Bridging the Gap: Evaluating AI Policies Through the Lens of Human Right

Research Report Authors:

Leshi Chen, Saeah Go and Vincent Liang

1. Introduction

1.1. Overview of the project

Artificial Intelligence (AI) is reshaping numerous sectors, from healthcare and education to governance and industry, with transformative potential. However, as these technologies become increasingly integrated into everyday life, concerns about their ethical implications and alignment with human rights frameworks have grown. This report examines AI policies from the United States (U.S.) Government, the European Union (EU), and the United Nations Educational, Scientific and Cultural Organization (UNESCO), evaluating their alignment with human rights principles.

Human rights are universal values designed to protect individuals' dignity, equality, and freedom. With AI impacting critical aspects of life, such as privacy, equity, and freedom of expression, assessing its regulation through this lens is crucial. This report adopts a comparative approach, analyzing how each organization integrates human rights into their AI policies, identifies gaps, and highlights best practices.

1.2. Methodology

To conduct this research, we evaluated official AI policies and frameworks from the U.S. Government, the EU, and UNESCO, with a focus on their protections for fundamental human rights. We analyzed specific cases where AI-related human rights violations, highlighting rights that were adequately protected, poorly addressed, or entirely overlooked in each framework. By contrasting these policies, we aim to provide insights into how they can be improved to prevent future violations and promote ethical AI development.

1.3. Motivations

The rapid growth of AI technologies presents both opportunities and risks. While AI holds the promise of improving lives, such as enhancing healthcare and streamlining governance, it also presents challenges. Bias in AI systems, the erosion of privacy, and the potential for misuse in surveillance or misinformation campaigns demonstrate how AI can harm individuals if improperly managed. As developers and researchers, we bear a responsibility to understand these risks and advocate for ethical development practices.

This evaluation is timely as global discussions on AI governance intensify, with organizations and governments striving to create policies that safeguard human rights while fostering innovation. By studying current AI policies and their alignment—or lack thereof—with human rights frameworks, we can identify pathways to develop more inclusive and fair AI systems.

1.4. Objective and Novelty

The goals of this project are threefold:

1. Highlight Violations: Identify cases where AI has led to human rights violations and analyze the underlying policy gaps.
2. Assess Alignment: Determine whether current AI policies align with established human rights frameworks.
3. Provide Recommendations: Offer insights into unprotected areas of human rights and propose ways to address them, using UNESCO's policies as a potential reference point for best practices.

This project provides comparative analysis of AI policies across three influential organizations and its focus on bridging gaps in human rights protections. Through this research, we aim to contribute to a more ethical and responsible approach to AI development, ensuring it serves humanity while respecting its fundamental rights.

2. Background and Related Work

The Universal Declaration of Human Rights (UDHR) serves as a foundational framework for promoting universal human rights, emphasizing equality, dignity, and freedom across all spheres of life. As AI technologies grow increasingly influential, organizations and governments have developed policies to align these advancements with human rights principles. UNESCO promotes ethical AI with a global framework prioritizing transparency and inclusivity. The EU focuses on regulating AI through risk-based policies and GDPR, while the U.S. emphasizes transparency and fairness in AI but lacks robust federal enforcement. Together, these policies reflect varying approaches to integrating human rights into AI governance while highlighting gaps that require attention.

3. Experiments and Analysis

3.1. Violations of human rights

This section examines various human rights articles violated due to the implementation or mismanagement of AI technologies. We provide examples and references for each article to illustrate the ethical implications.

- **Article 3 - Right to Life, Liberty, and Security**
  Mismanaged AI applications have directly impacted individuals' lives and safety.
  - An AI chatbot reportedly contributed to a 14-year-old boy's suicide by engaging with him inappropriately and exacerbating his mental health struggles. [1]
  - A man ends his life after an AI chatbot ‘encouraged’ him to sacrifice himself to stop climate change. [2]
  - Google AI Chatbot Tells Student to ‘Please Die’ [3]
- **Article 7 - Right to Non-Discrimination**
  AI-driven systems often perpetuate biases that lead to discrimination.
  - Bias in Automated Hiring Systems: In 2018, an investigation revealed that an AI recruiting tool used by a major tech company discriminated against women by penalizing resumes that included terms like "women's chess club captain."[4]
  - Discrimination in automated decision making: Money Bank received 800 applications for 20 financial analyst positions, using automated profiling in the recruitment process. Some candidates were unsuccessful despite having the advertised qualifications and experience, raising concerns about potential discrimination. [5]

- **Article 12 - Right to Privacy**
  AI technologies often infringe upon the right to privacy through unauthorized data collection or surveillance.
    - AI-powered surveillance technologies deployed in the U.S.-Mexico border has raised significant privacy concerns. These systems, including fixed AI-surveillance towers and facial recognition, monitor individuals without their consent, leading to potential human rights violations. [6]
- **Article 19 - Right to Freedom of Expression**
  AI technologies can limit freedom of expression by controlling or censoring online speech.
    - AI content moderation tools on social media platforms have occasionally censored legitimate speech, disproportionately affecting minority voices. This overreach can stifle free expression and limit the diversity of viewpoints available in public discourse. [7]
- **Article 25 - Right to Adequate Standard of Living**
  AI-driven credit scoring algorithms have shown evidence of bias, which restricts their access to financial resources.
    - AI-powered credit scoring algorithms used by financial institutions, like those implemented by Goldman Sachs with Apple's credit card, have shown evidence of gender bias. Women were reportedly assigned lower credit limits than men with similar profiles, limiting their access to financial resources. [8]

3.2. Alignment with human rights

The alignment of current AI policies with the UDHR varies across organizations and regions, with differing degrees of enforceability.

UNESCO

UNESCO has addressed most **UDHR** articles in its *Recommendation on the Ethics of Artificial Intelligence* [9]. The focus is on dignity, fairness, and non-discrimination throughout the lifecycle of AI systems. However, these are **non-binding suggestions** and lack enforcement mechanisms.

United States

AI-related policies in the U.S. address some UDHR articles through **initiatives**, but most lack federal enforcement:

- **Article 12 - Right to Privacy**
  The **California Consumer Privacy Act (CCPA)** [10] aims to protect individuals' privacy by regulating data collection and usage.
- **Article 2 - Right to Freedom**
  The **First Amendment** ensures freedom of expression, including digital communication, but there are few AI-specific regulations supporting this right.

European Union

The EU leads with **enforceable regulations** directly addressing human rights in AI applications:

- **Article 12 - Right to Privacy**
  The **General Data Protection Regulation (GDPR)** [11] enforces strict privacy laws to protect individuals' data rights.

- **Article 19 - Right to Freedom of Expression**
  The **EU AI Act** ensures AI systems support free speech and prevent censorship or discrimination.
- **Article 7 - Right to Non-Discrimination**
  The **EU AI Act** mandates fairness in AI systems, ensuring equality and preventing bias.

In summary, UNESCO provides a comprehensive ethical framework for AI but lacks enforcement. The EU, with the GDPR and AI Act, has implemented enforceable regulations to uphold human rights. The U.S., on the other hand, relies on fragmented initiatives like the CCPA and broad constitutional protections such as the First Amendment, which result in less cohesive alignment with UDHR principles.

3.3. Unprotected human rights

As mentioned in the last section, UNESCO's AI policies promote transparency, accountability, and inclusivity, but enforcement and global standardization remain challenges. The U.S. government and EU have strong regulatory frameworks, yet both the U.S. and the EU face challenges in fully addressing UDHR Articles 23, 25 and 26.

- **Article 23 - Right to work and equal pay**

AI-driven recruitment tools, increasingly used in the U.S. and EU, often exhibit biases that disadvantage minorities and women, violating the principle of equal opportunity. For instance, studies have shown that some algorithms in hiring processes systematically deprioritize resumes with name or educational backgrounds associated with underrepresented groups. This perpetuates workplace inequality and limits access to fair employment opportunities.

- **Article 25 - Right to an adequate standard of living**

AI applications in welfare and housing systems risk violating Article 25. For example, predictive algorithms in the U.S. have been used to manage social benefits, often inaccurately flagging individuals for fraud, leading to wrongful benefit denials. In the US, automated decision-making in housing allocations has raised concerns about discrimination against low-income and immigrant communities, further exacerbating economic inequality.

- **Article 26 - Right to education**

AI in education, such as adaptive learning platforms, has the potential to personalize learning but risks deepening existing inequities. In both the U.S. and EU, access to these technologies often depends on socioeconomic status, leaving underprivileged students without adequate resources to benefit from disproportionately impacting students from marginalized backgrounds.

3.4. Compare research result and analysis

Coverage of UDHR

The AI policies of UNESCO, the EU, and the U.S. reflect different approaches to human rights protection, with varying effectiveness. UNESCO emphasizes global inclusivity and comprehensive ethical principles, aligning well with the UDHR but lacks enforceability. The EU leads with enforceable regulations like GDPR and the AI Act, addressing privacy (Article 12), equality (Article 7), and freedom of expression (Article 19) through binding measures. In contrast, U.S. policies rely on state-level initiatives, such as the CCPA, and constitutional safeguards like the First Amendment, resulting in fragmented and less cohesive alignment with UDHR principles.

Strengths and Weaknesses

UNESCO's strength lies in its holistic and inclusive framework for ethical AI, yet the absence of binding mechanisms limits its impact. The EU excels with enforceable policies ensuring compliance, though regulatory implementation across member states can vary. The U.S. provides strong constitutional protections and state-level privacy laws but suffers from gaps due to the lack of a unified federal framework.

Effectiveness of Implementation

The EU leads in terms of enforceability, with clear and actionable regulatory frameworks ensuring a higher degree of compliance with human rights principles. Its GDPR and AI Act establish clear boundaries for AI applications and protect individual rights effectively. The U.S., however, struggles with implementation due to its decentralized and fragmented approach. While some states have made progress with privacy protections, the absence of a cohesive federal framework undermines the overall effectiveness of its policies. UNESCO, while influential in setting ethical guidelines, lacks mechanisms to enforce its recommendations, limiting its practical impact despite its global reach and emphasis on inclusivity.

Insights from Comparison

One of the most significant insights from this comparison is the distinction between initiatives and enforceable policies. Most AI-related policies across UNESCO, the EU, and the U.S. are initiatives rather than binding regulations, leaving technology companies with considerable autonomy and often resulting in inconsistent adherence to human rights principles. There is notable overlap among the organizations in addressing privacy (Article 12) and equality (Article 7), with all emphasizing transparency, fairness, and non-discrimination. However, there are clear divergences as well. UNESCO's broad, non-binding approach contrasts sharply with the EU's enforceable measures, while the U.S. focuses more on constitutional protections than AI-specific regulations.

4. Discussion

4.1. Recommendations for Aligning Human Rights

To ensure AI technologies align more comprehensively with human rights principles, governments and organizations should prioritize the following improvements:

1. **Establish Enforceable Global Standards**
   While frameworks like UNESCO's recommendations are valuable, they must be backed by enforceable global regulations. A unified international body could oversee AI development and ensure compliance with UDHR principles.
2. **Strengthen Federal Enforcement in the U.S.**
   Fragmented state-level initiatives, such as the **CCPA**, should be unified under comprehensive federal legislation to ensure consistent protections for privacy, freedom, and equity.
3. **Enhance Transparency and Accountability**
   AI systems should incorporate mechanisms for transparency, including detailed logs and open access to system decisions for public scrutiny. Accountability should include legal recourse for harm caused by AI decisions.

4. **Address Bias and Inequality**
   Mandate regular audits and fairness checks for AI systems to identify and mitigate biases, particularly in high-stakes applications like hiring, policing, and lending.
5. **Support Research and Education on Ethical AI**
   Governments and organizations should fund research and training programs on ethical AI practices to equip developers and decision-makers with the tools needed to design and govern AI responsibly.

4.2. Concern about our method

While conducting this research, several ethical considerations emerged:

1. **Selection Bias in Data Sources**
   The research relied on publicly available data, which may exclude lesser-known cases of AI misuse or successful implementations. This could skew the findings toward high-profile examples.
2. **Emphasis on Violations Over Successes**
   The report focused more on instances of human rights violations, potentially overlooking AI systems that have successfully upheld human rights.
3. **Comparative Limitations**
   Direct comparisons between the policies of UNESCO, the U.S., and the EU may not fully account for differences in governance structures, legal frameworks, and cultural contexts.
4. **Potential Overgeneralization**
   Simplifying complex AI systems and their impacts for clarity in the report risks misrepresenting technical or legal details.
5. **Ethical Responsibility to Avoid Fearmongering**
   While highlighting risks is important, it's equally critical to avoid overstating potential harms in a way that deters beneficial AI innovations.

5. Conclusion

As artificial intelligence continues to transform societies, its alignment with fundamental human rights remains critical. This report has explored AI policies from UNESCO, the United States, and the European Union, highlighting both strengths and gaps in their alignment with the UDHR.

UNESCO offers a comprehensive ethical framework addressing dignity, fairness, and non-discrimination. However, its lack of enforceability limits its impact. The European Union leads in setting global standards through enforceable laws like the GDPR and the AI Act, ensuring strong protections for privacy, equality, and freedom of expression. In contrast, the United States relies on fragmented state-level initiatives and constitutional safeguards, leaving gaps in enforcement and consistency.

Moving forward, it is imperative to establish unified global standards for AI governance, prioritize enforceable mechanisms, and address the ethical challenges posed by bias, inequity, and misuse. By taking proactive steps to align AI systems with human rights, we can ensure these technologies benefit all of humanity while safeguarding the principles of dignity, equality, and freedom.

This report underscores the need for ongoing dialogue, collaboration, and innovation to bridge the gap between technological advancement and ethical responsibility.

References

[1] https://www.cnn.com/2024/10/30/tech/teen-suicide-character-ai-lawsuit/index.html
[2] https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-
[3] https://aibusiness.com/chatbot/google-ai-chatbot-tells-student-to-please-die-
[4] https://www.google.com/url?q=https://www.onlyinfotech.com/2018/10/15/amazon-said-to-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-apps/&sa=D&source=docs&ust=1732589447569795&usg=AOvVaw1D0FzBboHnvi5aecHsAnL6
[5] https://www.lewissilkin.com/en/insights/discrimination-and-bias-in-ai-recruitment-a-case-study
[6] https://time.com/6979557/unregulated-border-technology-migration-essay/?utm_source=chatgpt.com
[7] https://www.lawfaremedia.org/article/assessing-impacts-of-ai-on-human-rights-it-s-not-solely-about-privacy-and-nondiscrimination
[8] https://www.bbc.com/news/business-50365609
[9] https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence
[10] https://oag.ca.gov/privacy/ccpa
[11] https://gdpr-info.eu/

Appendix

- Data-Informed Predictive Policing Was Heralded As Less Biased. Is It?
- Human Rights Watch - China's Algorithms of Repression
- Source: BBC - Apple's 'sexist' credit card investigated by US regulator
- https://www.cnn.com/2024/10/30/tech/teen-suicide-character-ai-lawsuit/index.html
- https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-
- https://www.amnesty.org/en/latest/news/2024/03/eu-artificial-intelligence-rulebook-fails-to-stop-proliferation-of-abusive-technologies/#:~:text=%E2%80%9CThe%20AI%20Act%20offers%20only,migrants%2C%20refugees%20and%20asylum%20seekers.
- The Deadly Digital Frontiers at the Borde
- Assessing Impacts of AI on Human Rights": It's Not Solely About Privacy and Nondiscrimination