

Forum: Promoting Science Committee
Agenda: On measures to establish international ethic standards and guidelines for AI development and usage
Student Officer: Billy Chan

Introduction

As AI emerges as one of the most influential innovations in the contemporary world, it is destined that such technologies will significantly reshape the global landscape. In fact, the emergence of AI has already proven itself as a dominant force in the future of humanity, with its impact evident in a vast multitude of sectors, industries, and cultures.

In the healthcare industry, AI has been extremely beneficial in interpreting medical images accurately, leading to more efficient and effective healthcare delivery. While in the education sector, the advancements in AI systems have been adapted to learning platforms, providing more tailored information to learners and creating a more inclusive educational environment. To a larger extent, AI is influencing cultural dynamics by integrating technology into cultural expressions and social interactions.

It is indeed because of the profound influence and potential that AI obtains, that the decisions that humans make on AI are extremely crucial, especially when recognizing the negative possibilities of AI.

When discussing the negative implications of AI, ethics—an essential aspect of humanity—emerges as one of the key priorities in the discourse. With more findings and discussions of academia on a reality basis, it is more evident that AI development and usage, without proper regulatory frameworks, could pose unprecedented challenges to human ethics. On a surface level, the rapid development of AI technology could be accompanied by enhancements of its capabilities when misused, including AI's potential to collect sensitive personal information and AI's capability of exacerbating cyberattacks. On a deeper level, AI technology can also change the current social landscape, leading to challenges of fundamental ethics, including equity and inclusivity.

Under this context of AI seemingly as a double-edged sword, which can both benefit society and pose ethical concerns, measures to establish international ethic standards and guidelines for AI development and usage play an essential role in AI governance.

While states and international organizations have recognized the urgent manner of the issue on hand, the current progress built on the issue is ultimately lacking, considering how recent and rapid AI has emerged. In this sense, there remains wide room for exploration regarding the possible regulatory frameworks and governance strategies on this topic. Generally, some possible solutions may include establishing an international network, creating communication campaigns for internet users, and setting monitoring and

accountability mechanisms. However, with the complexity of the issue on hand, these possible solutions all encompass a significant space for discussions and debates, and other strategies could be developed to create a comprehensive, feasible, and sustainable solution to the issues.

To promote science, bridge the divide, and secure the future, the pressing weight of this issue falls on the responsibilities of policymakers.

Key Terms

Artificial Intelligence (AI) - AI refers to a new form of technology with the capability of performing tasks that are commonly carried out by intelligent beings like humans. Their strength in reasoning, learning, and automating repetitive tasks can be extremely impactful towards humans' economic and social world, often leading them into the context of how corresponding regulations should be imposed. In terms of its integration into everyday reality, AI raises concrete concerns about privacy, accountability, and limitation of development.

Machine Learning- Machine learning is a field of AI that uses algorithms to make predictions based on given data. Machine learning enables AI to understand the content and extract patterns within the training data given, to provide a prediction resembling the training data given. Ensuring the quality of data given, the larger the amount of training data inputted into the machine learning algorithm, the more accurate the model will be.

Generative AI – Generative AI is a category of AI that is used in generating or creating new content, ranging in various forms including text, images, or videos. They are created through the process of deep learning, which involves training models for artificial neural networks to learn from large datasets. One well-known model of generative AI is ChatGPT, an AI chatbot created by OpenAI and launched in 2022, trained to provide detailed responses when given a prompt or instruction.

Regulatory Arbitrage: Regulatory arbitrage refers to a practice of corporations or firms to elude regulatory disadvantages by moving under different jurisdictions. In the context of AI development, this often refers to developers from countries with greater access to AI technology, but more regulations, to move to less developed countries with looser regulations. Such practices can contradict a cohesive framework established globally for responsible AI development and usage.

General Overview

Ethical concerns:

As AI technology rapidly proliferates, humans are exposed to both its remarkable benefits and potential drawbacks. In many scenarios, AI's qualities can be viewed as a double-edged sword, as most of AI's strength has high applicability, in both positive and negative aspects. For instance, AI's excellence in data collection can help increase efficiency, but it could also be misused to threaten individuals' privacy. In this dimension, ethical concerns arise, demanding careful consideration and governance.

Privacy concerns:

With AI technology being developed and integrated into a larger part of humans' lives, its alerting risks on privacy and personal information raises concerns. Typically, many AI systems utilize machine learning, which involves training based on large datasets. In reality, a large portion of these datasets are collected from generative AI's interaction with private users, which may contain sensitive information. A controversial example of this is the AI-powered facial recognition system, which gathers biometric information from users, potentially posing risks to user privacy. One driving factor to the privacy concerns of AI is the lack of consent in data collection and usage. As there is a lack of clear governmental regulatory frameworks on AI's data collection, many generative AI models lack explicit explanations to inform the users and ask for consent for data collection. The same issues apply to AI's data usage, where while users may understand that their data would be collected, they may not understand that their information will be included in the datasets and how they will be applied. Moreover, AI's risks of data leakage, an accidental leakage of sensitive data, further exacerbate AI's privacy issues. Incidents regarding data leakage not only have happened on small, proprietary AI models, but well-known AI models have also shown their vulnerability towards data leakage, as evidenced in ChatGPT's instance of leaking users' titles of conversation histories to other users.

Cybersecurity concerns:

While AI can potentially be beneficial to cybersecurity in certain ways, through its strength in security incident forensics automated response to threats, it can also raise cybersecurity concerns when improperly used. To illustrate, the issue with AI's potential in deepfake technology has been growing consistently as an everyday problem for many. Created by AI neural networks, these deepfakes range from images, audio, and videos. A notable incident caused by AI's deepfake technology happened in a Hong Kong bank recently, where a finance clerk claimed a loss of \$25 million to cybercriminals using deepfake technology to impersonate the company's chief financial officer. The misuse of such deepfake technology not only casts concerns about individuals' cybersecurity, but it also undermines public trust in authorities and other sources of information. Despite the social engineering attacks that AI could deliver, AI could potentially devastate cybersecurity in a more direct manner, typically through powering ransomware

attacks. A ransomware attack is a form of malware attack that encrypts a victim's data and demands payment for decryption, with its capability enhancing and evolving rapidly through the integration of AI. With certain AI machine learning algorithms, AI-powered ransomware can mimic normal system behavior, outrunning detection by security software. Moreover, machine learning models also help ransomware attacks by allowing ransomware to correspondingly adjust its encryption methods towards different systems, ensuring a greater challenge for personal decryption.

Social impact and job replacement:

AI, with its outstanding capability to automate repetitive tasks and increase productivity, may cause social issues when integrating into the current economic landscape, specifically by replacing a large portion of jobs. According to statistics from SEOAI, 14% of workers have experienced job displacement due to AI. In terms of industry, 35% of businesses have already experienced AI integration, while an additional 42% are in the process of exploring AI adoption. Taking a deeper look at AI's impact on job displacement, it can be seen that AI-driven autonomy may disproportionately impact certain groups of individuals. To illustrate, AI's ability in job displacement may pose excessive threats to those working in manufactory industries, receiving lower wages, and experiencing vulnerable employment situations. This phenomenon can be explained by AI's current role in the economy, where its primary impact stems from the automation of repetitive tasks. At scale, this drastically alters the manufacturing and economic landscape. In this aspect, AI's impact on job displacement is not limited to an economic level, but it also raises questions about its ethics. From an ethical perspective, workers with comparatively lower wages are not only exposed disproportionately, but they also experience more economic vulnerability, thus this raises questions regarding social equity, social mobility, and social justice.

Enforcement concerns:

While AI is a pressing concern internationally, with different stakeholders recognizing its necessity to be addressed, it is ultimately difficult to enforce corresponding regulations and guidelines, attributed to its novelty and complexity. On an international level, enforcement could proceed in a relatively consistent manner, considering the global nature of the issue on hand and preventing regulatory arbitrage.

Obligations of AI Developers:

In terms of enforcing regulations regarding the development and usage of AI technology, it is crucial to ask the question of whether a clear framework should exist on the obligations of AI developers in advance, and another on the importance of an impartial agreement between states and AI developers. Furthermore, one concrete question subsists: how much should states oblige AI developers to report their advances? This reflects the contrast of perspectives among AI developers and states. On one hand,

obliging AI developers to report their advances on a more frequent basis and nuanced manner can increase transparency, fostering more trust among stakeholders. Demanding more information regarding AI developers' progress in innovation would also be beneficial in terms of shared knowledge, leading to more collaborative advancements. On the other hand, requiring excessive reports from AI developers can potentially limit their competitive advantages, and in many legal frameworks, pose questions on the legitimacy of their intellectual property rights. In this sense, mandating comprehensive advancements in AI development not only could discourage innovations, but it also raises ethical considerations regarding developers' ability to protect their intellectual property. Thus, policymakers could potentially formulate a practical, impartial, and constructive framework regarding the obligations of AI developers to ensure a balance among different perspectives.

Distribution of AI technology:

The distribution of AI technology is another critical concern under the enforcement of AI frameworks, involving debates on practical and ethical levels. As AI continues to spread its influence among different sectors, questions are raised about whether countries with greater power, like MEDCs, should have access to more advanced AI technology. If more advanced AI technology is concentrated in countries with more wealth and resources, ethical concerns regarding equity and justice can be involved, as the inequality between countries can potentially widen. On the other side, if advanced AI technology is equally distributed to all nations, on a practical level, the inefficiency of AI usage can be a valid concern. Recognizing AI's profound potential, can developing nations that lack corresponding expertise and resources fully utilize AI? On a global scale, would this be a waste of opportunity to foster growth and development? To enforce constructive frameworks on AI governance, policymakers could take a deeper look at the complexity of distributing AI technology. Distributing AI in a fully meritocratic manner or equality-based manner may not necessarily be the only choice, policymakers could find a point of balance between the two approaches, or create an entirely new framework, to acknowledge different aspects of the issue.

Timeline of Events

In general, AI has been an innovation that spreads extremely rapidly, and different organizations have been actively involved in developing strategies accordingly. However, as humans are still in the process of exploring the complexities of AI, policymakers have been relatively hesitant in demanding comprehensive regulatory frameworks at this stage. Below is a timeline with several events that have happened in the past:

Date	Description of event
15 June 2020	<p>GPAI (Global Partnership on Artificial Intelligence) was officially launched with fifteen countries as founding members, as one of the first large-scale international collaborations focused on AI. It is devoted to fostering the responsible development of AI under the principles of human rights, inclusion, diversity, innovation, and economic growth. Specifically, GPAI fosters collaboration among different stakeholders, ranging from governments, and industries, to civil society, contributing to a diverse participation in managing AI development and usage.</p>
6 April 2021	<p>The European Union presented the proposed AI Act. This is the first published proposal to regulate AI in the European Union, setting a precedent for AI regulations. In a pioneering manner, the AI Act has a risk-based approach, with a classification system for AI systems to be categorized into different risk levels, including high risks, limited risks, and minimal or low risks.</p>
29 November 2022	<p>ChatGPT was launched by OpenAI, based on the GPT-4o large language model (LLM). The introduction of ChatGPT soon captured the public's attention and interest in the field of AI, leading to a quick adoption of generative AI. As public attention grew, ChatGPT's launch also attracted massive investment funds to companies focused on AI development, catalyzing the change in the AI industry.</p>
21 March 2024	<p>The UN General Assembly adopted the landmark resolution on AI. Recognizing AI's potential in enabling the progress of 17 SDGs, the UN urged all states to develop regulatory frameworks related to the responsible use of AI. Specifically, the resolution reinforces the ethical standards in AI technology by highlighting the necessity of human rights protection in developing AI systems, acknowledging that the same rights people have offline must also be maintained online.</p>

21 May 2024

The European Council formally adopted the EU AI Act. It is one of the first comprehensive frameworks adopted in AI regulation, creating a structured approach to be implemented. It not only addresses local concerns with the risk-based approach, but it also sets a standard for international collaboration on the topic of AI regulations, positioning the EU as one of the leaders in global AI management.

UN Involvement, Relevant Resolutions, Treaties and Events

Considering the novelty and complexity of AI, and the difficulty of comprehensively enforcing AI regulations with contrasting perspectives, the UN is still in the process of developing and implementing regulatory frameworks on AI. Thus, the UN currently has limited resolutions and treaties in establishing ethic standards and guidelines for AI development and usage.

- On 21 March 2021, the UN General Assembly officially adopted the landmark resolution on AI regulations, titled “Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development”. Co-sponsored by over 120 member states, the resolution led by a draft from the United States was adopted without a vote. The assembly sets ethical guidelines on the development and usage of AI, through emphasizing the promotion of "safe, secure, and trustworthy" AI systems. It focuses on ensuring AI's compliance with international human rights laws, closing the digital divide driven by AI within countries, and complementing AI strategies to future UN initiatives.
- On 16 October 2023, the resolution titled “AI in the military domain and its implications for international peace and security” was adopted by the UN General Assembly’s First Committee. With 165 votes in favor, two votes against, and six abstentions, this resolution co-led by the Netherlands and the Republic of Korea passed. The resolution continued in affirming that AI development and usage should follow international laws, including human rights laws and humanitarian laws, specifically in military applications of AI.
- In October 2023, the multi-stakeholder Higher Level Advisory Body on Artificial Intelligence (HLAB-AI) was formed as part of the UN, to analyze and recommend the international governance of AI. So far, the HLAB-AI has created multiple reports titled “Governing AI for Humanity” in regard to the responsible development and usage of AI. In the final report, it is

noted that “if AI risks intensify, member states may need to consider establishing a more robust international institution with monitoring, reporting, verification, and enforcement powers.” Moreover, it has provided guidance through specific recommendations, including the establishment of a global AI fund, the convention of a UN-led policy dialogue on AI governance, and the creation of an International Scientific Panel on AI.

Possible Solutions

While AI presents an unprecedented challenge to humanity, with proper mitigation strategies, the potential devastation of such repercussions could be reduced immensely.

Establishing an international network

To further explore AI’s complexities and how it can be leveraged by a political framework, establishing an international network on the subject can be a critical solution. Due to how recent and rapid AI has permeated into different aspects of our lives, international knowledge sharing could accelerate our exploration in understanding the potential of AI. Specific implementations of this could include periodically organizing conferences and workshops, for different nations or groups to share their recent thoughts, discoveries, or progress related to AI or corresponding ethic guidelines. Potentially, with a comprehensive international network established, different stakeholders, including governments, NGOs, industry leaders, and scientists, would have an opportunity to share their insights and perspectives on the topic. This could help in constructing a holistic approach based on diverse perspectives, ensuring that the potentials of AI can be equally shared among different groups, which reinforces the ethical concept of fairness and equity.

Communication campaigns to internet users

In reference to many ethical concerns of AI, to a large extent, the potential harm that AI can cause is particularly amplified when internet users have limited knowledge of AI systems. For such communication campaigns to be enforced, firstly, the public could be informed on the fundamental principles of AI systems, including how basic AI systems work, how the AI systems may interact with them, and how their interaction with AI could be applied. With knowledge about AI’s fundamentals, the public can be more conscious of their interaction with AI and facilitate their consent on the level of exposure to AI. Furthermore, such communication campaigns could be enforced by highlighting the ethical concerns related to AI, including their rights regarding privacy and cybersecurity. Ultimately, practices for protecting their personal information online could also be a part of the enforcement. By developing the public’s consciousness of their rights and practices for self-protection under the AI era, the public can be less susceptible to AI’s risks on an individual level. While the specific form of such

communication campaigns could vary, from social media to public education initiatives, the release of such campaigns, if effective, could be beneficial in fostering a positive landscape for ethical AI usage.

Establish monitoring and accountability mechanisms:

Establishing monitoring and accountability mechanisms could be essential in maintaining an effective and sustainable framework for the long term. As AI's potential to be misused grows, without such enforcement, not only that the public interest could be threatened on an individual level, but the power dynamics among different organizations and states could be more unstable and chaotic. On an implementation level, the first step to consider is whether regulatory oversight bodies should be established. While currently there are partnerships and organizations devoted to AI governance, an authoritative regulatory body that oversees international standards on the issue is still lacking. With such authorities established, the governance on the issue could potentially be more organized and comprehensive. Specifically, the questions about the obligations of the developers are raised again. Could states ask the developers to report their progress of advancements in AI technology? If so, at what frequency, and to what extent? In the same sense, audits on AI systems could also be part of the enforcement, assessing their progress on advancement and alignment with ethical standards, but the same questions could apply regarding the frequency and the nuance-level of such assessments. Furthermore, a report and redress system could be part of these mechanisms. This system could allow individuals to report concerns about AI applications, specifically when it involves ethical concerns. While this system could be effective when constructed thoughtfully, questions about its feasibility of it could arise regarding the structure of the report channels.

Bibliography

- “Ai Act.” Shaping Europe’s Digital Future, European Commission, digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai. Accessed 17 Dec. 2024.
- “Ai Replacing Jobs Statistics: The Impact on Employment in 2025.” AI Writer for SEO, seo.ai/blog/ai-replacing-jobs-statistics. Accessed 18 Dec. 2024.
- “Artificial Intelligence: High-Level Briefing.” Security Council Report, www.securitycouncilreport.org/whatsinblue/2024/12/artificial-intelligence-high-level-briefing.php. Accessed 19 Dec. 2024.
- Caballar, Rina. “10 Ai Dangers and Risks and How to Manage Them.” IBM Blog, 30 Aug. 2024, www.ibm.com/blog/10-ai-dangers-and-risks-and-how-to-manage-them/. Accessed 18 Dec. 2024.
- “European Approach to Artificial Intelligence.” Shaping Europe’s Digital Future, digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence. Accessed 21 Dec. 2024.

“General Assembly Adopts Landmark Resolution on Artificial Intelligence | UN News.” United Nations, United Nations, news.un.org/en/story/2024/03/1147831. Accessed 20 Dec. 2024.

“Global Partnership on Artificial Intelligence.” GPAI, GPAI, gpai.ai/. Accessed 21 Dec. 2024.

“Historic Timeline.” EU Artificial Intelligence Act, artificialintelligenceact.eu/developments/. Accessed 23 Dec. 2024. Accessed 21 Dec. 2024.

Pszenny, Chantel. “AI-Powered Ransomware: The next Generation of Damaging Cyberattacks.” Sotero, 16 Oct. 2024, www.soterosoft.com/blog/ai-powered-ransomware-the-next-generation-of-damaging-cyberattacks/. Accessed 19 Dec. 2024.

Yasar, Kinza, et al. “What Is Deepfake Technology?: Definition from TechTarget.” WhatIs, TechTarget, 13 Aug. 2024, www.techtarget.com/whatis/definition/deepfake. Accessed 19 Dec. 2024.

Contact Information:

Please contact the below person with any questions regarding the speech or report and good luck!

Billy Chan - President Chair

WeChat ID: billygenius123

Email: 24890@shenzhen.qsi.org

Cowen Cui – Secretary General

25ccui@student.uiszc.org

Chan Kim – Head of Chair

25kchan@student.uiszc.org

Gladys Ndunge Mutinda– Director of U2NESCO

gnm@uiszc.org