UNIDIR

Global Conference on
AI, Security and Ethics

Data Governance in Military AI: Transparency, Inclusion, and Security

Prof. Eduardo Migon, PhD
eduardomigon@gmail.com

Military Institute of Engineering
Brazilian Army

Abstract:
Data governance in military artificial intelligence (AI) is essential for global security, efficiency, and ethical compliance. The integrity, accuracy, and representativeness of datasets in AI-driven military systems directly impact decision reliability, risk mitigation, and battlefield effectiveness. This poster presents a strategic framework for responsible data management in military AI, integrating best practices, security protocols, risk assessment, and inclusion mechanisms, with insights from Brazil's operational landscape. We analyze how the absence of governance standards increases vulnerabilities, ethical concerns, and biases, potentially leading to destabilizing consequences. The proposal highlights key mechanisms such as continuous auditing, secure anonymization, cross-validation, and bias mitigation to ensure accountability and fairness. Additionally, it explores how emerging economies, particularly Brazil, can contribute to equitable AI regulations by ensuring diverse datasets, fostering resilience and interoperability in multinational defense strategies. By integrating AI into military decision-making, this approach seeks to enhance efficiency while addressing risks related to surveillance, misinformation, and adversarial manipulation. It outlines practical measures to promote an inclusive, ethical, and robust regulatory framework, ensuring multi-stakeholder participation, international cooperation, and compliance with best standards.
Keywords: Military AI, Data Governance, Transparency, Security, Ethical Compliance.

Resumo:
A governança de dados na inteligência artificial (IA) militar é essencial para a segurança global, eficiência e conformidade ética. A integridade, precisão e representatividade dos conjuntos de dados em sistemas militares baseados em IA impactam diretamente a confiabilidade das decisões, a mitigação de riscos e a eficácia no campo de batalha. Este pôster apresenta um modelo estratégico para a gestão responsável de dados na IA militar, integrando melhores práticas, protocolos de segurança, avaliação de riscos e mecanismos de inclusão, com insights do cenário operacional do Brasil. Analisamos como a ausência de padrões de governança aumenta vulnerabilidades, preocupações éticas e vieses, podendo levar a consequências desestabilizadoras. A proposta destaca mecanismos fundamentais, como auditoria contínua, anonimização segura, validação cruzada e mitigação de vieses, para garantir responsabilidade e equidade. Além disso, explora como economias emergentes, particularmente o Brasil, podem contribuir para regulamentações equitativas de IA, assegurando conjuntos de dados diversificados e promovendo resiliência e interoperabilidade em estratégias de defesa multinacionais. Ao integrar a IA na tomada de decisões militares, essa abordagem busca aprimorar a eficiência, abordando riscos relacionados à vigilância, desinformação e manipulação adversária. O estudo delineia medidas práticas para promover um marco regulatório inclusivo, ético e robusto, garantindo a participação de múltiplos stakeholders, cooperação internacional e conformidade com os melhores padrões.
Palavras-chave: IA Militar, Governança de Dados, Transparência, Segurança, Conformidade Ética.

PRÓ-PESQUISA

Artificial Intelligence and Quantum Technologies: Their Impact on the Readiness and Employment of Land Forces

https://www.eb.mil.br/
https://www.decex.eb.mil.br/
https://www.ime.eb.mil.br/

UNIDIR

Global Conference on
AI, Security and Ethics

The Human Element in Military AI: Perspectives from Brazil on Supervision, Responsibility, and Decision-Making

Prof. Eduardo Migon, PhD
eduardomigon@gmail.com

Military Institute of Engineering
Brazilian Army

Abstract:
The integration of Artificial Intelligence (AI) into military systems requires a critical balance between automation and human oversight. This poster proposes a governance model for military AI, emphasizing the importance of human supervision throughout the technology lifecycle, from development to field execution. From a Brazilian perspective, we examine how human oversight ensures ethical, strategic, and legally responsible decisions in AI-driven military operations, while addressing challenges such as data biases, lack of representativeness, and ethical risks. The study highlights mechanisms to maintain human intervention capacity in critical moments, including continuous auditing, cross-validation, and training for military operators to interpret and question AI recommendations. A decision-making framework combining AI with human judgment is proposed to mitigate biases and reduce risks of unintended automated actions. Brazil's unique cultural, geographic, and social diversity offers valuable insights for creating representative datasets and fostering inclusive AI governance. The country's experience in complex environments, such as the Amazon and urban areas, provides practical lessons for enhancing interoperability and resilience in multinational defense strategies. Finally, the poster suggests international policy guidelines to reinforce the human element in military AI, promoting transparency, interoperability, and common standards for operational security. By integrating Brazil's perspectives, this work contributes to the global debate on ethical and responsible AI use in defense, ensuring that technological advancements align with human values and security needs.

Resumo:
A integração da Inteligência Artificial (IA) nos sistemas militares exige um equilíbrio crítico entre automação e supervisão humana. Este pôster propõe um modelo de governança para a IA militar, enfatizando a importância da supervisão humana ao longo de todo o ciclo de vida da tecnologia, desde o desenvolvimento até a execução em campo. A partir de uma perspectiva brasileira, examinamos como a supervisão humana garante decisões éticas, estratégicas e juridicamente responsáveis em operações militares baseadas em IA, ao mesmo tempo em que aborda desafios como vieses de dados, falta de representatividade e riscos éticos. O estudo destaca mecanismos para manter a capacidade de intervenção humana em momentos críticos, incluindo auditoria contínua, validação cruzada e treinamento de operadores militares para interpretar e questionar recomendações geradas por IA. Propõe-se um modelo de tomada de decisão que combina IA com julgamento humano para mitigar vieses e reduzir os riscos de ações automatizadas não intencionais. A diversidade cultural, geográfica e social do Brasil oferece insights valiosos para a criação de conjuntos de dados representativos e para o fortalecimento da governança inclusiva da IA. A experiência do país em ambientes complexos, como a Amazônia e áreas urbanas, fornece lições práticas para aprimorar a interoperabilidade e a resiliência em estratégias de defesa multinacionais. Por fim, o pôster sugere diretrizes de políticas internacionais para reforçar o elemento humano na IA militar, promovendo transparência, interoperabilidade e padrões comuns para a segurança operacional. Ao integrar perspectivas brasileiras, este trabalho contribui para o debate global sobre o uso ético e responsável da IA na defesa, garantindo que os avanços tecnológicos estejam alinhados com valores humanos e necessidades de segurança.

PRÓ-PESQUISA

Artificial Intelligence and Quantum Technologies:
Their Impact on the Readiness and Employment
of Land Forces

https://www.eb.mil.br/
https://www.decex.eb.mil.br/
https://www.ime.eb.mil.br/

UNIDIR

Global Conference on AI, Security and Ethics

AI Without Borders: Harmonizing AI Governance Across the Gulf

Lara Arekat

MSc Graduate

CONTEXT

The Middle East, particularly Gulf Cooperation Council (GCC) countries, are emerging as key players in the global AI revolution. Gulf nations are increasingly integrating AI into their military operations, yet the novelty of AI leaves limited precedent for development of regulation. AI transcends borders, and so protective legislation must also be cross-border.
The GCC consists of Bahrain, Kuwait, Oman, Qatar, Saudi Arabia and the United Arab Emirates.

RENTIER THEORY

Rentier theory explains how states that derive a significant portion of their revenue from external rents—such as oil exports—develop their economic and political structures.
However, as Gulf nations diversify their economies - particularly by investing in new avenues of revenue like AI and technology - there is a growing need to develop regulatory frameworks that align with their economic transformation while maintaining stability and state control.

WHATS THE PROBLEM?

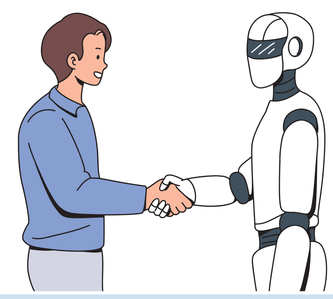
AI technology is advancing faster than legislation can keep up, creating regulatory gaps that pose security and ethical risks. In the Gulf Cooperation Council (GCC), where AI is increasingly integrated into military and commercial sectors, the absence of unified legal framework leads to fragmented national policies. Without coordinated regulation, the region risks inconsistencies in AI governance, potential misuse, and barriers to innovation.
A structured, cross-border approach is needed to ensure responsible AI development while maintaining regional stability and security.

ARTIFICIAL INTELLIGENCE TECHNOLOGY IS MOVING AT A FASTER PACE THAN AI LEGISLATION CAN KEEP UP...

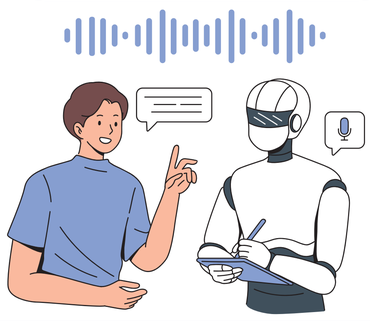
The solution?

Gulf Countries AI Council

The proposed Council will serve as a central body to unify and coordinate AI-related interests across the GCC by creating region-wide legislation to mitigate AI risks and address the fragmentation of individual national laws.
The Council will strike a balance between security and AI development. This initiative will not only ensure responsible AI use within the GCC but also set a global standard for coordinated AI governance.



GULF AI COUNCIL



WHO IT CONSISTS OF	WHAT THEY'LL DO
Gulf Countries	Work with fellow Gulf authorities to lead on and create region-wide legislation around AI use in military and commercial sectors. Develop GDPR/EU AI ACT style protective legislation.
Field Experts	Provide expertise on current debates in AI, from a technical standpoint. Offer insights on AI development, risks, and best practices, ensuring that regulations are both practical and forward-looking.
International org. eg UN	Provide clarity and guidance on existing military regulations, ensuring compliance with international law. Work alongside the council to develop new legislation specifically on AI use in warfare.
Business and trade	Ensure regulations support economic growth and innovation. Provide practical entrepreneurial advice, including a comprehensive handbook on best practices and regulatory compliance for start-ups.
Islamic Scholars	Integrate Islamic ethical principles and regional values in the overall work of the Council. Provide guidance on ethical issues such as fairness, accountability, and moral implications of AI.

HOW IT HITS THE SOLUTION THEMES

- **Building a knowledge base**: Shared AI legislation can create a shared knowledge base, since issues around AI governance are cross-border by nature of AI. A shared knowledge base would enable GCC countries to effectively address the challenges that arise with AI development and regulation.
- **Trust Building**: A unified AI framework builds trust, providing clarity for rational concerns around AI use in military, and tech businesses. Given the GCC's reliance on business under rentier theory, this trust fosters economic growth and innovation.

IMPACT AND LOOKING FORWARD

The proposed council will set a a global standard for coordinated AI governance, particularly in the Middle East, where AI regulation is scarce.
It would help GCC countries reduce reliance on oil by fostering a diversified economy. A unified AI regulatory framework would attract investment, support business growth, and enhance long-term economic stability.
Once frameworks of the council are set up, there is potential for non-GCC Middle East countries to either join (to create an even wider-reaching body), make their own council, or enact stronger AI legislation within their own borders.

ABOUT ME: LARA AREKAT

I'm Lara, a recent **MSc Middle East Politics grad** with a passion for international politics and research. Currently job hunting, I'm excited to find my next opportunity while staying curious about global issues and how they shape our world.

https://www.linkedin.com/in/lara-arekat/

laraarekat4@gmail.com

UNIDIR

Global Conference on AI, Security and Ethics

A Framework for Strategic Stability and Risk Mitigation in the Governance of Military AI in South Asia

Adeela Jawad

THE UNIVERSITY OF LAHORE

Ph.D. Candidate, International Relations, School of Integrated Social Sciences, The University of Lahore, Lahore, Pakistan.

ABSTRACT

In South Asia, where India and Pakistan Nuclear deterrent relationship is still very delicate, the development and deployment of Military Applications of Artificial Intelligence (AI), Cyber Capabilities, and Hypersonic Missiles into nuclear security initiatives poses serious threats to strategic stability. These new technologies raise uncertainties about unintentional escalation and miscalculations since they may speed up decision making, raise the possibility of first strike situations, and compromise crises management procedures. To mitigate risks, the poster gave a comprehensive governance structure that build on UNIDIR's multi-stakeholder approach to AI governance. It includes confidence building measures, to ensure openness and lower the possibility of misunderstanding, India and Pakistan crises communication channels would be strengthened to include talks on AI-driven support systems, cyber capabilities and hypersonic missiles. Defining explicit AI-nuclear thresholds includes monitoring hypersonic first strike capabilities to avoid destabilizing impact, preventing cyber interference in strategic systems, and AI automation for command and control. Establishing regional norms and accountability from global AI governance frameworks like UN's AI principles. South Asian code of conduct on Emerging technologies is being proposed to encounter the prudent use of military AI, Cyber restraint and the improvement of regional trust. In a fast changing technology world, this framework provided a specific way to improve strategic stability and lower the risks of conflict escalation by connecting South Asian security issues with more significant global AI governance initiatives.

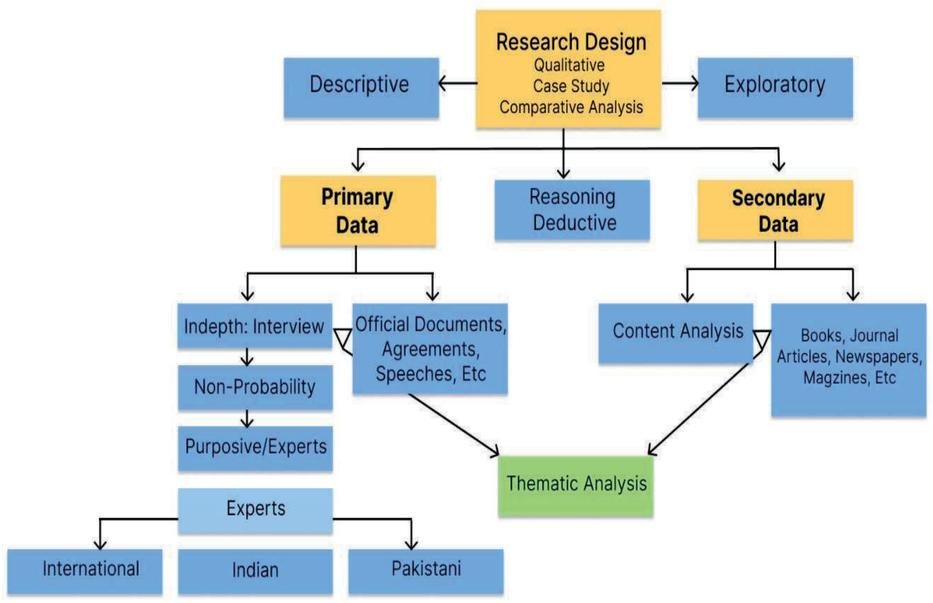
INTRODUCTION

India and Pakistan, both nuclear-armed nations, grapple with deep-rooted conflicts that pose significant challenges to resolution. Both have been involved in a series of nuclear conflicts and crises since 1998. These crises are concerning because tensions persist for a considerable time, increasing fears of escalation and military adventurism. Policymakers, practitioners, and researchers should thoroughly consider this South Asian context to ensure stability, prevent nuclear war, and uphold the taboo against nuclear weapon usage. In the 21st century, there is a noteworthy trend alongside the core threats and challenges: the increasing use of emerging technologies. The study intends to assess the impact of AI, cyber capabilities, and hypersonic missiles on strategic stability in South Asia. The objective is to develop a governance framework that mitigates the risks of miscalculation, crisis instability, and inadvertent escalation aligned with global AI governance principles. The lack of governance leads towards fragmented technological advancements and have grave consequences for regional and global security.

A Framework for Strategic Stability and Risk Mitigation

Confidence Building Measures (CBMs)	Defining Explicit AI-Nuclear Thresholds	Establishing regional norms and accountability

RESEARCH METHODOLOGY



Independent Variable: i)- Military applications of AI into decision making and NC2, ii)-Cyber offensive and defensive operations target nuclear infrastructure and iii)- hypersonic missiles leads towards first strike incentives. **Dependent Variable**: Strategic Stability between India and Pakistan (the probability of cries instability, inadvertent escalation)

Hypthesis: The integration of framework of governance of emerging technologies in India and Pakistan nuclear security initiatives is anticipated to effectively mitigate nuclear risks posed by military applications of AI, Cyber capabilities and hypersonic missiles thereby strengthen UN AI principles in South Asia.

Research Question: How can India and Pakistan work towards AI governance mechanisms and strengthen strategic stability in South Asia?

To mitigate risks, the poster gave a comprehensive governance structure that build on UNIDIR's multi-stakeholder approach to AI governance.

Confidence Building Measures: to ensure openness and lower the possibility of misunderstanding, India and Pakistan crises communication channels would be strengthened to include talks on AI-driven support systems, cyber capabilities and hypersonic missiles.

Defining Explicit AI-Nuclear Thresholds: includes monitoring hypersonic first strike capabilities to avoid destabilizing impact, preventing cyber interference in strategic systems, and AI automation for command and control.

Establishing Regional Norms and Accountability: from global AI governance frameworks like UN's AI principles. South Asian code of conduct on Emerging technologies is being proposed to encounter the prudent use of military AI, Cyber restraint and the improvement of regional trust.

Future Direction:

The proposed governance framework provides a region-specific approach to mitigate nuclear risks from emerging disrupting technologies in the third nuclear age. The research will have significant implications for policymakers, practitioners, and researchers in the fields of nuclear security, arms control, and international relations.

Global Conference on AI, Security and Ethics

The Vulnerability of AI Governance Depending on Safety Evaluations

Ashley Ferreira[1,2] (aferreira@cigionline.org)

1 Centre for International Governance Innovation

2 University of Waterloo

Abstract

Emerging research reveals critical limitations in current Artificial Intelligence (AI) safety evaluations, demonstrating that they frequently fail to accurately predict a model's trustworthiness and potential to cause harm. Despite this, AI governance initiatives, continue to rely on pre-deployment safety testing as a risk management strategy. This is not only an issue in the defense and security domain, but this is a domain it is especially consequential for. This poster critically examines the vulnerability that our overconfidence in these unreliable assessment methodologies poses to AI governance initiatives.

Relevance

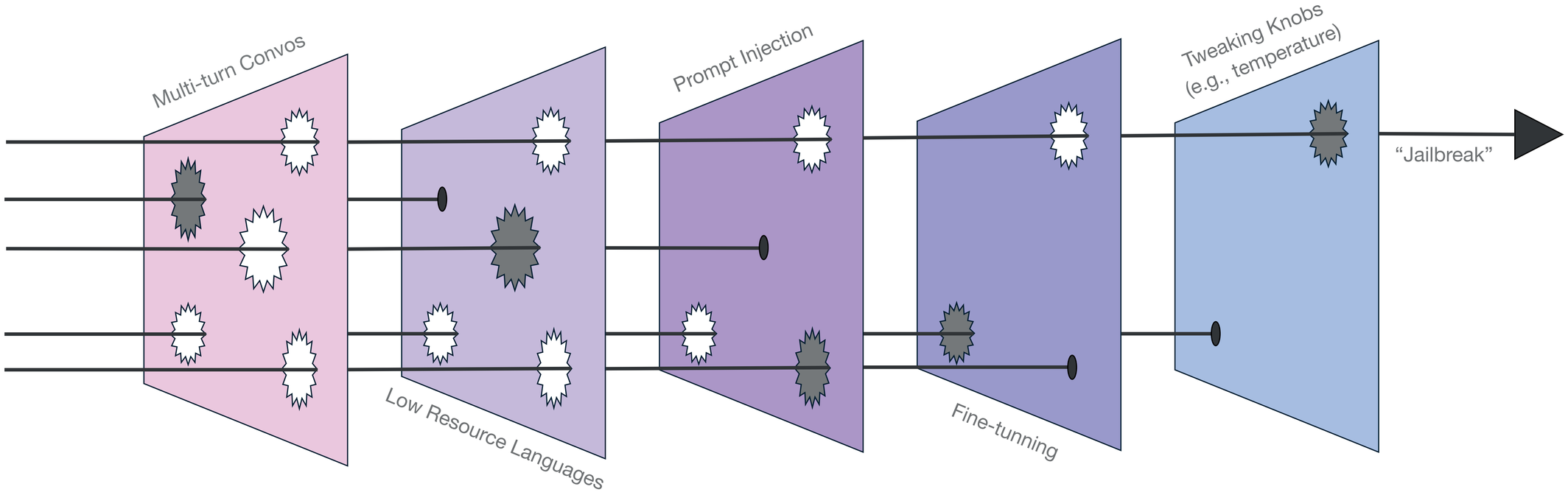
This research supports UNIDIR's Priority Area 6: Destabilization. By exposing a critical vulnerability in current AI governance approaches for defense and security, this work contributes to the recommended in-depth analysis of destabilization risks and encourages the development of concrete solutions for the mitigation and reduction of risk.

Early Career Researcher! Really open to feedback, please reach out and discuss!

Holes in AI Safety Measures

The pursuit of quantitative metrics to drive data-driven decision making for AI governance is commendable, however, there are currently blind spots in these metrics that include:

- Lack of uncertainties / error bounds
- No clear map from numbers to harm
- Hard to decouple from performance gains
- Impossible to measure all the variations…



Reliance on AI Safety Measures

While the following key international AI governance initiatives are largely fantastic documents, the table highlights their prominent mentions of AI safety evaluations and testing:

Example	Summit	Year	Direct Quotes	Additional Notes
Bletchley Declaration	AI Safety Summit	2023	"[…] AI systems which are unusually powerful and potentially harmful, have a particularly strong responsibility for ensuring the safety of these AI systems, including through systems for safety testing, through evaluations."	"The attendees of the Bletchley Park summit pledged to collaborate on testing frontier AI systems against various potential harms, such as those related to national security. They agreed that governments should play a major role in ensuring that safety testing regimens are fit-for-purpose." – Talha Burki, The Lancet Digital Health
Consensus Statement on AI Safety as a Global Public Good	International Dialogues on AI Safety (IDAIS) - Venice	2024	"A Safety Assurance Framework, requiring developers to make a high-confidence safety case prior to deploying models whose capabilities exceed specified thresholds." "Independent Global AI Safety and Verification Research, developing techniques that would allow states to rigorously verify that AI safety-related claims made by developers, and potentially other states, are true and valid."	Proposes to "consider setting up three processes to prepare for a world where advanced AI systems pose catastrophic risks." Two of the three (A Safety Assurance Framework, Independent Global AI Safety and Verification Research) focus on AI safety evaluations and are quoted in part. The third one (Emergency Preparedness Agreements and Institutions) which suggests the implementation of "model registration and disclosures, incident reporting, tripwires, and contingency plans."
Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy	Summit on Responsible AI in the Military domain (REAIM)	2023	"States should ensure that the safety, security, and effectiveness of military AI capabilities are subject to appropriate and rigorous testing and assurance within their well-defined uses and across their entire life-cycles."	Led by the US and signed by 58 states as of Nov'24, the declaration puts forward 10 measures of which the beginning of one is quoted. Additionally, the REAIM'23 Call to Action states "We recognise the need to assess the risks involved in the various types of current and future application of various AI techniques in the military domain […]"

Conclusion

- "Swiss cheese" model of AI safety guardrails is inadequate in protecting against known jailbreaks
- Existing safety evaluations cannot prove a model is safe under all conditions
- The appeal of safety metrics is undeniable, but they are unreliable and AI governance initiatives currently rely on these metrics
- This is a big vulnerability to defense & security

Next Steps

- Do not heavily rely on AI safety evaluations, it should not be assumed that it is possible to rigorously test and explain outputs from AI
- Assume dangerous AI will be deployed, potentially in the very near future
- Focus on emergency preparedness through continuous monitoring and rapid intervention protocols (enabled by hardware-level controls)

Learn More

- Heavily based on work from Peter Henderson & co
- Full references & more examples available

ashley-ferreira.github.io/AISE2025

27-28 March 2025
Palais des Nations,
Geneva, Switzerland

#AISE25

unidir.org

UNIDIR

Global Conference on AI, Security and Ethics

Establishing a Global Biosecurity Data Hub for Ethical AI adoption and cross-border collaboration

Delfina Hlashwayo, MSc, PhD[1]
Faculty of Sciences, Eduardo Mondlane University, Mozambique (delfina.hlashwayo@uem.ac.mz)

Why do we need this Hub?

The proposed centralization of epidemiological, genomic, and environmental data aims to facilitate real-time monitoring and timely detection of infectious disease outbreaks, with AI-driven analytics supporting evidence-based resource allocation and strengthening cross-border response efforts.

Concept and structure

1. Data collection
2. AI analysis
3. Early warnings
4. Policy recommendations

Data sources and role of AI

📊 **Sources:** [1] Epidemiological data (prevalence, incidence, and distribution of infectious diseases); [2] Genomic data of pathogens; [3] Environmental data (pathogens detected in environmental samples and information on disease vectors).

💻 **AI** would facilitate the harmonization of data across borders; enable predictive modeling and early warning systems.

Ethics and governance

The hub would follow international data protection regulations. Governance would involve a specialized international body, such as a United Nations agency, ensuring global oversight. An advisory committee of biosecurity, AI ethics, and international law experts would ensure ethical compliance, and equitable access to data.

Expected impact

- Facilitate cross-border collaboration in health data sharing.
- Improve early detection and response to infectious disease outbreaks.
- Support efficient resource allocation through AI-driven analysis.
- Strengthen global health security and preparedness strategies.

UNIDIR

Global Conference on AI, Security and Ethics

GC REAIM

GLOBAL COMMISSION ON RESPONSIBLE ARTIFICIAL INTELLIGENCE IN THE MILITARY DOMAIN

CG REAIM was launched by the Government of the Kingdom of the Netherlands following the Call to Action of the first REAIM Summit in 2023. The Government of the Republic of Korea followed up at the 2024 REAIM Summit with the adoption of the Blueprint for Action, which acknowledged the importance of GC REAIM's role.

Over the course of two years (*2024-2025*), GC REAIM has a clear and urgent mandate: to contribute to the search for actionable steps to solving some of the most pressing challenges associated with the integration of AI technologies in the military domain.

GC REAIM Commissioners and Experts form a globally diverse constituency with a wide range of expertise. Building on four workstreams, GC REAIM functions as an independent platform to foster mutual understanding and create a knowledge network among key stakeholders in government, military, industry, academia, and civil society. By linking dialogues between these communities, the Global Commission will contribute to an essential global task: supporting fundamental norm development and policy coherence.

Workstream Aims and Outputs

1 — Technological Foundations

Aim: to consolidate technical knowledge from AI and related fields to foster a better understanding of the technology and provide a coherent, agreed-upon basis for discussions.

Outputs: Taxonomy of current and future use cases of AI technologies in the military domain and debunking of common misunderstandings.

2 — Implications for Peace, Security, and Stability

Aim: to study the risks concerning the inherent scalable, dual-use, repurposable, and widely distributed nature of AI technologies in the military domain and beyond.

Outputs: Overview of risks from high-level strategic considerations to the level of the indivdiual and suggestions for risk-mitigation.

3 — Decision-making and Responsibility

Aim: to identify how some AI-based systems undermine conditions for human responsibility and develop a vocabulary that better captures the new complexities in AI-based military decision-making.

Outputs: Coherent and consistent normative vocabulary and mapping of current (counter) arguments on key concepts.

4 — Governance and Regulation

Aim: to clarify the meaning and application of both procedural and substantive aspects of international law and examine the role of international, regional and national institutions.

Outputs: Conceptual review of international law bodies, principles, instruments as well as governance frameworks.

The Commission
18 Commissioners

Responsible for agenda-setting, workstream guidance, substantive motor block.

The Expert Advisory Group
~40 Experts

Contribute through in-depth studies, consultation and revision.

The Secretariat
Provides administrative, logistical, and substantive support.

The Hague Centre for Strategic Studies

Timeline of Activities (2024-2025)

2024 Initial GC REAIM deliberations and meetings

Commission Meeting
Washington D.C., U.S.A / Brookings Institution (8-10th July. 2024)

Commission Meeting
Seoul, Republic of Korea / Seoul International Law Academy / **REAIM 2** (9-12th September, 2024)

Commission Meeting
Stellenbosch, South Africa / Defense AI Research Unit (11-14th November, 2024)

April-May 2025 Publication of 21 Expert Policy Notes

Expert Meeting
Edinburgh, Scotland / FCDO and University of Edinburgh (19-20th March, 2025)

Commission Meeting
Abu Dhabi, UAE / Trends Research and Advisory (6-8th May, 2025)

September 2025 Publication of Strategic Guidance Report

Commission Meeting
The Hague, The Netherlands / The Hague Centre for Strategic Studies (24-26th June, 2025)

REAIM 3, Spain (11-12th September, 2025)

GC REAIM's Strategic Guidance Report will consolidate the work of the four workstreams and provide actionable recommendations for practitioners throughout the AI lifecycle.

27-28 March 2025
Palais des Nations,
Geneva, Switzerland

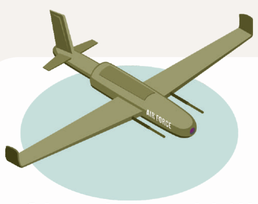
#AISE25

unidir.org

UNIDIR

Global Conference on AI, Security and Ethics

IMAGINE...



During an operation, autonomous drones glitch due to unique patterns on the traditional clothes of locals, and kill them. The drones were **extensively tested during review**, but the glitch was not discovered since the **hallucination only occurs when the patterns are viewed from a very specific angle. Some platoons reported similar incidents.**

A decision-support system (DSS) is used to recommend targets. **After extended use**, an NGO report reveals that many recommended targets were civilians. The DSS **worked perfectly during testing and prior operations**, but this time, officers had formulated their chat prompts agitatedly due to operational pressure, provoking the AI to recommend targets it was less confident about. **Even the AI's developers were shocked to learn this.**

HOW WOULD THESE INCIDENTS BE CHARACTERIZED?

OBSERVATION (1)

Using AI entails accepting **some unknown and unpredictable risks**

OBSERVATION (2)

These users are **good faith actors**, trying their best

OBSERVATION (3)

These are **structural issues**: the outcome will repeat under similar conditions

LIKELY CHARACTERIZATION

Genuinely unpredictable failure

+

Good faith & reasonable user

=

ACCIDENT

"Accidents and mishaps happen; it's a reality that we must accept"

ALTERNATIVE CONCLUSION

This is a structural problem!

+

Good faith & reasonable user

=

WE CAN DO BETTER

"We can and want to prevent this from happening again"

PROPOSAL :

ITERATIVE ASSESSMENT FOR MILITARY AI SYSTEMS

Accepts that AI systems can fail

Aims to Reduce Net Suffering without demanding reasonable decision-makers to know the unknowable

Does Not Prevent unknown failures from manifesting

BUT Enables swift and targeted mitigation once these manifest

complementary implementation

ITERATIVE REVIEW

TRADITIONAL REVIEWS

- Often done 1x
- Mostly ex ante

Re-review trigger:
- Factual change*
- Epistemic change*

ITERATIVE REVIEWS

- Done continuously
- Also ex post

Re-review trigger:
- Factual change*
- Epistemic change*

Adapt evaluation if necessary

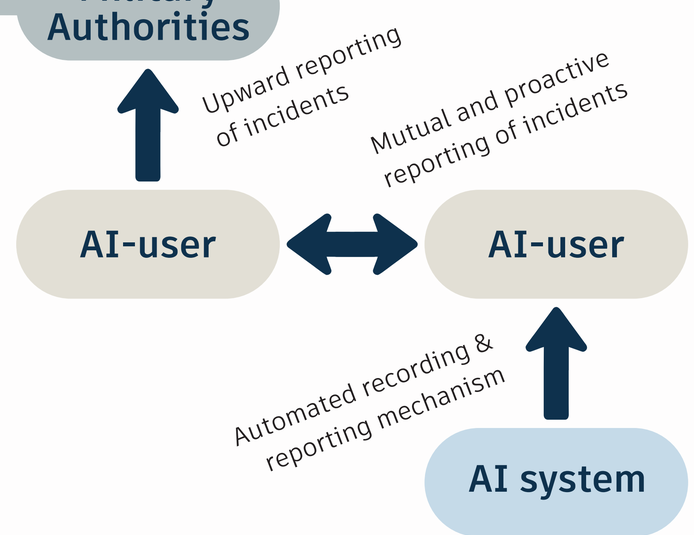
*__Factual change__ = The system or environment changes (e.g. software update, change in operational environment).

*__Epistemic change__ = New perception or understanding of existing behaviour/vulnerability that was previously unknowable (e.g. fabric pattern glitch, assertive prompt effect).

ITERATIVE ASSESSMENT IN DEPLOYMENT

Step 1: Iterative awareness

- **Monitor effects of AI use** using ISR assets after each operation
- Set up robust horizontal and vertical **reporting mechanisms**



Step 2: Proactive response

Commander/operator uses **discretion to determine most appropriate mitigation** depending on context

About the author:

Dr Jonathan Kwik is researcher in international law at the Asser Institute in The Hague. He specializes in techno-legal research on the military use of AI related to weapons, the conduct of hostilities, and operational decision-making.

Read the full article:

27-28 March 2025
Palais des Nations,
Geneva, Switzerland

#AISE25

unidir.org

UNIDIR

Global Conference on
AI, Security and Ethics

Technologiae Iuris Belli:

A Novel Juridical Approach for Governing AI in Security and Defense

Technological Distinction

Ensuring AI systems accurately differentiate combatants from civilians through legally mandated data validation and algorithmic precision.

Algorithmic Accountability

Defining legal responsibility among software developers, military operators, and states to establish a clear chain of accountability.

Technological Precaution

Implementing pre-deployment testing, certification, and regulatory oversight to ensure AI compliance with humanitarian law.

Algorithmic Transparency

Mandating traceability, auditability, and explainability of AI-driven military decisions to build trust and minimize bias.

Meaningful Human Control

Safeguarding critical military decisions (e.g., lethal force deployment) under accountable human supervision to prevent autonomous warfare escalation.

Jersain Llamas

Technologiae Iuris Belli presents a structured legal response to AI's growing role in military operations. By integrating law, ethics, and AI governance, this framework ensures responsible AI deployment, balancing security needs with fundamental human rights.

27-28 March 2025
Palais des Nations,
Geneva, Switzerland

#AISE25

unidir.org

UNIDIR

Global Conference on
AI, Security and Ethics

Operationalizing REAIM: A Context-Specific Assessment*

How to overcome the gap between overarching Principles of Responsible Use and their practical application within military operations?

Contact: DSCE@mindef.nl

Herwin Meerveld, Lonneke Peperkamp, Marie Šafář Postma, Roy Lindelauf

*Currently under review at *Ethics and Information Technology*

Problem

There is a gap between high-level frameworks, such as NATO's principles of responsible use (PRUs) of AI in the military domain, and the norms governing the practical use of AI in military operations.

MARC framework

The operationalization of PRUs requires a nuanced understanding of the military context in which AI is used. The Military AI Responsibility Contextualization (MARC) framework provides a structured yet adaptable approach.

Defining a military context

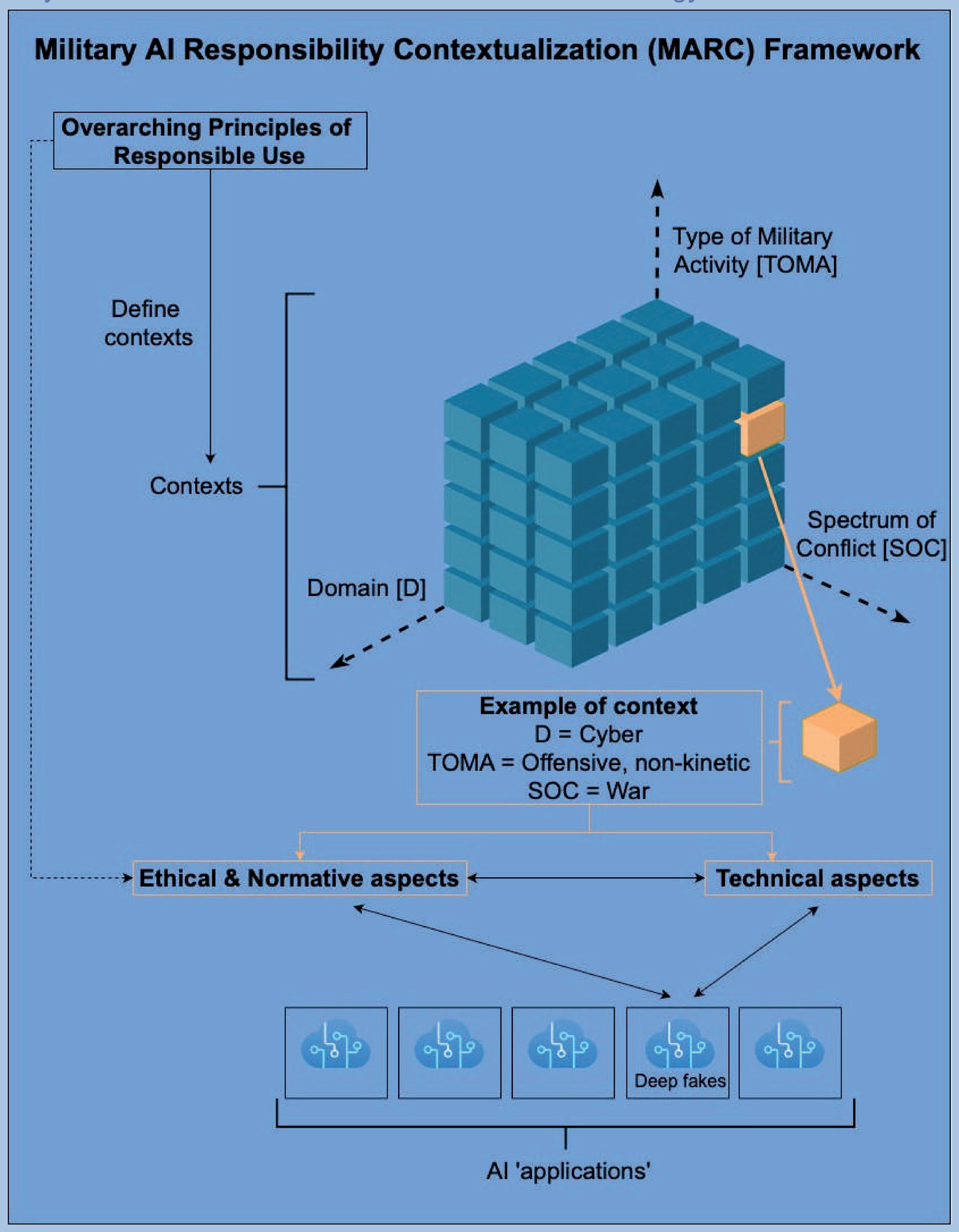
We use three key dimensions:

- Spectrum of conflict: peace, gray zone, war
- Operational domain: land, maritime, air, cyber, space
- Type of military activity: defensive kinetic, defensive non-kinetic, offensive kinetic, offensive non-kinetic, service & support

Interdisciplinary approach

Through interdisciplinary workshops we can assess all relevant aspects for each of the 75 distinct contexts within the MARC framework. Subject matter experts from diverse fields will be consulted depending on the specific context.

While this requires significant effort, a fully developed framework streamlines AI application analysis by clearly defining contexts, making requirements and guidelines immediately accessible.



Learning from incidents

- The MARC framework adapts by learning from AI incidents.
- An incident database can inform its refinement, integrating lessons to enhance ethical, normative, and technical guidelines.
- Continuous updates rely on international military and research collaboration.

UNIDIR

Global Conference on AI, Security and Ethics

Diplomacy in the AI Era: How AI Could Become a Game-Changer for Small Delegations in Multilateral Negotiations

Written by *Killian Foloppe*, Master's candidate at the Geneva Graduate Institute

Research question

To what extent can AI reduce power asymmetries in multilateral negotiations and enhance the effectiveness of small delegations?

Small delegations face:

- Asymmetric access to information
- Language & interpretation barriers
- Lack of expertise
- Limited financial & human resources

AI, through ML, DL, and NLP, can provide:

- Insights & data synthesis
- Document translation
- Virtual thematic counseling
- Predictive analysis of positions

Limits of the research

- Limited quantitative data
- Reliance on interviewees' perceptions
- Rapid evolution of AI

Armed conflicts, natural disasters, pandemics, the world is currently experiencing major crises that call into question the effectiveness of multilateralism.

By taking part in discussions on a potential reform of multilateralism, this study explores how AI could significantly facilitate the work of small delegations, helping to restore a real balance at the negotiating table.

Based on scholarly publications and interviews with permanent mission officials, this research presents the benefits of using AI both upstream and during negotiations, while also addressing its limitations and the ethical considerations that must be taken into account.

Expected results*

AI significantly helps small delegations achieve better results, particularly in complex negotiations, provided that data protection and ethical aspects are properly managed.

**Work in progress - Publication: Summer 2025*

killian.foloppe@graduateinstitute.ch

27-28 March 2025
Palais des Nations,
Geneva, Switzerland

#AISE25

unidir.org

UNIDIR

Global Conference on
AI, Security and Ethics

Military personnel, from general staff to troop level, urgently need immersive multidisciplinary training on the use of AI-based systems.

This can be achieved through wargaming.

Extended abstract:

Military training for Ethical AI Use

Background

There is broad consensus that training and education are key to the ethical and effective use of AI in the armed forces. UNIDIR's recent policy brief highlights the need for investment in AI literacy and capacity building (UNIDIR, 2024). While many advocate for continuous training covering social, political, institutional, and technical aspects (Klaus, 2024; Nadibaidze et al., 2024), there is still little guidance on how to structure and implement these programmes. In this spirit, this project aims to inform the design of training programmes and decision-making frameworks that foster safe and ethical use of AI in the military.

Initial findings

Wargaming has long been recognised as a useful training tool by militaries around the world, as evidenced for example by the many handbooks on the subject (BwCSC, 2006; UK MoD, 2017; USAWC, 2015). Its validity has also been widely argued for in the literature, including as a teaching tool for a wider range of topics (Combe, 2021; Curry, 2020; Fowler, 2024). Its immersiveness, flexibility, and familiarity to military personnel makes it perfectly suited to teach about the complex interactions of law, ethics, strategy, and tactics of AI systems. Although there are currently some military specific courses and projects on AI literacy such as the NPS or MIT/USAF programmes. Too often there is a temptation to adapt programmes from other environments, however these attempts do not produce ideal results (Abbe, 2021; Armendariz, 2023). To ensure their effectiveness it is crucial that these programmes are built from the bottom with the military context in mind.

Main reccomendations

Hands-on immersive learning

Cross-disciplinary approach

Military specific frameworks

Further research

Building on these findings, future research should focus on empirically testing wargaming's potential for AI literacy through game design and play. As well as producing a framework to implement these findings in military training practice. While recognising that developing tailored, immersive training tools for military personnel requires dedicated resources and access, this project ultimately seeks to bridge that gap.

Find out more here:

Michele C. Tripeni
University of Glasgow
Games and Gaming Lab

UNIDIR

Global Conference on AI, Security and Ethics

WEAPONS THAT THINK MUST NOT DECIDE: ENFORCING MEANINGFUL HUMAN CONTROL IN WARFARE

Adrian Klaits, Avi Perera, Heramb Podar, Patience Chepchirchir, Rujuta Karekar; *Stop Killer Robots Youth Network*

Meaningful human control (MHC) is key to ensuring autonomous weapons systems (AWS), (especially those capable of applying lethal force) remain accountable and compliant with **International Humanitarian Law**. Although a shared agreement seems to have emerged around the term's relevance among stakeholders, differences remain over the exact scope of **'meaningfulness'** concerning human control. This poster illustrates three actionable steps to implement MHC in order to **establish clear red lines of responsibility**, as well as **minimise risks to human life and dignity in situations of conflict**.

Distributed Decision-Making

Multi-person authorization and a responsible chain of command to prevent single points of failure.

MHC IN THE AWS LIFECYCLE

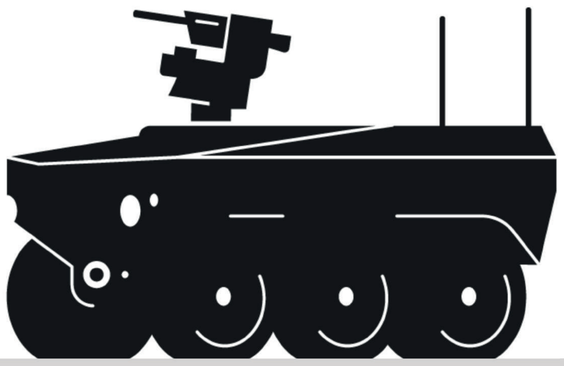
- **Development**
- **Testing and Training**
- **Deployment**
- **Post Deployment Assessment**

Human Override & Kill-Switch Mechanism

Operators must retain the ability to override or terminate AWS actions. States and manufacturers must be required to embed such an option into the system architecture.

Traceability of Decisions to a Human Source

Every AWS decision must be traced back to human oversight. This could be done through verifiable audit trails.



OPERATIONALIZING MHC

To ensure accountability, AWS deployment must integrate:

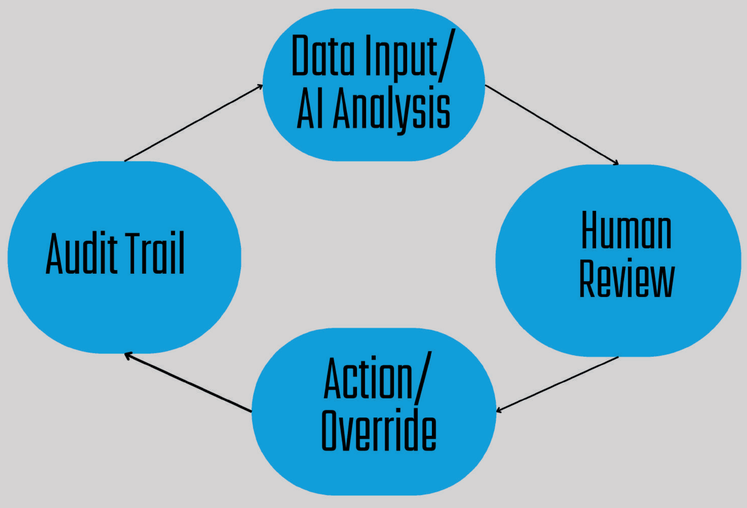
- **Multi-tiered authorization protocols** (e.g., distributed decision-making chains).
- **Pre-deployment ethical reviews** to assess risks to human dignity.
- **Real-time monitoring systems** for rapid human intervention.

FUTURE POLICY DIRECTIONS

Key recommendations for stakeholders:

- **Global MHC standards:** Harmonize definitions of "meaningfulness" in human control.
- **Transparency mandates:** Require public reporting on AWS testing and deployment.
- **Multi-stakeholder dialogue:** Foster collaboration between militaries, tech developers, and civil society.

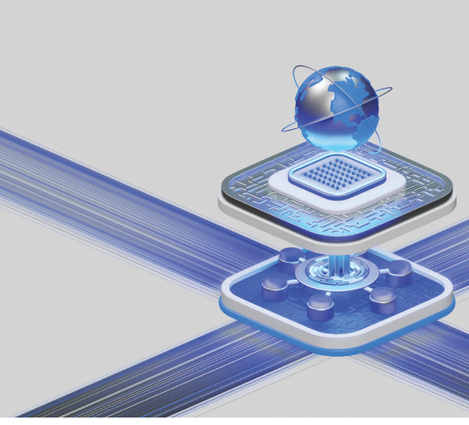
DECISION-MAKING PROCESS



27-28 March 2025
Palais des Nations,
Geneva, Switzerland

#AISE25

unidir.org



UNIDIR

Global Conference on AI, Security and Ethics

Artificial Intelligence (AI) and Cognitive Destabilisation in Hybrid Warfare

INTRODUCTION

This poster explores the idea of looking at AI systems as "cartographers" of perception. AI has come to play an active role in shaping cognitive landscapes and creating new realities. It influences what people see, believe, and even remember - way beyond merely distorting realities. Viewing AI this way highlights its potential to sow discord and fracture social cohesion. This emphasises the urgency to establish proper norms and protocols on AI use.

AUTHOR

Justin Huang
BA International Studies, Leiden University, The Netherlands
j.y.huang@umail.leidenuniv.nl
linkedin.com/justin-hyw

MAPS: THE ORIGINAL "CARTOGRAPHERS" OF PERCEPTION

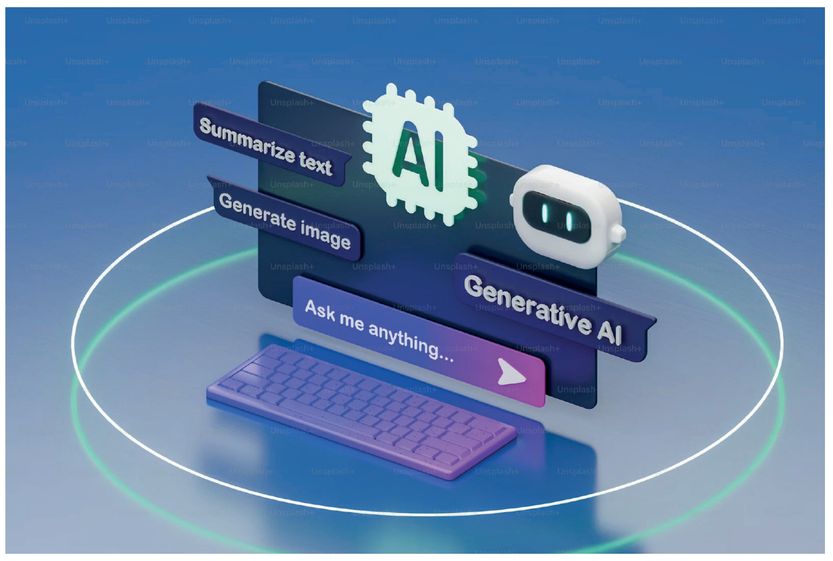
For Thongchai Winichakul, maps are not merely neutral mediums that help people make sense of spatial reality. Even viewing maps as distorting reality would be inaccurate. To him, maps in fact go so far as to create particular meanings and realities.

***"A map of a nation presupposes the existence of boundary lines. Logically this suggests that boundary lines must exist before a map, since a medium simply records and refers to an existing reality. Yet reality is a reversal of that logic. It is the concept of a nation...that requires having boundary lines clearly demarcated. A map may not just function as a medium, it could well be the creator of the supposed reality."* (Winichakul 1994, 56)**

AI: THE MODERN "CARTOGRAPHERS"OF PERCEPTION

Similarly, AI's ability to "map" perceptions extends beyond distortion. It does not simply filter content or reinforce biases - it can tailor entirely new ways of looking at the world. In other words, AI is capable of creating entirely new cognitive landscapes.

AI can filter and manipulate information, often without users' awareness. AI can also help manufacture deepfakes of trusted sources, eroding trust and sowing discord in our societies. Given enough computing power, AI can massively amplify the scale of such information operations.



ANALYSIS

This alternative conceptualisation of AI as a "cartographer" of perception helps us better grasp AI's true nature and capabilities. AI can (re-)define reality itself, or at least how we perceive reality. Its ability to manipulate of cognitive landscapes brings into question the autonomy of our own beliefs and perceptions.

Establishing norms and protocols for the use of AI in areas such as information warfare becomes a more pressing matter. It is critical to establish robust regional, cross-regional, and international partnerships to establish "rules of the road". Lastly, to preserve human agency in this age of AI, integrating and ensuring human oversight into AI systems is crucial. AI must be kept accountable and transparent.

Concerns over how war and conflict take place can also be considered. The increasing use of AI to optimise processes and improve efficiency, for example, may be "making the state and war incidental to warfare," as AI creates new realities in the context of war.

LITERATURE:

Afina, Yasmin, and Giacomo Persi Paoli. 2024. "Governance of Artificial Intelligence in the Military Domain: A Multi-Stakeholder Perspective on Priority Areas." UNIDIR Policy Brief.
Barrett-Taylor, Rupert. 2024. "How AI and Automation are Making the State and War Incidental to Warfare." Lecture posted on YouTube on November 15, 2024, by T.M.C. Asser Institute. https://www.youtube.com/watch?v=hI_V3DMkqUY.
Nordin, Astrid H.M., and Dan Öberg. 2015. "Targeting the Ontology of War: From Clausewitz to Baudrillard." *Millennium* 43 (2): 392-410. https://doi.org/10.1177/0305829814552435.
Pauwels, Eleonore. 2024. "Preparing for Next-Generation Information Warfare with Generative AI." Centre for International Governance Innovation. https://www.cigionline.org/static/documents/Pauwels-Nov2024.pdf.
Winichakul, Thongchai. 1994. *Siam Mapped: A History of the Geo-Body of a Nation*. University of Hawai'i Press.

IMAGES:

Donders, Timme H. et al. 2014. "Region-Specific Sensitivity of Anemophilous Pollen Deposition to Temperature and Precipitation." *PLoS ONE* 9 (8). https://doi.org/10.1371/journal.pone.0104774.
Oroni, Philip. 2024. Unsplash+. https://unsplash.com/photos/a-computer-keyboard-sitting-on-top-of-a-computer-mouse-AMAYQqzQYaI.

UNIDIR

Global Conference on AI, Security and Ethics

AI and Border Security:

Assessing Risks and Governance in Uttarakhand, India — A Critical Border Region with China and Nepal

Abstract:

Based on my experiences at the 2023 G20 Summit in India and the 2021 G20 Global Leadership program in South Korea, where I visited the Korean Demilitarized Zone (DMZ), this paper examines the increasing role of artificial intelligence (AI) in border security, particularly in Uttarakhand, a region bordering China and Nepal. All data used in this analysis is sourced from publicly available information. It discusses how AI-driven technologies, such as Lethal Autonomous Weapon Systems (LAWS), are being used to enhance military operational capabilities and efficiency. However, it also details the potential risks, such as escalation and misidentification, that might arise from the autonomous nature of these technologies in sensitive areas. The example of Barahoti, a high-altitude, contested area along the Sino-Indian border in the Chamoli district of Uttarakhand, highlights the challenges of potential deployment of AI technologies in disputed regions. The aim here is to spark discussions on the need for comprehensive AI governance frameworks that prioritize stability and conflict prevention, emphasizing international cooperation and the development of common standards to manage AI risks in military applications.

Global military expenditure has increased from 1.9 trillion USD in 2015 to 2.4 trillion USD in 2023 (SIPRI, 2025), and based on the current geopolitical situation, I estimate it will reach around 3.5 trillion USD by 2030, further accelerating the adoption of AI in military applications worldwide. My hypothesis is that while AI can enhance international border security efficiency, it also risks conflict escalation without stringent governance. By presenting a theoretical framework and encouraging meaningful dialogue, the analysis seeks to explore how AI might both strengthen and compromise regional security.

The policy recommendations include establishing an Indo-China AI Border Security Council (ICABSC) to oversee bilateral AI governance and deployment issues, aligning AI deployments with international norms and promoting regional peace. This initiative may pave the way for a global AI Safety Agency, similar to the International Atomic Energy Agency (IAEA), and lead to a treaty on safe AI usage, akin to the Non-Proliferation Treaty. Both countries should also consider joining the 'Responsible Military Use of AI and Autonomy' political declaration from REAIM 2023 to ensure ethical AI use in military operations. Further, prioritizing the development of domestic, specialized AI technologies specifically tailored for military applications is necessary. This strategic focus will enable more precise and secure advancements in border security capabilities, rather than diluting resources across the development of all-purpose large language models. Moreover, establishing tourist spots similar to Korea's Dora Observatory along the India-China border could foster peace and mutual understanding through educational tourism and cultural exchanges. Lastly, this paper connects with unique security challenges of the Asia-Pacific region, which can also be extended to other contested border areas globally.

JEL Codes: F52, O33, F55 **Keywords:** AI border security, LAWS, AI governance, Conflict prevention, Indo-China relations

I. Methodology and data sources:

I have used both quantitative and qualitative analyses. Data were collected exclusively from publicly accessible sources, including academic papers, government reports, and international publications.

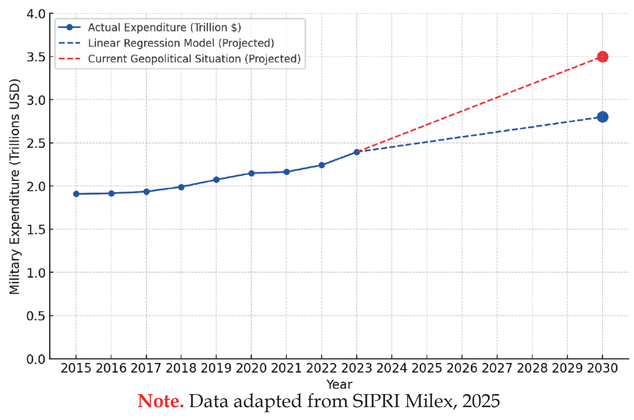
II. Contextual background and Analysis of the Current Scenario:

Situated in the heart of the Himalayas, Uttarakhand state in India is not just a region of immense natural beauty but also a crucible of geopolitical tensions due to its borders with China and Nepal. This unique positioning makes it an ideal case study for examining the implications of emerging technologies like Artificial Intelligence (AI) in border security. The proximity to international borders makes it a critical area for national security concerns and regional diplomacy. The state's terrain is predominantly mountainous, a feature that presents both challenges and strategic advantages in border security operations. For India, the Indo-Tibetan Border Police (ITBP) and the Indian Army secure India's border with China, while for China, it is the Ground Force of the People's Liberation Army (PLA).

Since the independence of India in 1947 and the establishment of China in 1949, both countries have grappled with border disputes along the Line of Actual Control (LAC), a legacy of colonial-era demarcations and intensified by the 1962 Sino-Indian War. The LAC witnessed severe escalations, the recent one being a deadly skirmish in June 2020 in the Galwan Valley where 20 Indian and 4 Chinese soldiers died. Despite some tactical pullbacks agreed upon in 2021, tensions flared again, with incidents reported by various newspapers, such as the transgression by PLA soldiers in Barahoti, Chamoli district, Uttarakhand. Barahoti is part of the 'middle sector' and is considered a demilitarized zone. This incident, along with Yangtse clash near Tawang, Arunachal Pradesh in 2022, highlights ongoing and unresolved tensions. Both countries continue to strengthen their military capabilities along this frontier, conducting several high-level military talks through China's Western Theater Command, which focuses on the region. These persistent disputes reflect the strategic importance both nations place on securing their perceived territories.

The 2025 Global Firepower Index ranks the USA, China, Russia, India, and South Korea as the top 5 countries respectively out of 145 in terms of military strength. This is complemented by the Lowy Institute Asia Power Index, which ranks countries based on military capability. The USA has the highest defense budget of $855.51 billion in 2025. India and China have almost the same population of around 1.4 billion people, but China's defense budget ($231 billion in 2024) far exceeds India's ($75 billion in 2025), demonstrating a significantly higher financial commitment to military expansion. With an estimated 2.04 million current active personnel, China also maintains a much larger standing force compared to India's 1.46 million, providing it with a greater operational capacity. However, while China's external debt stands at $2.54 trillion (as of June 2024), it is supported by China's massive GDP of $17.79 trillion (2023) making it more manageable. In contrast, India's external debt of $711.8 billion (as of September 2024), while smaller in absolute terms, represents a larger proportion of India's GDP ($3.57 trillion), potentially limiting its economic flexibility and long-term ability to sustain military spending. Also, India's trade deficit with China was $101.28 billion in 2022, adding another layer of economic challenge.

Figure 3. Global Military Expenditure with 2030 Projections (Constant 2021 US$)



Note. Data adapted from SIPRI Milex, 2025

Table 1. India vs China: A Comparative Analysis of Key Indicators

S.No	Category	India	China	Difference for India
1.	Population in Million (UNFPA, 2024)	1441.7	1425.2	16.5
2.	Defense Budget in Billion USD (Official Government data)	75 (2025)	231 (2024)	- 156
3.	Est. Active Military Personnel in 2025 (Global Firepower.com)	1,455,550	2,035,000	- 579,450
4.	External Debt in Billion USD (Official Government data)	711.8 (Sep 2024)	2545.3 (June 2024)	Not Applicable
5.	2023 GDP in current US$ Trillion (World Bank)	3.57	17.79	- 14.22
6.	2023 Per Capita GDP in current US$ (World Bank)	2480.8	12614.1	- 10133.3
7.	Trade Deficit in Billion USD in 2022 (Official Government data)	101.28	0	- 101.28
8.	Government AI Spending announced (Billion USD)	1.2 (2024-2029)	8.2 (from 2025)	- 7

Note. Data adapted from various sources

III. AI and its applications in the military domain

Artificial Intelligence has emerged as a potent tool, especially with the launch of large language models like ChatGPT in 2022 and China's DeepSeek in 2025. The US announced an investment of $500 billion through its Stargate project, the EU announced a £200 billion investment through its InvestAI program, and China has announced a $8.2 billion AI investment fund in 2025. India has launched the IndiaAI mission with a budget of $1.2 billion. These all are exclusive of independent AI investments by tech giants like Microsoft and Alibaba, as well as billions of dollars raised by startups. According to the 2024 Stanford University HAI Global AI Power Rankings, the Stanford HAI Tool ranks 36 countries in AI, with the top five being the US, China, UK, India, and UAE respectively. In 2022, China led in global AI patent origins with 61.1%, considerably outpacing the United States, which accounted for 20.9% of AI patent origins, and India, which accounted for 0.23%. However, from 2015 to 2023, the countries with the highest AI skill penetration rates were India (2.8) and the United States (2.2). In 2023, the number of newly funded AI companies by geographic area saw the US leading with 897, followed by China with 122, and India with 45. The 2024 Global Innovation Index by WIPO ranks Switzerland first, followed by Sweden and the USA. China is ranked 11th, significantly ahead of India in 39th place. China is also ranked ahead (68th) of India (109th) in the 2024 UN Sustainable Development Report and is one of the five permanent members of UN Security Council. These rankings highlight the important role of AI development in influencing global economic, military, and geopolitical dynamics. AI Safety Summits have been hosted by the UK, South Korea, and France, with India next in line, emphasizing a global commitment to responsible AI. The EU AI Act is the most comprehensive legal framework for AI regulation, but it excludes military applications.

As nations globally accelerate their adoption of AI, understanding its potential to enhance border security is crucial. According to the World Economic Forum's Global Risks Report 2025, cyber espionage and warfare is one of the top global risks in the next 2 years and 10 years. AI technologies at international borders can improve surveillance, monitoring, and data analysis capabilities through various advanced means, significantly increasing operational efficiency and response times. These technologies include drone-based automated surveillance systems that can monitor remote and rugged terrains, facial recognition systems that enhance the identification of persons of interest, and predictive analytics deployed via integrated software systems that help in anticipating security threats based on comprehensive data analysis. Embracing these technologies may strengthen border security but might also present challenges.

In 2018, the U.S. Department of Defense (DoD) released its Artificial Intelligence Strategy, warning that failing to adopt AI could make legacy systems obsolete, erode cohesion among allies, and reduce market access, leading to a decline in prosperity. Meanwhile, China's 2017 Next Generation Artificial Intelligence Development Plan aims for significant AI theory breakthroughs by 2025, positioning itself as the global AI innovation leader by 2030. In India, the Defence Artificial Intelligence Council (DAIC) and the Defence AI Project Agency (DAIPA) were established in 2019. In 2022, Ministry of Defence launched 75 AI technologies at the inaugural 'AI in Defence' (AIDef) symposium and recently hosted a seminar on AI's role in military strategies and emerging technologies, including Lethal Autonomous Weapon Systems (LAWS). This global AI arms race highlights the urgent need for nations to evaluate the security, ethical, and societal impacts of these technologies.

The integration of AI into border security introduces several complexities. The autonomous nature of these AI systems, which operate with minimal human oversight, raises important concerns about control and ethical use. Specifically, the deployment of Lethal Autonomous Weapon Systems (LAWS) — weapons that can select and engage targets without human intervention — in sensitive and geopolitically tense regions poses risks. These include the potential risks like:

- *Escalation:* Autonomous AI systems, especially those capable of defensive or offensive actions, could act on data without human oversight, potentially leading to unintended escalations. For example, an AI-driven system might misinterpret a benign activity as a threat and respond aggressively, prompting a disproportionate response from the other side.
- *Misidentification:* AI technologies such as facial recognition and movement pattern analysis rely heavily on algorithms that can sometimes produce errors. These errors could lead to the wrongful identification of individuals as threats, which can have serious consequences for individuals and diplomatic relations. Misidentification is particularly concerning in border areas where frequent civilian crossings occur.

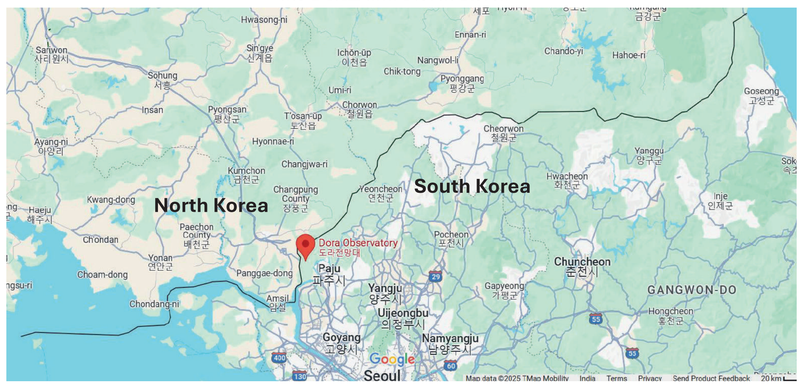
'The Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy' launched in February 2023 at the Responsible AI in the Military Domain Summit (REAIM 2023) is an international agreement that outlines guidelines for the ethical and responsible use of AI in military settings. It has been endorsed by 58 countries including the United States and South Korea as of November 2024, but India and China are not part of it. This highlights the need for robust international governance frameworks that build on these guidelines to ensure AI technologies are used effectively and safely, with clear human oversight to mitigate risks.

Figure 1. *Map showing the contested area of Barahoti along the Indo-China border in Uttarakhand, India*

Note. Map adapted from Google Maps, 2025

Disclaimer. These maps are for illustration purposes only. The boundaries and designations shown do not imply official endorsement by UNIDIR.

Figure 2. *Dora Observatory: A Strategic Viewing Point on Dorasan, Paju – Overlooking the DMZ and North Korea*



Note. Map adapted from Google Maps, 2025

Scan the QR Code for full access to the paper, which details the policy recommendations, theoretical governance framework, implementation roadmap, challenges, and mitigation strategies.

IV. Conclusion and Call to Action:

Quoting the Bible's Matthew 16:26 from the King James Version (KJV) where Jesus said, "*For what is a man profited, if he shall gain the whole world, and lose his own soul? or what shall a man give in exchange for his soul?*" - I believe AI has the potential to benefit humanity, but not if it comes at the cost of compromising human rights, individual freedom, or accountability for flawed algorithms and their potential negative outcomes. According to the 2024 State of Food Security and Nutrition in the World (SOFI) report, approximately 735 million people may have faced hunger worldwide in 2023, with projections suggesting nearly 582 million will be chronically undernourished by 2030. If even a fraction of the enormous military budgets were used to eradicate hunger, the world would be in a much better position today. The AI revolution should not take priority over humanitarian issues worldwide, including the urgent need to invest in achieving the UN Sustainable Development Goals (SDGs) by 2030.

Nonetheless, I believe AI is a double-edged sword, and its governance is of paramount importance, especially in the military domain, to prevent a potential World War III dominated by cyberwarfare. Further, qualitative methods like semi-structured interviews with stakeholders can be conducted to enhance our understanding of AI in border security.

V. References:

1. Afina, Y., & Persi Paoli, G. (2024). Governance of artificial intelligence in the military domain: A multi-stakeholder perspective on priority areas. United Nations Institute for Disarmament Research (UNIDIR). https://unidir.org/publication/governance-of-artificial-intelligence-in-the-military-domain-a-multi-stakeholder-perspective-on-priority-areas/
2. Ahluwalia, V. K. (2021, October). Boundary dispute of Barahoti Bowl in Central Sector: An assessment. Centre for Land Warfare Studies. https://www.claws.in/static/IB-305_Boundary-Dispute-of-Barahoti-Bowl-in-Central-Sector-An-Assessment.pdf
3. AI Index Steering Committee. (2024). Artificial intelligence index report 2024. Stanford Institute for Human-Centered Artificial Intelligence. https://aiindex.stanford.edu/report/
4. British Army. (2023, October). British Army's approach to artificial intelligence: A guide to accelerate the Army's adoption of AI and get the Army AI ready. https://www.army.mod.uk/media/24745/20231001-british_army_approach_to_artificial_intelligence.pdf
5. European Commission. (2025, February 11). EU launches InvestAI initiative to mobilise €200 billion of investment in artificial intelligence. https://ec.europa.eu/commission/presscorner/detail/en/ip_25_467
6. Global Firepower. (2025). 2025 Military Strength Ranking. https://www.globalfirepower.com/
7. Grand-Clement, S. (2023). Artificial intelligence beyond weapons: Application and impact of AI in the military domain. United Nations Institute for Disarmament Research. https://unidir.org/publication/
8. Hooda, D. S. (2023, February 16). Implementing artificial intelligence in the Indian military. Delhi Policy Group. https://www.delhipolicygroup.org/publication/policy-briefs/implementing-artificial-intelligence-in-the-indian-military.html
9. Kim, D. (2012). The Demilitarized Zone: Redrawing the 151 mile border between North and South Korea. Harvard University Graduate School of Design. https://cga-download.hmdc.harvard.edu/publish_web/Annual_Spring_Workshops/2012_Africa/posters/Kim_Dongsei.pdf
10. King, A. (2024). Digital targeting: Artificial intelligence, data, and military intelligence. Journal of Global Security Studies, 9(2), ogae009. https://doi.org/10.1093/jogss/ogae009
11. Manohar Parrikar Institute for Defence Studies and Analyses. (2025, January 28–30). India's approach to AI in military domain & emerging technologies in areas of lethal autonomous weapon systems (LAWS). https://www.idsa.in/idsa-event/indias-approach-to-ai-in-military-domain-emerging-technologies-in-areas-of-lethal-autonomous-weapon-systems-laws/
12. Ministry of Defence, India. (2022). Artificial intelligence in defence. AI in Defence Symposium & Exhibition, New Delhi. https://www.ddpmod.gov.in/publication/artificial-intelligence-defence
13. OpenAI. (2024). Announcing the Stargate project. https://openai.com/index/announcing-the-stargate-project/
14. Press Information Bureau, Delhi. (2024, March 7). Cabinet approves ambitious IndiaAI mission to strengthen the AI innovation ecosystem. Retrieved March 5, 2025, from https://pib.gov.in/PressReleaseIframePage.aspx?PRID=2012355
15. South China Morning Post. (2025). Tech war: China creates US$8.2 billion AI investment fund amid tightened US trade controls. Retrieved from https://www.scmp.com/tech/big-tech/article/3295513/tech-war-china-creates-us82-billion-ai-investment-fund-amid-tightened-us-trade-controls
16. Sustainable Development Report. (2024). Sustainable Development Goals (SDG) Index rankings. https://dashboards.sdgindex.org/rankings
17. The Hindu. (2021, October 3). Chinese transgressions testing India, say officials. https://www.thehindu.com/news/national/chinese-transgressions-testing-india-say-officials/article36808989.ece
18. The Indian Express. (2017, August 1). Sikkim border Doklam standoff: Barahoti a disputed area, no clear demarcation on which part belongs to China or India: Uttarakhand CM. https://indianexpress.com/article/india/sikkim-border-doklam-standoff-barahoti-a-disputed-area-no-clear-demarcation-which-part-belongs-to-china-or-india-uttarakhand-cm-4777817/
19. Tian, N., Lopes da Silva, D., Liang, X., & Scarazzato, L. (2024, April). Trends in world military expenditure, 2023. Stockholm International Peace Research Institute. https://www.sipri.org/sites/default/files/2024-04/2404_fs_milex_2023.pdf
20. U.S. Department of Defense. (2024). Annual report to Congress: Military and security developments involving the People's Republic of China 2024. https://media.defense.gov/2024/Dec/18/2003615520/-1/-1/0/MILITARY-AND-SECURITY-DEVELOPMENTS-INVOLVING-THE-PEOPLES-REPUBLIC-OF-CHINA-2024.PDF
21. U.S. Department of State. (2023, February). Political declaration on responsible military use of artificial intelligence and autonomy. Responsible AI in the Military Domain Summit (REAIM 2023), The Hague. https://www.state.gov/bureau-of-arms-control-deterrence-and-stability/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy
22. United States Institute of Peace. (2022, December). Another clash on the India-China border underscores risks of militarization. https://www.usip.org/publications/2022/12/another-clash-india-china-border-underscores-risks-militarization
23. Walker, N. (2025, February 21). Conflict in Ukraine: A timeline (current conflict, 2022–present). House of Commons Library. https://commonslibrary.parliament.uk/research-briefings/cbp-9847/
24. World Bank. (n.d.). GDP per capita (current US$). World Bank Open Data. Retrieved from https://data.worldbank.org/indicator/

Author: Naveen Kumar Samuel Kori, Policy Specialist | **E-mail:** *naveenkori.g20@gmail.com*
Note: *The views and opinions expressed in this paper are solely those of the author, including any errors.*

UNIDIR

Global Conference on
AI, Security and Ethics

AI-Powered Weaponization Implications and Capabilities Needs in Africa

Ernest TAMBO[1,2]&, Kennedy OKORIE[1], Clarence YAH[3] and Oluwasogo OLUWASOGO[4]

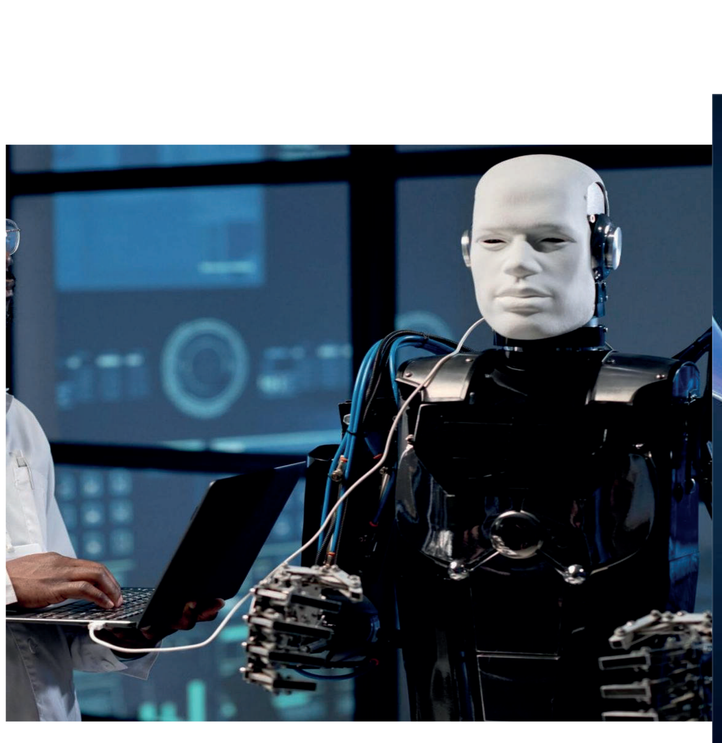
1Africa Disease Intelligence, Preparedness and Response, Cameroon
2University of Global Health Equity, Butaro/Kigali, Rwanda
3University of Pretoria, Pretoria, South Africa
4School of Public Health, Kwara State University, Malete,Nigeria
Corresponding author- tambo0711@gmail.com

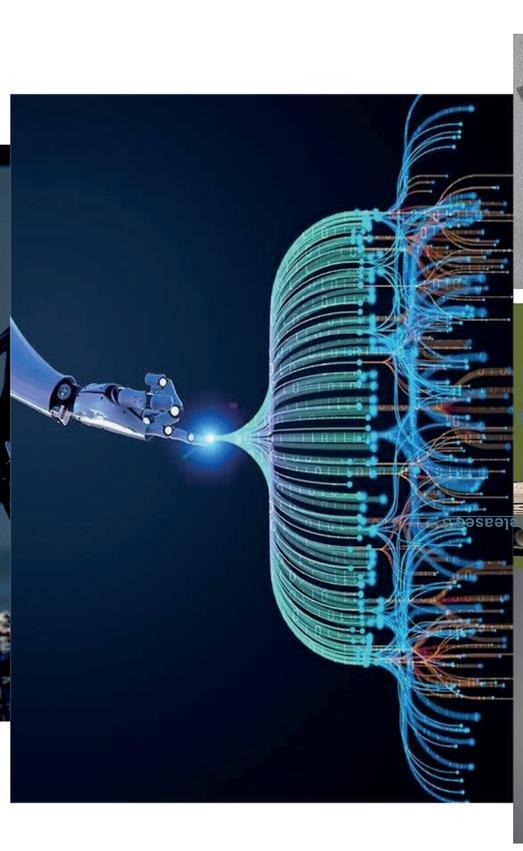
Introduction - The scale and sophistication of cyberattacks, threats, and cybercrime continue to drive the profitability of ransomware, intellectual property theft, and data breaches, raising significant business concerns. There is an urgent need to enhance cyber resilience and defense systems by prioritizing and investing in advanced AI-driven cybersecurity technologies. Governments and critical organizations must strengthen their cyber defense postures, as various complex systems and technologies are becoming increasingly vulnerable to cyber incidents and attacks. The weaponization of AI is a growing concern, with risks including AI-powered surveillance, deepfake-driven disinformation campaigns, autonomous weapons, human rights abuses, and cyberattacks. These threats could exacerbate existing conflicts, undermine democratic processes, and worsen political instability, particularly in Africa and other regions with weak governance structures. Despite the severity of these risks, little has been documented on the militarization of AI in warfare, particularly regarding autonomous weapons, cyberattacks, and conflict scenarios. Addressing these challenges requires collaboration among military end-users, AI developers, regulators, consumers, and affected communities across Africa.

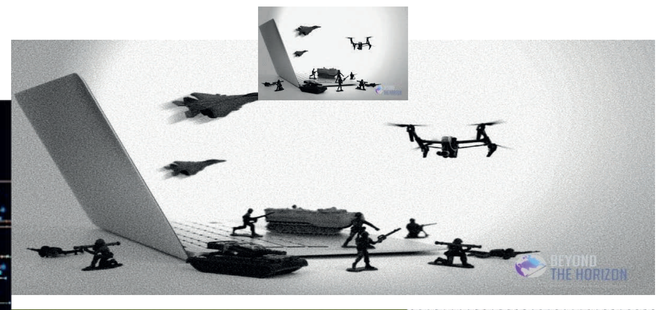
Objective - This article examines how AI-powered weapons and cybersecurity solutions affect defense and security. It analyzes cyber threats and vulnerabilities while exploring policy and ethical frameworks to reduce risks. The study also looks at how AI and machine learning can be used for positive development, particularly in critical infrastructure in the African region.

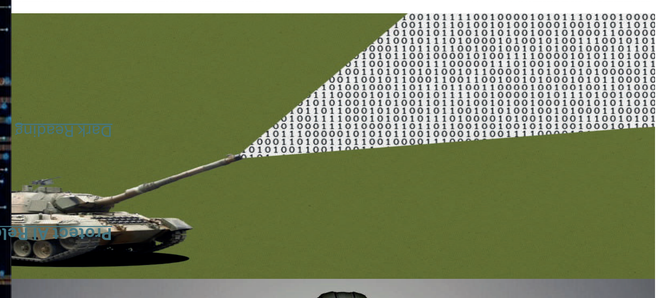
Methods - A generative AI and ML model was developed and trained to assess cyber threats in critical infrastructure and evaluate cybersecurity needs. The study employed a comprehensive approach using linear regression analysis, simulation, and optimization of AI and ML/NLP systems. The methodology incorporated support vector machines, random forests, artificial neural networks, and decision-making algorithms.

Results and discussions - Findings showed Africa experienced a 35% increase in cyberattacks per week per organization compared to the same period in 2020. The targets are mainly Government agencies, telecommunications companies and financial organizations (eg: banks) leading to huge financial losses and operational disruption and reputation damage. Malicious criminal actors leverage AI capabilities to automate, enhance, and personalize targeted cyber-attacks, making them more difficult to detect and defend against. Reported cyberattacks features included targeted phishing and blackmail (37%), deepfakes and fake email accounts (21%), social engineering (9%), malware (15%), and ransomware (19%). AI weaponization, particularly the development of Lethal Autonomous Weapons Systems (LAWS), raises serious ethical concerns, especially in conflict zones. Additionally, sophisticated AI-driven cyberattacks pose a significant threat to essential services, critical infrastructures, and national security. Other risks include AI-powered facial recognition surveillance, deepfake-based misinformation campaigns, and AI-generated fake news, which could undermine elections and social cohesion. Despite these challenges, AI also presents beneficial applications in healthcare, education, agriculture, and disaster management. However, several key vulnerabilities remain, including: lack of a security-enabling environment, limited awareness of cyber hygiene, inadequate cybersecurity infrastructure, low digital security awareness, shortage of skilled cybersecurity professionals. These gaps make data security an urgent priority, highlighting the need for robust cybersecurity measures, skilled professionals, and public education initiatives across Africa. Addressing AI weaponization and the digital divide requires greater transparency, ethical oversight, and regulatory measures. AI can enhance military capabilities by enabling faster decision-making, improving targeting accuracy, and optimizing resource allocation. However, without proper governance, these advancements could also intensify global security risks.









Conclusion - This article highlights Africa's rapid adoption of AI to enhance content creation, improve public service delivery, develop AI-enabled military capabilities, and streamline business processes. There however is an urgent need to leverage AI-powered cyber defense and cyber wellness solutions to combat increasingly sophisticated cyber threats, while strengthening database protection, security efficiency, and overall resilience both regionally and globally.

To ensure responsible AI deployment, investing in strong policies and regulatory frameworks for AI weaponization and cybersecurity is essential. Enforcing robust data protection laws and strengthening capacity-building efforts through awareness campaigns and coordinated partnerships will be crucial in fostering effective and responsible AI use. Additionally, proactive threat monitoring, predictive intelligence, and preventive cybersecurity measures must be prioritized. These strategies will help address Africa's cybersecurity vulnerabilities by enhancing training initiatives, collective defense efforts, and cooperative security programs, all while aligning with international laws and bolstering local and regional resilience.

The paper underscores the urgent need for expanded cybersecurity infrastructure and ethical AI governance to regulate access and usage. Ensuring secure communication systems, ethical decision-making frameworks, and knowledge-sharing mechanisms will be essential in mitigating risks, protecting communities, and fostering sustainable economic transformation across Africa and beyond.