Components of an Ethical Framework for Artificial Intelligence in Education January, 2025

Pierrette Renée Dagg, Ph.D.
Director of Research, Merit Network

Ethics Components

1. AI systems and policies should prioritize people, the environment, human rights, fundamental freedoms, and human interactions over corporations, profit, productivity goals, and human-machine interactions throughout the life cycle.
2. AI should be applied with a humanistic approach that focuses on an individual's worth and the value of human beings.
3. Peace, inclusivity, justice, interconnectedness, democracy, equity, and freedom of expression should be promoted throughout the AI life cycle. AI systems should not undermine freedom, safety, or turn individuals against each other.
4. Recognize that AI technologies do not ensure human flourishing.
5. AI systems should be applied in ways that protect and empower users.
6. No person or community should be subordinated, objectified, or undermined by the application of AI technology.
7. The transparency and explainability of AI systems are often essential preconditions to ensure the respect, protection, and promotion of human rights, fundamental freedoms, and ethical principles.
8. Develop mechanisms to instill the understanding of students' ethical use of AI tools in the learning process (e.g. ChatGPT).
9. Ensure that AI systems and policy support the tenets of Stoic philosophy and uphold the cardinal virtues of prudence, courage, temperance and justice. It is necessary that AIEd applications provide maximal benefit for all stakeholders at all times, maintain harmony, and are in accordance with nature.

Accountability Components

10. Remember that humans are ultimately responsible for educational outcomes, an AI system is not a replacement. Oversight of AI systems should come from humans, a governance group, and the public, as appropriate. Particular attention should be given to self-learning and autonomous AI systems.
11. Facilitate open debates on issues related to AI ethics, data privacy and security, and concerns about AI's negative impact on human rights and gender equality and related accountability metrics.
12. Develop accountability assessment measures and requirements of AI systems and tools before implementation.

13. Discuss within your governance group if additional accountability considerations for vulnerable or marginalized populations within your institution should be accounted for.
14. Continually monitor the achievement of educational objectives and intended impacts for AI systems. When goals are not met, or the use of AI systems result in harm, investigate potential causes, such as its implementation or design, and create a remediation plan.
15. Create independent oversight mechanisms, including whistleblowers' protections, for data governance, AI governance, system accuracy, and ethics to foster the benefits of AI and keep data and people safe. Ensure participation by those trained in AI ethics. Seek diverse, inter-sectoral, interdisciplinary and multistakeholder expertise.
16. Create enabling environments for stakeholders and the public to understand any benefits and negative consequences of AI's use in your institution.
17. Assess the broader societal impact of the AI system's use beyond the individual, including those indirectly affected by AIEd.
18. Assess and monitor the reproducibility or variability of decisions made by AI systems.

Transparency Components

19. The principle of transparency should be upheld whenever possible. Transparency contributes to democracy, inclusivity and prevents harms to human rights.
20. The principle of transparency should be followed as much as possible within institutional data applications, however there may need to be a balance between transparency, explainability, and other principles like privacy, safety, and security.
21. Data sets should be documented to the best possible standard to allow for traceability, auditability, and an increase in transparency while adhering to data privacy law. Consider the promotion of open data sets, when possible, to support the safe, fair and ethical sharing of data.
22. When the use of AI could be considered a surveillance of learners, discuss the justification of the direct or indirect benefits of this practice within your AI governance group.
23. Users and their caregivers should easily understand when they are interacting with AI. The AI system, even when imitating human behavior, should clearly signal that its behavior is simulated and that the system has no emotional capacities or feelings. Ensure this explanation is written in age-appropriate language.
24. Consider within your AI governance group whether the AI system should communicate to users that a decision, content delivery choice, advice, or outcome is the result of an algorithmic decision. When communication about such will not damage well-being, end users should be informed on the reasons and criteria behind AI system's outcomes. In cases such as those involving students with learning disabilities or for behavioral predictions such as dropout remediation may be more beneficial to well-being to remain obfuscated.
25. Set clear requirements for transparency.

Well-being Components

26. Promote continual learning and discovery on evidence-based insights into how AI can be used to support well-being and emotional development.
27. If AI assessment models are continuous, designate safe spaces where students are not assessed.
28. Seek evidence-based research and strategies to monitor and protect the physical and mental health of users in the context of AI systems.
29. Consider the application of AI technologies to monitor and improve the health and well-being of students.
30. Emphasize students' agency and social well-being in the process of integrating AI-based tools: Protect students' agency and motivation to grow as individuals; protect play and leisure time, social interaction, and school breaks.
31. Test and scale up evidence-based avenues of applying AI in learning; this includes fostering broad, transferable abilities including social-emotional skills, meta-cognition, collaboration, problem-solving, and creativity.
32. Monitor AI systems' impact on social relationships, social skills, and attachments.
33. Ensure that AI systems foster and support students' sense of purpose.
34. Develop guidelines and requirements for human-robot interactions and their impact on human-human relationships, based on research, with special attention to the mental and physical health of human beings.
35. Environmental protection should be promoted and examined throughout AI systems' lifecycles, including resource usage, energy consumption and supply chain impact.
36. When the risk of job loss or de-skilling is present, take action to counteract potential consequences.
37. Screen for technophobia in students and educators. When technophobia is present, enlist supportive resources and remediation strategies.

Autonomy Components

38. Establish whether AI's use could undermine or marginalize the authority of practitioners or educators, or disrupt accountability structures and take action based on the assessment.
39. Promote data agency whenever possible. Users should retain control of their own data and digital identities. When feasible within an institution, users should maintain the ability to access and erase their personal data in AI systems (while maintaining compliance with international law). Proactively monitor and evaluate data agency policies with your AI governance group.
40. Privacy is essential to agency and autonomy and should be protected throughout the AI lifecycle.
41. Lifestyle choices, beliefs, personal expression, and individual experiences should not be restricted during any phase of the AI lifecycle.

42. Ensure the AI system does not affect human autonomy by interfering with the user's decision-making process in an unintended way. Nudging behavior and choice architecting mechanisms should be weighed against potential harm or loss of agency.
43. AI systems should be used to increase the level of control that learners have over their learning and development, including the decision to opt out of their use. Alternative opportunities for students to explore topics that fall outside of the pathways to prescribed content in personalized learning applications should be provided.

Equity and Inclusion Components

44. Develop definitions for fairness, inclusion, and equity within your governance group.
45. Establish and monitor measurable targets to ensure inclusion, diversity, and equality in AI decision-making, teaching practices, systems, and data sets.
46. Encourage and facilitate the participation of individuals with diverse backgrounds and beliefs, marginalized people, girls and women, and students of all ages in AI leadership, governance, and decision-making roles within your organization.
47. Build pathways for diverse students to enter the AIEd field.
48. To the extent possible, investigate and work to mitigate experiences of the digital divide and disparities in technology device availability, home internet connectivity, and digital skills amongst learners.
49. Address the cultural impact of all AI systems within your institution, which could include leveraging AI to preserve culture, or examining the impacts of translation and voice assistants on language and expression.

Pedagogy and Teacher Training Components

50. Independently verify vendors' claims about AI's potentials based on pedagogical research, and ensure that student performance measures align with recognized evidence-based measurements.
51. Have clear educational goals when AI is used in curriculum or assessment.
52. Seek and use AI that supports diverse educational approaches and pathways for lifelong learning.
53. Use AI to assess and recognize a broad range of learners' talents and competencies.
54. Develop education programs that support the technical and soft skills needed to be successful in future AI-driven environments in addition to the acquisition of “prerequisite skills” for AI education, such as basic literacy, numeracy, coding and digital skills, and media and information literacy.
55. Identify and develop AI literacy curriculum for students.
56. Review and adjust curricula to reflect necessary changes in pedagogy and assessment as a result of AI's development and adoption within and external to educational environments (eg: ChatGPT).

57. Analyze potential changes in teachers' roles in facilitating knowledge transfer, higher order thinking, and human values.
58. Analyze potential changes and best practices concerning the teaching and learning of human values, such as love, tolerance, and respect.
59. Consider a self-assessment for teachers to assess and develop their AI awareness and skills.
60. Develop the skills needed by educators to select, use, and critically evaluate AI tools in their pedagogical practices and professional development.

Information Technology Risk Management Components

61. Understand that AI systems may pose negative impact, including those which may be difficult to anticipate and measure. Adopt adequate measures to mitigate these risks when appropriate.
62. Periodically assess compliance with laws to ensure the use of pupil data for AI purposes is permitted.
63. Consider necessary balances between privacy and the use of data to achieve organizational and educational goals.
64. Require ongoing testing of AI systems for safety and security, including the consideration of vulnerabilities such as data pollution, physical infrastructure, and cyber-attacks. Develop organizational policies for risk prevention and mitigation.
65. Leverage AI systems to promote user safety, when applicable.
66. Develop a disaster recovery plan in the event of AI system breach or failure.
67. Consider an insurance policy for AI systems.
68. Implement a strong AI information technology training program to build stakeholders' understanding of safety and security protocols.
69. Ensure AI systems are properly constrained to the places they can write data to; dynamic decisions made by AI can impact systems related or connected to AI tools.

Administration Considerations Components

70. Set up a structure for policy governance and coordination of AIEd.
71. Promote, encourage, and participate in AI in education research initiatives to enhance opportunities and mitigate challenges.
72. Support knowledge transfer and capacity building on AI and its use in your educational environment for top management, administration, and decisionmakers.
73. Encourage investment and funding for AI technologies in your institution or district.
74. Create and implement technology adoption and change management strategies for all AI systems and policies.
75. Develop institutional guidelines for human-robot interactions and their impacts on human-human relationships, in consideration of the mental and physical health of users.

76. Mechanisms and protections for stakeholders to raise concerns about the application of AI systems should be developed.
77. Identify, monitor, and evaluate the ways AI can increase organizational capacity and improve processes and educational information management while respecting human relationships.
78. Consider the values of current practices before automating them with AI.
79. Identify staff roles at risk for displacement by AI and develop job transition plans.

Supplier Scrutiny Components

80. Insist that suppliers explain how their AI resource achieves the desired objectives and impacts, including information relating to algorithmic decision-making.
81. Require that suppliers make explicit whether there were any trade-offs between accuracy and explainability in the design of the AI application or system.
82. Require confirmation that AI resources were not designed to coerce learners. Insist that suppliers demonstrate that when AI is used for behavioral influence, this influence is positive and supported by scientific evidence.
83. Require confirmation that AI systems were not developed to encourage addictive behaviors.
84. Confirm that appropriate measures were taken to mitigate against biases. This includes the application of diverse data sets.
85. Suppliers must demonstrate that systems and tools have been designed and proven to be effective for users with additional cognitive or physical needs.
86. Evaluate with suppliers that solutions are relevant to local needs in terms of educational best practices, diversity, context, cultures, and objectives.
87. Investigate the auditability of AI tools provided by suppliers, which includes the assessment of algorithms, data, and design processes.
88. Insist that suppliers conduct periodic internal investigation to ensure products are achieving outcome claims and not introducing harmful or unintended consequences.
89. Data ownership should be actively examined. It is critical to understand which entities own the data that is generated by students as a result of interacting with AI systems, and the rights retained by the data owners. This includes knowledge regarding the owner's rights to sell data in the future, how that data will be used to make decisions about students outside of an educational setting, and the ways in which user data could inform suppliers' product development and enhancement.
90. Consider the institutional development of a supplier evaluation that includes aspects of security, transparency, and ethics.