THE STANFORD JOURNAL
AN INTERNATIONAL JOURNAL
THE STANFORD JOURNAL ISSN: 3048-8923
(AN INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH) IJIN VALUE/
IMPACT FACTOR : 8.3
Volume–2, Issues– 3 July- September: 2025

IMPORTANCE OF ETHICS, PRINCIPLES & FRAMEWORK OF ARTIFICIAL INTELLIGENCE (AI) ON COLLEGE STUDENTS

DR. SURESH KUMAR
Assistant professor
M.L.R.S. college of education,
Charkhi dadri (Haryana)
EMAIL-sureshdhakla1234@gmail.com

ABSTRACT:
Artificial Intelligence (AI) technology has brought about significant advancements in various areas surrounding India and worldwide. "This paper will focus on Studying, Raise Awareness, Identify Challenges and Opportunities, Promote International Educational Cooperation and Decision Making in Policy Makers, Industry Leaders and Educationist with the help of Ethics, Principles and Framework of development of Artificial Intelligence (AI)". AI technology will reinforcement in areas such as Facial Recognition, Medical Diagnostics, and Autonomous Vehicles (Mission Chandrayaan-2 & Tesla). Artificial Intelligence (AI) offers substantial potential to drive economic progress, foster social development, and enhance human safety and general well-being. Despite these benefits, challenges such as limited Explainability, Data Biases, Privacy Concerns, Security Risks, and Ethical Dilemmas present considerable threats to users, developers, and societies at large. As AI continues to evolve, addressing its ethical and moral principles implications becomes increasingly crucial. Although the idea of "Machine Ethics" emerged around 2006, AI ethics remains a developing discipline. This paper focuses on studying ethical and principles concerns in AI implementation, with frameworks often designed by standards organizations, technology companies, research analysts, and governmental bodies. With AI already exerting a profound societal impact(On College Students), this paper seeks to investigate whether a global consensus exists on ethical AI in education (AIED) by examining and mapping international organizations' policies and guidelines.

Keywords: *Artificial Intelligence, Social Ethics & Principles, Societies.*

A-INTRODUCTION:

The application of artificial intelligence (AI) in education has been featured as one of the most pivotal developments of the century. Despite the rapid growth of AI for education (AIED) and the surge in its demands under the COVID-19 impacts, little is known about what ethical principles should be in guiding the design, development, and deployment of ethical and trustworthy AI in education. And even if those are addressed, the depth and breadth to which contemporary ethical and regulatory frameworks are able to capture the impacts of AI evolvement remain unfolded.

The complexity and "intelligence" of this technology have led to potentially extensive ethical threats that trigger a pressing need for risk-intensive procedures to ensure the quality of delivery. Indeed, a sense of flexibility that acknowledges human values within the developing momentum of AI is vital to fostering sustainable innovations. In the wake of such demand, UNESCO launched global standards for AI ethics which were agreed and signed by its 193 member countries on November 25, 2021. The document, whilst recognizing the "profound and dynamic" influences of AI, also highlights related flourishing dangers to the cultural, social, and ecological diversity (United Nations Educational, Scientific and Cultural Organization. Recently, researchers and international organizations have specifically examined the ethics of AI in education. Despite there being some overlaps and common agreements among these ethical guidelines and reports, no previous study has systematically assessed a global consensus on ethics for AIED.

B.Ethical Principles for Artificial Intelligence:

Ethics is defined as the branch of philosophy that systematically studies human actions, determining their rightfulness or wrongfulness by examining the moral principles that government behaviour and conduct. It is a system of moral principles that influences our personal and professional decisions, helping us to determine concepts like good and evil, right and wrong, and what constitutes a good life and a just society.

There are continued calls for substantial ethical guidelines and open communications with beneficiaries: educators, students, parents, AI developers, and policymakers stated that more emphasis is necessary to enforce ethical guidelines for AI systems to better align with societal values. Safeguard measures and human oversight are required to oversee how these AI systems are designed, how they function and evolve. The knowledge of behavioural science, equipped with self-awareness and empathy at the fore, is argued to intrinsically motivate AI developers to develop more trustworthy and responsible AI.

We conducted thematic analysis on relevant ethical guidelines and reports related to AIED found from international organizations, including UNESCO Ethics AI (AHEG, 2020), UNESCO Education & AI, Beijing Consensus, OECD (Organization for Economic Co-operation and Development,), European Commission, and European Parliament Report AI Education (2021). The paper focused on identifying and developing a set of main principle themes by using inductive analysis, based on thematic analysis process. The analysis consists of initial familiarization with the ethical guidelines and reports. This involved re-reading of reports and noting down patterns, such as similar use of words, points of discussion, and definitions. This is followed by an open coding approach where the terms and definitions are meaningfully categorized, followed by labelling each category with a code. This resulted in a total of 39 codes. Next, these codes were examined and collated into patterns of broader meaning, resulting in 7 themes i.e. principles.

1. **Principle of Governance and Stewardship**
2. **Principle of Transparency and Accountability**
3. **Principle of Sustainability and Proportionality**
4. **Principle of Privacy**
5. **Principle of Security and Safety**
6. **Principle of Inclusiveness**
7. **Principle of Human-Cantered AIED**

1. Principle of Governance and Stewardship:

A recurring theme across AI policies is the issue of governance and stewardship of AIED. For example, the 2021 UNESCO Education & AI 2021 asserted the need to "set up a system wide organizational structure for policy governance and coordination". This is further acknowledged in other papers such as OECD recommendation for "Principles for responsible stewardship of trustworthy AI". AIED governance and stewardship declares and manages how AI should be employed in education and relevant mechanisms to assure the compatibility between the role of the technology being deployed and its designed purposes, to optimize educational stakeholders' needs and benefits. AI principle of governance has been formally defined as "the practice of establishing and implementing policies, procedures and standards for the proper development, use and management of the infosphere." According to OECD Principles for responsible stewardship of trustworthy AI, there are five complementary principles relevant to all stakeholders:

(i) Inclusive growth, sustainable development and well-being;
(ii) Human-centred values and fairness;
(iii) Transparency and explainability;
(iv) Robustness, security and safety; and
(v) Accountability.

While the first and second principles sought to attain the inclusiveness and human-centredness in AIED, the later three OECD principles share several common intersections with data ethics and physical safety in using AIED. Accordingly, we propose that the governance and stewardship of AIED should accomplish all ethical aspects of relevant domains.

2. Principle of Transparency and Accountability:

Data ethics emphasized the need for transparency in data usage in AIED. AI tools have been gradually applied quite extensively in education to enhance learning and teaching practices, but the challenge remains unaddressed regarding the transparency of the data generated. Cope & Kalantzis highlighted that this ethical principle is essential to teachers and students as data visualization represents learner behaviour, and accentuates additional support that educators could provide. It should be noted that the transparency lies in what the data itself is, where it is collected, what it shows, what happens to it, and how it is used (Digital Curation Centre, 2020). These questions could be answered once data ownership, accessibility, and explainability are sustained.

The concept of explainability in AI and data is closely linked to the transparency of the AI system and data generated. Indeed, data should feature the ability to explain some predictions from a technical viewpoint of a particular human. AI explainability underscores the insights in how AI systems functioned and made a decision should be well informed and explicable to the stakeholders, though the explicability relies on their technical expertise and role (Kazim & Koshiyama, 2021; UNESCO, 2019). The opaque nature of AI often poses many challenges for stakeholders to fathom the logic of this "black box" behind its decision-making. For instance, the absence of explainability could result in teachers being unable to use AIED effectively and timely detect the problems related to students' behaviour and learning performance .Therefore, this ethical concern centre's on the intelligibility of the operation and outcomes of AI educational systems.

3. Principle of Sustainability and Proportionality:

Similar to other technology advances, the development and deployment of AI should also take into account environmental concerns to the extent that is referenced (AHEG, 2020; OECD, 2021). Particularly, sustainability calls for the design, development, and use of AIED to consider optimizing energy efficiency and minimizing its ecological footprint (European Commission, 2019). Accordingly, regulations of AIED are required to create policies ensuring these considerations are accomplished throughout the processes of developing and deploying AIED. Moreover, regulation of AIED must consider other sustainable domains, including economic and societal aspects such as employability, culture, and politics. For instance, regulation of AIED should consider ensuring policies supporting accountability of potential job losses and to leverage challenges as an opportunity for innovation (UNESCO, 2019). Careful deliberations of sustainability and proportionality will make AIED more approachable and beneficial to all.

4.Principle of Privacy

Personal privacy also emerged as a critical ethical concern in the implementation of AIED. Privacy, by nature, could be defined as "the right to be left alone", which underscores the right of having personal information being protected .This digital revolution in education, particularly the use of AI and learning analytics in the field of education, entails a massive amount of personal data generated, captured, and analyzed to optimize learning experiences. The personal data of teachers and learners may run the risk of privacy breaches. For instance, in respect of agent-based personalized education, personal information of learning performance accumulated in the past could be utilized for future prediction. However, this is considered against the will of many students.

To protect and support the right of learners' privacy and social well-being while learning in the context of increasingly knowledgeable machines and computer agents, AIED developers need to assess the views of teachers and students to decide how AI should be deployed in the classroom. For instance, an ethical concern may arise from a real-time facial expression recognition system used to predict the affective state or attendance of the learners without their. Developers and educators should embed transparency and visibility to AIED-related threats while explaining potential ramifications to students' learning, careers, and social lives. The objective is to cultivate trust among learners and provide them with insights to leverage their skills across contexts while maintaining control of their respective data and digital identities.

In most cases, when AIED tools are used to engage users in a particular learning activity, users are assumed to give consent, by which they would agree on terms of use of technology and how their personal data is collected, managed, and processed. Aligned with the principle of transparency and accountability, consent must be well informed as a pragmatic approach to building trust among students since the consent demonstrates their ease with the use of data by teachers to enhance their own learning performance also highlighted the significance of this consent to show respect towards students and reinforce their autonomy and freedom of choice. Once data is garnered, questions arise about how data management works, where and how long their personal information should be stored, and to whom the rights of accessibility should be granted.

5.Principle of Security and Safety:

One of the main functions of educational learning systems is to collect data of users, from which predictions about the learning behaviors and performance of users will be made. However, it is inevitable to envisage a scenario when the data is probably manipulated or corrupted by another party, or even worse, by cybercriminals.

The concept of incorruptibility in AIED traces its root from incorruptibility in AI, or robustness against malicious manipulation by external factors. Bostrom & Yudkowsky pointed out that AI systems must be "robust against human adversaries deliberately searching for exploitable flaws in the algorithm". Therefore, it can be stated that the incorruptible nature and integrity of the data go hand in hand with data security. It is essential to protect the personal data of stakeholders, including students, teachers, and schools, to prevent any misuse or violation. The protection of data privacy and security is even more essential in the current context of normalizing virtual learning, and it requires concerted effort and the self-awareness of all the stakeholders.

Whereas learning analytics are governed by data ethics, many AIs are forms of intelligence expressed by some artefact that interacts with humans at various levels, such as robots and self-driving cars. This raises a universe of technical safety concerns regarding AI operation throughout its lifecycle in normal use, especially in harsh conditions or where other agents (both human and artificial) can interfere with the system.

As a result, it is pivotal that AIED developers take great care to design, train, pilot test, and validate the safety of AI systems. Multistakeholder Groups, Including Product Developers, Educators, and Public Authorities, Should Establish Appropriate Oversight, Assessment, and due diligence mechanisms to ensure accountability and robustness throughout the AI lifecycle. This group should produce detailed guidelines and ensure that AI users (educators and learners) receive adequate training to operate the system safely within the defined environment.

6.Principle of Inclusiveness:

Previous ethical discourse suggested that AI systems should contribute to global justice and be equally accessible to all. Accessibility is vital to allow society to gain significant benefits from these systems. The exclusion of any individual is a violation of human rights. It is, hence, paramount that accessibility entails affordability, user-friendly designs catering to individuals of different demographics, cultures, and particularly those with disabilities. As highlighted in the European Commission Report 2021, inclusion and fairness of access to AI-powered education stress the basic needs and availability for internet coverage, followed by next-generation digital infrastructure.

7.Principle of Human-Cantered AIED:

In recognition of autonomy as a modern moral and political value the development and regulation of AIED need to adopt a human-centric approach that safeguards and empowers human autonomy This principle emphasizes the importance of supporting learners in developing their own potential.

Human autonomy, **According to Deci and Ryan (2020),** refers to the capacity to live one's life according to one's own motivation that is not the result of deception or manipulation. AI assistants today serve a variety of functions, generally intending to provide and assist individuals with some recommendations. In a sense, these can be considered external factors that affect an individual's cognitive bias and emotions undermining or manipulating one's intrinsic motivation. The design and operation of AI must thus, avoid misleading information, compromising users' autonomy in developing independent thoughts, or negatively affecting users' emotions and social well-being.**C-Pedagogical appropriateness**

On its face, the adjective "pedagogical" seems straightforward. Derived from the Greek pedagogue or "the one who leads the child (to and from school)", it is now commonly employed to describe anything to do with teaching, schooling and/or instructional technique. However, the term pedagogy has a more complex lineage and usage. In

Europe, for example, pedagogy is sometimes distinguished from didactics. Didactics refers to the mechanics of teaching subject matter, whereas pedagogy encompasses a larger host of educational concerns such as curriculum, student-teacher relations, teaching and learning philosophies and values, and teachers' professional Thus the term "pedagogical" includes but also extends beyond the "what, when, why, and how to teach" As an ethic, "pedagogy involves … distinguishing actively and/or reflectively what is good or right and what is life enhancing, just, and supportive from what is not good, wrong, unjust, or damaging in the ways we act, live, and deal with children"

D-Final Remarks and Future Directions:

The education system faces a paradox of artificial intelligence. Though regarded as vital for AI generation of high-quality educational outcomes, AIED and related large-scale collection and analysis of personal data about learners are of considerable concern to human-rights advocates. This paper contributes to the discussions of the benefits of AI in education, and at the same time, raises concerns for the adverse impacts on fundamental issues surrounding human rights. The intricacy of AI necessitates a holistic and applicable set of ethical principles for AI in the educational context. By systematically analysing well-documented general AI principles, we propose a set of ethical tenets for AIED as a starting point to engage and spark further debates on the robustness of these guidelines, followed by actionable and shared policies to ensure the AIED systems developed are essentially ethical by design. The proposed set of ethical principles should be considered when developing and implementing ethical and trustworthy AI systems for education.

Overall, this paper contributes to the ongoing discussions of the benefits of AI in education, and at the same time, raises concerns for the adverse impacts on fundamental issues surrounding human rights. The intricacy of AI necessitates a holistic and applicable set of ethical principles for AI in the educational context. By systematically analyzing well-documented general AI principles, we propose a set of ethical tenets for AIED as a starting point to engage and spark further debates on the robustness of these guidelines, followed by actionable and shared policies to ensure the AIED systems developed are essentially ethical by design.

References:

- Akgun, S., & Greenhow, C. (2021). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. AI and Ethics, 1–10. https://doi.org/10.1007/s43681-021-00096-7
- Araújo, D., Couceiro, M., Seifert, L., Sarmento, H., & Davids, K. (2021). Artificial Intelligence in Sport Performance Analysis. Routledge. https://doi.org/10.4324/9781003163589

- Cope, B., & Kalantzis, M. (2019). Education 2.0: Artificial intelligence and the end of the test. Beijing International Review of Education, 1, 528–543. https://doi.org/10.1163/25902539-00102009
- Dhanrajani,S. 2018 3 Ways To Human Centric AI.https://www.forbes.com/sites/cognitiveworld/2018/12/12/3-ways-to-human-centric-ai/?sh=495e42804a38
- Müller, V. C. (2020). Ethics of Artificial Intelligence and Robotics. https://plato.stanford.edu/entries/ethics-ai/
- Reiss, M. J. (2021). The use of AI in education: Practicalities and ethical considerations. London Review of Education, 19(1), 5, 1–14. https://doi.org/10.14324/LRE.19.1.05