structures—becomes a primary determinant of political outcomes, operating alongside but increasingly independent of the human intentions that classical IR theory treats as the fundamental unit of analysis. This represents a significant departure from the instrumentalist treatment of technology that has characterized realist, liberal, and constructivist scholarship to date.

### *7.2. Theoretical Implications*

The theoretical implications of these findings are substantial. For realism, the paper suggests that the distribution of capabilities among states—Waltz's (1979) central structural variable—must be expanded to include computational capability as a distinct and increasingly consequential dimension of power. The capacity to develop, deploy, and control advanced AI systems is not reducible to traditional measures of material power such as military expenditure or industrial output; it encompasses data infrastructure, algorithmic expertise, and the institutional capacity to integrate AI into governance and security architectures. Mearsheimer's (2001) framework of great-power competition must similarly be updated to account for the ways in which AI-driven information advantages reshape the competitive landscape, not only in military domains but across the full spectrum of international interaction.

For liberalism, the paper challenges the assumption that technological connectivity inherently fosters cooperation and mutual benefit. Keohane and Nye's (2001) framework of complex interdependence recognized that technology could create new forms of vulnerability, but it did not anticipate a condition in which the tools of interdependence themselves—the algorithms that process trade data, allocate humanitarian resources, and model climate risk—could become instruments of structural domination. The liberal expectation that institutions can moderate power asymmetries through rules and norms is complicated by the finding that algorithmic systems can embed power asymmetries within the technical infrastructure of institutions themselves, rendering them invisible to conventional accountability mechanisms.

For constructivism, the findings confirm and extend the tradition's core insight that the structures of international politics are socially constructed, while demonstrating that the agents of construction now include non-human algorithmic systems. When an AI model deployed by an international organization defines what counts as a refugee, a climate-vulnerable state, or a creditworthy borrower, it participates in the construction of international reality with a scope and speed that no individual human norm entrepreneur could achieve. This extension of the constructivist project to account for algorithmic agency is, as noted in the theoretical framework, one of the paper's central contributions.

The analysis also engages with and extends emerging scholarships at the intersection of AI and international security. Johnson (2019) has argued that artificial intelligence will fundamentally alter the character of warfare and strategic competition, introducing new dynamics of speed, complexity, and unpredictability that challenge established theories of deterrence and escalation management. The findings of this paper corroborate Johnson's analysis and extend it beyond the military domain

to encompass the full range of diplomatic and governance interactions in which AI systems are now embedded. Similarly, Taddeo and Floridi (2018) have warned that the absence of effective international regulation of AI in the military domain risks triggering a new arms race conducted not in nuclear warheads but in algorithmic capabilities—a concern that the nuclear command and control analysis in Section 4 substantiates with specific reference to the destabilization of MAD.

Maas (2019), drawing on lessons from nuclear arms control, has argued that international governance of military AI is viable but requires adapting verification and compliance mechanisms to the distinctive characteristics of software-based systems. The analysis presented here supports Maas's call for institutional innovation while highlighting the additional challenge that AI governance must address not only military applications, but the broader spectrum of algorithmic systems embedded in international governance, trade, and climate institutions. The governance challenge, in other words, is not confined to the security domain but pervades the entire architecture of multilateral diplomacy.

### *7.3. Connecting AI to Sustainable Development and Global Equity*

The findings on algorithmic bias and computational power asymmetries connect directly to broader debates about AI and global development. Vinuesa et al. (2020) have demonstrated that AI has the potential to both advance and undermine progress toward the United Nations Sustainable Development Goals (SDGs), depending on how it is developed, governed, and distributed. Their analysis reveals that AI could serve as an enabler for most SDG targets, including those related to poverty reduction, health, education, and climate action, but that without deliberate governance interventions, AI deployment risks exacerbating inequalities along precisely the North–South axis documented in this paper. The finding that algorithmic systems in international organizations systematically disadvantage Global South Populations thus has direct implications for the achievement of the 2030 Agenda and for the credibility of the multilateral development framework more broadly.

Bostrom's (2014) analysis of existential risk from advanced AI, while focused on hypothetical future superintelligence, underscores a more immediate and practical concern that this paper's findings illuminate: the challenge of maintaining meaningful human control over AI systems whose decision logic is opaque, whose operational speed exceeds human cognitive capacity, and whose embedded priorities may diverge from the intentions of their deployers. In the nuclear domain, this challenge is existential in the most literal sense; in the governance and negotiation domains, it is existential in a political sense, threatening the foundational principles of accountability, transparency, and sovereign equality on which the legitimacy of the international order depends.

### *7.4. Limitations*

Several limitations of this analysis must be acknowledged. First, the classified nature of AI integration in nuclear command and control systems means that the analysis in Section 4 necessarily relies on publicly available assessments, academic analyses, and informed inference rather than direct access to operational systems. Although the sources used—such as RAND Corporation, Chatham House, and top security journals—offer strong unclassified evidence, AI integration in nuclear systems may differ significantly from what is publicly known.

Second, the analysis of computational power asymmetries in multilateral negotiations (Section 5) is necessarily illustrative rather than exhaustive. A comprehensive empirical mapping of AI deployment across all major multilateral negotiation contexts would require access to proprietary modeling tools and classified intelligence capabilities that are, by their nature, inaccessible to academic researchers. The paper's analysis identifies structural dynamics and illustrative cases rather than claiming comprehensive empirical coverage.

Third, the theoretical framework proposed—machine-mediated international relations—is necessarily preliminary. The concept is intended to open a research agenda rather than to close one, and its further development will require both more granular empirical research and more sustained

engagement with the Science and Technology Studies (STS) literature on algorithmic agency, the critical data studies literature on data justice, and the emerging interdisciplinary scholarship on AI and global governance.

Fourth, the paper focuses primarily on state actors and international organizations and gives less attention to the roles of non-state actors, including technology corporations, civil society organizations, and epistemic communities—in shaping the dynamics of algorithmic diplomacy. While the political economy of AI production is addressed through engagement with Crawford (2021), Couldry and Mejias (2019), and Acemoglu and Johnson (2023), a more comprehensive treatment of corporate power in the algorithmic diplomacy landscape is an important direction for future research.

## 8. ConclusionS

This paper has argued that the integration of AI and algorithmic decision-making into the practice of international relations constitutes a transformative development that existing IR theory has inadequately conceptualized, and that the consequences—for global equity, nuclear stability, and the fairness of multilateral negotiations—are already materializing in ways that demand urgent scholarly and policy attention. Three core findings emerge:

First, the deployment of AI systems in international governance institutions is not a neutral efficiency enhancement but a politically consequential act that systematically reinforces existing inequalities along the North–South axis. The training data, design assumptions, and optimization objectives of AI systems deployed in humanitarian, developmental, and sanctions contexts encode the political economies of their production, which are overwhelmingly situated in the Global North. When these systems determine refugee resettlement outcomes, development finance eligibility, or sanctions targeting, they do so with a structural bias that is invisible precisely because it is embedded in technical architecture rather than expressed in explicit policy.

Second, the integration of AI into nuclear command and control architectures is generating new forms of strategic instability that Cold War deterrence theory did not anticipate and that current arms control frameworks are structurally unequipped to address. The compression of decision timelines, the opacity of algorithmic reasoning, and the vulnerability of AI systems to adversarial manipulation collectively threaten the deliberate, human-controlled decision-making processes on which nuclear stability has always ultimately depended.

Third, differences in computational power between advanced and developing nations create an overlooked form of international influence. This often turns the formal equality of states in trade and climate talks into a hidden, practical hierarchy shaped by algorithmic systems. Together, these findings support the case for a new theoretical framework in International Relations: the concept of machine-mediated IR, which treats AI systems not as neutral tools but as constitutive forces that shape the information environment, the distribution of power, and the normative frameworks of international politics. The development of this framework is not merely an academic project. It is a prerequisite for designing governance institutions capable of ensuring that the AI revolution serves the interests of the global community rather than entrenching the advantages of those who were already dominant when the revolution began.

The words of Kissinger, Schmidt, and Huttenlocher (2021, p. 5) carry a particular resonance for the international community: AI systems perform tasks that "humans have heretofore defined as requiring human thought," yet they do so without consciousness, context, or values. The task now before diplomats, scholars, and international institutions is to ensure that the consequential international decisions shaped by AI reflect not merely the values embedded in code by engineers in a small number of technologically dominant capitals, but the values of a genuinely pluralistic world order committed to equity, stability, and the dignity of all its inhabitants. Advancing this agenda requires a fundamental overhaul of global governance for the algorithmic era, not just minor tweaks or reforms. International organizations, national governments, and civil society must collaborate to develop robust frameworks for transparency, accountability, and oversight of algorithmic systems.

This includes not only setting standards for the ethical deployment of AI but also ensuring meaningful participation from states and communities that have historically lacked access to these technologies and the power they confer.

As this paper has argued, the risks and opportunities of AI in international relations are distributed unevenly. Without deliberate efforts to bridge computational divides and embed principles of algorithmic equity, there is a real danger that technological advancements will deepen global inequalities and undermine the legitimacy of international institutions. Conversely, a proactive approach to algorithmic diplomacy—one grounded in pluralism, inclusivity, and shared responsibility—offers a path toward more just and resilient global governance.

The challenges are formidable, yet the stakes could hardly be higher. As AI becomes increasingly central to decision-making in domains ranging from climate policy to nuclear security, the imperative for principled and participatory governance will only intensify. By embracing the concept of machine-mediated international relations and committing to reforms that place equity, transparency, and human dignity at the heart of global governance, the international community has an opportunity to shape an AI future worthy of its highest ideals.

In sum, the integration of AI into the fabric of international relations is not a distant prospect—it is a present reality, reshaping power, norms, and the prospects for peace. The way forward will be determined by the collective choices of policymakers, technologists, and global citizens alike. It is in their hands to ensure that the rise of algorithmic governance marks not a new age of inequity and instability, but the dawn of a more inclusive and just world order.

## 9. Recommendations

This paper presents a set of policy-oriented recommendations targeting states, international organizations, and the broader multilateral community. These recommendations, informed by empirical analysis, are intended to address governance gaps, equity issues, and strategic risks associated with algorithmic diplomacy in the international arena.

Firstly, international organizations should implement mandatory algorithmic impact assessments for all AI systems utilized in significant governance applications, such as refugee resettlement, development finance, sanctions enforcement, and climate modeling. Impact assessments must consider differential outcomes for populations in the Global South, mandate disclosure of training data sources and optimization goals, and undergo independent review processes including representation from affected groups. Fjeld et al. (2020) emphasize that consensus on impact assessment exists within institutional frameworks; the next step is to codify these principles into binding operational requirements.

Secondly, the international community should advance negotiations for a multilateral agreement addressing AI in nuclear command and control, leveraging insights from established arms control protocols and adapting them to software-based systems. Such an instrument should prohibit fully automated nuclear launch decisions, establish transparency measures for integrating AI into C3I architectures, and include verification protocols tailored to algorithmic technologies. As highlighted by Maas (2019), success will depend on creative adaptations of arms control models to fit the unique attributes of AI, underscoring the urgency of immediate action.

Thirdly, substantive investment in computational capacity-building for developing countries is essential. This initiative should be positioned as integral to equitable multilateral governance, encompassing the transfer of open-source modeling tools, training datasets, and technical expertise to low- and middle-income nations. The aim is to facilitate meaningful participation in AI-enabled negotiations concerning trade, climate, and development. Cihon (2019) notes that cooperation on technical standards and shared infrastructure can foster confidence and support extensive governance collaboration.

Fourthly, inclusive procedures for developing multilateral AI governance frameworks are imperative to ensure effective participation from states and communities in the Global South. Given that frameworks created by OECD, G7, and EU may reflect the priorities of advanced economies,

balancing these interests requires robust institutional support and monitoring mechanisms, as exemplified by the UNESCO Recommendation on the Ethics of Artificial Intelligence (2021). Risse (2019) suggests that embedding human rights impact assessments into international AI governance could enhance legitimacy and protect vulnerable populations.

Fifthly, academia should pursue interdisciplinary research on algorithmic diplomacy, integrating international relations theory, computer science, ethics, and area studies. The preliminary theoretical framework outlined in this paper demands further empirical investigation across governance domains, as well as continued engagement between scholars and practitioners. Funding agencies, universities, and international research bodies ought to establish dedicated programs, recognizing the significance of algorithmic diplomacy challenges for future global governance.

### 9.1 Directions for Future Research

Several promising avenues for further study arise from the present analysis. Empirical research is needed to examine the deployment of specific AI systems in international governance contexts—including refugee resettlement, sanctions screening, and climate modeling—focusing on how algorithmic outputs affect political decision-making and distributive impacts. Methodological progress should center on developing specialized frameworks for algorithmic impact assessment in international settings, building upon existing domestic regulatory approaches (Risse, 2019); interdisciplinary teams are particularly well-positioned to contribute in this regard. Theoretical work must continue to elaborate the concept of machine-mediated international relations through engagement with actor-network theory, posthumanism in IR, and ongoing discussions of digital sovereignty and data governance in the Global South. The intersection between algorithmic diplomacy and platform governance, digital trade regulation, and cyber norms also sustained scholarly attention.

In conclusion, the findings underscore the critical importance of interdisciplinary collaboration among international relations experts, computer scientists, ethicists, and practitioners. Addressing the complexities of algorithmic diplomacy necessitates integrative scholarship, modeled in this paper, to meet the pressing challenges ahead.

**Funding Statement:** This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2602).

**Conflicts of Interest:** The author declares no conflicts of interest.

**Transparency:** The author confirms that the manuscript is an honest, accurate, and transparent account of the study, that no vital features of the study have been omitted, and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

## References

1. Acemoglu, D., & Johnson, S. (2023). *Power and progress: Our thousand-year struggle over technology and prosperity*. Basic Books. ISBN 978-1541702530
2. Acton, J. M. (2018). Escalation through entanglement: How the vulnerability of command-and-control systems raises the risks of an inadvertent nuclear war. *International Security*, 43(1), 56–99. [https://doi.org/10.1162/isec\\_a\\_00320](https://doi.org/10.1162/isec_a_00320)
3. AI Safety Summit. (2023, November). *The Bletchley declaration by countries attending the AI Safety Summit, 1–2 November 2023*. UK Government. <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration>
4. Allison, G. (2017). *Destined for war: Can America and China escape Thucydides's trap?* Houghton Mifflin Harcourt. ISBN 978-0544935273
5. Altmann, J., & Sauer, F. (2017). Autonomous weapon systems and strategic stability. *Survival*, 59(5), 117–142. <https://doi.org/10.1080/00396338.2017.1375263>



6. Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., & Weinstein, J. (2018). Improving refugee integration through data-driven algorithmic assignment. *Science*, 359(6373), 325–329. <https://doi.org/10.1126/science.aao4408>
7. Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press. ISBN 978-1509526390
8. Bjola, C., & Holmes, M. (Eds.). (2015). *Digital diplomacy: Theory and practice*. Routledge. ISBN 978-1138792074
9. Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press. ISBN 978-0199678112
10. Brodie, B. (1959). *Strategy in the missile age*. Princeton University Press.
11. Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., hÉigeartaigh, S. Ó., Beard, S., Belfield, H., Farquhar, S., ... Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. Future of Humanity Institute, University of Oxford. <https://arxiv.org/abs/1802.07228>
12. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1–15. <http://proceedings.mlr.press/v81/buolamwini18a.html>
13. Cihon, P. (2019). *Standards for AI governance: International standards to enable global coordination in AI research & development*. Future of Humanity Institute, University of Oxford. [https://www.fhi.ox.ac.uk/wp-content/uploads/Standards\\_-FHI-Technical-Report.pdf](https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf)
14. Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press. ISBN 978-1503609822
15. Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press. ISBN 978-0300209570
16. Cummings, M. L. (2017). *Artificial intelligence and the future of warfare* [Research Paper]. Chatham House.
17. <https://www.chathamhouse.org/sites/default/files/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings.pdf>
18. Dafoe, A. (2018). *AI governance: A research agenda*. Future of Humanity Institute, University of Oxford. <https://www.fhi.ox.ac.uk/wp-content/uploads/AI-Governance-Research-Agenda.pdf>
19. Deibert, R. (2020). *Reset: Reclaiming the internet for civil society*. House of Anansi Press. ISBN 978-1487007003 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=>
20. Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press. ISBN 978-1250074317 [REF14]([URL14])
21. European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L. [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L\\_202401689](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689)
22. Farrell, H., & Newman, A. (2023). *Underground empire: How America weaponized the world economy*. Henry Holt and Company. ISBN 978-1250877413 [REF15]([URL15])
23. Fearon, J. D. (1995). Rationalist explanations for war. *International Organization*, 49(3), 379–414. <https://doi.org/10.1017/S0020818300033324>
24. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI* (Berkman Klein Center Research Publication No. 2020-1). Berkman Klein Center for Internet & Society, Harvard University. <https://doi.org/10.2139/ssrn.3518482>
25. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
26. G7. (2023). *Hiroshima Process international guiding principles for advanced AI systems*. G7 Leaders' Summit. <https://www.mofa.go.jp/files/100573471.pdf>
27. Geist, E., & Lohn, A. J. (2018). *How might artificial intelligence affect the risk of nuclear war?* (PE-296-RC). RAND Corporation. <https://doi.org/10.7249/PE296>

28. Horowitz, M. C. (2018). Artificial intelligence, international competition, and the balance of power. *Texas National Security Review*, 1(3), 36–57. <https://tnsr.org/2018/02/artificial-intelligence-international-competition-balance-power/>
29. Jervis, R. (1978). Cooperation under the security dilemma. *World Politics*, 30(2), 167–214. <https://doi.org/10.2307/2009958>
30. Johnson, J. (2019). Artificial intelligence & future warfare: Implications for international security. *Defense & Security Analysis*, 35(2), 147–169. <https://doi.org/10.1080/14751798.2019.1600800>
31. Kania, E. B. (2017). *Battlefield singularity: Artificial intelligence, military revolution, and China's future military power*. Center for a New American Security. <https://www.cnas.org/publications/reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power>
32. Keohane, R. O., & Nye, J. S. (2001). *Power and interdependence* (3rd ed.). Longman. ISBN 978-0321048580
33. Kissinger, H. A., Schmidt, E., & Huttenlocher, D. (2021). *The age of AI: And our human future*. Little, Brown and Company. ISBN 978-0316273800
34. Lee, K.-F. (2018). *AI superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin Harcourt. ISBN 978-1328546395
35. Lewis, P., Williams, H., Pelopidas, B., & Aghlani, S. (2014). *Too close for comfort: Cases of near nuclear use and options for policy*. Chatham House. [https://www.chathamhouse.org/sites/default/files/field/field\\_document/20140428TooCloseforComfortNuclearUseLewisWilliamsPelopidasAghlani.pdf](https://www.chathamhouse.org/sites/default/files/field/field_document/20140428TooCloseforComfortNuclearUseLewisWilliamsPelopidasAghlani.pdf)
36. Maas, M. M. (2019). How viable is international arms control for military artificial intelligence? Three lessons from nuclear arms control. *Contemporary Security Policy*, 40(3), 285–311. <https://doi.org/10.1080/13523260.2019.1576464>
37. Mearsheimer, J. J. (2001). *The tragedy of great power politics*. W. W. Norton & Company. ISBN 978-0393349276
38. Nye, J. S. (2011). *The future of power*. PublicAffairs. ISBN 978-1610390699
39. O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishers. ISBN 978-0553418811
40. OECD. (2019). *Recommendation of the Council on artificial intelligence* (OECD/LEGAL/0449). Organisation for Economic Co-operation and Development. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
41. Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press. ISBN 978-0674368279
42. Risse, M. (2019). Human rights and artificial intelligence: An urgently needed agenda. *Human Rights Quarterly*, 41(1), 1–16. <https://doi.org/10.1353/hrq.2019.0000>
43. Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C., Ng, A. Y., Hassabis, D., Platt, J. C., ... Bengio, Y. (2022). Tackling climate change with machine learning. *ACM Computing Surveys*, 55(2), Article 42. <https://doi.org/10.1145/3485128>
44. Russell, S. (2019). *Human compatibility: Artificial intelligence and the problem of control*. Viking. ISBN 978-0525558613
45. Sagan, S. D., & Waltz, K. N. (2003). *The spread of nuclear weapons: A debate renewed* (2nd ed.). W. W. Norton & Company. ISBN 978-0393977479
46. Scharre, P. (2018). *Army of none: Autonomous weapons and the future of war*. W. W. Norton & Company. ISBN 978-0393608984
47. Schlosser, E. (2013). *Command and control nuclear weapons, the Damascus accident, and the illusion of safety*. Penguin Press. ISBN 978-1594202278
48. Suleyman, M., & Bhaskar, M. (2023). *The coming wave: Technology, power, and the twenty-first century's greatest dilemma*. Crown. ISBN 978-0593593950
49. Taddeo, M., & Floridi, L. (2018). Regulate artificial intelligence to avert cyber arms race. *Nature*, 556(7701), 296–298. <https://doi.org/10.1038/d41586-018-04602-6>
50. UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>

51. UNHCR. (2023). *Global trends: Forced displacement in 2022*. United Nations High Commissioner for Refugees. <https://www.unhcr.org/global-trends>
52. Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11, Article 233. <https://doi.org/10.1038/s41467-019-14108-y>
53. Waltz, K. N. (1979). *Theory of international politics*. McGraw-Hill.
54. Wendt, A. (1999). *Social theory of international politics*. Cambridge University Press. ISBN 978-0521469609
55. Zegart, A. B. (2022). *Spies, lies, and algorithms: The history and future of American intelligence*. Princeton University Press. ISBN 978-0691147130
56. Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs. ISBN 978-1610395694

## Author Bio

Dr. Safran Safar Almakaty is a Professor at Imam Mohammad ibn Saud Islamic University (IMSIU) in Riyadh, known for his contributions to communication, media studies, and higher education in Saudi Arabia and the Middle East. He holds a master's from Michigan State University and a PhD from the University of Kentucky. His research focuses on media evolution, technology, and sociopolitical influences shaping public discourse.

Dr. Almakaty consults on communication strategy and policy for government, corporate, and non-profit organizations, with expertise in media literacy and educational reform. He has published extensively, contributed to international forums, and advanced Saudi Arabia's Vision 2030 through research on hybrid conference formats and strategic events. Committed to mentoring future scholars, Dr. Almakaty encourages academic innovation and excellence across the region.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.