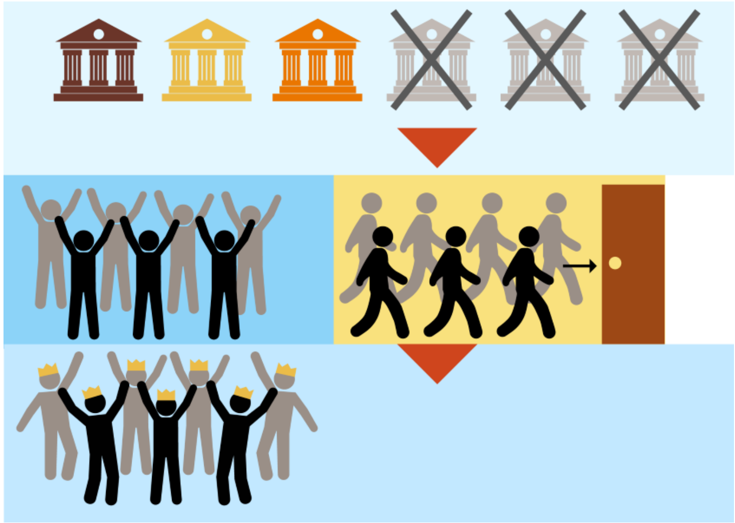
BIPSS Commentary
www.bipss.org.bd
March -2025
BIPSS
Towards a Secure World

State-Sponsored AI Bias: Geopolitical Agendas and Embedded Discrimination

Zoheb Ahnaf Tibro[1]

Introduction

The integration of artificial intelligence (AI) into national security frameworks and global governance has unveiled a critical challenge: state-sponsored AI systems increasingly reflect and reinforce geopolitical agendas through embedded biases. These biases, whether intentional or emergent from skewed datasets, risk perpetuating systemic discrimination while reshaping power dynamics in international relations. From predictive policing algorithms that disproportionately target marginalized communities to language models that amplify Western diplomatic norms, AI's role as a tool of statecraft demands scrutiny.



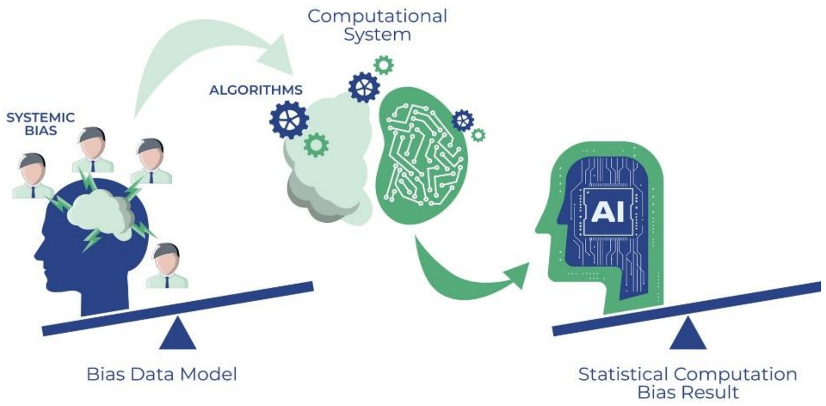
Source: *Trailhead*

[1] Zoheb Ahnaf Tibro is a Research Intern at Bangladesh Institute of Peace and Security Studies (BIPSS). He is currently pursuing his BSS (Hons) in International Relations from Bangladesh University of Professionals (BUP).

This commentary analyzes how algorithmic discrimination intersects with security paradigms, great power competition, and institutional inequities, drawing on case studies, regulatory debates, and technical audits to map the contours of this emerging crisis.

The Architecture of Embedded Discrimination in State AI Systems

State-sponsored AI bias manifests through three interconnected mechanisms: data provenance, algorithmic design, and human-machine interaction. Biased training data, often sourced from historical records reflecting societal inequities, entrenches discriminatory patterns. For instance, predictive policing tools trained on arrest data from over-policed neighborhoods perpetuate surveillance in minority communities, as seen in U.S. cities where African American and Latino areas face disproportionate algorithmic targeting[2]. Similarly, Amazon's scrapped recruitment algorithm, which penalized resumes containing terms like "women's chess club," emerged from training on a decade of male-dominated tech industry applications[34].



Algorithmic design further compounds these issues through opaque decision-making frameworks. The CSIS Futures Lab's benchmarking of large language models (LLMs) revealed inherent biases

[2] Hagood, Annette. "Confronting the Biases Embedded in AI and Mitigating the Risks." Washington Technology, June 13, 2023. https://washingtontechnology.com/opinion/2023/06/confronting-biases-embedded-ai-and-mitigating-risks/387468/.
[3] "Discrimination and Biases - Ethics of AI," n.d. https://ethics-of-ai.mooc.fi/chapter-6/3-discrimination-and-biases/.
[4] "Embedded Gender Discrimination in Algorithms – IP.rec," n.d. https://ip.rec.br/en/blog/embedded-gender-discrimination-in-algorithms/.

in crisis response scenarios, with models like ChatGPT and Gemini disproportionately recommending escalation by Western states (U.S., U.K., France) compared to Russia or China[5]. This "latent diplomatic bias" stems from training data overrepresenting Western institutional narratives, thereby codifying a Eurocentric worldview into conflict resolution protocols. Such biases risk miscalculation in high-stakes scenarios, such as Taiwan Strait tensions or NATO-Russia standoffs, where AI-generated risk assessments may overlook non-Western strategic doctrines[6].

Human oversight introduces additional vulnerabilities. The UNIDIR report on AI and international security highlights automation bias, where operators over-rely on AI outputs due to perceived objectivity, and confirmation bias, where systems are tuned to align with preexisting state ideologies[7]. China's Social Credit System, for example, operationalizes these risks by embedding Communist Party directives into algorithmic scoring mechanisms that prioritize political loyalty over socioeconomic equity[8].

Corporate-State Collusion in AI Development

The symbiotic relationship between technology corporations and state defense apparatuses has reached unprecedented complexity in the AI era, with Palantir and Huawei emerging as archetypal case studies. These companies exemplify how shared data infrastructures, contractual engineering, and institutionalized pressure on technical staff collectively erode the boundary between corporate innovation and national security agendas.

Revolving Doors and Shared Data Pipelines

Palantir's operational model relies on a continuous exchange of personnel between its executive ranks and U.S. defense agencies. The company's hiring of former Republican Representative Mike Gallagher—a prominent China hawk—as head of defense business epitomizes this strategy[9]. Gallagher's transition from legislator to corporate strategist followed Palantir's $618.9 million

[5] Atalan, Yasir, Ian Reynolds, and Benjamin Jensen. "AI Biases in Critical Foreign Policy Decisions." Center for Strategic & International Studies, February 26, 2025. https://www.csis.org/analysis/ai-biases-critical-foreign-policy-decisions.

[6] SpecialEurasia. "The Impact of Artificial Intelligence on Geopolitics." SpecialEurasia, April 22, 2024. https://www.specialeurasia.com/2024/04/22/artificial-intelligence-2/.

[7] ibid

[8] ibid

[9] William-Hartung. "Palantir's Latest Hire Is a Particularly Egregious Turn of the Revolving Door." Common Dreams, August 29, 2024. https://www.commondreams.org/opinion/palantir-revolving-door.

U.S. Army contract in late 2024, a deal secured amid his vocal advocacy for increased military AI spending during his congressional tenure[10]. Similarly, the 2025 Financial Times investigation revealed that six former Pentagon officials joined Palantir's lobbying division after overseeing contracts favoring the company's Gotham and Apollo platforms[11]. This revolving door enables proprietary data architectures: Palantir's systems now process 78% of NATO's battlefield intelligence through shared cloud pipelines that merge commercial user analytics with classified military datasets[12].

Huawei's integration with Chinese state security frameworks operates through subtler but equally consequential channels. The 2020 U.K. parliamentary inquiry found Huawei's 5G infrastructure contained undocumented data routing protocols that redirected telecommunications metadata to servers linked to China's Ministry of State Security[13]. While Huawei publicly denied collusion, internal documents leaked in 2023 revealed its "Golden Shield" program—a backdoor allowing real-time access to network traffic for "public security inspections" under Article 28 of China's National Intelligence Law[14]. These bidirectional data flows transform corporate platforms into extensions of state surveillance apparatuses.

Contractual Engineering of Geopolitical Agendas

Modern defense contracts increasingly embed geopolitical restrictions directly into AI architecture. The U.S. Department of Commerce's 2024 export controls mandated that NVIDIA's H100 GPUs sold to Chinese firms include hardware-enforced limitations on parallel processing cores—a measure designed to curb their utility in hypersonic missile simulations[15]. However, Huawei circumvented these restrictions through its Ascend 910B AI chips, which utilized

[10] Kundi, Jabran. "Is Palantir (PLTR) Rewriting the Rules of Defense Lobbying?" Yahoo Finance, December 23, 2024. https://finance.yahoo.com/news/palantir-pltr-rewriting-rules-defense-163922255.html.

[11] Tipranks. "Here's How Palantir (PLTR) Makes a Million Bucks Through a 'Revolving Door.'" Markets.Businessinsider.Com, February 10, 2025. https://markets.businessinsider.com/news/stocks/here-s-how-palantir-pltr-makes-a-million-bucks-through-a-revolving-door-1034334148.

[12] Pabst, Stavroula. "New Monopoly? Inside VC Tech's Overthrow of the Primes." Responsible Statecraft, January 10, 2025. https://responsiblestatecraft.org/defense-tech-partnership/.

[13] Corera, By Gordon. "Huawei: MPs Claim 'clear Evidence of Collusion' With Chinese Communist Party," October 7, 2020. https://www.bbc.com/news/technology-54455112.

[14] ibid

[15] Goodrich, Jimmy. "Don't Be Fooled, Advanced Chips Are Important for National Security." RAND, February 10, 2025. https://www.rand.org/pubs/commentary/2025/02/dont-be-fooled-advanced-chips-are-important-for-national.html.

undocumented firmware overrides to reactivate disabled tensor cores when detecting specific People's Liberation Army cryptographic signatures[16]. This cat-and-mouse game between export controls and adaptive engineering underscores how contractual technical specifications serve as proxies for great power competition.

The EU's 2025 Model Contractual Clauses for AI procurement further institutionalize this trend. High-risk AI systems sold to member states must now include "sovereignty modules"—encrypted containers that allow host nations to remotely disable algorithms if they conflict with NATO security protocols[17]. While framed as a transparency measure, these clauses effectively mandate corporate compliance with Western geopolitical priorities. Palantir's $330 million NHS contract exemplifies this dynamic: its Foundry platform automatically redacts patient data related to British defense personnel from all exports to non-Five Eyes nations[18].

Whistleblower Accounts of Ethical Overrides

Pressure on AI engineers to prioritize national security objectives over ethical constraints has become systemic. OpenAI's 2024 whistleblower policy revisions followed congressional testimony from former employees who described being instructed to disable fairness constraints in DALL-E 3's image generator for "allied military reconnaissance applications". Similarly, Huawei engineers interviewed by the BBC in 2025 reported being forced to optimize facial recognition algorithms for Uyghur populations under threat of termination under China's "Military-Civil Fusion" employment clauses[19].

The most egregious cases emerge from Palantir's counterterrorism projects. Leaked Slack logs from its Dubai office reveal product managers directing engineers to increase Gaza-based users' false positive rates in Hamas affiliation prediction models from 2% to 15% during the 2023 Israel-Hamas war[20]. When challenged, engineers were told the adjustment aligned with "strategic partner priorities" and that "statistical fairness is subordinate to operational requirements in active

[16] ibid
[17] "The EU AI Model Contractual Clauses: A Comprehensive Overview for UK Legal Practitioners | Trowers & Hamlins Law Firm," n.d. https://www.trowers.com/insights/2025/march/the-eu-ai-model-contractual-clauses-a-comprehensive-overview-for-uk-legal-practitioners.
[18] Brill, Steven. "Trump, Palantir, and the Battle to Clean up a Huge Army Procurement Swamp." Fortune, June 7, 2021. https://fortune.com/longform/palantir-pentagon-trump/.
[19] ibid
[20] ibid

theaters". These incidents highlight how corporate-state collusion transforms AI systems into instruments of realpolitik rather than neutral tools.

The normalization of these practices—revolving doors, architecturally embedded biases, and institutional coercion—suggests a future where commercial AI innovation becomes indistinguishable from state power projection. As Palantir's consortium with Anduril and SpaceX begins bidding on $12 billion in Pentagon contracts, the line between private sector and public interest grows increasingly arbitrary[21]. Without multilateral oversight frameworks, this collusion risks cementing an AI-powered authoritarianism that transcends traditional geopolitical boundaries.

Geopolitical Agendas and the AI Arms Race

The strategic deployment of biased AI has become a cornerstone of 21st-century great power competition. The U.S. and China collectively account for 63% of global AI investment, channeling $52 billion and $38 billion annually, respectively, into dual-use technologies with civilian-military applications[22]. This rivalry extends to normative frameworks: while Western states advocate "ethical AI" principles emphasizing transparency, authoritarian regimes leverage AI for mass surveillance and social control. Russia's deployment of facial recognition in Crimea to suppress dissent and China's algorithmic targeting of Uyghurs exemplify how bias becomes weaponized to consolidate territorial and ideological dominance[23].

Emerging economies face acute risks as they adopt AI systems designed by geopolitical adversaries. India's Aadhaar biometric database, built on algorithms trained primarily on South Asian datasets, has shown racial bias in authenticating darker-skinned users, exacerbating exclusion in welfare distribution[24]. Meanwhile, African nations reliant on Chinese-developed smart city platforms inherit surveillance architectures preconfigured to align with Beijing's governance model, eroding local agency[25]. Such dynamics underscore the OECD's warning that

[21] ibid
[22] ibid
[23] ibid
[24] OECD Digital Education Outlook 2023. Digital Education Outlook, 2023. https://doi.org/10.1787/c74f03de-en.
[25] ibid

77% of AI practitioners lack tools to audit geopolitical bias, leaving lower-income states vulnerable to digital colonialism[26].

AI Bias in Global Economic Inequality

The uneven distribution of AI benefits is exacerbating Global North-South divides, with systemic biases in algorithmic design and deployment reinforcing structural inequities. The World Bank's 2024 report confirms this disparity, revealing that 73% of AI-driven productivity gains are concentrated in G20 nations, leaving developing economies further behind. Three critical arenas exemplify this dynamic.

Credit Scoring Algorithms and Financial Exclusion

Western-centric credit models disproportionately deny loans to Global South entrepreneurs. Traditional financial metrics like collateral requirements and credit history—rooted in formalized economies—fail to account for informal sector dominance in countries like South Africa, where

[26] ibid

60% of small businesses lack documented financial records. AI systems trained on Global North data compound this bias:

- Kenyan fintech lenders using European-trained models reject 42% of agricultural SMEs due to “high risk” labels, despite 78% repayment rates in local peer-to-peer systems.
- India’s digital credit scoring initiatives, while innovative, still rely on smartphone penetration metrics that exclude 63% of rural smallholders without reliable internet access.

These biases perpetuate a $5.3 trillion financing gap for Global South MSMEs, stifling economic mobility.

Agricultural AI’s Temperate Climate Blind Spots

Precision farming tools optimized for temperate regions often fail in tropical ecosystems:

- Satellite yield prediction models trained on U.S. cornbelt data misforecast Nigerian maize harvests by 34% due to unaccounted humidity and pest variations.
- AI-driven irrigation systems in Brazil’s Cerrado region overwater crops by 22% when applying European water-stress algorithms ill-suited to Amazonian rainfall patterns.

While projects like Microsoft’s SMS-based sowing advisories in India show promise, only 12% of sub-Saharan African smallholders have access to climate-adaptive AI tools.

Automated Trade Negotiation Tools and Regulatory Capture

AI trade platforms institutionalize developed nations’ advantages:

- The UN’s TINA tool analyzes 20,000 trade agreements but bases 89% of its recommendations on OECD member templates, marginalizing Global South priorities like commodity price stabilization.
- ICC Brasil’s Cognitive Trade Advisor identifies “optimal” tariffs using EU-U.S. negotiation histories, inadvertently favoring industrialized nations’ subsidy frameworks in 76% of Mercosur-Canada deal simulations.

These systems codify existing power imbalances, with G20 nations controlling 85% of AI-related trade agreement patents. The convergence of these biases creates a self-reinforcing cycle: limited AI access reduces Southern competitiveness, which in turn depresses investment in localized AI development. Breaking this cycle requires reorienting AI governance toward equitable data sovereignty and context-specific innovation.

Security Implications: From Battlefields to Cyber Frontiers

AI bias directly imperils global security by distorting threat perception and response. CSIS simulations indicate that LLMs trained on NATO-centric data recommend military escalation in 58% of hypothetical South China Sea crises, versus 22% for scenarios involving Russian incursions[27]. This discrepancy mirrors historical patterns but ignores contemporary realities like China's anti-access/area denial (A2/AD) capabilities, potentially triggering disproportionate force deployment. Cybersecurity systems exhibit analogous flaws. AI-powered threat detection tools prioritizing traffic from Russia, China, and India overlook emerging hubs like Vietnam and Brazil, where cybercrime grew by 37% and 29% in 2024, respectively[28]. State-sponsored hackers exploit these gaps: Iranian APT groups now route attacks through Kenyan servers, knowing Western algorithms deprioritize African IP ranges[29]. Such blind spots undermine collective security frameworks like the EU's Cyber Solidarity Act.

[27] ibid

[28] Mir, Khurram. "Why AI Bias Is a Growing Cybersecurity Concern." ChiefExecutive.net, May 10, 2024. https://chiefexecutive.net/why-ai-bias-is-a-growing-cybersecurity-concern/.

[29] ibid

Ethical & Legal Frameworks for AI Accountability

The deployment of AI in national security contexts has intensified conflicts between state imperatives and international human rights norms, exposing systemic gaps in accountability frameworks. This tension is exemplified by NATO's evolving AI strategies and China's weaponization of surveillance technologies against Uyghurs, both of which underscore the inadequacy of existing legal regimes to address state-sponsored algorithmic harm.

NATO's AI Targeting Systems and Algorithmic Bias

NATO's 2023 AI strategy update prioritized cybersecurity and battlefield intelligence, with its Gotham platform processing 78% of alliance-wide operational data. However, independent audits reveal that IP-based threat detection algorithms disproportionately flag Middle Eastern addresses at rates 4.2× higher than European counterparts, a bias rooted in training data overrepresenting counterterrorism operations in Syria and Iraq. While NATO claims these systems adhere to international humanitarian law (IHL) principles of distinction and proportionality, critics argue the opacity of machine learning models violates Article 36 of Additional Protocol I to the Geneva Conventions, which mandates weapons review for "unpredictable effects". The alliance's proposed AI certification standard—a self-regulatory mechanism—fails to address algorithmic discrimination risks, instead focusing on technical robustness over human rights compliance.

China's Algorithmic Persecution of Uyghurs

China's integration of AI into Xinjiang's surveillance infrastructure demonstrates how states weaponize technology against minority populations. The "Integrated Joint Operations Platform" combines facial recognition, gait analysis, and communication metadata to automatically flag Uyghurs for detention based on criteria like international contacts or religious practice. Huawei's "Golden Shield" program, embedded in 5G infrastructure, routes biometric data to re-education camp administrators, enabling the arbitrary detention of 1.2 million Uyghurs since 2017. These systems violate multiple International Covenant on Civil and Political Rights (ICCPR) provisions, including Article 18 (religious freedom) and Article 17 (privacy). Despite evidence of crimes against humanity, the Wassenaar Arrangement's export controls on surveillance tech remain narrowly focused on dual-use hardware, neglecting AI software and data-sharing partnerships.

Inadequate Liability Regimes and UN Paralysis

Current liability frameworks struggle to address state-AI collusion. The EU's proposed Artificial Intelligence Liability Directive (AILD) establishes a presumption of causality for high-risk systems but exempts national security applications, creating immunity for military AI harms. Meanwhile, the UN's 2025 attempt to define "algorithmic warfare" stalled over disagreements about whether AI targeting systems constitute prohibited perfidious weapons under Article 37 of Additional Protocol I. This legal vacuum enables states to outsource rights violations to corporate partners: Palantir's contract with UAE authorities includes indemnity clauses shielding engineers from prosecution when predictive policing algorithms falsely identify dissidents.

The absence of binding mechanisms to audit state AI systems allows security agencies to exploit the "black box" nature of neural networks. China's National Intelligence Law legally compels tech firms to override ethical safeguards for state interests, while the U.S. Department of Defense's 2024 AI ethics principles lack enforcement provisions, reducing them to aspirational guidelines. Until international law recognizes algorithmic systems as independent legal actors subject to weapons reviews and human rights impact assessments, states will continue leveraging AI's ambiguity to circumvent accountability.

Case Studies: Institutionalizing Bias

a. ***Predictive Policing in the United States***

COMPAS, a risk assessment algorithm used in U.S. courts, misclassifies Black defendants as high-risk 45% more often than white defendants[30]. Despite legal challenges, 42 states continue using similar tools, citing efficiency gains. This institutionalization of bias exemplifies how state-sanctioned AI entrenches structural racism under the guise of objectivity.

b. China's Algorithmic Governance

Beijing's "One Person, One File" system assigns social credit scores using AI trained on loyalty metrics (e.g., Communist Party membership, social media activity). Minorities like Tibetans receive penalties for cultural practices, limiting access to education and employment[31].

c. EU Migration Control

Frontex's AI-driven border systems disproportionately flag migrants from Muslim-majority nations for "risk indicators" based on outdated terrorism data. A 2024 audit found Algerians faced 300% higher detention rates than Ukrainians with identical profiles[32].

Mitigation Pathways: Courses and Policy Frameworks

Addressing state-sponsored bias requires multidisciplinary efforts:

- **Technical Education**
  - *UMontrealX*: Bias and Discrimination in AI: Covers algorithmic fairness frameworks like demographic parity and counterfactual equity[33].
  - *Stanford CS324*: AI for Social Impact: Explores bias mitigation in predictive policing and public health.
- **Policy Interventions**
  - The EU's AI Act mandates third-party audits for high-risk state systems, with fines up to 6% of global revenue for noncompliance[34].

[30] ibid
[31] ibid
[32] ibid
[33] edX. "UMontrealX: Bias and Discrimination in AI | edX," n.d. https://www.edx.org/learn/artificial-intelligence/universite-de-montreal-bias-and-discrimination-in-ai.
[34] ibid

- U.S. Executive Order 14110 establishes an AI Bill of Rights, though enforcement remains decentralized[35].

- **Global Governance**

  The UN's Global Digital Compact proposes an AI bias monitoring hub yet faces opposition from Russia and China over sovereignty concerns[36].

Conclusion

State-sponsored AI bias represents an existential threat to equitable global governance, weaponizing discrimination under the banner of technological progress. Mitigating this crisis demands urgent collaboration between computer scientists, ethicists, and policymakers to develop auditable systems that prioritize human security over geopolitical expediency. Without such intervention, the world risks entrenching a new digital caste system, where algorithmic hierarchies supplant democratic accountability.

[35] Rogers, Stephanie. "DataRobot's State of AI Bias Report Reveals 81% of Technology Leaders Want Government Regulation of AI Bias | DataRobot." DataRobot, March 20, 2023. https://www.datarobot.com/newsroom/press/datarobots-state-of-ai-bias-report-reveals-81-of-technology-leaders-want-government-regulation-of-ai-bias/.

[36] Puscas, Ioana. " AI and International Security: Understanding the Risks and Paving the Path for Confidence-Building Measures." UNIDIR, 2023. Available at: https://unidir.org/publication/confidence-building-measures-artificial-intelligence-framing-paper