CASE STUDY

Analyzing the guiding principles of AI ethics: A framing theory perspective on the communication of ethical considerations in artificial intelligence (AI)

Asifa Younas[*]

M. Phil-HRM, Superior University, Lahore, Pakistan

[*]**Correspondence:**
Asifa Younas,
asifayounas12@gmail.com

Various organizations have created AI ethics standards and protocols in an era of rapidly expanding AI, all to ensure ethical AI use for the benefit of society. However, the ethical issues raised by AI's societal applications in the actual world have generated scholarly debates. Through the prism of framing theory in media and communication, this study examines AI ethics principles from three significant organizations: Microsoft, NIST, and the AI HLEG of the European Commission. Institutional AI ethics communication must be closely examined in this rapidly changing technical environment because of how institutions frame their AI principles.

Keywords: Artificial intelligence, AI ethics, AI principles, framing theory, TRUST, AI framings

1. Introduction

A new era marked by artificial intelligence's (AI) ubiquitous influence across many sectors has begun with AI technology's rapid growth. Many industries, including healthcare, aerospace, banking, entertainment, and many more, have been impacted by this technological transformation, which is sometimes referred to as the "fourth industrial revolution." These businesses are all trying to increase productivity and efficiency while cutting costs. In this sense, artificial intelligence (AI) describes computer programs that mimic human intelligence processes, carrying out or even surpassing human performance (1).

The application of AI technology is challenging, though. Biases from training data are known to be inherited by AI systems, which can have unforeseen repercussions and promote inequality in a variety of domains. Examples of this problem include instances of gender bias in research publishing and racial prejudice in healthcare projections (2). These biases have sparked questions regarding the reliability of AI systems and their opaque decision-making procedures, especially because sophisticated AI technologies like deep learning are still difficult for people to understand (3).

To ensure responsible use and shape the development of AI technology, it is imperative to define ethical rules and guidelines in light of these challenges. Notably, leading technology corporations have taken action to regulate their AI endeavors, such as Microsoft with its Responsible AI framework (4). Recognizing the strategic significance of AI for innovation, equity, and security, the US government has also joined the AI standards and regulatory space through the National Institute of Standards and Technology (NIST) (5). Furthermore, through its High-Level Expert Group on AI (AI HLEG), the European Union has been actively involved in creating ethical standards for AI, with an emphasis on an approach to AI ethics that is human-centric (6).

These many pioneering institutions' conceptualization of these institutional ethical principles for AI technology provides insights into regulating AI's social and technological advancement (7). Understanding the guiding concepts behind AI development and deployment is essential to ensure that these technologies remain reliable, open, and consistent with human values as they become increasingly integrated into our daily lives (4).

The proliferation of ethics guidelines by multiple organizations has fractured the debate on AI ethics, making it difficult to fully understand the field and making the pursuit of equitable implementation more difficult (8). Many organizations, such as user groups, government agencies, and developers, have published AI ethics principles (9). As a result, there are a lot of similarities and discrepancies between their efforts to create practical rules for the benefit of society (10). There needs to be a broad agreement on normative frameworks and standard norms for AI ethics (11). The central question is how to define "common good" and "social benefit" in an increasingly globalized and digitalized world (12). This calls for clear definitions of justice, human rights, and widely acknowledged values, as well as how to identify potential risks in AI applications that have the potential to support or contradict these values in various social and economic contexts (4).

This research is important because it offers a semi-systematic overview of governance, legislation, and ethics in AI and sheds light on how the area of AI ethics is developing (13). It tackles ethical issues and conflicts in formulating and disseminating ethical AI principles by classifying AI guidelines and pointing out institutional overlaps and omissions (14). As AI technology continues to advance in societal use cases, research helps to bring hidden tensions, fresh viewpoints, and tech-business social agendas to the fore (15). This promotes conflict resolution and progress. By offering insightful information for regulatory strategies and assurance services, this study adds to the continuing conversation on AI ethics (16). It guarantees stakeholders' comprehension of AI technology's performance, risk, and compliance (17). Additionally, by using framing theory to study institutional AI ethics principles and norms, it highlights the crucial roles that trust and understanding play in sophisticated AI technologies and their communication (18).

1.1 Literature review

1.1.1 Framing theory literature: a viewpoint for research and instrument for communicating AI ethics

One of the first academics to define the term "framework" was (19), who described frames as "schemata of interpretation" for understanding what has happened (20) frames assist in bringing seemingly unrelated occurrences into coherent wholes. The intricacy of framing was emphasized by pointing out that there might be frames inside frames (21). According to (3) framing is the process of choosing which parts of reality to highlight in a communication to support particular problem definitions, causal interpretations, moral assessments, or therapeutic suggestions (22).

The conceptualization and communication of climate change in Swedish agriculture were examined by (11), emphasizing the discrepancy between farmers' perceptions and media portrayals of the issue. (20) used framing analysis to examine how the news media covered the IPCC Fifth Assessment Report on climate change to find dominating frames (23).

Research on framing in political science and sociology looks at the words, pictures, sentences, and ways that news items are presented, as well as the processes that shape them (24). Diverse theoretical and methodological approaches to framing have been given by many scholars [(25), Matthes, 2009 #629].

While framing and agenda-setting are similar, framing concentrates on the substance of issues rather than particular subjects (10). Discourse analysis and the idea of the explanatory theme are connected to framing (20). Four framing processes were distinguished by (26):

- Frame creation
- Frame Placement
- The consequences of frames at the individual level
- The audience role of journalists

1.1.2 TRUST framings serve as the study's academic framework

Transparent and understandable AI systems are required to solve the "black box problem" in AI (4). To reduce dangers and improve confidence in AI decision-making processes, academics and organizations are developing technological and moral regulation strategies (9).

AI development and application heavily depend on the public dissemination of AI principles and guidelines (25). These published AI ethics principles do, however, have some notable distinctions, similarities, and conflicts (9). The project's goal is to find important TRUST framings in texts, including AI concepts and guidelines (27).

AI principles and guidelines writings that incorporate issues such as interpretability, transparency, comprehensibility, and explainable AI are called transparent and understandable AI framing (The Royal Society, 2019, 28).

Safe and Reliable AI Framing: Covers safety management procedures, public reporting of issues and future goals, and reliability (4).

Human augmentation, user control, autonomy, and consent are the main topics of the User Control and Autonomy Framing [(4, Endsley, 2018 #634].

Data security, privacy, and the requirement for secure AI systems are all covered under the "Secure and Privacy AI Framing" (29).

Changing narratives surrounding the complexity, risks, and issues surrounding artificial intelligence, such as ethical conundrums, human resources, employment, rights, accessibility, fairness, non-discrimination, justice, inclusion,

diversity, solidarity, accountability, whistleblowers, and AI audits; additionally, hidden costs associated with AI and responsible research funding (30). These scholarly frameworks provide a basis for comprehending the various facets of communication on AI ethics (31).

1.1.3 Research questions

RQ1: What kinds of frameworks are included in the text of the selected organizations' AI principles and guidelines?
RQ2: How much do the framings that these institutions use correspond with or mimic the TRUST framings that are explained in this study? These frameworks include The Other Framings, User Control and Autonomy, Secure and Privacy AI, and Transparent and Comprehensible AI.

2. Methodology

The goal of the study is to examine AI ethics communication in the context of leading AI organizations' AI principles and guidelines—Microsoft, NIST, and AI-HLEG, in particular—and to distinguish different framings in their communication about AI ethics. The TRUST is used to identify these framings—Framings from the AI literature review that were developed in the preceding part. The selection of these AI firms for analysis was done with great care to reduce the possibility of author bias. Other prominent AI organizations were not included in the analysis because of unclear institutional approaches to AI research, innovation, and self-regulation, ongoing ethical disputes that have been covered in the media recently (like Google's Project Maven), or past ties to the author. The processes for gathering textual data and the researcher's approach to locating frames in the AI messages of the selected universities are described in the part that follows.

Phase 1: The researcher gathered the text data from the open-access AI principles and standards published on the websites of the chosen three institutions. **Table 1** contains the source links for this text data.

2.1 Data sources

Phase 2: As Matthes (2009) noted in their methodical examination of media framing studies published in prestigious communication journals, frame analysis is an essential technique for closely examining the selection and prominence of particular components of a problem {(Guenther, 2023) #745}. The framings within the textual data were manually identified using the (3) concept of framing and the academic sources cited in the literature study. The framings included in the AI principles language of the chosen firms were identified using inductive and deductive methods (7). Based on the qualitative paradigm of frame analysis, which holds that frames are visible through particular words, this study explores framings using direct quotations taken from the selected AI pioneers' recently developed and published AI principles and guidelines, making connections with different aspects of the current scholarly debate on AI ethics (32). During the textual study of Microsoft, NIST, and AI-HLEG's AI principles and guidelines, the identification of frames was led by the systematic processes described by (22) in 'Frames in Communication'.

Describing the process for identifying certain framings is crucial before providing the research and findings (33). "When researchers employ computer programs for analyzing large volumes of text, they must identify the universe of words that signal the presence of a frame," according to guidelines (34). This study's academic framing literature review phase found theme words indicative of the identified framings in the sample text on AI principles and guidelines. It is important to remember that identifying "frames in communication" entails being aware of the important points highlighted in a speaking act. In the methodology, which lacks uniform measuring standards, persuasive communication research adheres to four essential steps: (1) Identifying a particular problem, occasion, or person; these components define communication frames. (2) Isolating particular attitudes to understand how frames shape

TABLE 1 | Artificial Intelligence (AI) principles data for textual analysis as downloaded in Dec 2021.

AI Principles	Microsoft	AI-HLEG	NIST
Published Document Source Links	https://www.microsoft.com/en-us/ai/principles-and-approach	https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf	https://www.nist.gov/system/files/documents/2020/08/17/NIST%20Explainable%20AI%20Draft%20NISTIR8312%20%281%29.pdf
Active Web Links	https://www.microsoft.com/en-us/ai/responsible-ai	https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai?view=azureml-api-2 https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence	https://www.nist.gov/artificial-intelligence
Document Length	13 full-length webpages with text on AI approach (7 video transcripts and 6 additional AI guideline blog entries) and 1 training module with 9 units	24 pages (August 2020) Draft NISTIR 8312 and website updates on AI principles.	36 pages (additionally 1 page mentioning High-Level Expert Group members) of Deliverable 1 (Ethics Guidelines for Trustworthy AI) and web links to Deliverables 2, 3, 4.

public opinion (32). (3) Determining an issue's starting set of frames inductively to create a coding scheme. (4) Using the original set of frames that have been identified to select the content sources for analysis.

All of the methods above for finding framing were followed, except the second stage, which examined how frames influence public opinion, given the study's goals and scope {(Mhlanga, 2020) #744}. Previous sections identified and explained specific topics, pertinent events, examples, AI actors, and the chosen sample institutions. The academic framing literature review portion identified and elaborated on an initial set of framings corresponding to the concerns covered. Regarding the last phase (32), the study's introductory part detailed the textual selection of AI principles and guidelines taken from three institutional sources for analysis.

3. Results or finding

As was already mentioned, every institution's AI principles should encourage risk reduction and problem-solving related to this new technology. This insight is related to Goffman's person-role formula, which states that an AI actor's social role is closely related to Its type. The framings of the AI principles and guidelines are soft (because there is no legal obligation) but strong (as they take into account each position's/society's role's priorities) (35). The following two research problems are addressed by the AI ethics principles and guidelines text analysis:

RQ1: What framings can be found in the AI principles and guidelines text of the chosen institutions?

The High-Level Expert Group on Artificial Intelligence (AI HLEG) was established by the European Commission to foster trust in the AI system's entire life cycle (from development to deployment, from planning and communication to policy and investment recommendations). They produced a comprehensive guiding document that is currently influencing Europe's overall AI approach to empower, benefit, and safeguard European citizens (18). In addition to the guidelines, which are referred to as the "Ethics Guidelines for Trustworthy AI," the expert group produced three other deliverables: the AI Ethics Guidelines document itself included Sectoral Considerations on the Policy and Investment Recommendations, Assessment List for Trustworthy AI (ALTAI), and Policy and Investment Recommendations for Trustworthy AI. The AI ethics standards serve as the cornerstone upon which more comprehensive texts are constructed. Following the foundation chapter on Ethics Guidelines, each extension above receives a full chapter treatment.

RQ2: Which of the institutional framings are the same as or similar to TRUST framings explained in this study? (Where TRUST Framings indicate Transparent and Comprehensible AI Framing, Reliable and Safe AI Framing, User Control and Autonomy Framing, Secure and Privacy AI Framing, and The Other Framings).

The principles that underpin the guidelines drafted by the AI high-level expert group are rooted in Ethics in Science and New Technologies and the Fundamental Rights Agency (36). These three components are adhering to legal requirements, upholding ethical principles, and providing assurance of "robustness" (specifically, "technical robustness" combined with safety measures for humans, animals, and the environment in a variety of settings, as well as fallback plans)—all taken from AI HLEG's EU documents and assessment list for trustworthy-AI.

According to (9), the standards specify essential requirements that are not legally binding. Although the seven conditions don't impose any new legal duties, they offer developers and stakeholders thorough guidance in persuading them to comply (6). Developing and implementing AI systems that meet the seven specified characteristics of AI HLEG would create reliable AI systems. The guidelines state that if AI applications respect the following:

1. Human agency and oversight.
2. Technical robustness and safety.
3. Privacy and data governance.
4. Transparency.
5. Diversity, non-discrimination, and fairness.
6. Societal and environmental well-being.
7. Accountability, then they will be considered trustworthy.

The guidelines' text and their communication to the European Parliament (18) are related to the study's other framings (diversity, non-discrimination and fairness, accountability) as well as the following: transparent and understandable AI framing, reliable and safe AI framing, user control and autonomy framing, secure and privacy AI framing, and user control and autonomy (26). **Table 2** provides some sample quotes from chosen AI principles and guidelines data documents linked to the TRUST framings of this study (37). Refer to Appendix A, **Tables 3**, **4** in the ensuing sections for further AI ethical language framing examples from Microsoft, the EU's AI HLEG, and NIST's AI principles and guidelines (38).

Transparent and Comprehensible AI Framing Because advanced artificial intelligence (AI) systems in social settings can be complicated, NIST, a federal non-regulatory agency under the U.S. Department of Commerce whose goal is to foster innovation and industrial competitiveness in the

TABLE 2 | Examples of Identified Framings in the Institutional AI Ethics Principles and Guidelines Text data (EU's AI HLEG, Microsoft, NIST).

Framing	Identifying Word/Phrase	Examples
Transparent and Comprehensible AI Framing	Transparency, Explainability, Interpretability, Comprehensibility	"Per-decision explanations provide a separate 370 explanation for each decision. . .Self-explainable models of machine learning systems themselves can be used as global explanations (since the models explain themselves). Likewise, many global explanations (including self-explainable models) can also be used to generate per-decision explanations." (NISTIR 8312, 2020, p.8)
Reliable and Safe AI Framing	Reliability, Management Practices directed toward Safety, Public reports of Problems/Failures/Misses /Future plans, Oversight Boards	"ORA [Office of Responsible AI] puts Microsoft principles into practice by setting the company- wide rules for responsible AI through the implementation of our governance and public policy work. It has four key functions." "Aether [AI, Ethics and Effects in Engineering and

country, places a strong emphasis on "transparency" in its AI principles (39). The transparency of AI systems and their understandability by human recipients of the information are the foundations of three of the four NIST AI principles (40). NIST's AI principles, which elaborate on the kinds, meanings, and precision of explanations, support The Royal Society (2019) assertion that there are several explainability approaches, which are covered under the Transparent and Comprehensible AI Framing in this study's literature review. The principles of NIST reaffirm that the nature and specifics of an explanation would differ based on the application in question and the kind of AI

TABLE 3 | Identified Framings in the Institutional AI Ethics Principles and Guidelines Text Data.

Microsoft	NIST	AI-HLEG
Transparent and Comprehensible AI Framing	Transparent and Comprehensible AI Framing (Explainability)	Transparent and Comprehensible AI Framing (Explicability)
Reliable and Safe AI Framing	Reliable and Safe AI Framing	Reliable and Safe AI Framing
NA	NA	User Control and Autonomy Framing
Secure and Privacy AI Framing	NA	Secure and Privacy AI Framing
The Other Framings (Fairness, Inclusiveness, Accountability)	The Other Framings (Accountability) Knowledge Limits Principle	The Other Framings (Fairness, We foster accountability in societal and environmental situations and encourage inclusivity for marginalized or historically underprivileged populations.
Avoid being ableist when creating, refining, or evaluating AI systems.	Prejudice, resiliency, and unjust, hurtful, or misleading results are avoided.	Promoting well-being, reducing harm, evaluating threats to democracy, the human condition, the rule of law, and distributive justice principles.

technique created and implemented in a social context (41). The text under AI principles in Microsoft's published case studies and video transcripts covers three AI framings: Secure and Privacy (words: Privacy and Security), Fairness, Inclusiveness, and Accountability, and Transparent and Comprehensible (words: Transparency and Explainability) (42). These are discussed in the academic frames section of this study's literature review (for data examples, refer to **Tables 3, 4**)

3.1 Safe and dependable AI framing

AI ethical guidelines published by an organization are considered soft law or non-legislative policy tools with persuasive language but no legal force behind them (9). Through its three offices/committees—the Office of Responsible AI (ORA), the Aether Committee (which stands for AI, Ethics, and Effects in Engineering and Research), and the Responsible AI Strategy in Engineering (RAISE)— Microsoft operationalizes its AI principles, which it has dubbed "Responsible AI." While the Aether Committee advises Microsoft's senior leadership on responsible AI issues, technologies, processes, and best practices, RAISE is an initiative and engineering team designed to facilitate the implementation of Microsoft's responsible AI rules and processes across its engineering groups (44). In summary, committees that advise Microsoft's leadership, engineering, and all other teams inside the organization provide direction as it implements its responsible AI principles. Thus, the text's six key AI principles come first.

4. Discussions

The debate highlights the significance of word choices and framing within AI principles and standards when examined through the prism of framing theory. The results of this study support the notions put out by (19) and (3) on the existence of frames within frames by showing that these frames might function as "signs of priorities" within these documents. For instance, Microsoft prioritizes some framings by partner

TABLE 4 | Examples of identified framings in the institutional AI ethics principles and guidelines text.

TRUST Framings	NIST	Microsoft
Transparent and Comprehensible AI Framing	NA	NA
Transparency	NA	"At Microsoft, we've recognized six principles that we believe should guide AI development and use — fairness, reliability and safety, privacy and security, inclusiveness, transparency, and accountability" (microsoft.com/en-us/ai/)
Explainability	"As the fundamental qualities of explainable AI systems, we provide four key principles for explainable artificial intelligence (AI). These guidelines were developed with the diverse fields of computer science, engineering, and psychology in mind while discussing explainable AI. We realize the need for various explanations to meet the specific needs of different users, realizing that no one explanation fits all circumstances. We also present an overview of explainable AI ideas and identify five explanations" (p.i).	NA
Interpretability	NA	NA
Comprehensibility	NA	NA
Reliable and Safe AI Framing	NA	"It is important to recognize that new intelligent technology has advantages but also unexpected and unintended consequences as it develops and spreads throughout society. Some of these effects are harmful and have significant ethical ramifications. As a result, we must proactively foresee and mitigate these unexpected repercussions resulting from the technology we bring into the world using deliberate actions." "The establishment of guiding principles for responsible AI is strategic planning and continuous oversight. Aether, ORA, and RAISE lead a concerted project to create responsible AI throughout Microsoft. These three organizations—Aether, ORA, and RAISE—work closely with our teams to ensure Microsoft's responsible AI concepts are incorporated into their day-to-day operations. (from Microsoft.com, about the company-wide adoption of responsible AI)"
Reliability	The "Knowledge Limits" notion, as explained on page 4, suggests that systems can identify circumstances in which they are assigned tasks that they were not designed or permitted to carry out or in which their replies are unreliable.	"AI systems must operate consistently, safely, and reliably in expected and unexpected circumstances to build confidence.
Management practices directed towards Safety	NA	NA
Public Reports of Problems/Failure/Misses/Future Plans	NA	Within 24 hours of user interactions, Tay, an AI chatbot, changed from nice software to a hate speech platform. This emphasizes the necessity of designing AI systems with the human aspect in mind and preparing for novel attacks on learning datasets, especially for AI systems with the capacity for autonomous learning.
Oversight Boards	"The National AI Initiative Office and the President will receive advice on AI-related issues from the inaugural National Artificial Intelligence Advisory Committee (NAIAC) members, which consists of 27 experts. The first public webcast meeting of the NAIAC is set for May 4, 2022.	NA
User Control and Autonomy Framing	NA	AI systems with autonomous learning capabilities were equipped with sophisticated content filters and human supervisors in reaction to new assaults that affected learning datasets and to stop the Tay problem from happening again.
Autonomy	NA	NA

(Continued)

TABLE 4 | (Continued)

TRUST Framings	NIST	Microsoft
User Control	NA	NA
Augmentation	NA	NA
Human Understanding	Modeling issues arise from various elements influencing meaningful interactions between AI and humans. Computational and human aspects must be considered by systems that provide meaningful explanations. Additionally, explanations may need to be modified over time as users' judgments of meaningfulness shift with experience.	NA
Secure and Privacy Framing	NA	AI will undoubtedly affect decision-making, data security, privacy, and worker skills; therefore it's important to think about how to take use of its benefits while protecting privacy. Unit 3 of Identify guiding principles for responsible AI module, Section: Societal implications of AI)
Security and Safety (w.r.t data collection, processing, access, share, consent, data subject to AI decision making)	This process of identifying and recognizing knowledge boundaries protects against making decisions that might not be appropriate.	
The Other Framings	NA	NA
Ethical Dilemma and Moral Framing	NA	NA
Human Resource, Employment, Rights and Accessibility Framing	NA	NA
Fairness, Non-discrimination, and Justice Framing	" The Knowledge Limits Principle can increase trust in a system by preventing misleading, dangerous, or Unjust decisions or outputs." (43)	"Microsoft worked with a significant financial lending organization to create a risk assessment system for loan applications. When the system was audited, it turned out that even though it only authorized low-risk loans, all of the accepted borrowers were men. Before the system was implemented, this transparency allowed us to identify and address the historical prejudice among loan officers in favor of male applicants.
Accountability and AI Audits Framing	The first step in combating prejudice is for people to become aware of the limitations and repercussions of AI recommendations and forecasts. Ultimately, people must supplement AI results with sound human judgment and take ownership of important decisions affecting others.	"The first step in combating prejudice is for people to understand the limitations and ramifications of AI recommendations and forecasts. Ultimately, people must supplement AI conclusions with sound human judgment and take accountability for important decisions that affect others. Microsoft and a well-known financial lending organization worked together to develop a risk assessment system for loan applications. We used the customer's data to train a well-known industry model. We discovered a prejudice during our system audit, indicating a past predilection among loan officers whereby all authorized loans were given to male applicants. Through this analysis, we addressed the bias before system deployment.
Inclusion, Diversity, Solidarity, Protection of Cultural Differences and Whistleblowers Framings	NA	NA
AI Education, Science policy, and Public Awareness Framing	NA	"As we learn more and work with consumers, other digital businesses, researchers, civic society, and other stakeholders, we anticipate these principles will evolve and change. This module's summary and resources section will provide an overview of these concepts.
Responsible Research funding, Hidden AI Costs, Field Specific Deliberations Framing	NA	NA

needs. Still, it withholds the weight given to these framings across different industries, creating a lack of transparency in the process of deciding how certain settings will turn out. Contrastingly, the approach taken by the European Union, as described in the AI ethics document by (18), treats all framings equally. The research also emphasizes how convincing these documents are, despite not having legal force behind them, and how they add to the conversation about global AI ethics, governance, and legislation [(9), (18) #639].

The conversation emphasizes how international AI stakeholders must come together to create a single database with ethical norms and principles unique to AI. According to (13) this convergence is necessary to handle the difficulties and possible conflicts that may occur when giving particular AI principles, like fairness and priority. As it prepares the way for the creation of formal AI norms and laws for various societal scenarios, convergence in the framing of AI ethics principles is essential for building faith in the technology's transformative potential (16). This talk emphasizes the importance of framing theory in understanding how AI ethical discourse impacts our future and the necessity for convergence to protect the common good in the setting of a global digital society.

This study, which focused on pioneering organizations like the European Commission and NIST in developing AI principles and standards, was confined to AI ethics draft texts available until December 2021. However, many actors from many sectors—including enterprises, academic institutions, national and international organizations, and more—are involved in the quickly changing field of artificial intelligence and are working on reports and frameworks related to AI ethics (45). Future studies should, therefore, take into account the dynamic field of AI ethical principles and delve further into the implications of these frames at the personal level (45). It should consider the difficulties that come with putting these ideals into reality and the diversity of values that exist among various socioeconomic classes and geographic regions. The three components of this research framework—developing AI ethics principles, applying them in particular contexts, and examining their effects on individuals and society as a whole—can greatly support moral behavior and just (14).

5. Conclusion

To sum up, this research explores the quickly changing field of AI ethics standards and principles, concentrating on trailblazing organizations like NIST and the European Commission. The study's limitations, which only included draft texts accessible through December 2021, draw attention to the necessity for continued research in this rapidly developing sector. The significance of examining developing AI ethics frameworks is highlighted by the spread of

AI technology and its interactions with diverse industries and societies. Future ethical studies in AI should consider the varied values found in various social groups and geographic areas, in addition to monitoring modifications to guiding principles and guidelines and investigating their consequences at the individual level.

Furthermore, since these are only the first steps, it is crucial to address the difficulties that come up when putting AI ethics concepts into practice. The present study underscores the significance of a thorough research methodology that encompasses three fundamental domains:

- Devising ethical guidelines for AI
- Executing them in particular situations or settings
- Examining their influence on individuals and society as a whole

In an AI environment that is always evolving, such research can substantially contribute to moral behavior and the fair application of AI ethics concepts.

In the end, as AI technology continues to change society, it will be vital for everyone to work together to create, modify, and apply AI ethics principles to make sure that AI upholds ethical standards, advances justice, and respects a variety of values.

References

1. Rich E. *Artificial intelligence*. New York, NY: McGraw-Hill, Inc (1983).

2. Obermeyer Z, Mullainathan S. Dissecting racial bias in an algorithm that guides health decisions for 70 million people. *Paper presented at the Proceedings of the conference on fairness, accountability, and transparency*. New York, NY (2019).

3. Entman RM. Framing: Toward clarification of a fractured paradigm. *J Commun.* (1993) 43:51–8.

4. Nagar N. *Framing TRUST in Artificial Intelligence (AI) Ethics Communication: Analysis of AI Ethics Guiding Principles through the Lens of Framing Theory*. Rochester: Rochester Institute of Technology (2022).

5. Sivan-Sevilla I. Complementaries and contradictions: National security and privacy risks in US federal policy, 1968–2018. *Policy Internet.* (2019) 11:172–214.

6. Parviala T. *EU Entering the Era of AI: A qualitative Text analysis on the European Union's Policy on Artificial intelligence*. Brussels: European Commission (2019).

7. de Greeff J, de Boer MH, Hillerström FH, Bomhof F, Jorritsma W, Neerincx MA. The FATE System: FAir, Transparent and Explainable Decision Making. *Paper presented at the AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering*. New York, NY (2021).

8. Sarwar H, Ishaq MI, Amin A, Ahmed R. Ethical leadership, work engagement, employees' well-being, and performance: a cross-cultural comparison. *J Sustain Tour.* (2020) 28:2008–26.

9. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell.* (2019) 1:389–99.

10. Arowolo SO. Understanding framing theory. *Mass Commun Theory.* (2017) 3:4.

11. Asplund T. *Climate change frames and frame formation: An analysis of climate change communication in the Swedish agricultural sector*. London: Linköping University Electronic Press (2014).

12. Benefo EO, Tingler A, White M, Cover J, Torres L, Broussard C, et al. Ethical, legal, social, and economic (ELSE) implications of artificial intelligence at a global level: a scientometrics approach. *AI Ethics.* (2022) 2:667–82.

13. Binns R. Fairness in machine learning: Lessons from political philosophy. *Paper presented at the Conference on fairness, accountability and transparency.* New York, NY (2018).

14. Holton R, Boyd R. 'Where are the people? What are they doing? Why are they doing it?'(Mindell) Situating artificial intelligence within a socio-technical framework. *J Sociol.* (2021) 57:179–95.

15. Caliskan A. Beyond Big Data: What Can We Learn from AI Models? Invited Keynote. *Paper presented at the Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security.* New York, NY (2017).

16. Friedler SA, Scheidegger C, Venkatasubramanian S. The (im) possibility of fairness: Different value systems require different mechanisms for fair decision making. *Commun ACM.* (2021) 64:136–43.

17. Caplar N, Tacchella S, Birrer S. Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nat Astron.* (2017) 1:0141.

18. Hleg A. *Ethics guidelines for trustworthy AI. B-1049 Brussels.* Brussels: European Commission (2019).

19. Goffman E. *Frame analysis: An essay on the organization of experience.* Cambridge, MA: Harvard University Press (1974).

20. O'Neill S, Williams HT, Kurz T, Wiersma B, Boykoff M. Dominant frames in legacy and social media coverage of the IPCC Fifth Assessment Report. *Nat Clim Change.* (2015) 5:380–5.

21. Carabantes M. Black-box artificial intelligence: an epistemological and critical analysis. *AI Soc.* (2020) 35:309–17.

22. Aftab J, Sarwar H, Kiran A, Qureshi MI, Ishaq MI, Ambreen S, et al. Ethical leadership, workplace spirituality, and job satisfaction: moderating role of self-efficacy. *Int J Emerg Mark.* (2022) doi: 10.1108/IJOEM-07-2021-1121 [Epub ahead of print].

23. Chien S, Doyle R, Davies AG, Jonsson A, Lorenz R. The future of AI in space. *IEEE Intell Syst.* (2006) 21:64–9.

24. Chong D, Druckman JN. Framing theory. *Annu Rev Polit Sci.* (2007) 10:103–26.

25. D'angelo P. News framing as a multiparadigmatic research program: A response to Entman. *J Commun.* (2002) 52:870–88.

26. Scheufele DA. Framing as a theory of media effects. *J Commun.* (1999) 49:103–22.

27. Ingram K. AI and ethics: Shedding light on the black box. *Int Rev Inf Ethics.* (2020):28.

28. Xu W. Toward human-centered AI: a perspective from human-computer interaction. *Interactions.* (2019) 26:42–6.

29. Harris J, Anthis JR. The moral consideration of artificial entities: a literature review. *Sci Eng Ethics.* (2021) 27:53.

30. Hagendorff T. The ethics of AI ethics: An evaluation of guidelines. *Minds Mach.* (2020) 30:99–120.

31. Hernández D, Cano J-C, Silla F, Calafate CT, Cecilia JM. AI-enabled autonomous drones for fast climate change crisis assessment. *IEEE Internet Things J.* (2021) 9:7286–97.

32. Druckman JN. The implications of framing effects for citizen competence. *Polit Behav.* (2001) 23:225–56.

33. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform.* (2021) 113:103655.

34. Došilović FK, Brčić M, Hlupić N. Explainable artificial intelligence: A survey. *Paper presented at the 2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO).* New York, NY (2018).

35. Datzov NL. The Role of Patent (In) Eligibility in Promoting Artificial Intelligence Innovation. *UMKC L Rev.* (2023) 92:1.

36. Hugosson B, Dinh D, Esmerson G. *Why you should care: Ethical AI principles in a business setting: A study investigating the relevancy of the Ethical framework for AI in the context of the IT and telecom industry in Sweden.* Brussels: European Commission (2019).

37. Saetra HS, Coeckelbergh M, Danaher J. The AI ethicist's dilemma: fighting Big Tech by supporting Big Tech. *AI and Ethics* (2022) 2:15–27.

38. Schnack H. *Bias, noise, and interpretability in machine learning: From measurements to features Machine learning.* London: Elsevier (2020). p. 307–28.

39. Shneiderman B. Human-centered artificial intelligence: Reliable, safe & trustworthy. *Int J Hum Comput Interact.* (2020) 36:495–504.

40. Siau K, Wang W. Artificial intelligence (AI) ethics: ethics of AI and ethical AI. *J Database Manage.* (2020) 31:74–87.

41. von Eschenbach WJ. Transparency and the black box problem: Why we do not trust AI. *Philos Technol.* (2021) 34:1607–22.

42. Warner R, Sloan RH. Making artificial intelligence transparent: Fairness and the problem of proxy variables. *Crim Just Ethics.* (2021) 40:23–39.

43. Whittlestone J, Nyrup R, Alexandrova A, Cave S. The role and limits of principles in AI ethics: Towards a focus on tensions. *Paper presented at the Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society.* New York, NY (2019).

44. Pitney AM, Penrod S, Foraker M, Bhunia S. A systematic review of 2021 microsoft exchange data breach exploiting multiple vulnerabilities. *Paper presented at the 2022 7th International Conference on Smart and Sustainable Technologies (SpliTech).* New York, NY (2022).

45. Wilson N. *Understanding the Battle for AI in Warfare through the Practices of Assemblage: A Case Study of Project Maven.* Brussels: European Commission (2020).