# AI Human Rights Literacy

**Saah Agyemang-Badu**
*University of Connecticut*

**Felisa Tibbitts**
*Department of Law, Economics and Governance, Utrecht University, Utrecht, The Netherlands and Visiting Professor, Nelson Mandela University*

**Sage Phillips**
*University of Connecticut*

*The treatment of artificial intelligence (AI) in the field of education has so far been typically characterized by (a) information about how AI can assist educators in carrying out their work, and (b) concerns about the misuse of AI by learners, for example, concerning plagiarism. The links between AI and ethics within the field of education are much more complex. Beyond the concerns about the organization of teaching and learning with the rise of AI—and the associated rights to privacy and safety—there are legitimate needs for instructors and learners to understand how AI affects their daily lives. What are the wider ethical considerations for using AI, particularly from the perspective of human rights norms? This paper critically analyzes some of the human rights at stake regarding the use of AI and its implications for the organization and content of formal education (K-12 and higher education). The human rights perspective on AI's dynamic and changing field—AI human rights literacy—is critical to convey to instructors and learners as they navigate these new technological developments. This paper overviews human rights relevant to everyday encounters with AI in education. It proposes an AI Human Rights curriculum to help both learners and educators become critically aware of these human rights implications.*

*Keywords: AI, human rights, human rights education, AI literacy.*

## Introduction

The rapid advancement of artificial intelligence (AI) technology presents some of the most profound opportunities and challenges for humanity in the 21st century. AI systems are being integrated across nearly every domain of society, from healthcare and transportation to finance, media, and education. While advocates are enthusiastic about AI's potential to enhance efficiency, automation, and data-driven policy decision-making, a rising chorus of voices has begun to sound the alarm about the risks of AI posing threats to human rights (Zuboff, 2019), and the endurance of democratic values such as non-discrimination and equality (Noble, 2018). AI is not neutral; it operates within frameworks shaped by the biases and inequalities embedded in the data it processes and the systems it supports (Noble, 2018; Eubanks, 2018). These dynamics

make it important to consider AI's impact on human rights worldwide. This is particularly relevant in the educational sector given how education has emerged as a high-stakes arena for upholding human rights.

Human rights, broadly defined, are the fundamental rights inherent to all individuals, regardless of nationality, ethnicity, gender, or socioeconomic status. These rights are codified in international legal frameworks such as the Universal Declaration of Human Rights (UDHR) (1948) and the Convention on the Rights of the Child (CRC) (1989). Within these frameworks, the right to education (Article 26 of the UDHR) is considered a cornerstone for the development of individuals and communities. Human rights education (HRE) is a field of practice specifically focused on fostering knowledge, skills, and values to promote dignity, equality, justice, and democratic engagement (Tibbitts, 2002). As AI becomes increasingly integrated into educational systems, human rights can offer a critical lens for assessing the opportunities and risks of using AI in education. HRE can provide a guiding framework to prepare educators and learners to maximize AI's benefits while addressing its risks (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2021).

Two areas of focus have typically characterized the treatment of AI in education: (a) information about how AI can assist educators in carrying out their work, and (b) concerns about learners' misuse of AI, for example, concerning plagiarism. The links between AI and ethics within the field of education are much more complex. In addition to concerns about the organization of teaching and learning, and the associated rights to privacy and safety, there are legitimate needs for instructors and learners to understand how AI affects their daily lives inside and outside educational settings. In this paper, we examine the intersection of AI and HRE, addressing two interrelated questions: *What human rights are at stake with AI integration in education?* And *in what ways can HRE support educators and learners to engage critically with AI?* We argue that addressing these challenges is a moral imperative for educators and policymakers, ensuring that the transformative potential of AI benefits all learners equally.

We undertook a literature review to identify key debates about AI and human rights implications within the field of education. This literature review involved research using the online database EBSCO, open-access scholarship, and grey literature on Google Scholar. Based on these results, we explore key human rights at stake when using AI in formal education: privacy, equity, and the right to education. Following this literature review, we propose key AI human rights literacy aims for educators and learners that can be contextualized to a wide range of educational settings. The explicit link between AI and HRE in our paper advances aims to foreground the moral obligations when using AI in learning environments. The final section of this article proposes an AI human rights literacy curriculum designed to equip educators and learners with tools to navigate these challenges.

**Definitions of AI**

AI scientist John McCarthy coined and defined the term 'Artificial Intelligence' in 1956 as "the science and engineering of making intelligent machines" (McCarthy et al., 1955). This definition encapsulates the dual aspects of AI: the scientific pursuit of understanding intelligence and the engineering challenge of creating machines that exhibit such intelligence. AI can be categorized into different types based on the capabilities and functionalities exhibited by systems. The earliest approaches in the 1950s and 1960s focused on general AI attempting to demonstrate human cognition but faced challenges scaling to handle complex real-world tasks. This evolved into specialized *Narrow AI* typically focused on excelling at a single application area like translating languages (Hernández-Orallo & Dowe, 2010). Narrow AI powers many modern technologies by automatically extracting insights from patterns in data using machine learning without explicitly programmed rules. Meanwhile, advanced systems are beginning to incorporate different narrow AI capabilities like computer vision, speech recognition, and natural language processing in an integrated manner, demonstrating some properties of general intelligence across multiple domains. Narrow AI or Weak AI focuses on performing specific tasks, like recognizing your voice when you chat with Siri. This type of AI has a limited context, meaning it doesn't have a broader understanding or awareness. Currently, Narrow AI is the most advanced form of AI we have.

*Artificial General Intelligence (AGI),* sometimes called *Strong AI*, refers to AI that can perform any intellectual task that a human can. It can understand, learn, adapt, and apply knowledge in a broad range of tasks, much like a human being. Current systems still rely heavily on training data rather than conceptual knowledge and remain brittle outside fixed contexts. However, innovative approaches keep expanding the boundaries of how algorithms can efficiently learn complex functions, plan using internal models, communicate via language, and transfer expertise between related tasks – bringing science closer to realizing Artificial General Intelligence (Goertzel & Pennachin, 2007). Theories about AGI posit that future systems will self-improve and design ever smarter generations of AI potentially leading to superintelligence surpassing humans across all cognitive faculties. This is *Superintelligent AI*, which goes beyond AGI and is, theoretically, smarter than humans in all aspects (Bostrom, 2014).

The AI revolution has likely only begun scratching the surface of AI's transformative impact. More sophisticated systems integrating multiple AI capabilities like computer vision, speech recognition and natural language processing in general AI architecture could one day match or exceed human-level intelligence across diverse domains. Despite skepticism around the possibility of self-improving superintelligent AI or technological singularity, the accelerating pace of AI breakthroughs hints at paradigm shifts in short order (Marcus & Davis, 2020). The integration of AI into education often focuses on narrow AI technologies, such as personalized learning systems or automated grading tools. These systems have the potential to improve educational access and

outcomes by offering tailored learning experiences and real-time feedback. However, this increased reliance on AI raises fundamental questions about ethics and governance in education. The educational sector must grapple with how to leverage AI's capabilities while safeguarding learners' rights and ensuring equitable outcomes.

**Human Rights at Stake with AI**

Human rights frameworks, such as the UDHR, establish education as a universal right, emphasizing principles of equitable access, privacy, and dignity. The integration of AI into education presents significant opportunities to advance these rights but simultaneously raises profound ethical concerns. This section examines three critical rights that are at stake in the age of AI, with a specific focus on implications for education settings: the right to privacy; freedom from bias, discrimination and right to equality; and the right to education. Berendt et al. (2020) underscore the need to highlight essential human rights as a starting point for implementing AI in educational systems. These rights-related concerns are critical for evaluating the use of AI in education and serve as a foundation for designing curriculum content to foster AI Human Rights Literacy. The final section of this paper will expand on this potential curriculum, offering insights into how AI can be approached through a human rights lens, empowering students to understand and navigate its implications.

*The Right to Privacy*

The integration of AI into education offers the promise of enhanced learning personalization and operational efficiency. However, it also raises significant concerns about the fundamental human right to privacy. Machine learning systems routinely collect vast quantities of data, including students' academic performance, behavioral analytics, and sensitive personal or biometric information. While often justified as necessary for improving educational outcomes, this data collection poses critical risks to students' autonomy and dignity, particularly given the opacity of these technologies (Park & Humphry, 2019). Without robust safeguards and democratic oversight, these practices can lead to significant privacy violations, undermining trust in educational institutions.

These AI systems collect and analyze user data through diverse methods such as online activities, purchases, location tracking, and surveillance technologies. Within educational contexts, learning management systems (LMS) and adaptive learning platforms use similar methods to optimize learning pathways, gathering granular details about student interactions (Mayer-Schönberger & Cukier, 2014). While these tools enhance personalization, the lack of transparency regarding how data is used or who has access to it raises ethical concerns. Students, parents, and educators are often unaware of the extent to which these systems collect, store, and utilize data, which limits their ability to make informed decisions about participation. Studies underscore the limited awareness students have of data privacy risks. For example, research by Selwyn (2019)

found that over 70% of students using AI-based learning platforms were unaware of the extent of data collected about them. This lack of awareness undermines their ability to give informed consent and leaves them more susceptible to exploitation by opaque systems.

One of the most concerning aspects of AI in education is its potential to exacerbate existing power imbalances. Students in marginalized communities are particularly vulnerable to privacy violations. Many of these students lack the digital literacy or institutional support needed to understand their data rights, leaving them disproportionately exposed to potential misuse (Slade & Prinsloo, 2013). Schools serving under-resourced populations often rely on free or low-cost AI tools with inadequate privacy protections or unclear data ownership policies. This dynamic perpetuates systemic inequities, disproportionately exposing vulnerable communities to exploitation.

Children, as a demographic, are uniquely at risk. Their limited capacity for informed decision-making about privacy and data sharing makes them particularly vulnerable. The normalization of data collection in educational settings erodes their expectation of privacy from an early age. Scholars like Zuboff (2019) caution against the unchecked expansion of surveillance systems, which shift societal norms toward control and monitoring rather than empowerment. In education, this shift not only jeopardizes individual rights but also undermines the trust necessary for effective learning environments.

The lack of regulatory clarity further compounds these issues. While frameworks like the European Union's General Data Protection Regulation (GDPR) establish principles for data consent, minimization, and transparency, many of these systems in education fail to adhere to these standards consistently. Human rights advocates argue that the rapid pace of technological change necessitates updated regulations emphasizing both data protection and user empowerment (European Council, 2023). Current practices, which often prioritize optimization over transparency, fall short of these ideals.

To address these challenges, educational institutions and policymakers must ensure that privacy is a core principle in the design and implementation of these AI systems. Data collection should be minimized and strictly limited to what is necessary for educational purposes. Anonymization techniques must be rigorously applied to ensure data cannot be traced back to individual students (UNESCO, 2021). Systems must operate with full transparency, explaining how data is collected, stored, and used. Students and their families should have the right to opt out of invasive practices without forfeiting access to educational benefits. These measures are crucial for fostering trust and ensuring AI in education empowers learners rather than exploiting them.

Safeguarding the right to privacy in educational AI demands a paradigm shift prioritizing student empowerment over technological efficiency. This shift requires

rigorous oversight, robust regulatory frameworks, and an ethical commitment from all stakeholders to uphold the dignity and autonomy of students. Without such measures, the promise of AI in education risks being overshadowed by its potential to normalize surveillance, undermining the very rights it should protect.

### *Freedom from Bias and Discrimination and the Right to Equality*
The integration of AI into educational systems raises critical ethical and practical concerns regarding bias, discrimination, and systemic inequality. AI, often perceived as neutral and objective, operates within frameworks shaped by historical data and algorithmic design, frequently amplifying pre-existing societal inequities. This section explores the structural roots of bias in AI, its manifestations within educational contexts, and its broader implications for the right to equality as enshrined in international human rights frameworks.

Bias in AI systems stems from their dependence on historical data and algorithmic processes, both of which often reflect entrenched social inequities. Data used to train AI systems frequently encode biases present in society, leading to discriminatory outcomes. For instance, hiring algorithms trained on data from male-dominated industries have consistently favored male candidates, perpetuating gender disparities (Dastin, 2018). In educational contexts, similar biases emerge in automated grading systems and predictive analytics, which disproportionately penalize students whose cultural or learning styles deviate from majority norms.

A particularly concerning example of bias in AI is its racial dimension. Facial recognition technologies deployed in schools, for instance, have been shown to misidentify individuals with darker skin tones at significantly higher rates than those with lighter skin tones (Buolamwini & Gebru, 2018). This misidentification not only undermines the reliability of such technologies but also exposes students of color to heightened surveillance and punitive measures, mirroring broader societal patterns of over-policing and systemic marginalization.

Algorithmic discrimination is further evident in predictive analytics, which often relies on socio-economic markers associated with lower academic performance without accounting for structural inequities. Tools designed to identify "at-risk" students can inadvertently reinforce stereotypes and limit opportunities for marginalized populations. Eubanks (2018) highlights how such systems, while intended to support interventions, often perpetuate cycles of disadvantage by lowering expectations for certain student groups. This underscores the pressing need to integrate ethical training and bias mitigation strategies into AI development and deployment.

The principle of equality, as articulated in Article 1 of the UDHR, asserts that "all human beings are born free and equal in dignity and rights." However, the deployment of AI in education risks undermining this cornerstone of human rights by entrenching systemic inequities. The reliance on historical data within AI systems often results in the

replication and amplification of social biases, particularly in areas such as student placement, scholarship allocation, and grading systems. These biases have profound implications for the educational experiences of marginalized communities. Automated grading systems, for example, tend to prioritize surface-level correctness, often penalizing the creativity and critical thinking skills exhibited by students from diverse cultural or linguistic backgrounds (Benjamin, 2019). Similarly, predictive analytics used to monitor student performance often classify students from underprivileged backgrounds in ways that reinforce negative stereotypes, ultimately limiting their opportunities for growth and advancement. This cycle of bias not only perpetuates existing inequalities but also fundamentally undermines the goal of equitable education.

The digital divide constitutes a significant barrier to the equitable deployment of AI in education. Access to AI-driven tools is predicated on reliable internet connectivity, technological infrastructure, and digital literacy, resources that are unevenly distributed across socio-economic lines. Rural and underfunded schools often lack the means to adopt advanced AI technologies, leaving their students at a significant disadvantage compared to their peers in wealthier institutions. A study by the World Economic Forum (2020) found that students in low-income households were twice as likely to lack basic digital literacy skills compared to those from higher-income households. This disparity exacerbates educational inequities, limiting the ability of underserved communities to benefit from the transformative potential of AI.

The intersection of AI and equality is particularly concerning for students with disabilities or those from minority linguistic and cultural groups. AI systems, such as speech recognition tools, frequently perform poorly with diverse speech patterns or accents, alienating students whose needs are not represented in the training data (Noble, 2018). Similarly, adaptive learning platforms often operate on assumptions about 'typical' learning trajectories, further marginalizing students and requiring alternative approaches to education.

Tackling bias and promoting equality in educational AI systems requires deliberate, multifaceted interventions. The development of AI systems must prioritize the use of diverse and representative datasets that capture the experiences of all learners, particularly those from historically marginalized groups. Scholars such as Buolamwini and Gebru (2018) advocate for inclusive data collection practices as a foundational step toward equitable AI design. Regular audits and fairness assessments are critical to identifying and mitigating biases before they result in harm. These processes should be integrated throughout the lifecycle of AI systems to ensure accountability and transparency. Participatory policymaking is equally essential, involving educators, students, and community stakeholders in decisions about AI deployment to ensure that the voices of those most affected are centered.

Investments in digital infrastructure and literacy programs must be prioritized to bridge the digital divide. Policymakers and educational institutions should provide

underfunded schools with the resources necessary to adopt AI technologies while equipping students with the skills to navigate these tools responsibly. Finally, transparency mechanisms must be embedded in AI systems to explain decision-making processes and establish avenues for redress in cases of unfair outcomes. Ethical training programs for technologists and educators can further ensure that the deployment of AI aligns with principles of equity and human dignity.

### *Right to Education*

The right to education is recognized as a fundamental human right essential for empowering and developing individuals and societies. This right emphasizes not only access but also the quality and inclusivity of education, aiming to foster the full development of human potential and promote social equality UNESCO (2021). The introduction of AI into educational systems presents both opportunities to enhance this right and significant challenges that could undermine it if left unaddressed.

AI technologies have demonstrated considerable potential to enhance education. Personalized learning platforms, for example, adapt content and pacing to individual students' needs, enabling more tailored and effective educational experiences. Intelligent tutoring systems can provide real-time feedback, helping students overcome specific challenges and facilitating more efficient learning processes (Luckin et al., 2016; Kulik & Fletcher, 2016). Such innovations hold promise to address some of the longstanding gaps in education, particularly for students requiring additional support. However, these advancements are unevenly distributed, often exacerbating existing disparities and raising critical questions about the universal realization of the right to education.

Over-reliance on AI in education can narrow the scope of learning to metrics that are easily quantifiable, potentially sidelining broader educational goals such as critical thinking, creativity, and social-emotional development. Automated grading systems, for instance, often prioritize surface-level correctness over more nuanced skills like problem-solving or collaborative reasoning (Luckin et al., 2016). This emphasis risks reducing education to a transactional process, undermining its role as a transformative force that nurtures holistic personal and intellectual growth.

The integration of AI also presents risks to teacher roles and autonomy. While AI can serve as a valuable supplementary tool, excessive reliance on it may erode the essential human elements of education (Selwyn, 2019). Teachers play a critical role in mentoring, fostering empathy, and cultivating relationships that contribute to students' emotional and social well-being. AI, no matter how advanced, cannot replicate these human interactions. Instead, they must be designed to support educators, amplifying their capacity to address diverse learning needs without replacing their unique contributions.

The right to education is fundamentally tied to the principle of inclusivity. AI systems often fail to adequately accommodate the needs of students with disabilities, those from linguistic minorities, or those with non-standard learning trajectories (Kleinberg et al.,

2020). For example, many adaptive learning platforms rely on language models that do not account for regional dialects or non-native speakers, alienating students who may already face barriers to learning. Inclusive design, informed by diverse user inputs, is therefore essential.

To align AI-driven education with the right to education, deliberate policies and practices are required. Policymakers must prioritize investments in digital infrastructure to bridge the technological divide, ensuring that all students—regardless of their socioeconomic background—have access to the tools necessary for effective learning. Equally important is the need for teacher training programs that equip educators with the skills to integrate AI into their teaching while preserving the human-centered values of education. AI should be viewed as a complement to, rather than a replacement for traditional educational practices (Heffernan & Heffernan, 2014).

Governance frameworks must place equity and inclusivity at the center of AI adoption in education. This includes regular audits of AI systems to identify and address biases, participatory policymaking that involves marginalized communities in decision-making, and clear accountability mechanisms to ensure that these systems align with the broader goals of education (UNESCO, 2022). Without such measures, the promise of AI in education will remain unrealized for many, leaving the right to education as an unfulfilled aspiration rather than a lived reality.

**Human Rights Education and Critical Engagement with AI**
After acknowledging some key human rights at stake with the use of artificial intelligence in education, we now turn our focus on how educators and learners can critically engage with AI through HRE. Children, youth, and adults all need to have their human rights protected as AI becomes integrated into education systems. In this section, we first examine what educators and administrators need to know about AI from a human rights perspective in terms of the use of AI in schools and educational systems. Second, we explore what learners need to know about AI from a human rights perspective in terms of their functioning in wider society. We then overview the implications for curriculum development and training of teachers and administrators.

*What do administrators and educators need to know about AI use in the classroom from a human rights perspective?*
To ensure that education is protected, upheld, and promoted through a human rights lens when incorporating AI into school curricula and culture, educators and administrators must be vigilant about potential human rights violations. Earlier in this paper we mentioned ways that teachers are using AI in the classroom. Administrators are also using AI in their roles, such as in educational monitoring, where AI replaces the manual inputting and generation of information and data. Administrators may also

interact with AI to predict funding needed in their schools. However, school leaders must not lose sight of the risks that come with these new technologies.

Scholars are concerned that the biggest winners of AI in education are technology companies who collect data from AI educational initiatives (Berendt et al, 2020). Administrators require critical training to understand the essential ways that AI can be used in the classroom and school, and how to mitigate any threats to children's human rights. Administrators should support educators in their classrooms through education and training on the responsible use of AI technologies, as well as access to the AI technologies themselves. Educators need to understand the AI applications and software they are using and be critically aware of the strengths and weaknesses of AI in learning, to be "empowered – not overpowered – by technology" (Holmes et al., 2022, p. 11).

Many areas within AI need to be researched and addressed concerning K-12 education, such as machine learning integration, teaching assistant systems, as well as computational processes. However, the main human rights concerns are privacy, data protection, and collection surrounding AI to ensure that there is an effort to address human rights at stake with AI in the classroom.

One of the key worries about educators using AI is the use of predictive systems. Here, the generation of data used to monitor student performance and generate tasks to benefit their learning on behalf of the teacher is effective insofar as the data being used has not been produced with an influence of biases (Berendt et. al, 2020). With the increased reliance on machine learning to generate predictive data, the direct monitoring of student progress by the teacher in the classroom is greatly reduced, with potentially less oversight by the teacher on predictive data that has been influenced by bias. Another potential negative side effect is that teachers will have less personal interaction with students.

Educators must be aware of AI technologies such as natural language processing (NLP) that can produce "auto-journalism" (Miao et. al, 2021, p. 9) and can prove to be a risk to their critical thinking development. We have seen NLP perform with ChatGPT, producing paragraphs or whole essays that students attempt to use as a final project. There is another layer we must analyze with not only NLP's potential to complete a student's task(s) but also the chance that with such a resource, students may inevitably hinder their ability to critically think and develop solutions to problems on their own without any awareness of their doing so.

### *What do learners need to know about AI's presence and be responsible users of AI from a human rights perspective?*

Returning to the risks at stake with the right to education, learners need to be made aware of how AI collects their data for algorithm and academic purposes and how it can affect their learning. Students should know that their right to privacy is at stake if AI tools are collecting their data. Berendt et. al (2020) discuss "the ability to opt-out" of data

collection and state that students who are facing obligatory participation may face "a reduction of fundamental human rights, including the right to exercise autonomy and to make choices" (p. 5). However, opting out may influence the overall quality of a data set, through the underrepresentation of learners coming from groups who have opted out.

One way forward lies in the hands of policymakers to ensure this can happen through students taking action to protect their rights, as noted by UNESCO, "facilitate the participation of student representatives in countrywide initiatives that promote new competencies in the curriculum" (Miao et. al, 2021, p. 34). Learners should be able to choose when and where their data and records of achievement are utilized, but those with less privilege might have less access to support on how to make these decisions in ways that help them in the future (Berendt et al, 2020).

As AI continues to permeate educational environments, formulating and implementing proactive policies that uphold principles of diversity, interpretability, and universal inclusion is imperative to prevent the inadvertent instrumentalization of inequality. Human rights frameworks emphasizing participation, transparency, and non-discrimination should serve as guiding principles to ensure that educational AI empowers diverse learners equitably (UNESCO, 2022).

**Overview of AI and Human Rights Literacy Curriculum**
The human rights perspective on the dynamic and changing field of AI, AI human rights literacy, is critical to convey to both educators and learners as they navigate these new technological developments in their everyday lives. UNESCO's *AI and Education: Guidance for Policymakers* reports "future learning and training systems must equip all people with core AI competencies, including understanding of how AI collects and can manipulate data, and skills to ensure safety and protection of personal data" (Miao et. al, 2021, p.1). Arguably, before students are allowed to utilize AI tools in the classroom, they must be made aware of this phenomenon and understand that if granted access to AI alongside learning, it cannot be trusted one hundred percent. That is, AI is generated through data input, and users would be remiss to assume that that data is completely trustworthy and accurate. AI has much to offer the next generation of learners, however, the responsibility to identify the risks does not fall solely on policymakers. Educators and students must engage in dialogue about the consequences of fully entrusting AI technologies before assuming that they can solely rely on them.

Empirical evidence underscores the importance of integrating AI ethics into education. A study by UNESCO (2021) found that students who participated in AI ethics workshops demonstrated a 25% increase in their ability to identify biases in algorithmic outputs and improved digital literacy by 30%. Similarly, programs at Stanford University focusing on ethical AI use have shown that students who engage with real-world case studies are better equipped to critically analyze the societal implications

of AI technologies (Li, 2018). These findings highlight the potential of such curricula to empower learners and mitigate risks associated with AI integration.

The AI, Human Rights, and Education module of six lessons is a collaborative effort between Human Rights Education Associates (HREA) and Pedagog.ai.[1] This curriculum assumes that learners and educators need to know both the basis of AI as well as human rights. Although this curriculum is not able to address all of the human rights and AI topics presented in this article, it offers a gateway into understanding the human rights at stake and what are the areas requiring not only vigilance but good governance.

This module will be open access as of January 2025 on both websites. Below is an overview of the lesson themes, which are eligible to be used by human rights education organizations as well as educational institutions and schools interested in addressing basic AI literacy through the lens of human rights. These lessons will ideally all be taught, and sequentially. However, educators might opt to use a fewer number of lessons, according to their interests and opportunities to implement.

*Lesson 1: The Basics of AI and Ethical Frameworks*
In this lesson, learners will explore the basic definitions of artificial intelligence and examples of how AI is already being used by governments, schools, corporations, and other institutions in ways that are impacting lives around the world. Learners will then dive into the ethical and human rights implications of AI use through short scenarios and small group discussions. Learners will be able to:

- Define artificial intelligence and name common AI technologies
- Identify the benefits and harms of various uses of AI technology
- Identify frameworks that human society uses to codify ideas of right and wrong including religion, ethics, and human rights
- Apply these frameworks to a specific AI use case.

*Lesson 2: Critical Analysis of AI and Human Rights*
In this lesson, learners will learn about international human rights standards and how these apply to the area of generative AI. Learners will then work in small groups to investigate various examples of AI use and discuss both positive and negative human rights implications. Learners will be able to:

- Describe the origins and key features of human rights
- Critically analyze examples of AI use from a human rights perspective.

---

[1] HREA is an international non-governmental organization that supports human rights education; the training of human rights defenders and professional groups; and the development of educational materials and programming. (See hrea.org) Pedagog.ai is a platform that empowers educators with AI-driven tools and resources to create engaging lesson plans, assignments, assessments, and more. The curriculum was developed by Felisa Tibbitts, Nina Bamberg and Jay Pier.

### *Lesson 3: AI Governance from a Human Rights Perspective*

In this lesson, learners explore the various actors responsible for writing principles and policies related to AI. Learners will understand the current AI policy landscape and will explore a key resource in AI governance to discern how such documents are developed to deduce the major themes and human rights considerations relating to AI. Learners will be able to:

- Explain the nature and purposes of AI governance documents
- Identify institutional actors with a responsibility to create AI Governance and defend the special role of international organizations
- Define AI Governance and describe potential mechanisms, such as principles, policies, and laws
- Analyze common themes found within AI governance documents from a human rights perspective.

### *Lesson 4: Impacts of AI on the Right to Education – Opportunities and Risks*

This lesson begins with a discussion on the right to education, as it relates to access and quality. Learners then work in small groups to discuss the ways that AI-supported Personalized & Adaptive Learning might enhance or impede the right to education. The class concludes with a general discussion of ways that any negative consequences of this, and other AI, technologies might be minimized. Learners will be able to:

- Understand the fundamental principles of the right to education and its importance
- Identify the potential opportunities that AI can provide in enhancing educational access, personalized learning, and teaching effectiveness
- Recognize that unequal access to AI technology exacerbates educational inequalities
- Analyze the risks and ethical concerns associated with the use of AI in education, such as privacy, bias, and the digital divide
- Identify recommendations for leveraging the benefits of AI while mitigating its risks in the educational context.

### *Lesson 5: AI and Human Rights Literacy Building*

In this lesson, learners will be introduced to the ways that biases and factual inaccuracies can be present in outputs from AI tools like chatbots. Learners will work in small groups to analyze key differences between AI-generated and human-written text. Learners will be able to:

- Articulate how AI chatbot tools can produce human-like writing
- Compare and contrast human-written and AI-generated texts
- Discuss the qualities of a reliable source
- Discuss the potential impacts of AI tools that can produce human-like writing on society.

*Lesson 6: Researching AI and Human Rights in Public Services*

This lesson prepares learners to do a research project in which they choose an AI tool being used in a public sector (e.g., healthcare, the justice system) and research how it might be used to promote HR, how it might have HR challenges, and how to promote responsible use. These results are presented in a physical poster or an online digital environment. Learners will be able to:

- Identify and research the functions of one or more AI tools/technologies in use in a public sector setting
- Critically assess the outcomes of the use of these AI tools and technologies from the perspective of human rights
- Propose strategies for reducing or removing negative human rights impacts of the use of these AI technologies in a public sector setting
- Communicate these findings in a poster for public viewing.

**Conclusion**

As the integration of artificial intelligence into education continues to accelerate, the imperative for comprehensive governance frameworks that uphold fundamental human rights has never been more urgent. The analysis presented in this paper has underscored the profound risks AI poses to core principles - like equality, privacy, freedom of expression, and democratic participation within learning environments - if deployed without rigorous safeguards. Conversely, this technology also holds immense promise to enhance access, personalization, and efficiency in education, provided it is harnessed in service of empowering students rather than optimizing for narrow metrics. Ultimately, establishing a balanced, rights-respecting approach to AI in schools will require collaborative policymaking that centers the voices of diverse stakeholders - from technical experts and educators to marginalized community members directly impacted. Robust transparency, accountability, and redress mechanisms must be enshrined to ensure algorithmic decision-making is fair, contestable, and aligned with human dignity. Moreover, comprehensive digital and AI literacy programs empowering both students and teachers are crucial to cultivating their agency and critical thinking in an increasingly automated world.

As AI's disruptive power continues to reshape fundamental structures of teaching and learning, the moral imperative to place human rights at the center of this technological revolution has never been clearer. By proactively defining ethical frameworks and inclusive governance models, policymakers and stakeholders can help ensure AI becomes a great equalizer empowering the next generation, rather than an instrument of oppression and control. The stakes are high, but the opportunity to create a more just, enlightened future through technology-enabled education remains within our grasp.

**Saah Agyemang-Badu** *is a researcher specializing in human rights and culturally responsive education frameworks. Her academic focus bridges global human rights principles and local cultural epistemologies, advancing equitable education systems. She holds advanced degrees in Public Administration and Human Rights and has extensive experience in international and cross-cultural educational initiatives.*

**Felisa Tibbitts** *is UNESCO Chair in Human Rights and Higher Education and Chair in Human Rights Education at the Human Rights Centre of Utrecht University (Netherlands). She is also a Visiting Professor at Nelson Mandela University (South Africa). Felisa is the co-founder and director of Human Rights Education Associates (www.hrea.org).*

**Sage Phillips** *is a proud Penobscot Nation citizen. Her main focus is Indigenous education rights as human rights, specifically equitable solutions for Indigenous youth seeking access to higher education institutions, primarily land-grant institutions.*

## References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L.  (2016, May 23). *Machine Bias*. ProPublica. www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Akyuz, Y. (2020). Effects of Intelligent Tutoring Systems (ITS) on Personalized Learning (PL). *Creative Education*, *11*(6). https://doi.org/10.4236/ce.2020.116069

Benjamin, R. (2019). *Race after technology: Abolitionist Tools for the New Jim Code.* John Wiley & Sons.

Berendt, B., Littlejohn, A. & Blakemore, M. (2020). AI in education: learner choice and fundamental rights. *Journal of Educational Media: The Journal of the Educational Television Association*, *45*(3), 312–32. https://doi.org/10.1080/17439884.2020.1786399

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Buolamwini, J. & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, *81*(81), 77–91. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, *5*(2), 153-163. https://doi.org/10.1089/big.2016.0047

Dastin, J. (2018, October 10). *Insight – Amazon scraps secret AI recruitment tool that showed bias against Women*. Reuters. https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* St. Martin's Press.

European Council (2023, December 11). *The General Data Protection Regulation*. www.consilium.europa.eu. https://www.consilium.europa.eu/en/policies/data-protection-regulation/

Goertzel, B. & Pennachin, C. (2007). *Artificial General Intelligence* (Vol. 2). Springer.

Heffernan, N. T. & Heffernan, C. L. (2014). The ASSISTments Ecosystem: Building a Platform That Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching. *International Journal of Artificial Intelligence in Education*, *24*(4), 470-497. https://doi.org/10.1007/s40593-014-0024-x

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N. & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, *29*(6), 82-97.

Holmes, W., Persson, J., Chounta, I.-A., Wasson, B. & Dimitrova, V. (2022). *Artificial Intelligence and education: A critical view through the lens of human rights, democracy and the rule of law*. Strasbourg: Council of Europe.

Kleinberg, J., Ludwig, J., Mullainathan, S. & Sunstein, C. R. (2020). Algorithms as discrimination detectors. *Proceedings of the National Academy of Sciences*, *117*(48), 30096–30100. https://doi.org/10.1073/pnas.1912790117

Kulik, J. A. & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, *86*(1), 42-78. https://doi.org/10.3102/003465431558142

Li, F.-F. (2018). *How to Make A.I. That's Good for People*. The New York Times. https://www.nytimes.com/2018/03/07/opinion/artificial-intelligence-human.html#

Luckin, R. & Holmes, W. (2016). *Intelligence Unleashed: An Argument for AI in Education.* UCL Knowledge Lab. London, UK. [Report]. UCL Knowledge Lab. https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/innovation/Intelligence-Unleashed-Publication.pdf

Marcus, G., & Davis, E. (2020). *Rebooting AI: Building artificial intelligence we can trust.* Vintage Books.

Mayer-Schönberger, V. & Cukier, K. (2013). *Big Data: A Revolution that Will Transform how We Live, Work, and Think*. Houghton Mifflin Harcourt.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Dartmouth College. http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

McCarthy, J., Minsky, M. L., Rochester, N. & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12–12. https://doi.org/10.1609/aimag.v27i4.1904

Miao, F., Zhang, H., Huang, R. & Holmes, W. (2021). *AI and Education: Guidance for Policymakers*. UNESCO. ttps://unesdoc.unesco.org/ark:/48223/pf0000376709

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York University Press.

Park, S. & Humphry, J. (2019). Exclusion by design: intersections of social, digital and data exclusion. *Information, Communication & Society*, 22(7), 934-953. https://doi.org/10.1080/1369118X.2019.1606266

Selwyn, N. (2019). *Should Robots Replace Teachers? AI and the Future of Education.* Polity Press.

Slade, S. & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist*, *57*(10), 1510-1529. https://doi.org/10.1177/0002764213479366

Tibbits, F. (2002). Understanding what we do: Emerging models for human rights education. *International Review of Education, 48*(3–4), 159–171.

https://doi.org/10.1023/A:1020338300881

United Nations. (1948). *Universal Declaration of Human Rights*. United Nations. https://www.un.org/en/about-us/universal-declaration-of-human-rights

United Nations. (1966). *International Covenant on Economic, Social, and Cultural Rights*. United Nations. https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights

United Nations. (1989). *Convention on the Rights of the Child*. United Nations. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child

United Nations Educational, Scientific and Cultural Organization [UNESCO]. (1970). *Artificial Intelligence in Education*. UNESCO. https://www.unesco.org/en/digital-education/artificial-intelligence

UNESCO. (2021). *UNESCO Publications*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000380063

UNESCO. (2022). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137

UNESCO. (2023). *Global report on teachers: addressing teacher shortages*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000387400

World Bank. (2018). *World Development Report 2018: Learning to Realize Education's Promise*. World Bank Group. www.worldbank.org/en/publication/wdr2018.

World Economic Forum. (2020, October 20). *The Future of Jobs Report 2020.* World Economic Forum. https://www.weforum.org/reports/the-future-of-jobs-report-2020

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (1st ed.). PublicAffairs Books.