

GOVERNING AI FOR HUMANITY: REFLECTIONS FROM THE UNITED NATIONS AI ADVISORY BODY

Alondra Nelson*

Abstract: Artificial Intelligence (AI) holds immense potential for societal advancement, yet it also poses significant risks. This paper examines global challenges and opportunities in AI governance, highlighting insights from the United Nations High-Level Advisory Board on AI. It discusses the need for inclusive, transparent, and ethical governance frameworks to foster public trust and ensure AI technologies serve the public good. Recommendations include establishing an international scientific panel, promoting trustworthy AI, and enhancing global cooperation.

1. INTRODUCTION

As AI technologies rapidly evolve, their impact on societies worldwide is profound and multifaceted. While AI promises benefits in science, healthcare, accessibility, and global development, it also raises ethical, social, and security concerns. This paper effective governance. An adaption of the Organisation for Economic Co-operation and Development (OECD) definition suggests that AI is a machine-based system that infers, from the input it receives, how to generate outputs including “predictions,” recommendations, decision-making suggestions, content (e.g., text, image, sound), “reasoning.” AI systems vary in their levels of autonomy and adaptiveness after deployment; some are static, others iterative.

Challenges and Risks of AI

AI poses significant challenges, including:

- **Labor Market Disruption:** Concerns over AI replacing human jobs.
- **Security Threats:** Potential misuse in bioweapons, financial fraud, and misinformation.
- **Social Harm:** Risks to vulnerable populations, particularly children, from

explores the global governance of AI, focusing on building trust, ensuring public safety, and guiding AI development for collective benefit.

Understanding AI: Clarifying Definitions

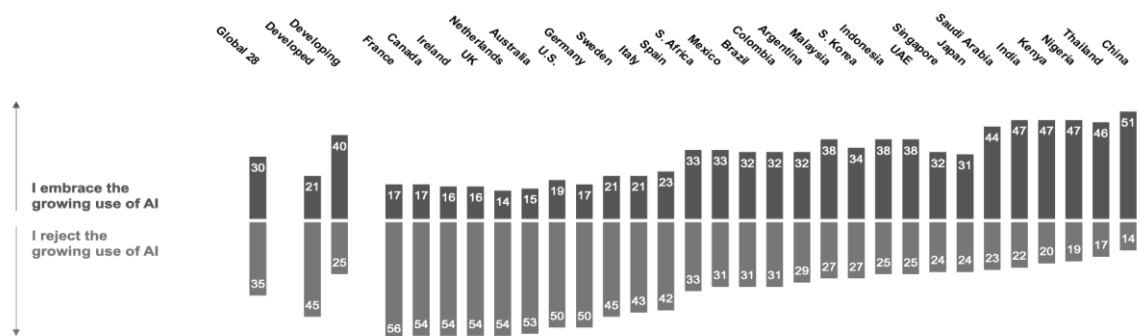
Clear definitions of AI are crucial for harmful AI applications like deep fakes and addictive content.

Public Perception and Trust in AI

Global surveys reveal growing public concern about AI. A 2023 Pew Research Center study found that half of Americans feel more concerned than excited about AI's integration into daily life. These concerns are important because if the public feels tentative or resistant then it will be much harder to harness the huge potential of AI and the desired outcomes if there is not a social situation in which people feel comfortable and trusting to use these tools. The Edelman Trust Barometer (2024) highlights trust disparities between developed and developing countries, with scepticism prevalent in the U.S., U.K., Australia, France, and Canada. Building public trust and resilience is essential for widespread adoption of beneficial AI technologies.

* Prof., Member of the National Science Board, Member of U.N. High-Level Advisory Body on Artificial Intelligence, Professor at the Institute for Advanced Study, Princeton (USA), Former Deputy Assistant to President Joe Biden, Former Acting Director of the White House Office of Science and Technology Policy (OSTP)

Percent who say



2024 Edelman Trust Barometer, CHG_TEC_COM. How would you characterize your feelings about each of the following? 5-point scale; bottom 2 box, resistant or hesitant = reject; top 2 box, enthusiastic or passionate = embrace. General population, 28-nat avg., and by developed and developing countries.

Figure 1: Resistance to Ai Stronger in Developed Markets

The Promise of AI for Societal Good

AI's transformative potential spans various sectors including:

- **Science:** AI models like AlphaFold have revolutionized protein structure prediction, advancing drug development and bio manufacturing.
- **Accessibility:** Technologies such as Waymap aid visually impaired individuals in navigation, exemplifying universal design principles.
- **Agriculture:** AI enables real-time crop monitoring, contributing to food security and sustainable farming.

However, these outcomes are not guaranteed. There is nothing inherent in this deep learning technology that make such benefits inevitable. Rather, there is a need to steward these technologies through effective governance, ethical design, and public engagement, if AI's positive potential is to be realized.

The Role of Global Governance

Recognizing these challenges, the United Nations established the High-Level Advisory Board on AI to offer some recommendations to the broader world about how it might govern AI for humanity. The board's 2023 interim report emphasized principles such as inclusivity, public interest, adaptive governance, and alignment with international

human rights and the Sustainable Development Goals (SDGs).

Key Recommendations for AI Governance

1. AI should be governed **inclusively**, by and for the benefit of all
2. AI must be governed in the public **interest**
3. AI governance should be built in step with data governance and the promotion of **data commons**
4. AI governance must be **universal, networked** and rooted in adaptive **multi-stakeholder collaboration**
5. AI governance should be anchored in the **UN Charter, International Human Rights Law**, and other agreed international commitments such as the Sustainable Development Goals

Implementing Governance Frameworks

Whilst AI regulation can be perceived as punitive or negative, AI governance potentially leads to more effective and beneficial conversations in this landscape. Governance incorporates standards and standard making bodies, norms and professional norms, for the way that AI researchers is conducted, alongside policies which contribute to the clear articulation of desired future states. To address gaps in AI governance identified by the UN AI Advisory Body, the body has identified seven recommendations for governance functions

which could be carried out by individual institutions or a network of institutions.

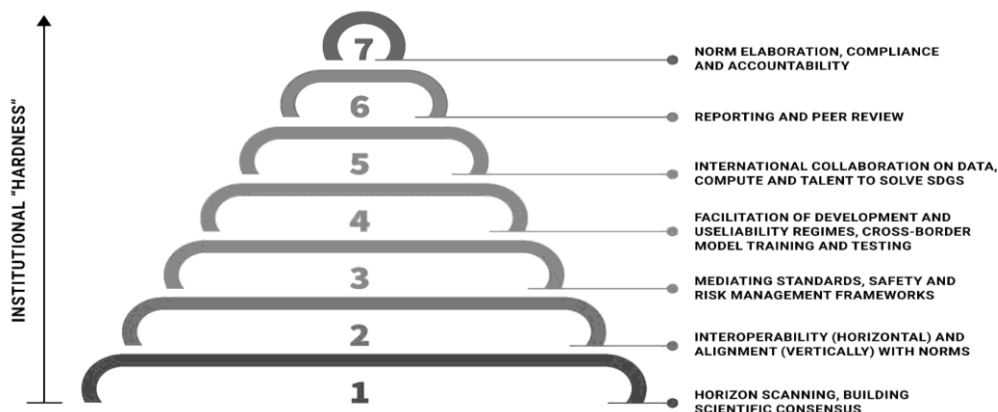


Figure 2: AI Governance Functions

The United Nations' Global Digital Compact (GDC) has begun integrating these recommendations. Key initiatives include the launching of the International Scientific Panel on AI and facilitating global dialogues on AI governance. Crucially, however, as the implementation of government frameworks are taking place, further research and public feedback will be key ingredients for success, alongside policy innovation and creative ideas about how to engage the public community around a technology that is perceived as too complicated for many to understand, yet the implications of it are so profound, that it demands that ways are found to ensure everyone is engaged.

2. CONCLUSION

AI governance must balance innovation with ethical responsibility. Through inclusive frameworks, international collaboration, and proactive stewardship, AI technologies can be guided to serve humanity's best interests. Continued engagement from policymakers, researchers, and civil society is essential to building a trustworthy and beneficial AI future.

This article is adapted from Professor Alondra Nelson's presentation at the SIU Shaping Futures Conference – January 11, 2025

GOVERNING AI FOR HUMANITY: REFLECTIONS FROM THE UNITED NATIONS AI ADVISORY BODY

Alondra Nelson*

Tóm tắt: Trí tuệ nhân tạo (AI) mang trong mình tiềm năng to lớn để thúc đẩy tiến bộ xã hội, song đồng thời cũng đặt ra những rủi ro đáng kể. Bài viết này xem xét những thách thức và cơ hội toàn cầu trong quản trị AI, đồng thời chia sẻ các khuyến nghị từ Ban Cố vấn Cấp cao của Liên Hợp Quốc về AI. Tác giả nhấn mạnh sự cần thiết của các khuôn khổ quản trị bao trùm, minh bạch và có đạo đức nhằm xây dựng niềm tin của công chúng và bảo đảm các công nghệ AI phục vụ lợi ích chung của nhân loại. Những đề xuất trọng tâm bao gồm việc thành lập hội đồng khoa học quốc tế, thúc đẩy các tiêu chuẩn AI đáng tin cậy, và tăng cường hợp tác toàn cầu để hiện thực hóa các mục tiêu phát triển bền vững.

* GS., Thành viên Hội đồng Khoa học Quốc gia Hoa Kỳ, Cố vấn cấp cao về Trí tuệ nhân tạo cho Tổng Thư ký Liên Hợp Quốc, Giáo sư Viện Nghiên cứu Cao cấp - Princeton; Nguyên Phó trợ lý Tổng thống Hoa Kỳ Joe Biden và Giám đốc Chính sách Khoa học và Công nghệ Nhà Trắng (OSTP)