# Evolution of Norms for a Trustworthy AI Society and Our Responsibilities and Roles

**Ji-Hun Lim, Jeong-Eun Seo and Hun-Yeong Kwon**

School of Cybersecurity, Korea University, Seoul, Republic of Korea
ljh89719@naver.com, sje5279@korea.ac.kr, khy0@korea.ac.kr

**Abstract**

Artificial Intelligence (AI) holds promising opportunities for enhancing human life, yet its pervasive influence on society also introduces unintended negative consequences. Throughout the 2010s, a growing recognition of ethical concerns in AI led to the formulation of principles emphasizing ethical research and development. Various global entities have since endeavored to establish norms for AI ethics. Recently, beyond the declaration of ethical principles, concrete practical guides and codes of conduct are emerging, and in some countries, they are developing into specific legal regulatory discussions. This paper provides a detailed examination of the evolutionary trajectory of AI norms and further explores the responsibilities and roles of diverse stakeholder groups, including developers, suppliers, and users, in ensuring the trustworthiness of AI. The analysis aims to move beyond abstract ethical principles to actionable norms, emphasizing the need for a holistic societal response.

## The Importance of Social Norms for AI

### Why do we talk about AI ethics?

The discussion on how to minimize the negative impacts of technological advancements while providing various benefits to human society has been a crucial topic in academic communities, and AI is no exception. Being a powerful technology, AI presents potential risks, and unexpected problems have already arisen in various fields incorporating AI technology due to bias, opacity, and unclear responsibility. Therefore, it is imperative to proactively address the ethical issues that AI technology may cause, considering the potential harm it may bring (Russell and Norvig 2021; Leslie 2019).

The risks posed by AI are diverse, encompassing operational risks such as errors in automated judgments, bias, and opacity, as well as security risks like algorithm manipulation, cyber-attacks, and privacy issues. Control-related risks include sudden malfunctions, malicious abuse, and concerns related to automated weapons. Ethical risks involve dilemmas like the trolley problem, issues of equality, and the gap between AI and human value alignment. Additionally, socio-economic risks include unemployment resulting from AI development and disparities in AI access (Cheatham, et al. 2019; Anand 2018).

To tackle these risks and promote the development and use of socially beneficial AI, numerous scholars, policymakers, think tanks, and civil society are actively engaged in discussions (Cath, et al. 2018; Galanos 2019; Ulnicane, et al. 2021a; Ulnicane, et al. 2021b; Vesnic-Alujevic, et al. 2020).

### What is the difference between AI ethics and traditional engineering ethics?

AI ethics shares similarities with the discussion of traditional engineering ethics, but notable differences exist. The discourse on AI ethics has not solely been shaped by ethicists; rather, key personnel from AI development companies, who have direct investments, actively contribute with specific opinions beyond a mere academic interest. For instance, Demis Hassabis of DeepMind imposed conditions on the company's sale to Google in 2014, including restrictions on military use, and advocated for AI ethics by proposing the establishment of an AI ethics committee at Google (Hsu 2017). Recently, over 1,000 technology leaders and researchers, including Elon Musk of Tesla, have called for a pause on AI, citing significant risks to society (Metz and Schmidt 2023).

The actions taken by engineers in this context may be challenging to comprehend from the standpoint of traditional engineering ethics. Interestingly, leaders of AI companies are earnestly considering ethical issues, even amidst their busy involvement in technology development and management. Traditional engineering ethics typically grapples with the dilemma between benefits and ethics, as

complete adherence to safety and fairness can pose challenges to productivity and benefits. However, the conflict between benefits and ethics takes on a distinct form in AI.

In some instances, AI technology can only progress and yield business benefits if ethical issues associated with the technology are adequately addressed. For example, achieving a fully autonomous vehicle driven by AI requires the resolution of ethical dilemmas related to specific traffic situations. In essence, ethical considerations transcend being merely a compromise between technology development and ethics; they emerge as a pivotal factor influencing various fields, including industry, management, and science and technology. As ethical challenges are addressed, they not only lead to the advancement of AI technology but also contribute to the sustainability of businesses (Research Society of AI and Value 2021).

## The Evolution of Social Norms for AI

### AI Ethics Discussions Led by Technology Experts

As mentioned earlier, the discourse on AI ethics originated from technical experts involved in the development of AI technology. In the early 2000s, Eliezer Yudkowsky, an AI researcher and writer, advocated for the concept of 'Friendly AI.' His argument emphasized aligning the purpose of AI with the goals of society, asserting that friendliness should be integrated into the design from the inception of AI development. In 2014, research initiatives were undertaken with the objective of 'Beneficial AI,' spearheaded by the US Future of Life Institute. These movements prompted AI experts to formulate various AI ethics principles and foster the adoption of responsible AI policies within leading IT companies at the forefront of AI development, such as Google, Microsoft, and IBM.

**Asilomar AI Principles**

In 2017, the Future of Life Institute unveiled 23 AI principles during a conference held in Asilomar. Prominent AI research experts and endorsers, including Demis Hassabis of DeepMind, Elon Musk of Tesla, Ilya Sutskever of OpenAI, Yoshua Bengio of the University of Montreal, and Stuart Russell of the University of California, Berkeley, participated. More than 90% of the attendees reached an agreement on the proposed ethical principles.

These principles were categorized into three groups: Research Issues, Ethics and Values, and Longer-term Issues. Research Issues focus on developing AI that is beneficial to humans, emphasizing the need for funding to ensure the constructive use of AI. The principles also advocate for positive exchanges between AI researchers and policymakers, the establishment of a culture of cooperation, trust, and transparency between researchers and AI developers, and collaboration to uphold safety standards in AI development.

Ethics and Values encompass 13 principles, including Safety, Failure Transparency, Judicial Transparency, Responsibility, Value Alignment, Human Values, Personal Privacy, Liberty and Privacy, Shared Benefit, Shared Prosperity, Human Control, Non-subversion, and AI Arms Race.

Longer-term Issues consist of five principles: Capability Caution, Importance, Risks, Recursive Self-Improvement, and Common Good (the Future of Life Institute 2017).

**IEEE, Ethically Aligned Design**

Since 2016, IEEE has been preparing 'Ethically Aligned Design,' a document that encompasses both technical and socio-humanitarian considerations, and it was completed in 2019. The document delves into the implementation of values in Autonomous and Intelligent Systems, ethical design methodologies, personal data access control, economic and humanistic issues, Policy and Law, Classical Ethics, and Well-being from various perspectives. The development of these ethics guidelines engaged experts from diverse fields worldwide, including researchers, educators, and business professionals. The collaboration extended beyond researchers and educators in the humanities, social sciences, and natural sciences to encompass specialists in electrical and electronic engineering, such as AI (The IEEE Global Initiative 2019).

**IT Companies' AI Policies**

The establishment of AI norms is not only a subject of discussion among researchers, as mentioned earlier, but is also actively pursued by leading IT giants in the field of AI technology development.

In 2018, Google announced its AI principles, outlining seven goals for AI applications and four criteria that should not be pursued in AI applications. These principles emphasize the importance of AI being socially beneficial, preventing unfair bias, being developed and tested safely, being accountable to people, and complying with privacy design principles. Google also aims to maintain a high level of scientific excellence and commits not to use AI technology for weapons or human rights violations. To put these principles into practice, 'Responsible AI' practices were introduced (Pichai 2018).

Microsoft, too, has declared its commitment to 'Responsible AI' and outlined principles of Fairness, Reliability and safety, Privacy and security, Inclusiveness, Transparency, and Accountability. The Office of Responsible AI is tasked with creating rules and governance related to responsible AI across the company. In 2017, Microsoft established the Aether Committee, which stands for AI, Ethics, and Effects in Engineering and Research, to advise the company's management on the challenges and opportunities presented by AI innovation (Askell, et al. 2019).

IBM emphasizes that simply implementing AI as a performance-oriented paradigm is insufficient for AI design. IBM places a strong emphasis on Building trust in AI,

stressing the need to learn how to build, assess, and monitor trust. They are developing various approaches to achieve this, including addressing explainability, fairness, robustness, transparency, and privacy (Rossi 2018). IBM has distributed various toolkits to ensure adherence to ethical principles during development and created a system for monitoring and management through 'Watson OpenScale.' Also, 'AI FactSheets 360' helps consumers understand how AI models and services are created, enabling them to make informed decisions about suitable AI for their specific situations and enhancing the overall reliability of AI.

### National-level Ethical AI Principles

The AI ethics discussions initiated by researchers and companies have resulted in the development of national AI ethics principles led by governments. This signifies that governments are actively addressing the risks posed by AI, recognizing its significance as a crucial technology for national industrial competitiveness and social development. AI is no longer solely a concern for companies and researchers.

Each country acknowledges the inescapable impact of AI ethics-related issues on AI technology and the necessity for regulatory control. Consequently, governments are proactively taking steps to anticipate and control these impacts, establishing standards to enhance national AI technology competitiveness. In the historical context between 2016 and 2020, a period marked by a surge in the enactment of AI ethics principles, 42 government-led AI ethics principles were established as numerous countries sought to formalize AI ethics norms (AlgorithmWatch 2020).

Examining the ethical principles of these countries, the overarching goal is to create an AI society centered around Human Centricity and Public Benefits, with fundamental ideology. The core principles supporting this vision include Accountability, Fairness, and Transparency. To ensure these core principles, detailed tasks such as Privacy, Safety & Security, Diversity, Reliability, Auditability, and Contestability are emphasized. Ultimately, it is emphasized that all elements of ethical principles must ensure Sustainability for these values to effectively function in society (Lim and Kwon 2021).

### Formation of International Ethical Norms

The enactment of AI ethics principles by individual countries has sparked discussions about the need for a common AI ethics norm in the international community. However, due to variations in each country's AI ethics principles, focusing on different items and varying in content based on individual circumstances, establishing a unified AI norm at the international level has proven challenging (Nakagawa 2020).

Nevertheless, in 2018, the G7 Innovation Ministers' Statement on AI was released (G7 Innovation Ministers 2018), urging the international community to strive for a vision of human-centered AI. This vision considers addressing social challenges, promoting economic growth, ethical considerations of AI, and biases in data sets.

In Europe, in April 2019, the EU Commission's High-level expert group on AI announced the 'Ethics Guidelines for Trustworthy AI,' providing a universal set of AI principles within Europe. Trustworthy AI integrates human-centered approaches to AI technology, alignment with human social value systems, stability and fairness, and transparency and explainability of systems. After the announcement of the guidelines, over 350 organizations provided feedback, and in February 2020, an AI white paper was issued, emphasizing that AI research and development in Europe should be approached based on excellence and reliability. The Joint Research Centre monitors Europe's industry, technology, and research and shares the results through the AI Watch portal (European Commision 2024).

In May 2019, the OECD proposed the first AI recommendations with international-level consensus. The Values-based principles emphasized Inclusivity and Sustainability, Human Values and Fairness, Transparency and Explainability, Robustness, Security and Safety, and Accountability. Policy recommendations included investment in R&D, creating a digital ecosystem, fostering a flexible policy environment for innovation, human capacity building, responding to job transformation, and the need for international cooperation (OECD 2019).

In November 2021, UNESCO released the 'Recommendation on the Ethics of AI,' a global standard for AI ethics. This recommendation, adopted by all 193 member states, establishes a common AI ethics norm globally. The cornerstone of the recommendation is the protection of human rights and dignity, based on fundamental principles like transparency and fairness. It emphasizes the importance of human supervision of AI systems and calls for extensive policy implementation to translate core values and principles related to AI into action (UNESCO 2022). As of 2024, UNESCO is hosting the Global Forum on the Ethics of AI, working on concrete and practical solutions beyond principles, and developing a system for comprehensive and multidimensional assessment of global AI ethics preparedness.

## From Ethics to Law

### Limitations of Ethical Principles and the Need for Legal Regulation

Currently, AI ethics is faltering in numerous instances. It lacks a mechanism to enforce norms, resulting in a lack of consequences even when ethical principles are violated. Moreover, companies and institutions often exploit ethics as a mere marketing strategy. Ultimately, ethical guidelines

have minimal impact on the decision-making of AI developers (Munn 2023; Kaspersen and Wallach 2021). Consequently, many countries are striving to move beyond AI ethics and establish effective norms.

In any society, certain issues cannot be solely addressed by ethical norms and require legal enforcement. However, it is also true that not all concerns can be regulated by law. Accurately defining AI, conducting a normative evaluation of all potential consequences of AI technology, and clearly regulating it in a singular code of law is virtually impossible. Therefore, legal regulation of AI will encompass only the absolute minimum basic requirements that cannot be overlooked. The determination of which matters remain in the realm of ethics and which transition to the realm of legal norms will depend on the values pursued by each society and the speed at which social consensus is achieved.

## AI Regulation Trends

### EU AI Act

The European Union has been actively discussing and developing a system to regulate AI, aiming to protect the values of European citizens while simultaneously enhancing the EU's competitiveness in AI. This effort is primarily led by the Commission, and several key milestones have been achieved in this regard.

In 2018, the 'European AI strategy' was announced, followed by the development of AI guidelines in 2019 and the creation of a trustworthy AI assessment list in 2020. Notably, the 'AI white paper' issued in 2020 presented a clear vision for Europe's AI, emphasizing the need for rules to ensure that AI systems are safe, transparent, ethical, unbiased, and controllable by humans (European Commission 2020). As a culmination of these efforts, on April 21, 2021, a draft of the 'AI Act' was proposed (European Commission 2021), and approved by the European Parliament on 14 March 2024.

The AI Act is recognized as the world's first law on AI and is considered a sophisticated and innovative piece of legislation. A key characteristic of this law is the classification of AI based on the degree of risk into categories such as Unacceptable risk, High risk, General purpose and generative AI, and Limited risk. The level of regulation is then differentiated according to the degree of risk.

Unacceptable risk includes AI that poses a clear threat to people's safety, living, and rights, and systems that manipulate behavior while interfering with people's free will. The production and use of such AI are fundamentally prohibited.

High-risk AI, which is subject to the most highlighted regulation in the Act, is considered a high-risk system if it negatively affects safety or fundamental rights. For example, AI systems used in products subject to the EU's product safety legislation and AI systems falling within specific areas, such as critical infrastructure, education, and law enforcement agencies, undergo a life cycle assessment before being released to the market. Essential requirements include building a risk management system, applying data governance, record-keeping and labeling obligations, automatic logging, transparency and information provision, human supervision, accuracy, robustness, cybersecurity assurance, and the installation of a quality management system.

General purpose and generative AI, including models like ChatGPT, are obligated to comply with transparency requirements. This involves disclosing that the content was generated by AI, designing models to prevent the generation of illegal content, and publishing summaries of copyrighted data used for training. Highly influential general-purpose AI models, such as GPT-4, must undergo a thorough evaluation and report any serious incidents to the European Commission.

Regarding limited risk, AI must comply with minimum transparency requirements to allow users to make informed decisions.

### AI legislation in the United States

In November 2020, the Office of Management and Budget (OMB) released 'Guidance for Regulation of AI Applications' aimed at developing legal regulations for AI (Vought 2020). The guidance underscores the significance of AI in maintaining the nation's economy and national security, focusing on considerations for regulations or policies related to AI applications. It emphasizes the evaluation of AI-related risks and the importance of implementing risk management based on those evaluations.

In April 2019, the 'Algorithmic Accountability Act of 2019' was introduced to strengthen the accountability of companies using AI algorithms. The bill proposed that organizations utilizing, storing, and sharing personal information conduct impact assessments for automated decision-making systems and data protection (Quandary Peak Research 2020). However, the bill faced criticism for potential hindrances to AI innovation and overregulation.

In February 2022, a new bill, the 'Algorithmic Accountability Act of 2022,' was introduced, incorporating feedback from various stakeholders, including addressing criticisms (Bartlett, et al. 2020). The goal is to prevent new risks arising from errors or biases as the use of algorithms expands. The focus is on providing a pre-safety device called impact assessment, requiring companies to develop and distribute algorithms to evaluate the impact of automated decision-making systems or critical decision-making processes on consumers. The bill grants the Federal Trade Commission (FTC) various powers and obligations to ensure the effectiveness of these impact assessments, including establishing specific regulations on impact assessments and publishing annual reports.

On October 30, 2023, the Biden administration announced its first executive order regulating AI (The White House 2023). The order mandates safety checks by a

government-verified team of experts (AI Red Team) for AI models that could threaten the security, health, and safety of the United States, with developers required to submit the results to the government. It recommends that the National Institute of Standards and Technology (NIST) establish the highest standards for AI technology safety. Additionally, the order establishes content authentication standards, requiring AI developers to mark content with identifiable watermarks to prevent the spread of false information using AI. Lastly, it addresses privacy protection, particularly for generative AI like ChatGPT, emphasizing guidelines to regulate the illegal use of personal information in AI development and training due to the extensive collection of such data for algorithm training.

**AI legislation in China**

While China does not have a comprehensive AI regulation comparable to the EU AI Act, it has implemented specific regulations for certain AI technologies. Notable examples include the Deepfake Regulation and the Generative AI Regulation.

The Cyberspace Administration of China (CAC) has been enforcing the Provisions on the 'Administration of Deep Synthesis of Internet Information Services' since January 10, 2023, which regulates deepfake technology (Hine and Floridi 2022). According to these regulations, providers of deepfake services and content creators must visibly indicate when they have used the technology to create content and attach a digital watermark to track the original. Consent must be obtained if the technology is used to edit someone's image or voice, and if used to report news, the original must come from a government-approved media outlet.

On July 13, 2023, the 'Interim Measures for the Management of Generative AI Services (Interim GAI Measures)' were promulgated, and on August 15, 2023, regulations on generative AI technology, generated content, and products were implemented. The Interim GAI Measures require that generated content reflect core socialist values and not disturb economic and social order. Discrimination against certain races and genders must be prevented in algorithm design, model generation and upgrading, and service provision. Service providers must take measures to prevent intellectual property rights infringement, trade secret leakage, and the generation of false information. It is also mandatory to submit security evaluation results to the CAC before providing services, comply with relevant laws such as the Cybersecurity Law, Data Security Law, Copyright Law, and Personal Information Protection Law, report non-compliant content to the competent authority, and suspend the service. Responsible use is emphasized by requiring users to register for the service under their real names (China Law Translate 2023).

## Roles and Responsibilities of Stakeholders

The societal norms governing AI are evolving beyond declarative ethics and are now progressing toward mandatory compliance. AI regulation entails imposing obligations on various stakeholders in the field of AI technology to enhance aspects such as fairness, transparency, safety, and accountability, which have been recurring themes in the overarching topic of AI ethics.

Present ethical guidelines and legal frameworks predominantly focus on regulating AI technology developers and providers of AI services. Additionally, numerous ethical principles emphasize the need for users to engage in ethical and responsible practices. Beyond this, considerations must also be extended to the roles and responsibilities of government and regulatory authorities, as well as organizations such as academia and civil society.

### Compliance requirements for developers and providers

As mentioned earlier, recent AI regulation adopts a risk-based approach. Consequently, developers and providers must ascertain whether the AI they develop and commercialize is legally permissible or requires a conformity assessment before product launch.

Specifically, developers and providers of AI with high risks are obligated to possess the capability to assess the suitability and social impact of the AI. Certain requirements must be met in the development and operation of the system, which include:

1. Risk management: This entails identifying and analyzing foreseeable risks before the system launch and systematically managing risks through monitoring after the system launch to ensure safety and robustness.
2. Data governance: This involves ensuring security through data protection and data security, guaranteeing fairness through the review of data bias, and ensuring accuracy through the identification of data gaps and deficiencies.
3. Compliance with regulations and cooperation with authorities: This requires adherence to national and international standards related to AI, ensuring accountability through the preparation of technical documentation, and establishing reporting systems.
4. Record keeping and transparent information provision: This encompasses monitoring and record-keeping throughout the lifecycle and providing users with information about the system to ensure transparency.

### Compliance requirements for users

While there aren't many direct regulations for users, the EU AI Act imposes obligations on users to comply with instructions and monitor system operation.

1. Compliance with instructions: This entails users using the system ethically and appropriately in line with its intended purpose. Users need to enhance their understanding and ability to use AI technology responsibly.
2. Monitoring system operation: This involves users promptly notifying developers and providers of any significant errors or accidents when using AI and taking appropriate action. Additionally, users should assume responsibility for safeguarding personal information and security when using AI technology and should be mindful of the potential consequences of their actions.

### Role of government and regulatory authorities

The approach of each country to regulating AI varies. Some countries, like the EU, have comprehensive laws, while others individually regulate specific AI technologies, such as deepfakes and generative AI. In terms of regulation, some countries opt for government-led approaches, while others pursue self-regulation. However, regardless of the form, the role of government in addressing the negative effects of AI is likely to be similar.

Governments must clarify what aspects need regulation to ensure ethical and social responsibility in AI. They need to establish standards for verifying and evaluating AI and develop related policies (de Almeida, et al. 2021). Particularly, careful consideration and monitoring are required to ensure that regulations do not excessively hinder AI research and development processes, undermine the competitiveness of existing businesses and startups, and disrupt market order.

### Role of academia and educational institutions

Creating a trustworthy AI society cannot be achieved solely through laws or guidelines. Ultimately, all members of society, from developers to users, must be capable of empathizing with the ethical and social responsibilities associated with AI. In this context, the role of academia and educational institutions becomes more crucial than ever (Borenstein and Howard 2021). Their educational efforts should encompass literacy enhancement for users and ethics education to reinforce responsibility among developers and providers. Additionally, they should actively contribute to the formulation of policies by governments and regulatory authorities through research on ethical and social responsibility in AI.

### Role of civil society organizations

Civil society organizations play a crucial role in establishing social norms for AI. While the government's responsibility is to establish AI regulations, civil society organizations have a critical role in critically examining whether these regulations sufficiently consider the infringement of users' rights. They play a vital role in urging improvements from the perspective of the people. As AI regulation must strike a proper compromise between social and economic development, civil society organizations need to actively contribute to social consensus-building within civil society. Furthermore, there is a need to monitor and evaluate the ethical and social responsibility of both government and corporate AI (Buhmann and fieseler 2021).

## Conclusion

The development of AI technology is progressing faster than ever before. With the rapid development of AI, discussions on AI ethics have deepened, and recently, there has been a move to directly regulate AI. Although various types of AI regulation have been proposed, these regulations have not yet been fully implemented. At this stage, it is very important to conduct a realistic analysis and evaluation of the risks and dysfunctions posed by AI for AI regulation to achieve its intended effect, and to identify the obligations and roles of all stakeholders to prevent them (Miller 2022).

Once AI regulation is applied, developers and providers are expected to have very specific obligations and compliance requirements. They must be familiar with the regulatory compliance requirements for the entire lifecycle of AI, from planning and development to after launch, and have the capability to assess the social impact of their AI. However, enforcing the law should be done with minimal intervention in areas that require regulation, and should be clear to avoid confusion among regulated parties. It also requires careful attention to ensure that it does not hinder technological progress and social prosperity. To this end, governments and regulatory agencies will need to carefully review the content of regulations and continuously monitor regulatory improvements in line with technological trends, from the adoption of regulatory methods. Furthermore, the implementation of regulations and policies necessary for our society will be possible when the monitoring function of civil society organizations works well in the performance of the roles of the government and the corporate sector.

In addition, the implementation of a ‘Trustworthy AI society’ is not the sole responsibility of developers and providers. As AI is used in more areas of everyday life, the responsibility of users has become very important. Users can be legally regulated, but what is most important is that users use AI ethically and correctly. For this, the role of the education system is needed. It should not only foster professionals in AI technology but also lead the public to acquire digital citizenship.

Now we are in the battle against the dysfunctions of AI in earnest. The problem can only be solved when we can properly control AI technology. To do this, a holistic approach is needed from all stakeholders, not just technologists and developers, and the whole society should be in an environment where it can control AI.

## References


AlgorithmWatch 2020. AI Ethics Guidelines Global Inventory, https://inventory.algorithmwatch.org/

Anand S. R. 2018. Responsible AI & National AI Strategies, European Union Commission.

Askell, A., Brundage, M., and Hadfield, G. 2019. The role of cooperation in responsible AI development. *arXiv preprint arXiv:1907.04534*.

Bartlett, R.; Morse, A.; Stanton, R.; and Wallace, N. 2020. Algorithmic Accountability -A Legal and Economic Framework.

Borenstein, J., and Howard, A. 2021. Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, 1: 61-65.

Buhmann, A., and Fieseler, C. 2021. Towards a deliberative framework for responsible innovation in artificial intelligence. *Technology in Society*, 64: 101475.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., and Floridi, L. 2018. Artificial intelligence and the 'good society': the US, EU, and UK approach. *Science and engineering ethics*, 24: 505-528.

Cheatham, B., Javanmardian, K., and Samandari, H. (2019). Confronting the risks of artificial intelligence. *McKinsey Quarterly*, 2(38): 1-9.

China Law Translate 2023. Interim Measures for the Management of Generative Artificial Intelligence Services 2023 (Translation).

de Almeida, P. G. R., dos Santos, C. D., and Farias, J. S. 2021. Artificial intelligence regulation: a framework for governance. *Ethics and Information Technology*, 23(3): 505-525.

European Commision 2024. A European approach to artificial intelligence. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

European Commission 2020. White Paper on Artificial Intelligence－A European approach to excellence and trust, Pub. L. No. COM(2020) 65 final.

European Commission 2021. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, Pub. L. No. COM(2021) 206 final.

G7 Innovation Ministers 2018. Annex B:G7 Innovation Ministers' Statement on Artificial Intelligence, Montreal, Canada.

Galanos, V. 2019. Exploring expanding expertise: artificial intelligence as an existential threat and the role of prestigious commentators, 2014−2018. *Technology Analysis & Strategic Management*, 31(4): 421-432.

Hine, E., and Floridi, L. 2022. New deepfake regulations in China are a tool for social stability, but at what cost?. *Nature Machine Intelligence*, 4(7): 608-610.

Hsu, J. 2017. Tech Leaders Are Just Now Getting Serious About the Threats of AI. WIRED, 2017.01.27. https://www.wired.com/2017/01/tech-leaders-are-just-now-getting-serious-about-the-threats-of-ai/

Kaspersen, A., and Wallach, W. 2021. Why Are We Failing at the Ethics of AI?. *Carnegie Council, Artificial Intelligence & Equality Initiative, Nov, 10*.

Leslie, D. 2019. *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. The Alan Turing Institute.

Lim, J. H., and Kwon, H. Y. 2021. A Study on the Modeling of Major Factors for the Principles of AI Ethics. In *DG. O2021: The 22nd Annual International Conference on Digital Government Research*: 208-218.

Metz, C., and Schmidt, G. 2023. Elon Musk and Others Call for Pause on A.I., Citing 'Profound Risks to Society.' The New York Times, 2023.03.29. https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.html

Miller, G. J. 2022. Stakeholder roles in artificial intelligence projects. *Project Leadership and Society*, 3: 100068.

Munn, L. 2023. The uselessness of AI ethics. *AI and Ethics*, 3(3): 869-877.

Nakagawa H. 2020. Focuses of Recent AI Ethics Guidelines and Personal AI Agent. *Journal of Information and Communications Policy*, 3(2): 1-24.

OECD 2019. Recommendation of the Council on Artificial Intelligence. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Pichai, S. 2018. AI at Google: our principles. The Keyword, 7, 1-3.

Quandary Peak Research 2020. On Our Radar: The Algorithmic Accountability Act. https://quandarypeak.com/2020/03/on-our-radar-the-algorithmic-accountabi

Research Society of AI and Value 2021. AI Ethics: Multidisciplinary Approach. Pakyoungsa: Seoul.

Rossi, F. 2018. Building trust in artificial intelligence. *Journal of international affairs*, 72(1), 127-134.

Russell, S. J., and Norvig, P. 2021. *Artificial intelligence a modern approach*, 4th US ed. University of California at Berkeley.:981-982.

the Future of Life Institute 2017. Asilomar AI Principles, https://futureoflife.org/open-letter/ai-principles/

The IEEE Global Initiative 2019. Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, first edition. https://standards.ieee.org/content/ieee-standards/en/industryconnections/ec/autonomous-systems.html

The White House 2023. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. October 30, 2023. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

Ulnicane, I., Eke, D. O., Knight, W., Ogoh, G., and Stahl, B. C. 2021a. Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies. *Interdisciplinary Science Reviews*, 46(1-2): 71-93.

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., and Wanjiku, W. G. 2021b. Framing governance for a contested emerging technology: insights from AI policy. *Policy and Society*, 40(2): 158-177.

UNESCO 2022. Recommendation on the Ethics of Artificial Intelligence. Adopted on 23 November 2021. United Nations Educational, Scientific and Cultural Organization, 7, place de Fontenoy, 75352 Paris 07 SP, France.

Vesnic-Alujevic, L., Nascimento, S., and Polvora, A. 2020. Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, 44(6): 101961.

Vought, R. T. 2020. Re: Guidance For Regulation Of Artificial Intelligence Applications.