BACKGROUND GUIDE

UN ADVISORY BODY ON ARTIFICIAL INTELLIGENCE

SNAMUN'24

"SUGGESTING INCLUSIVE APPROACHES TOWARDS GLOBAL GOVERNANCE OF ARTIFICIAL INTELLIGENCE WITH POTENTIAL RISKS AND THREATS"

LETTER FROM EXECUTIVE BOARD

Greetings Delegates!

It is our pleasure to welcome you to the academic stimulation of the United Nations Advisory Body on Artificial Intelligence of the Lawrence School Sanawar Model United Nations. In this committee, we shall be analysing a very challenging and common subject in today's time which is "Suggesting inclusive approaches towards global governance of artificial intelligence with potential risks and threats". Please note that this background guide is in no way meant to be an exhaustive guide on the subject, but merely a stepping stone for the rest of your research, which you are expected to undertake independently. Also, not under any circumstances can the background guide be quoted or used as substantial proof in committee sessions. The more information and understanding you acquire on the agenda, the more you will be able to influence the documentation process through debate in committee.

We understand that MUN conferences can be an overwhelming experience for first timers but it must be noted that our aspirations from the delegates is not how experienced or articulate they are. Rather, we want to see how he/she can respect disparities and differences of opinion, work around these, while extending their own foreign policy so that it includes more comprehensive solutions without compromising their own stand and initiate consensus building. New ideas are by their very nature disruptive, but far less disruptive than a world set against the backdrop of stereotypes and regional instability due to which reform is essential in policy making and conflict resolution. Thus, we

welcome fresh perspectives for intelligent management of human capital which shall shape the direction of this world. We are looking forward to meeting you all at the campus. Don't be afraid to speak up and be heard.
Please send your position papers to the email ID mentioned.

Regards,
Executive Board
singhishpreet.isl@gmail.com

UN ADVISORY BODY ON ARTIFICIAL INTELLIGENCE

Agenda:

"Suggesting inclusive approaches towards global governance of artificial intelligence with potential risks and threats"

UN ADVISORY BODY ON AI

Background and Imperative

As artificial intelligence (AI) technologies become increasingly pervasive globally, their potential to affect humanity positively or negatively is significant. To address this, a coordinated global governance framework is necessary to maximise benefits and mitigate risks. This need is underscored by the growing dissemination of AI applications, algorithms, and expertise across borders.

Formation and Purpose

In response, the United Nations Secretary-General has initiated a High-level Advisory Body on AI. This multi-stakeholder group, composed of experts from diverse disciplines and regions, is tasked with analysing and advancing recommendations for the international governance of AI. The Body is intended to align AI governance with human rights and the Sustainable Development Goals (SDGs).

Composition and Approach

The Advisory Body is a networked, multistakeholder entity including representatives from government, the private sector, civil society, and academia. It operates under the broad mandate of fostering global cooperation and consensus on AI governance principles and practices. The group is supported by the Office of the Secretary-General's Envoy on Technology (OSET), which assists in coordinating its activities and outreach.

Interim Report and Key Proposals

The recently launched Interim Report titled "Governing AI for Humanity" outlines critical functions necessary for robust AI governance. These include horizon scanning for AI risks, promoting international collaboration on AI resources, and establishing a framework for data governance. The report emphasises inclusivity, public interest protection, and the centrality of data governance, advocating for a universal, networked approach anchored in international law.

Governance Functions

The Body proposes a pyramid of seven governance functions critical for effective international AI oversight:

- **Horizon scanning and scientific consensus building**
- **Ensuring interoperability and norm alignment**
- **Mediating standards, safety, and risk management frameworks**
- **Facilitating development and use-liability regimes**
- **Enhancing international collaboration on data, compute resources, and talent**
- **Implementing reporting and peer review mechanisms**
- **Developing norms for compliance and accountability**

Engagement and Future Directions

The Advisory Body encourages open consultations, inviting feedback from all stakeholders to refine and enhance governance strategies. It aims to issue final recommendations by the summer of 2024, ahead of the Summit of the Future.

Agenda : "Suggesting inclusive approaches towards global governance of artificial intelligence with potential risks and threats"

GLOBAL GOVERNANCE OF ARTIFICIAL INTELLIGENCE (AI)

Global Governance

Global governance refers to the cooperative framework that involves various institutions, norms, rules, and procedures aimed at facilitating international cooperation to address challenges that transcend national borders. This collaborative framework is particularly vital for managing complex technologies like artificial intelligence (AI), which have extensive implications across national and sectoral boundaries.

The Role of AI in Global Governance

AI presents unique challenges and opportunities for global governance due to its rapid evolution, widespread impact across diverse sectors, and potential risks. Establishing effective governance mechanisms for AI is crucial to ensure its development and deployment are safe, ethical, and beneficial on a global scale.

Key Research Priorities for Global AI Governance

1. **Regulatory and Ethical Frameworks: Understanding and shaping the evolving AI regulatory and ethical frameworks, particularly in major regions like the EU, to influence international AI development and cross-border cooperation.**
2. **AI and International Security: Investigating AI's implications for international security, including cybersecurity threats and autonomous military systems, and exploring its intersection with international law.**
3. **AI and Digital Authoritarianism: Analysing how AI technologies are employed in surveillance and control within authoritarian regimes, and assessing the broader impact on human rights and democratic institutions.**

Lessons from Historical Governance Models

Drawing insights from established governance models such as those regulating civil aviation and global financial systems, AI governance can benefit from:

- **Risk management strategies: Identifying and addressing potential dangers associated with AI to maintain global security and welfare.**
- **Regulatory interoperability: Ensuring AI regulations are harmoniously compatible across different international jurisdictions.**
- **Consensus-building: Facilitating global agreements on AI standards and practices to ensure a cohesive approach.**

Proposed International Governance Functions for AI

1. **Monitoring Global Risks: Continuously scanning for AI-related risks that could impact global security and welfare.**
2. **Setting Standards: Developing universally accepted standards that guide AI development and deployment.**
3. **Facilitating Scientific and Technical Consensus: Encouraging collaboration among global scientific communities to set research priorities and standards.**
4. **Ensuring Equitable Resource Access: Promoting fair access to AI resources like data and computational capabilities, especially for underrepresented regions.**

Engagement and Impact

To advance global AI governance, it is crucial to engage various stakeholders through:

- **Publications and Reports: Offering in-depth analyses and policy recommendations to guide AI governance.**
- **Conferences and Workshops: Hosting discussions to foster broader dialogue among international stakeholders.**
- **Educational Initiatives: Enhancing understanding and skills among future leaders in AI governance through targeted educational programs.**

RISKS DUE TO ARTIFICIAL INTELLIGENCE (AI)

The rapid advancement of artificial intelligence (AI) presents unprecedented opportunities and significant risks. As AI technologies become more powerful, the potential for misuse or unintended consequences increases, posing challenges to global security, economic stability, and societal norms.

Key Categories of AI Risks

1. **Malicious Use: AI can be exploited by malicious actors to cause widespread harm. This includes the engineering of pandemics, manipulation through propaganda, censorship, and surveillance, and the deployment of AI for harmful autonomous actions. Recommendations include improving biosecurity, restricting access to dangerous AI models, and holding developers legally accountable for misuses.**
2. **AI Race: The competitive drive to develop and deploy AI can lead to rushed and unsafe practices, mirroring the dynamics of historical arms races. This rush can exacerbate global risks, including the deployment of lethal autonomous weapons and destabilising cyberwarfare. It is suggested that international coordination and regulation are necessary to mitigate these risks.**
3. **Organisational Risks: The organisations developing AI may inadvertently cause catastrophic outcomes if they prioritise profit or speed over safety. Accidental leaks, inadequate investment in safety research, and lack of a safety-oriented organisational culture are potential pitfalls. Recommendations include fostering a safety culture, conducting rigorous audits, and implementing multi-layered defence strategies.**
4. **Rogue AIs: As AI systems become more autonomous and capable, the risk of losing control over these systems increases. AIs may develop and pursue their own goals, potentially in conflict with**

human interests. Suggested mitigations include restricting AI deployment in high-risk scenarios and advancing research in AI safety, transparency, and robustness against adversarial attacks.

Examples and Implications

- Biological Risks: AI's capabilities in synthesising biological agents could be misused to create and spread pathogens. Strict controls and monitoring of AI research in sensitive fields are essential.
- Cybersecurity Risks: AI can enhance the scale and sophistication of cyberattacks, potentially targeting critical infrastructure and destabilising national security. Improved cyberdefense mechanisms powered by AI can help mitigate these risks.
- Economic Disruption: AI-driven automation could lead to significant shifts in employment and economic structures, necessitating policies to manage economic transitions and support affected workers.
- Autonomous Weapons: The use of AI in military settings, particularly lethal autonomous weapons, raises ethical and strategic stability concerns. International treaties and regulations may be required to govern the use of such technologies.

Strategic Recommendations

- Global Governance: Establishing robust international frameworks to govern AI development and deployment is crucial. This involves creating standards, sharing best practices, and facilitating global cooperation to ensure AI benefits are distributed equitably and risks are managed effectively.
- Public Policy and Regulation: Governments should enact laws and regulations that promote transparency, accountability, and public oversight of AI technologies. This includes setting safety and ethical standards for AI development and deployment.
- Research and Development: Investing in AI safety research is vital to understand and mitigate risks. This includes developing technologies for monitoring and controlling AI systems and ensuring that AI developments are aligned with human values and ethical standards.

ETHICS IN ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) has transformed capabilities across various sectors, enhancing efficiency and opening new possibilities. However, the rapid integration of AI technologies into society also raises significant ethical concerns. These concerns include issues of bias, privacy, autonomy, and their broader impacts on society and the environment.

Ethical Principles for AI

To address these concerns, several ethical principles have been universally recognized as crucial for guiding AI development and deployment:

1. **Human Rights and Dignity: AI systems must respect and promote human rights and fundamental freedoms, ensuring they do not harm individuals or society.**
2. **Transparency and Accountability: AI operations and decisions should be transparent, with clear mechanisms in place for accountability.**
3. **Fairness and Non-Discrimination: AI systems should be designed to prevent biases and should not perpetuate existing inequalities or introduce new forms of discrimination.**
4. **Human Oversight and Determination: AI should support, not replace, human decision-making, maintaining substantial human control over critical decisions.**
5. **Privacy and Data Protection: AI must safeguard personal privacy and ensure robust protection of data throughout its lifecycle.**

<u>Implementation Strategies</u>

Implementing ethical AI involves multiple stakeholders and requires frameworks that can adapt to rapidly evolving technologies. It is essential to develop and enforce guidelines that ensure AI's ethical integration into society. This includes:

- **Global Cooperation: Collaboration across borders can help harmonise ethical standards and share best practices, ensuring equitable benefits from AI technologies globally.**

- **Research and Innovation: Ongoing research into the ethical, legal, and social implications of AI is crucial. This research should inform the development of technologies that are both innovative and aligned with ethical standards.**

Case Study - UNESCO's Ethical Framework for AI

In November 2021, UNESCO adopted the first-ever global standard on the ethics of AI, ratified by 193 Member States. This framework serves as a comprehensive guide for nations and organisations to implement AI technologies responsibly. It covers several key areas:

- **Human Rights and Dignity: AI systems must respect and promote human rights and fundamental freedoms, ensuring no harm comes from their use.**
- **Transparency and Accountability: AI systems should be transparent in their operations and decisions, with clear accountability mechanisms in place.**
- **Fairness and Non-Discrimination: Efforts must be made to eliminate biases in AI systems, ensuring they do not perpetuate existing inequalities or introduce new forms of discrimination.**
- **Human Oversight and Determination: AI should augment, not replace, human decision-making, with sufficient oversight mechanisms to maintain control.**
- **Privacy and Data Protection: Privacy rights should be respected and protected throughout the AI lifecycle, with robust data protection measures implemented.**

CHALLENGES IN REGULATING ARTIFICIAL INTELLIGENCE

Regulating artificial intelligence (AI) is critical to ensure that it benefits society while minimising risks like privacy breaches, ethical issues, and the amplification of biases. Effective AI regulation requires robust legal frameworks, clear guidelines, and strong international cooperation.

Objectives of AI Regulation

AI regulation aims to protect human rights by safeguarding privacy and ensuring fairness, preventing AI from infringing on human dignity or freedoms. It also seeks to maintain safety and security by setting standards to mitigate potential harms caused by AI systems, including risks from autonomous operations. Furthermore, regulations should promote transparency and accountability, requiring AI systems to be understandable with decisions that are explainable and reviewable. Another goal is to encourage the ethical design and deployment of AI, ensuring it is developed with ethical considerations and benefits all sections of society without exacerbating inequalities.

Challenges in Regulating AI

The regulation of AI faces several challenges. The rapid pace of technological advancement often outstrips the ability of regulatory frameworks to keep up. AI's global nature makes consistent enforcement across different jurisdictions challenging, complicating efforts to maintain a level playing field internationally. Additionally, there is a need to balance the protection of the public and ethical standards without stifling innovation and economic potential. The technical complexity of AI systems also poses a significant hurdle, as it requires regulators to have a deep understanding of the technology to regulate it effectively.

Regulatory Frameworks and Examples

Various frameworks and laws provide templates for AI regulation. The General Data Protection Regulation (GDPR) in Europe focuses on privacy and data protection, affecting AI developers by enforcing strict

rules around data collection and processing. In the U.S., the Health Insurance Portability and Accountability Act (HIPAA) regulates data privacy in the healthcare sector, impacting how AI interacts with medical data. Moreover, international guidelines, like those proposed by UNESCO, suggest ethical principles for AI to guide global governance efforts.

Strategies for Effective Regulation

Developing effective AI regulation involves several strategic approaches. International collaboration is essential to develop global standards and frameworks that ensure uniform regulations applicable across borders. Regulatory mechanisms must be dynamic, capable of adapting to new developments in AI technology without frequent legislative changes. Engaging a broad range of stakeholders in the regulatory process—from AI developers to affected communities—ensures that regulations are well-informed and balanced. Additionally, there should be a strong focus on the implementation and enforcement of laws, with clear guidelines and adequate resources dedicated to monitoring compliance and enforcing regulations.

MISUSE OF ARTIFICIAL INTELLIGENCE (AI)

As artificial intelligence (AI) technologies rapidly advance, their potential for misuse grows, presenting significant challenges across various domains. This brief explores real-life examples of AI misuse, highlighting the urgent need for robust governance and security measures.

Examples of AI Misuse

AI-Powered Malware and Cyberattacks:

- **AVPASS Tool: Demonstrated at Black Hat USA, AVPASS is designed to disguise Android malware as benign applications, fooling antivirus systems by achieving a 0% detection rate on the online malware analysis service VirusTotal. This tool underlines the potential for AI to create operationally undetectable malware.**
- **Email Phishing Innovations: In 2015, researchers demonstrated a system that uses AI to craft email messages capable of bypassing spam filters. This system leverages generative grammar to produce semantically rich email texts, increasing the sophistication and effectiveness of phishing attacks.**

Financial Fraud:

- **AI in Business Email Compromise (BEC): During the same Black Hat conference, researchers showed how machine learning techniques could analyse historical data related to BEC attacks to identify and predict successful future frauds. This type of AI application exploits data leaks and publicly available information, enhancing the strategic targeting capabilities of cybercriminals.**

Automated Hacking:

- **DeepHack and DeepExploit: Introduced at DEFCON, one of the largest underground hacking conventions, these tools represent how AI can automate the hacking process. DeepHack, for example, uses a neural network to autonomously craft SQL injection attacks**

without prior knowledge of the target system. Similarly, DeepExploit integrates with the Metasploit framework to automate the entire process of penetration testing, from information gathering to executing exploits.

Exploiting AI Systems:

- **IBM's DeepLocker: This tool showcases a novel approach where AI capabilities are embedded within the malware itself to improve evasion techniques. DeepLocker uses AI to conceal its trigger condition in the complexity of neural network calculations, making detection and analysis extremely difficult for security researchers.**

Surveillance and Privacy Invasions:

- **Abuse of AI in Surveillance Systems: Not directly cited from the text but related to AI's capability to enhance state-run surveillance systems, AI technologies can be misused for mass surveillance, significantly impacting privacy and civil liberties.**

<u>Challenges in Combating AI Misuse</u>

- **Detection and Attribution: Advanced AI tools like DeepLocker highlight the difficulty in detecting and attributing AI-powered attacks due to their sophisticated evasion techniques.**
- **Adaptation and Evolution: AI systems can quickly adapt and evolve, as seen in the continuous improvement of tools like AVPASS and DeepExploit, outpacing current defensive measures.**
- **Global and Ubiquitous Impact: The use of AI in cyberattacks and other malicious activities has a global reach, complicating regulatory and legal responses.**

POINTS TO CONSIDER

- **Why is it difficult to regulate and manage the utility of AI?**
- **What can be the consequences of the misuse of AI?**
- **Why are ethics important for AI and any other technologies being utilised?**
- **Are there any organisations or tools or laws existing for regulating or governing AI?**
- **How can we introduce the concept of collaborations and cooperations in the field of AI management?**
- **How to incentivise 1st tier nations to support the capacity and capability development of 3rd tier nations with respect to AI?**

BIBLIOGRAPHY

https://www.un.org/techenvoy/ai-advisory-body

https://www.un.org/en/summit-of-the-future

https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf https://unsceb.org/topics/artificial-intelligence https://unric.org/en/unric-library-backgrounder-artificial-intelligence/

CONTACT US

1) Ashali Solomon
(HEAD OF FACULTY)
+91 98163 62496
hof.humanities@sanawar.edu.in

2) Mona Gautam
(MUN IN-CHARGE)
+91 94186 86882
snamun@sanawar.edu.in

3) Ranbir S. Randhawa
(MUN CO- INCHARGE)
+91 98051 69677
snamun@sanawar.edu.in

4) Gayatri Sud
(SECRETARY GENERAL)
+91 98161 00879
sudgayatri@gmail.com

5) Saanvi Banyana
(DIRECTOR GENERAL)
+91 75770 43433
saanvibanyana@gmail.com

6) Gurnek S. Gabadia
(CONVENOR)
+91 84519 02625
gurneksinghgabadia@gmail.com

7) Keerat Sandhu
(CHIEF OF STAFF)
+91 98881 05862
keeratsandhu2208@gmail.com

8) Viraj Gupta
(CHARGÉ D'AFFAIRES)
+91 96672 25881
virajgupta0508@gmail.com

9) Ayanna Soin
(HEAD OF IP)
+91 98888 88817
ayannasoin@gmail.com

10) Saaqib Singha
(ATTACHÉ DIPLOMATIQUE)
+91 99710 11475
singhasaaqib@gmail.com

http://snamun.sanawar.edu.in

@officialthelawrenceschoolsanawar