



**How does the use of Artificial Intelligence in global governance create risks for
diplomacy and negotiations between great powers?**

Yerkezhan Serikbolova



Annotation

This research examines the impact of artificial intelligence (AI) on great-power diplomacy, i.e., how its advent would shift power equations in negotiations and global leadership. Although much research has been undertaken on AI in military uses, cyber defence, and ethics, relatively fewer research studies have been conducted on how it can shift trust-building, communication, and power equations in diplomacy. The research aims to address the gap below by comparing how great powers—the European Union, China, Russia, and the United States—go about searching for AI in their policy agenda and where policies diverge or meet when negotiating abroad.

Comparative literature review and theory analysis are utilized in the methodology. Realist, liberal institutionalist, and constructivist theses are utilized to apply to government utilization of AI: realists examine competitive arms races, liberals examine the need for multilateral agreements, and constructivists examine new practices and norms. Theoretical bridging theses to diplomatic practice, the research examines both potentialities and risks of AI in bargaining.

Transparency, credibility, verifiability, bias in algorithms, and amplifying crises through AI-driven decision-making or disinformation are some of the most important issues being argued. The report also mentions risks subject to further investigation, including the potential for AI to manipulate diplomatic messaging, enhance power imbalance, and undermine negotiating equity.

The research suggests that while AI can accelerate the spread of information and crisis control, it also holds the potential to introduce elephant-sized complexities to great-power diplomatic trust and institutional solidity. The article adds to AI scholarship with its diplomatic dimension map of AI governance, presenting an analytical model of its effects on great-power relations, and making a call for governance frameworks balancing innovation and integrity guarantees in diplomacy.

Keywords: artificial intelligence, global governance, diplomacy, negotiations, great powers.

Introduction

Artificial intelligence (AI) has changed very quickly from a technology to an unprecedented force transforming world affairs. Governments and global institutions are now contemplating the dual role of AI: improving the efficacy of governance and challenging traditional modalities of security and diplomacy. Research available focuses primarily on military uses, cyber threats to security, and ethical concerns like algorithmic bias. Though insightful, these views tend to underappreciate the influence of AI on great-power diplomacy—the subtle art of negotiation, building trust, and institutional crafting that facilitates global governance.

This research seeks to bridge this gap by examining how AI can transform negotiation dynamics between the major powers: the United States, China, Russia, and the European Union. These powers not only produce AI at scale but also frame norms for global governance. Their strategies are indicative of various avenues: The U.S. emphasizes innovation and private leadership; China integrates AI with state-led governance and international diplomacy; Russia

seeks AI as a strategic pressure tool; and the EU emphasizes ethical standards and the rule of law. One must gain some understanding of these distinctions to forecast cooperation and competition in AI diplomacy.

The study draws on three principal international relations theories to explain AI in government. Realism sees AI as a cause of strategic rivalry and emerging arms races. Liberal institutionalism spotlights the work of multilateral institutions in averting perils. Constructivism identifies the establishment of norms and socialised practice that shapes diplomatic action. All these offer a framework for studying the impact of AI on great-power diplomacy.

By connecting theoretical imagination and actual diplomatic challenges, this study aims to explain the potential, risk, and policy imperative involved in employing AI for global relations with emphasis on its less-widely discussed role towards remaking diplomacy itself.

Aims

Artificial Intelligence (AI) is becoming a crucial element of international governance as well as international relations. While it opens the doors to efficiency and innovation, it also threatens to undermine diplomacy as well as great power negotiations. The objective of this research paper is to analyze the threats of the use of AI in diplomacy and assess their potential effects on international governance.

The first hypothesis is that AI diplomacy poses new threats to trust, transparency, and cooperation between states.

The second hypothesis is that technological disparity, variously known as the "AI gap," can potentially shift power relationships and disadvantage poorer states in negotiations.

The third hypothesis is that algorithmic decision-making accelerates crisis management, but possibly also increases the risk of miscommunication or escalation.

The fourth hypothesis is that reliance on AI could change the future of diplomacy and multilateral governance by reducing human judgment and discrediting the traditional mechanisms of negotiation.

These hypotheses will enable us to determine and examine the most critical risks of AI for diplomacy and negotiations and to analyze their effects on the stability of global governance.

Literature review

1. Gaps in the literature

The current body of scholarship establishes AI as a transformative force across global security, governance, and diplomacy. Kolade (2024) emphasizes AI's operational benefits in predictive analytics, anomaly detection, and cyber-defense, while also pointing to geopolitical tensions and fragmented governance frameworks as obstacles to cooperation. Horowitz et al. (2018) stress AI's broad applications for defense, intelligence, and statecraft, warning about vulnerabilities to manipulation and adversarial use. Robinson (2025) offers a design for an International AI Agency (IAIA) in a comparative model of the International Atomic Energy Agency consensus

model. Similarly, Garrido Rebolledo (2025) discusses how AI puts governance architectures under strain and necessitates multilateral and adaptive ones.

However, throughout these books, a trend is evident: research is tackling military-strategic risk and governance architecture patchily relative to the diplomatic process itself. The literature amply covers autonomous weapons, espionage, cyberwar, and great-power rivalry (Rauf & Iqbal, 2023), but there are relatively few (Mostafaei et al., 2025) that begin to untangle how AI reconfigures the dynamics of negotiation, agenda-setting, and trust-building. Even in these exceptions, AI is depicted more as an efficiency instrument (e.g., crisis prediction, simulations) and less as a disruptive force that can erode mutual trust, fix asymmetries, or recalibrate inaccurate expectations between states.

Thus, an underlying gap exists: little systematic exploration of how AI reformulates the micro-dynamics of great power diplomatic bargaining. For example, how does bargaining strategy differ under the leadership of predictive models? Could excessive dependence on AI-generated forecasts heighten miscommunication or produce false confidence in unwarranted demands? Such concerns, though important, are poorly examined.

2. Theoretical framing: linking IR theory to AI in government

International relations (IR) theory offers useful interpretive frameworks for understanding AI's impact on diplomacy and governance.

Realism. Realist scholarship interprets AI as an accelerator of strategic rivalry. Rauf and Iqbal (2023) describe AI as a core driver of the US–China arms race, particularly in autonomous weapons and cyber-capabilities. From a realist perspective, AI becomes a new source of relative gains: states that achieve AI superiority secure decisive strategic advantages. This reasoning anticipates that great powers will oppose binding governance structures, favoring unilateral progress and hedging against the rise of their competitors. Threats of escalation, concealment, and distrust are therefore inevitable corollaries.

Liberal institutionalism. In contrast, liberal institutionalists underscore AI's transnational risks, which cannot be managed unilaterally. Robinson (2025) advocates for an IAIA as an inclusive global framework modeled on existing IGOs, while Garrido Rebolledo (2025) emphasizes adaptive, cross-sector collaboration as the foundation for effective governance. From this lens, AI governance requires rule-based cooperation, information-sharing, and joint verification to prevent escalation and maintain stability.

Constructivism. Constructivist approaches focus on the normative dimensions of AI. Mostafaei et al. (2025) highlight how AI reshapes diplomatic norms by the consequences for new governance actors and digital diplomacy practice. Garrido Rebolledo (2025) contends that the absence of collectively agreed-upon norms is a threat to fragmentation in regulation. Constructivism thus concentrates on how norm entrepreneurs—whether the EU, the UN, or civil society—frame the use of AI as legitimate or illegitimate, shaping the behavior of states through social more than material pressures.

Together, these theoretical frameworks illustrate the multidimensionality of AI: it is simultaneously a security dilemma (realism), an issue of cooperation (liberalism), and a battle of norms (constructivism).

3. Comparative and thematic analysis of great-power approaches

Trust and Verification.

Kolade (2024) notes AI's capacity to secure critical infrastructure, yet emphasizes that global consensus remains elusive due to divergent national interests. Similarly, Horowitz et al. (2018) warn that AI-generated information—such as anomaly detection in financial systems or image verification—can both enhance and undermine trust. Verification is the biggest challenge: while the US and EU move forward with open AI systems, China historically resisted invasive control mechanisms (Rauf & Iqbal, 2023). Asymmetry here makes it hard to bargain because there is no criterion to verify AI-derived claims to intelligence.

Transparency of Algorithms.

Garrido Rebolledo (2025) refers to the dispersed ecosystem of frameworks, with the UN noting that "hundreds of guides, frameworks and principles have been adopted, yet none... truly global in reach." Algorithmic opacity entrenches suspicion in diplomatic affairs: if a state bases its decisions on AI predictions, it can be delegitimized by others as unprovable or biased. Robinson (2025) thinks that credible control can be provided by only a legitimised central authority (such as the IAEA) to great powers. Escalation Risks.

Horowitz et al. (2018) caution that AI compresses decision loops, reducing deliberation time and heightening risks of unintentional escalation. Mostafaei et al. (2025) caution that AI simulations can displace human judgment and lead to premature or excessively rigid bargaining positions. The US–China 2024 statement that "humans, not AI, should control nuclear arms" (Robinson, 2025) acknowledges these risks, but enforceability is uncertain, especially with both great powers persisting in heavy investment in AI-defense.

Bias and Fairness in Negotiations.

Artificial intelligence technologies reinforce structural inequalities.

As Garrido Rebolledo (2025) shows, the emergence of "digital empires" (US, China, EU) remaps power inequalities since these actors monopolize datasets, computer capacity, and technical standards. Developing nations may be left behind and lack the capacity to challenge or countercheck AI-generated evidence. Rauf & Iqbal (2023) note that over 70 states now use AI surveillance, yet leadership ambitions still revolve around three actors: China, Russia, and the US. Such bifurcation of capability can deflect diplomatic negotiation away from open-ended discussion toward dominance by technologically leading states. Institutional Responses.

Robinson (2025) and Garrido Rebolledo (2025) recommend institutional solutions—IAIA, EU norms, or UN coordination—but profound disagreements persist.

The US prefers multi-stakeholder agreements with corporations, and China prefers governance led by the state. The EU is promoting norm leadership founded on ethical AI and human rights, and Russia is primarily applying AI to expand surveillance and asymmetric cyber conflict

(Horowitz et al., 2018). Such divergences render consensus elusive, echoing the realist refrain that great-power rivalry will restrain multilateralism.

4. Critical evaluation: risks, trade-offs, and underestimated issues

Across the literature, scholars acknowledge AI's dual nature: it can enhance efficiency, prediction, and cooperation, but also amplifies mistrust and miscalculation. However, several underestimated risks stand out.

Diplomatic loss of trust. There is little examination of how AI dependence could actively undermine trust. If actors presume algorithms are neutral, they will uncritically accept results, even biased ones. On the other hand, distrust of AI-generated intelligence will delegitimize legitimate claims, undermining negotiation credibility.

Information asymmetries. The literature underrates the extent to which asymmetrical access to data sets and computer capacity can alter bargaining power. AI-dominant nations can make the rules, and weaker nations don't have verification capacity, with structural dependence.

Corporate power. Robinson (2025) contends that Big Tech has power, yet the majority of IR scholarship still examines states as the singular unit of analysis. Corporations, in fact, control AI research, patents, and standards in the physical world. This is the privatized power, and this brings with it the problem of accountability and whether or not states can oversee negotiation agendas driven by corporate technologies.

Deepfakes and disinformation. While Mostafaei et al. (2025) mention deepfakes, the broader literature is more likely to view this as a media problem instead of a diplomatic one. Indeed, manipulated audio or video could sabotage negotiations by generating artificial adversarial quotes, irrevocably shattering trust.

Over-dependence on AI. AI-based simulations and predictions can foster determinism, leaving little room for compromise. Negotiators may over-depend on algorithmic guidance at the cost of human creativity, empathy, and spontaneity.

The deal is extreme: AI may speed up crisis communication and resource mobilization (Kolade, 2024; Garrido Rebolledo, 2025), yet exactly that speed risks short-circuiting traditional backchannel diplomacy, where ambiguity and delay have a defusing effect on crises.

5. Contribution: bridging the gap

This study adds value by productively focusing on the diplomatic aspect of AI governance, an emphasis that has been downplayed in the existing literature. Whereas other studies address security threats (Horowitz et al., 2018), institution-building (Robinson, 2025), and great-power rivalry (Rauf & Iqbal, 2023), they do not analytically assess how AI influences negotiation dynamics, trust-building, and agenda-setting between states.

By combining insights from realism, liberalism, and constructivism, this research accentuates how AI exacerbates arms race politics (realism), creates demands for multilateral management (liberalism), and undermines norms of openness, legitimacy, and trust (constructivism).

In effect, this study places AI not only as a technological or strategic issue but as a fundamental challenge to the social architecture of diplomacy. Such an examination of this understudied dimension is critical to establishing governance architectures that preserve human-centered negotiating norms and practices while responsibly embracing technological tools.

Analysis

Artificial intelligence is reshaping the pillars of diplomacy by transforming the way states negotiate, building confidence, and dealing with uncertainty. While a great deal of focus has been given to the military and strategic aspects, the diplomatic process itself endures equally dramatic shifts.

Second, AI creates a trust problem. Algorithmic tools can enhance openness in the guise of predictive modeling and crisis forecasting, on the one side. Their black box character can substitute for credibility during negotiations because states can't always check or replicate AI-produced outcomes, on the other side. This creates a paradox: AI reinforces the diplomatic relationship at the same moment it drains it.

Second is the challenge of information asymmetry. Nations with advanced AI capabilities not only accumulate material power but also bargaining power. This asymmetry risks creating a two-tier diplomatic order where technologically advanced states exercise influence while weaker states are locked out, not even capable of questioning or verifying algorithmic intelligence on offer.

Third, corporate power complicates things. Technology companies, and not states, have the most powerful AI systems. They set the agenda and standards for negotiation, and create accountability and sovereignty issues in a domain that has traditionally belonged to states. Diplomats can be constrained by means and structures beyond their control.

Fourth, AI alters the tempo and pitch of diplomacy. By compressing decision cycles, it closes the space for procrastination, ambiguity, and backchannel communications—strategies that have hitherto cooled down crises. Speed is not always security; in diplomacy, it can shorten deliberation and heighten the risk of escalation.

And finally, overdependence on AI presents an insidious but essential risk to diplomacy. Overdependent diplomats lose the flexibility, creativity, and humanness necessary to negotiate. Negotiation becomes a data-guided and not a dialogue-guided process, and in doing so, it loses the essence of diplomacy.

Short of forgoing diplomacy in the face of emerging threats, AI reinvents diplomacy through re-inventing negotiation and trust at its core. The challenge is to make technology serve as a tool for human judgment, rather than a substitute for it. A balance has to be found between

maintaining human control of decisions, putting AI into responsible contexts, and retaining the slow, uncertain, and relational character of diplomacy that cannot be replaced.

Conclusion

In conclusion, this study sought to analyze how the use of artificial intelligence in global governance threatens great power diplomacy and negotiations. Utilizing existing literature and applying international relations theory, the study listed key areas where AI can break trust, equity, and communication processes that maintain multilateral diplomacy.

The preliminary finding of this study is that while much of the literature traces the origin of security and military risk from AI, it fails to accord sufficient significance to the diplomatic sphere. In this study, it has been found that tools based on AI possess the power to undermine negotiation efforts by providing algorithmic obscurity, creating power disparities among technologically advanced and developed nations, and fueling suspicion by allowing for the easy spread of disinformation and deepfakes. This affirmatively corroborates the hypothesis that AI poses unique threats to transparency and cooperation in global affairs.

The second result shows how the powers differ in their attitudes. The United States and China conceive of AI first and foremost as a strategic asset, Russia as a tool of information control and sovereignty, and the European Union in ethical and regulatory terms. Rival such approaches complicate negotiations, confirming the hypothesis that technological divergence — much-hyped "AI gap" — makes competition an element of geopolitical rivalry and potentially an obstacle to cooperation.

The third general finding is that AI has the potential to speed up crisis management by destabilizing it by generating unrealistic expectations and over-reliance on simulation. This finding aligns with the hypothesis that algorithmic decision-making can reshape both the pace and the direction of crisis diplomacy in ways that can erode tried verification and trust-building mechanisms.

Overall, the research highlights that dangers presented by AI in diplomacy have been less examined compared to its possible applications in war. In closing the gap, the research provides contributions towards a deeper awareness of the mechanisms through which AI will destabilize great power norms of cooperation and negotiation. One of the answers to the primary research question — "How does the use of AI in global governance threaten great power diplomacy and negotiations?" — is that AI, in delivering efficiency and swiftness, simultaneously undermines the pillars of diplomacy: trust, transparency, and fairness.

References

1. Ilcic, A., Fuentes, M., & Lawler, D. (2025). Artificial intelligence, complexity, and systemic resilience in global governance. *Frontiers in Artificial Intelligence*, 8.

<https://doi.org/10.3389/frai.2025.1562095>



-
2. Manfredi-Sánchez, J. L., & Morales, P. S. (2024). Generative AI and the future for China's diplomacy. *Place Branding and Public Diplomacy*.
<https://doi.org/10.1057/s41254-024-00328-7>
 3. *Mechanism design, behavioral science, and artificial intelligence in international relations*. (n.d.). Google Books.
<https://shorturl.at/CBKD5>
 4. *The role of artificial intelligence in global politics and governance*. (2025). ScienceDirect.
<https://www.sciencedirect.com/science/article/pii/S2666188825000565>
 5. *Title of the Article*. (Year). *International Affairs*, 101(4), 1483-_____.
<https://academic.oup.com/ia/article/101/4/1483/8141294>