

Responsible AI literacy: A stakeholder-first approach

Big Data & Society
July–December: 1–15
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517231219958
journals.sagepub.com/home/bds



Daniel Domínguez Figaredo¹  and Julia Stoyanovich² 

Abstract

The need for citizens to better understand the ethical and social challenges of algorithmic systems has led to a rapid proliferation of AI literacy initiatives. After reviewing the literature on AI literacy projects, we found that most educational practices in this area are based on teaching programming fundamentals, primarily to K-12 students. This leaves out citizens and those who are primarily interested in understanding the implications of automated decision-making systems, rather than in learning to code. To address these gaps, this article explores the methodological contributions of responsible AI education practices that focus first on stakeholders when designing learning experiences for different audiences and contexts. The article examines the weaknesses identified in current AI literacy projects, explains the stakeholder-first approach, and analyzes several responsible AI education case studies, to illustrate how such an approach can help overcome the aforementioned limitations. The results suggest that the stakeholder-first approach allows to address audiences beyond the usual ones in the field of AI literacy, and to incorporate new content and methodologies depending on the needs of the respective audiences, thus opening new avenues for teaching and research in the field.

Keywords

Responsible AI, AI education, ethical AI, AI literacy, AI fairness, AI accountability

Introduction

As the number of automated decision systems (ADS) based on artificial intelligence (AI) grows, so does the interest of citizens and regulators in the ethical and legal challenges they pose (Barocas et al., 2019). Data and algorithms that form part of ADS are subject to biases and other deficiencies that may compromise the responsible use of these systems. In addition, generative AI makes it possible to process existing content, such as text, audio files and images, to create new and original content that may confuse or misinform the audience. It is therefore clear that greater awareness on the part of citizens, technology managers and regulators would help to identify and ultimately mitigate the risks that may arise.

Educational efforts to provide citizens with that understanding are at the core of AI literacy. This is a movement that aims to equip citizens with the cognitive tools necessary to distinguish social processes governed by algorithms, and to be aware of the risks and benefits they entail. Many definitions of AI literacy have been formulated in recent years (Ali et al., 2019; Laupichler et al., 2022; Long and Magerko, 2020; Ng et al., 2022; Steinbauer et al., 2021; Yi, 2021; Zhang et al., 2022), and many projects have been promoted in other fields with similar goals, such as data literacy (D'Ignazio, 2017; Koltay, 2015; Mandinach and Gummer,

2013; Wolff et al., 2016), data ethics (Kerr et al., 2020; Richterich, 2018; Zwitter, A. 2014), and responsible data science (Getoor, 2019; Lewis and Stoyanovich, 2022). Finally, recent meta-studies have analyzed progress in the field by using systematic literature review methods (Laupichler et al., 2022; Ng et al., 2021).

Conceptually, AI literacy is about citizens understanding the social implications of AI. First and foremost, citizens are interested in being aware of how the operation of ADS poses ethical challenges regarding the transparency and fairness of the social processes in which these systems are involved. However, when analyzing the practice of AI literacy, the literature reports that most of the experiences present some limitations that move them away from the original conceptualization. This paper focuses on two of the features of AI literacy found in the literature, listed below:

¹Department of Educational Theory and Social Pedagogy, Universidad Nacional de Educacion a Distancia, Madrid, Spain

²Tandon School of Engineering and Center for Data Science, New York University, New York, USA

Corresponding author:

Daniel Domínguez Figaredo, Universidad Nacional de Educacion a Distancia, Madrid, Spain.
Email: ddominguez@edu.uned.es



- Regarding the target audience, currently, most educational practices in the field of AI literacy are aimed at K-12 students (Kim et al., 2021; Leaton, 2020; Lee et al., 2021; Mertala et al., 2022; Ng et al., 2022; Su et al., 2022; Touretzky et al., 2019). While this is an important audience, it is also necessary to involve the adult population and higher education students. The usual spectrum of most literacy proposals in new emerging fields —i.e., media literacy, digital literacy, web literacy, etc.— is lifelong learning (Belshaw and Hilliger, 2022; Potyrała and Tomczyk, 2021; Rasi et al., 2021), with practices taking place not only in formal settings, but also in informal spaces (Eynon, 2021; Rayendra et al., 2020) and under the umbrella of open learning methods (Domínguez, 2017; Long et al., 2021; Sangrà et al., 2019; Smyth and Breshears, 2017). The fact that AI literacy is preferentially focused on K-12 students narrows the scope of the initiatives and distances them from other important populations, including users of technology, decision-makers, and the public at large.
- Regarding the content of the learning experiences, currently most efforts focus on teaching about the technical aspects of AI systems (Lee et al., 2021, 2022; Williams, 2021), with students spending most of their time practicing with simulations on how the data pipeline is organized to generate decisions (Ali et al., 2019; Ng et al., 2022; Williams et al., 2019; Yannier et al., 2020). While this hands-on approach is appropriate for fostering deep learning about how an AI system works, it is also necessary to teach about the ethical, legal, and social implications of AI.

In this article, we address these two limitations of AI literacy by proposing an approach to designing educational projects from the field of responsible AI. Responsible AI education targets a broader range of audiences in formal and non-formal education —from people in the digital industry to citizens— and focuses more on the social and ethical implications of AI systems. The suggested proposal is embodied in a theoretical-practical formulation of a “stakeholder-first approach”, which consists of specializing instruction according to the stakeholder it addresses and the context in which learning takes place. The assumption is that by adopting a stakeholder-first approach, AI literacy initiatives can increase the diversity of target audiences and highlight the social issues of AI systems.

We describe the stakeholder-first approach and present a study that applies this approach to analyze examples of responsible AI education initiatives, to discover common patterns in instructional design. Our findings can be valuable in guiding the development of AI literacy initiatives. The study was guided by two research questions that were related to addressing the two weaknesses that have been previously identified in the literature on AI literacy:

1. Regarding the target audience of AI literacy, how can a responsible AI education approach contribute to broadening the typology of participants and adapting learning strategies to the needs of each group?
2. Regarding the inclusion of social/ethical issues, how can a responsible AI education approach broaden the corpus of AI literacy by addressing the social implications of AI?

What follows is the path to addressing these questions. Section 2 provides an analysis of the evolution of the field of AI literacy and of the gaps that motivate this study. Then, Section 3 presents the contribution of responsible AI to educating about the social impact of ADS. Next, Section 4 describes the pedagogical framework of the stakeholder-first approach. In Section 5, the elements of the framework are used as a reference to analyze the case studies that explain the educational practice of responsible AI, responding to the research questions. Section 6 concludes with a summary of the stakeholder-first approach and of the findings.

Overview of AI literacy

The definition of “literacy” has evolved from the origin of the word, which refers to the ability to read and write, to the most recent conceptualizations, which define literacy as the ability to identify, understand, interpret, create, communicate, and calculate using printed and written materials in various contexts (UNESCO, 2004). However, as the concept has evolved, there has been an increasing emphasis on the need to acquire a set of technical skills in order to be literate.

The conceptualization of “digital literacy” has followed a similar trajectory. Early definitions of digital literacy referred to the ability to access, manage, evaluate, integrate, create, and communicate information in a digital and networked environment (Dobson and Willinsky, 2009; Karpati, 2011; Reddy et al., 2020). Subsequent interpretations have developed the formulation of comprehensive and increasingly technical frameworks for teaching and assessing digital literacy (Antoninis, 2019; Vuorikari et al., 2022; OECD, 2021). For example, the European DigCom, one of the most widely used frameworks, divides digital competence into five key dimensions —information and data literacy, communication and collaboration, digital content creation, safety, and problem solving — where each dimension is further divided into a number of sub-dimensions and assessment indicators (Vuorikari et al., 2022). Similarly, UNESCO’s Digital Literacy Global Framework identifies seven domains of digital literacy: fundamentals of hardware and software, information and data literacy, communication and collaboration, digital content creation, safety, problem solving, and career-related competences (Law et al., 2018).

The case of “AI literacy” can be appropriately treated as a subset of digital literacy (Kateryna et al., 2020; Pegrum et al., 2018; Yang, 2022). (Note that the terms “AI literacy” and “data literacy” are often conflated in the literature.) Initiatives in the field of AI literacy are more recent than in digital literacy: for example, Ng et al. (2021) document the first academic references to AI and data literacy in 2014; and Laupichler et al. (2022) begin their literature review in the context of higher education in 2016. However, it should be noted that research on the relationship between AI and learning has a longer history, and in particular the scientific field of AI in education (AIED) — which is directly related to AI literacy (Wilton et al., 2022) — that has been active for the past 30 years (Chen et al., 2020).

According to the literature, most researchers working on AI literacy come from data science and computer science, and generally conduct research on specific intervention cases aimed at teaching the basics of AI (Wang, 2020; Zawacki-Richter et al., 2019). Unlike what happens with digital literacy, where most pedagogical proposals are based on a competency-based learning model, AI literacy still lacks pedagogical designs based on outlines of skills that students should acquire. Most projects are specific approaches that combine the proposal of AI content to be acquired by students, the definition of competencies, and methodological schemes based on different learning theories (Bartolomé et al., 2018; Hew et al., 2019; Chen et al., 2020; Ng et al., 2023).

The lack of a broader pedagogical base leads to interpretations of AI literacy that rely primarily on the content/competencies that make up each educational experience. Recent systematic literature reviews analyzing the evolution of AI literacy initiatives reveal a diversity of definitions and educational experiences from different approaches, as is typical of a nascent field (Koltay, 2017; Laupichler et al., 2022; Ng et al., 2021; Wang, 2020). A summary of this can be found in Appendix A, which presents the main definitions of AI literacy, and the resulting content/competencies that are subsequently applied in educational projects.

Regarding the educational levels and target audiences of AI literacy experiences, Ng et al. (2021) reviewed existing research and concluded that most focused on primary and secondary school students, with only a few cases applying to adult populations, university students, and teachers. A similar approach was used by Laupichler et al. (2022) to review AI literacy cases at the higher education level and across the lifespan, with far fewer cases than at the K-12 level, and with most proposals targeting undergraduate students, teachers and administrators, and some experiences targeting AI literacy researchers or, in a few other cases, open to anyone interested in the topic.

Finally, we highlight the key findings of recent systematic reviews that are of most interest to this paper in terms of

identifying the construct of AI literacy and educational initiatives in this area. First, the most common conclusion is that the term AI literacy is not yet well defined. Second, there is consensus that AI literacy has more to do with the usefulness of AI technology in everyday life than with the technological development of AI applications. Third, AI literacy is considered to be primarily about understanding AI and not about learning to design AI systems, which in turn points to the need to overcome reliance on programming exercises as the primary means of instruction (Laupichler et al., 2022; Long and Magerko, 2020; Ng et al., 2021).

Conceptualization of responsible AI

Like AI literacy, the field of responsible AI has evolved as AI systems have become more prominent in society. Responsible AI is closely associated with an ethical approach to the design and development of AI systems, which has led to fewer educational experiences and has mostly focused on the corporate and professional settings (Arrieta et al., 2020; Conradie et al., 2022; Dignum, 2019; Gupta et al., 2021; Lu et al., 2022; Wang et al., 2020). In contrast, AI literacy is strongly influenced by the field of AIED, which has a long history of educational practice in primary and secondary education (Wilton et al., 2022).

Jobin et al. (2019) presented a systematic review that leads to three main approaches to the initial foundations of responsible AI: The first, more superficial, approach recommends acting with “integrity” and clarifying the attribution of responsibility and legal liability in contracts, as well as focusing on remediation of negative impacts. The second approach proposes to go deeper and suggests focusing on the underlying reasons, and design and implementation processes, that may lead to potential harm from AI systems. The third approach emphasizes the responsibility of whistleblowing to identify potential harms, aims to promote diversity among members of technical teams designing AI, and introduces ethics in science and technology education.

The focus of this paper is on two most recent trends: the ethical control and evaluation of AI systems, and the social awareness of their risks and benefits. First, with respect to addressing the social issues involved in the design of AI systems, the field of responsible AI seeks to identify the ethical dimensions and criteria to be considered for acceptable system performance. Most authors agree on a set of essential ethical criteria to be taken into account in AI systems, such as: transparency and privacy; equitable outcomes for stakeholders; welfare of users, customers and employees; bias and fairness of algorithms; transparency and explainability; and reliability and safety (Schiff et al., 2020). Following this scheme, many countries and institutions have proposed criteria along the same lines

to guide and eventually regulate the responsible development and use of AI systems (United States Congress, 2022; Artificial Intelligence Committee, 2018; Beijing Academy of Artificial Intelligence, 2019; European Commission, 2020; Government of Canada, 2021; Montreal Declaration Responsible AI, 2018; OECD, 2019; UNESCO, 2021). Still, some disagreements remain—mainly about how these principles should be understood and what other elements there may be (Ghallab, 2019; Jobin et al., 2019; Wang et al., 2020), and how the criteria should be applied in practice (Birhane et al., 2022; Hagendorf, 2020; Peters et al., 2020; Schiff et al., 2020). Table 1 summarizes an essential scheme proposed by Murad (2022).

Dividing the principles into two groups—first-order and second-order—helps clarify the variety of governance orientations that responsible AI frameworks suggest to those responsible for designing ADS. According to Murad (2022), first-order principles represent the ideals of responsible AI use, and concern “what” should be evaluated in the design and deployment of algorithmic systems. Second-order principles address “how” to ensure that the first set of principles is met.

Closely related to the principles of responsible AI governance are the actors to whom these guidelines are addressed, and the stakeholders in the system deployment

Table 1. Principles for the responsible use of AI systems (Murad, 2022).

First-order principles	Second-order principles
<ul style="list-style-type: none"> Fair and non-discriminatory: Actively assesses, monitors, and mitigates bias; aims to produce properly calibrated fairer outcomes and decisions. Explainable: Able to produce interpretable justification for the decisions produced. Secure: Enacts effective controls to protect systems from threats; actively flags and mitigates vulnerabilities. Robust: Consistently meets accuracy and performance requirements and is robust to perturbations. Upholds data privacy rights: Protects data privacy rights and conforms to existing data laws for both direct and indirect users. Safe: Avoids harm for impacted users and aims to promote human wellbeing. 	<ul style="list-style-type: none"> Ensuring transparency of the system: At a basic level, transparency translates to system visibility, and at a more sophisticated level, it reflects the system's performance on first-order principles. Ensuring accountability of the system: This refers to the ability of system owners to explain their actions (and failings) and take responsibility for them. Preserving human agency and possibility for recourse: When a system fails and adversely impacts an individual, the concerned individual should have a clear recourse process to follow in order to rectify the error.

process (Clarke, 2019). In this context, the term stakeholder refers to the users of AI systems, as well as any individual, group, or organization that can influence or be influenced by AI systems (Deshpande and Sharp, 2022). The literature review by Jobin et al. (2019) identifies the collectives that are primarily responsible for AI actions and decisions: developers, designers, industry, policymakers, and institutions. And Murad (2022) rearranges and clarifies the role of these stakeholders in ensuring the responsible design and use of AI as follows:

- Civil society. These are the groups of individuals affected by AI systems, as well as public interest groups that could be involved in the design and use of the systems, particularly those who are at high risk of harm.
- Public entities. These are organizations responsible for regulating, assessing and holding owners of AI systems accountable.
- System owners. These include individuals, groups and organizations responsible for the implementation, design, development, and maintenance of an AI system, which would be responsible for establishing both purchase requirements and vendor obligations based on responsible use principles.

The involvement of civil society as a stakeholder in responsible AI actions is an important trend that directly links this area to that of AI literacy. Civil society as a stakeholder refers to the idea of raising public awareness about the risks and benefits of AI, and is mainly implemented through educational initiatives. Educational initiatives are also important for other stakeholders. For example, Wang et al. (2020) analyzed a set of practices representative of responsible AI implementation in companies, and grouped them into four categories: data governance, ethical design solutions, human-centered oversight/risk control, and training and education. Focusing on the group of educational practices of interest here, according to Wang et al. (2020), the role of educational programs in companies is to provide managers and employees with a deeper understanding of the ethical use of AI and data. Courses, mentoring systems, cross-functional team-based training, and self-learning are the most common educational practices that help employees develop the mindset and culture of ethical AI.

Finally, regarding the learning methods implemented in responsible AI educational initiatives outside the corporate space, authors such as Dignum (2021), Jobin et al. (2019), and Luckin and Cukurova (2019) focus on how a responsible and trustworthy view of AI relates to and influences research on education and learning. Their proposals aim to deepen the relationship between education studies on the one hand, and AI research and development on the other hand, with the goal of mutual benefit: both to increase the understanding of

educational dynamics among AI developers, and to help teachers and students understand and trust the use of AI.

A stakeholder-first approach to responsible AI education

The practice of responsible AI education points to a methodological pattern that puts the stakeholders of AI systems at the center, mainly due to the importance of different types of stakeholders in promoting responsible principles, as discussed above. In order to sequence the type of instructional design that results from applying a primarily stakeholder-based approach, we have defined a methodological scheme that begins with stakeholder identification —i.e., to whom is the instruction addressed, what is the profile of the stakeholders and the context in which they are involved—and has implications for content —i.e., what is to be learned—and didactic methodology —i.e., how the learning will take place, in terms of the learning resources involved and the learning strategies to be applied— (Figure 1).

Learner-centered frameworks have strong pedagogical foundations and are standard in many educational approaches. Within this umbrella, two of the most important current trends in instructional design stem from the theories of student-centered learning and competency-based learning. Both are clearly related to the stakeholder-first approach, but they also have certain differences: the student-centered approach revolves around the idea of adapting the methodology to the needs of the students (Hoidn and Reusser, 2020), while the competency-based approach focuses more on learning outcomes and their assessment (Gervais, 2016; Voorhees, 2001). However, the main similarity with the framework presented here is that the content to be taught is secondary to the learning design, in the sense that it does not determine the instructional process.

How this framework works in the context of responsible AI educational projects is explained below, and the table in Appendix B summarizes all the dimensions involved:

(1) To whom the learning experiences are addressed. As seen, the large impact of AI systems on society means that the number of stakeholders affected by these systems is very large. (Clarke, 2019) gives examples that can be

classified into the groups of system owners, public bodies, and civil society, such as: drivers and occupants of a self-driving vehicles, pedestrians, and those responsible for directing traffic (self-driving cars example); workers, contractors, insurance companies, and other workers excluded in AI-based contracting processes (AI-assisted hiring and employment); or patients, healthcare professionals, and health insurers (implanted medical devices). Adapting responsible AI education projects to these stakeholders involves not only identifying the target audiences, but essentially tailoring the entire education project to each profile, taking into account the background of each one, the role they play in the process of using/evaluating AI systems, and the actions they can take. For example, in some cases, it may only be of interest that the stakeholders targeted by the educational project have information about the operation of the system that affects them. In other cases, it may be more appropriate for them to know the regulatory framework in which the system is involved. In other cases yet, stakeholders may need access to resources to claim damages caused by the automation of a certain process.

(2) What content is proposed. We share the view that content on “responsibility” should be included in all phases of the learning cycle in the field of responsible AI (Lewis and Stoyanovich, 2022). This, together with the focus on the stakeholder profile and the context of the experiences, implies the need to approach the concept and principles of responsible AI from very different angles to meet the needs of each target group addressed by the educational experiences. Thus, the content on responsibility in stakeholder-oriented education can take different registers, such as responsibility in the design of AI systems, in data management, in the interpretation of results, in the analysis of social impacts, in legal implications, etc.

(3) How learning is supported. Two major dimensions of didactic methodology are addressed here:

- Learning resources should be relevant to the participants and appropriate to their profile and educational level. For example, to address the challenge of teaching data interpretability to undergraduate and graduate students, one may use as a resource “objects-to-interpret-with”—a pedagogical construct inspired by Papert’s (1980) “objects-to-think-with”—so that students can acquire

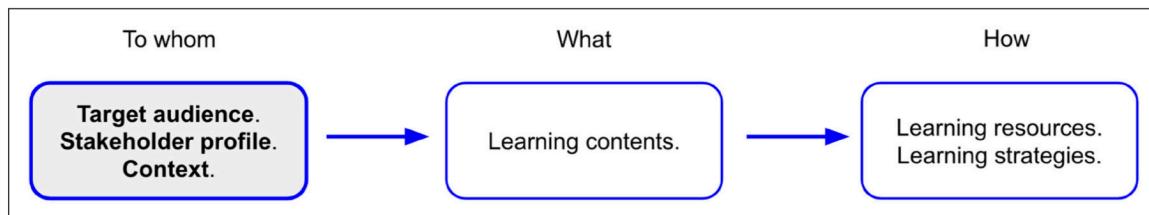


Figure 1. Generic instructional design framework of the stakeholder-first approach.

heuristic and contextualized knowledge. For introductory courses on responsible AI targeting the general public, supplementary resources like videos, comics, simulations, etc., can help simplify complex terminology and reduce cognitive load (Lewis and Stoyanovich, 2022).

- Learning strategies can be adapted to the following two scenarios and their combinations:

- People with technical background. This is the case of the courses on responsible AI for computer science and data science students in higher education settings. These types of courses provide technical knowledge about the process of introducing responsibility in the design of AI systems, as well as ethical learning opportunities applied in specific social contexts once the system is running. In these cases, it is proposed to rely on methodologies with a constructivist pedagogical foundation to maximize student engagement through hands-on programming and visualization, connecting theory through applied activities and ongoing research in practice (Barnes et al., 2017; Hundhausen et al., 2002).
- People with non-technical background. This may be the case for courses aimed at the general population in informal education contexts, or for educational initiatives aimed at a broad spectrum of people affected by AI. The approach here is also based on constructivist theory, which promotes active learning, starting from the student's interest and through sequential processes. But it also includes other simplified methods that facilitate the holistic understanding of responsibility in uninitiated audiences through the use of simulations, case studies, and discussion-based learning. In these open and informal learning situations, it is equally important to include social dynamics in the group, such as learning circles, simulations of AI ethical challenges, or peer-to-peer learning.

Analysis of case studies of responsible AI education practices

Case study descriptions

The cases presented here are drawn from responsible AI education projects, and their analysis seeks to identify the distinctive characteristics of the implementation of the stakeholder-first approach. These cases represent a limited sample of current responsible AI education efforts, and collectively include a variety of instructional design elements. Notably, not all cases have all the characteristics of the stakeholder-first approach.

The following selection criteria were applied: (1) represent educational practices focused on teaching the principles of responsible AI and designed with the target

audience as a priority; (2) show initiatives from different contexts and with different audiences, with a focus on higher education and lifelong learning; (3) are designed as a course, not as a set of resources or groupings of content; (4) the approach/curriculum and learning practices are open access to allow for information analysis and data contrast. The case studies are summarized in this section, with details in Appendix C.

Case 1: Responsible AI, Law, Ethics & Society (Gal et al., 2023; Hod et al., 2022). This course is offered to undergraduate and graduate students with diverse disciplinary backgrounds from Boston University (USA), Tel Aviv University (Israel), the Technion (Israel), and Bocconi University (Italy). The course discusses how the use of AI systems raises challenges and concerns in key areas such as accountability, responsibility, fairness, transparency, and privacy. It takes an interdisciplinary approach when presenting ways to address these challenges.

The methodological approach also emphasizes interdisciplinarity and active work of the participants. It is proposed to deconstruct the issues involved throughout the life cycle of AI systems—design, development and implementation—from the perspective of different disciplines, and to reconstruct solutions with an integrated mindset, from principles and practices located between data science, ethics and law. To this end, students from different disciplinary backgrounds work in teams and perform joint tasks in a series of in-class sessions, including lectures and discussions.

Case 2: Data: Past, Present, and Future (Jones and Wiggins, 2023). This is a Columbia University (USA) course designed for undergraduate students from a variety of backgrounds, including engineering and applied sciences, general studies, arts, and sciences. This course introduces students to both critical thinking and practice in understanding how data-driven algorithms shape our professional, personal, and political realities. In addition, the course reviews the key concepts in “small data” statistics and introduces students to recent trends in computational data exploration.

The methodology proposes that students be divided into two tracks based on their backgrounds: students with less technical knowledge do more technical work, which includes problem solving, while students with a more technical background do more humanistic work, which includes longer writing assignments.

Case 3: Responsible Data Science (Stoyanovich, 2023). This is a New York University (USA) course for undergraduate and graduate students in computer science, data science, and information technology. The course focuses on the “second wave” of data science, which deals with ethics and responsibility in data-driven systems, including legal compliance, data quality, algorithmic fairness and

diversity, data and algorithm transparency, privacy, and data protection. The goal is to address the design of data-driven algorithms by considering their impact on individuals, population groups, and society at large.

The methodology follows that of a typical technical course, augmented with critical reading, writing, and discussion. Learning activities consist of a series of lectures, hands-on labs, individual homework assignments, and a course project, conducted in small teams.

Case 4: We are AI (Corbett and Stoyanovich, 2023). This is an open course developed by the Center for Responsible AI (New York University, USA), and offered to the public in 2021 in online and in-person modalities, with the support of Peer-2-Peer University, a public education non-profit, and the Queens Public Library. The course explains the basics of AI, including how AI-based systems—which are often invisible to the public—“learn” from data to make decisions. The primary focus is on helping learners understand how AI is affecting the way we do things on the Internet, and how it is driving decisions in critical areas like hiring, education, and law enforcement.

The course uses a methodology based on learning circles, which can be face-to-face or online, and consist of facilitated study groups for people who want to meet regularly and learn about a topic with others. In a learning circle, there are no teachers or students: it is a group in which everyone learns together. The facilitator sets the meeting schedule, keeps the group on task during the meetings, and supports each learner’s participation and goals. Group activities include watching short instructional videos, participating in guided discussion, and completing short learning tasks. As an additional resource, a series of comics is provided to help supplement learning in the circles.

Case 5: The Algorithmic Transparency Playbook (Bell et al., 2022, 2023). This is an open course, aimed at technology industry practitioners, organizational decision makers, policy makers, and regulators. The course aims to raise awareness of the importance of improving the levels of transparency of ADS among those who are designing and overseeing the use of these systems. The course provides a playbook, detailing how to influence change and implement algorithmic transparency for ADS in organizations. It discusses guidelines, best practices, and recommendations for algorithmic transparency to avoid potential risks and mitigate harm.

The course is offered online and is based on self-paced study, providing participants with resources that explain what algorithmic transparency is, guiding them through case studies, explaining the transparency best-practices that can be used to effect change at the organizational level, and offering technical and design guidelines for implementing transparency in practice.

Case 6: Principles of Data Science Ethics (Bruce et al., 2023). This is an open online course on the EdX platform,

aimed at practitioners and managers in the field of digital technologies. The course aims to mitigate the harmful effects of machine learning algorithms and AI models by training the developers and implementers of these algorithms in the field of data science ethics. It provides guidance and practical tools to build better models, as well as an audit process to follow to review them. Ultimately, participants will be able to establish a responsible data science framework for their projects.

The methodology is self-paced and includes watching instructional videos, participating in forums, and providing case studies along with Python code.

Case 7: Data Science Ethics (Jagadish, 2023). This is an open online course offered on the Coursera platform for the general public interested in digital technologies and AI. The course addresses ethical considerations related to the privacy and control of consumer information and big data. It provides a framework for analyzing these concerns by examining the ethical and privacy implications of the collection and management of big data. Emphasis is placed on the impact of the field of data science on society and on the principles of fairness, accountability, and transparency.

The methodology is based on students’ self-regulated learning and includes watching instructional videos, participating in discussion forums, and completing assignments at the end of each module.

Findings

Several findings emerged from the application of the stakeholder-first approach as an analytical framework to identify regularities in the case studies (Blyth and Velissaratou, 2019). We analyzed how the pedagogical elements of the case studies vary according to their target audience, taking into account the variables involved in responsible AI education projects. The analysis consisted of three exercises. First, the information from each case was operationalized according to the pedagogical dimensions of the stakeholder-first approach described in Section 4 and summarized in Appendix B, and organized according to the general instructional design sequence outlined in Figure 1 (see Table 2). Next, the cases were grouped according to the type of stakeholder they target, i.e., civil society, public entities, and system owners. Finally, a thematic analysis was conducted to describe the resulting pedagogical patterns (Clarke and Braun, 2014).

The purpose of the study is to provide a basis for investigating how the instructional design of educational projects differs based on the target stakeholders. The goal is to identify patterns that can be applied to other AI literacy educational contexts.

The results of the analysis of how the pedagogical elements of the cases vary according to their target audience

Table 2. Case study data grouped into the dimensions of the stakeholder-first approach to responsible AI education.

Cases	(1) To whom: Stakeholders			(2) What: Learning topics			(3) How: Approach and methods		
	Systems owners	Public entities	Civil society	Responsible AI principles	Contextualization of AI	Technical profile	Non-technical profile		
Case 1 Gal et al., 2023; Hod et al., 2022	- Undergraduate and graduate students with diverse disciplinary backgrounds.	- Liability & robustness. - Discrimination & Fairness. - Privacy. - Transparency. - Explainability.	- AI & Us. - Integration: Content Moderation	- Approach: Students join teams to work together on assignments in in-class sessions. - Methods: Case-studies, lectures, discussions and group work.	- Approach for humanistic background: Students with less technical majors will do more technical work. - Methods: Problem sets.	- Approach for technical background: Students with more technical background will do more humanistic work. - Methods: Longer writing assignments.	- Approach for humanistic background: Students with less technical majors will do more technical work. - Methods: Problem sets.	- Approach: Lectures, labs, and accompanying assignments. - Methods: hands-on technical work, critical reading and interpretation, a course project.	
Case 2 Jones and Wiggins, 2023	- Undergraduate students with diverse disciplinary backgrounds.	- History of human use of data. - Functioning of data-driven systems. - Impacts of data-driven systems. - Biases of data-driven systems. - Key concepts of “small data”, statistics. - Trends in computational data.							
Case 3 Stoyanovich, 2023	- Undergraduate and graduate students with technical backgrounds.	- Fairness. - Data Science Lifecycle. - Data Protection. - Transparency & Interpretability.							

(continued)

Table 2. Continued

Cases	(1) To whom: Stakeholders			(2) What: Learning topics			(3) How: Approach and methods		
	Systems owners	Public entities	Civil society	Responsible AI principles	Contextualization of AI	Technical profile	Non-technical profile		
Case 4 Corbett and Stoyanovich, 2023			– Citizens.		– What is AI? – Learning from data. – Who lives, who dies, who decides? – All about that bias. – We are AI.		– Approach: “Learning circles” meetings (online/ face-to-face). – Methods: Course resources and facilitators support.		
Case 5 Bell et al., 2022; Bell et al., 2023;	– Tech industry practitioners. – Decision makers in organizations.		– Policy- makers. – Regulators.		– Algorithmic transparency.		– Approach: Playbook, self-regulation learning. – Methods: Discuss instructions, best-practices, recommendations.		
Case 6 Bruce et al., 2023	– Practitioners and managers.				– Legal issues. – Transparency. – Responsible Data Science.	– Landscape of harm.	– Approach: Self-regulation learning: – Methods: Video-lectures, case-studies, discussion forums.		
Case 7 Jagadish, 2023	– Citizens.				– What is Ethics? – Privacy. – Anonymity. – Data Validity. – Algorithmic Fairness. – Societal Consequences.		– Approach: Self-regulation learning: – Methods: Video-lectures, case-studies, discussion forums.		

are presented below. The analysis consisted of two exercises: first, the cases were grouped according to the type of stakeholder targeted —i.e., civil society, public entities, and system owners— and then the resulting pedagogical patterns were described by organizing the information according to the general instructional design sequence of the stakeholders-first approach described in Section 4 and in Figure 1.

Focus on civil society

Case 1 (Gal et al., 2023; Hod et al., 2022), Case 2 (Jones and Wiggins, 2023), Case 4 (Corbett and Stoyanovich, 2023), and Case 7 (Jagadish, 2023) are aimed at civil society, with an important distinction between learners with technical and non-technical profiles (learning scheme in Figure 2). For the technical profiles, the focus is on ethics, legal issues, and principles of responsible AI. For the non-technical profiles, the focus is on the context in which the system is situated, along with a social approach to the principles of responsible AI. The methodology in these cases is aligned with the learning objectives, which are also different for the two profiles:

- For technical profiles, the interest is in providing practical skills related to the implementation of responsible AI. The learning strategies are mainly based on case studies, hands-on laboratory work (programming assignments), and applied projects.

- For non-technical profiles, the aim is to teach about general functioning of these systems and to raise awareness of the risks/benefits of their applications in society. Therefore, the learning strategies are based on socialization, through group work or learning circles, based on real-life examples, such as the use of AI in hiring, health-care or education.

Focus on public entities

Case 5 (Bell et al., 2022, Bell et al., 2023) is aimed specifically at agents responsible for overseeing the use of ADS (see instructional scheme in Figure 3). It focuses on the principle of transparency, the importance of which is widely recognized in the field of responsible AI. It is an initiatory approach that seeks to explain in a simple way the main elements to be taken into account when it comes to analyzing algorithmic systems. But it also offers some methodological resources that can be applied to ensure transparency, which could eventually be the goal of regulation.

It therefore follows a didactic scheme based on providing information on what the principle of transparency consists of, together with a series of assumptions that exemplify real situations in which its compliance can be audited. The learning strategy is based on guidance to lead the discussion on the elements of transparency and is also supported by learning resources such as case studies, good practices and recommendations for implementation and auditing.

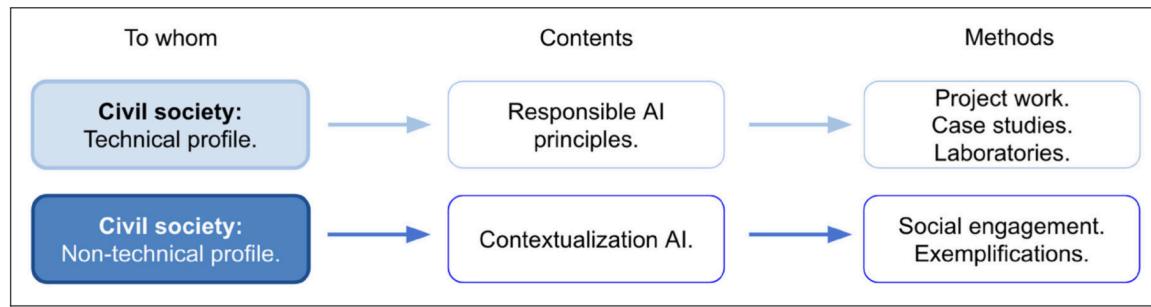


Figure 2. Stakeholder-first approach to responsible AI education with a focus on “civil society”.

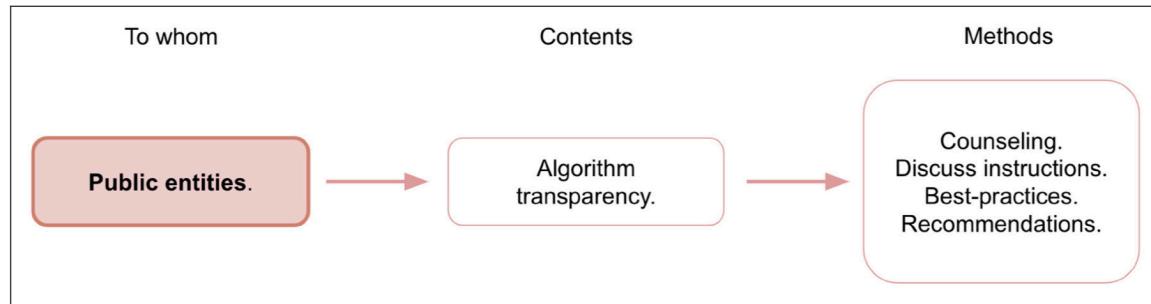


Figure 3. Stakeholder-first approach to responsible AI education with a focus on “public entities”.

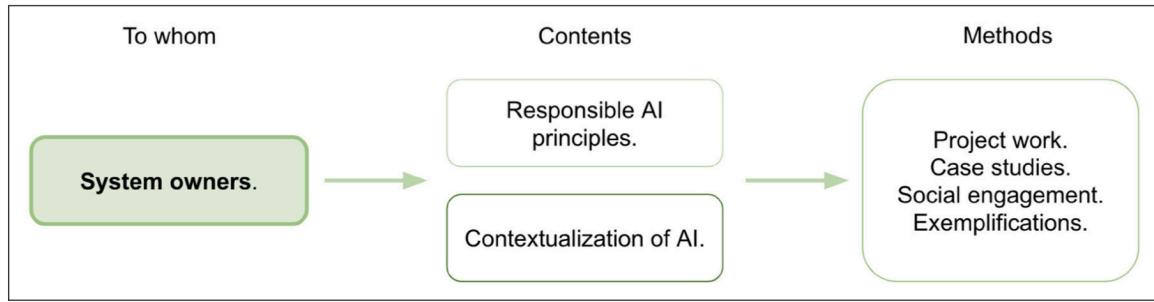


Figure 4. Stakeholder-first approach to responsible AI education with a focus on “system owners”.

Focus on system owners

The cases addressed to the system owners are Case 3 (Stoyanovich, 2023) —focusing on future data scientists, i.e., practitioners—, Case 5 (Bell et al., 2023; Bell et al., 2022) —this one together with a focus on public entities— and Case 6 (Bruce et al., 2023). This audience includes a variety of profiles, such as technology industry professionals, organizational leaders and managers, who are expected to have different types of technical knowledge. As a result, the learning objectives for them are set at two levels: to apply the principles of responsible AI to the design of algorithmic systems, and to engage in the day-to-day management and oversight of these systems. It is assumed that there are a number of technical requirements to make an AI system responsible, and that contextual knowledge about the social implications of AI is also required to ensure their social and ethical appropriateness. Thus, the educational content encompasses the dual sense of responsible AI principles and social contextualization of AI. Likewise, the methods group together the set of strategies seen for the two profiles —technical and non-technical—of civil society actors (see Figure 4): project work, case studies, social engagement, and exemplifications.

The format of the courses aimed at system owners focuses on student self-learning, in one case with a massive open online course (MOOC) model, and in the other case in the form of a playbook without a social platform or teachers per se. In both cases, however, there is a common need for guidance for those responsible for designing and evaluating the technologies: either directly (in the case of the MOOC) through tutoring or forums, or indirectly (in the case of the playbook) through counseling materials and guides.

Conclusions

This paper discusses the characteristics of the “stakeholders first” approach, a framework based on the design of educational initiatives in the area of responsible AI, to address two limitations identified in the literature on AI literacy projects: (1) an emphasis on K-12 students, and the lack of

responsible AI offerings for citizens and university students; and (2) an emphasis on the technical components of algorithmic systems and the lack of attention to the social implications of AI. The stakeholder-first framework was used to analyze seven cases of responsible AI education projects in order to extract pedagogical regularities that emerge from different learning contexts and with different target groups. The analysis showed that responsible AI education projects have a wide scope, expanding the target groups beyond the usual ones in the field of AI literacy and including new content and methodologies according to the needs of stakeholders.

On a methodological level, the stakeholder-first approach proposes to shift the focus of AI literacy from the content to be learned to the target audience and the contexts in which AI systems are used. This has several advantages for civic AI literacy:

- It allows a better analysis of the situation of the students, the level and the type of profile targeted by each educational experience. The analysis of the cases shows that citizens can acquire responsible AI competencies through a variety of learning strategies, provided that the strategies are adapted to each type of audience.
- By focusing first on the recipients of the training, it is possible to propose more appropriate content. The cases showed that the learning content was divided into two groups, used according to the target audiences: content dealing with the basic principles of responsible AI, aimed at technical audiences; and content on contextual aspects of the ethical use of AI, aimed at non-technical audiences.
- It allows the adaptation of the methodology. For technical university audiences, projects, laboratories, and lectures are proposed; for professionals, guidance and practical implementation cases; and for civil society, contextualized examples and peer support to encourage participation and engagement.

Most of the responsible AI cases analyzed here have citizens as the target audience, with the following features to consider in order to improve the AI literacy approach:

- As with the other audiences, both technical and non-technical profiles can be found in the civil society group. Adapting instructional design for technical profiles consists of focusing on aspects that allow AI systems to be adapted to legal and ethical requirements from an applied point of view, for example by referring to good practices, with laboratory experiments, or by working on projects. And in the case of non-technical profiles, adaptation consists of interpreting AI systems in the contexts where they are used (e.g., employee selection, leisure in social networks, or education system), using examples from everyday problems (e.g., with case studies or practical situations), and supporting learning about how AI works in participatory dynamics to make it meaningful and to understand the real effects of the technologies (e.g., group work, learning circles, or discussion in forums).
- The learning content for civil society focuses more on the contextual aspects of AI than on the technical realization of these principles. This means that when the target audience is citizens, the most relevant questions about AI shift from the technological aspects to a more social dimension. Issues related to the social impact of AI that fall into this category include bias management, content moderation, ethical issues in automated decision making, data ownership, and informed consent.

Regarding the presence of computational skills in IA educational programs, the results of the study suggest that it is preferable to focus first on how different audiences can better appropriate the topics to be taught. For example, educational initiatives that incorporate programming skills may be most appropriate for those audiences who want to enter the field of computer science or data science (e.g., the technical university profile) or those who are directly involved in implementing the systems (e.g., public institutions). That being said, all audiences —both technical and non-technical— will have a better understanding of AI if they develop critical thinking skills, reflection on the impact of technologies, or ethical evaluation skills of digital applications.

Overall, the results of the study have implications for both the practice of AI literacy and responsible AI education. The proposed framework can be a useful tool for designing educational projects in fields with a strong focus on literacy practice, such as non-formal education, open learning, and adult education. The main implications of applying the stakeholder-first approach to other realities of AI education can be summarized as shifting the focus from the content to be taught/learned to the audience and the contexts in which AI systems are used, while emphasizing understanding and reflection on the impact of AI on everyday life. These premises are crucial from a pedagogical perspective and provide a pathway for further research in the field of educating society about the

implications of AI. They also provide an opportunity to empirically validate the core assumptions of the framework.

Acknowledgments

This research was supported in part by National Science Foundation (NSF) Awards No. 1916505, 1922658, 2312930, and 2326193.

The authors are grateful to Armando Lewis, Andrew Bell, and Rodrigo Kreis de Paula for their feedback on this article.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Science Foundation, (grant number 1916505, 1922658, 2312930, 2326193).

ORCID iDs

Daniel Domínguez Figaredo  <https://orcid.org/0000-0002-7772-1856>

Julia Stoyanovich  <https://orcid.org/0000-0002-1587-0450>

Supplemental material

Supplemental material for this article is available online.

References

Ali S, Payne BH, Williams R, et al. (2019) Constructionism, ethics, and creativity: Developing primary and middle school artificial intelligence education. In: International Workshop On Education In Artificial Intelligence k-12 (eduai'19). pp.1–4.

Antoninis M (2019) Digital literacy skills: From a framework to a measure. In: UNESCO Institute for Statistics Blog.

Arrieta AB, Díaz-Rodríguez N, Del Ser J, et al. (2020) Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58: 82–115.

Artificial Intelligence Committee (2018) AI in the UK: Ready, willing and able? Report of the session 2017–19. *House of Lords AI Committee*.

Barnes T, Boyer K, Hsiao SI-H, et al. (2017) Preface for the special issue on AI-supported education in computer science. *International Journal of Artificial Intelligence in Education* 27: 1–4.

Barocas S, Hardt M and Narayanan A (2019) *Fairness and machine learning: Limitations and opportunities*. Available at: <https://fairmlbook.org/> (accessed 1 June 2023).

Bartolomé A, Castañeda L and Adell J (2018) Personalisation in educational technology: The absence of underlying pedagogies. *International Journal of Educational Technology in Higher Education* 15(1): 1–17.

Beijing Academy of Artificial Intelligence (2019) *Beijing AI principles*.

Bell A, Nov O and Stoyanovich J (2022) Think About the Stakeholders First! Towards an Algorithmic Transparency Playbook for Regulatory Compliance. *arXiv preprint arXiv:2207.01482*.

Bell A, Nov O and Stoyanovich J (2023) The algorithmic transparency playbook. Available at: <https://dataresponsibly.github.io/algorithmic-transparency-playbook/> (accessed 2 March 2023).

Belshaw D and Hilliger L (2022). *Web literacy map. A map of the skills and competencies people need to read, write and participate effectively on the web*. Available at: <https://wiki.mozilla.org/Learning/WebLiteracyStandard> (accessed 2 March 2023).

Birhane A, Ruane E, Laurent T, et al. (2022) The forgotten margins of AI ethics. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. pp. 948–958.

Blyth A and Velissarou J (2019) *Analytical framework for case study collection. Effective learning environments [Working paper]*. OECD Publishing. Available at: <https://www.oecd.org/education/effective-learning-environments/Analytical-Framework-for-Case-Study-Collection.pdf> (accessed 1 June 2023).

Bruce P, Fleming G and Carlin V (2023). Principles of data science ethics. Available at: <https://www.edx.org/es/course/principles-of-data-science-ethics> (accessed 2 March 2023).

Chen X, Xie H, Zou D, et al. (2020) Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence* 1: 100002.

Clarke R (2019) Principles and business processes for responsible AI. *Computer Law & Security Review* 35(4): 410–422.

Clarke V and Braun V (2014) Thematic analysis. In: Michalos AC (ed) *Encyclopedia of Quality of Life and Well-Being Research*. Springer, pp. 6626–6628.

Conradie N, Kempt H and Königs P (2022) Introduction to the topical collection on AI and responsibility. *Ethics and Information Technology* 24(3): 1–4.

Corbett E and Stoyanovich J (2023) We are AI. Taking control of technology. Available at: <https://dataresponsibly.github.io/we-are-ai/> (accessed 2 March 2023).

Deshpande A and Sharp H (2022) Responsible AI systems: Who are the stakeholders? In: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society. pp. 227–236.

D'Ignazio C (2017) Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal* 23(1): 6–18.

Dignum V (2019) *Responsible artificial intelligence: how to develop and use AI in a responsible way*. Springer.

Dignum V (2021) The role and challenges of education for responsible AI. *London Review of Education* 19(1): 1–11.

Dobson T and Willinsky J (2009) Digital literacy. In: Olson D and Torrance N (eds) *The Cambridge Handbook of Literacy*. New York: Cambridge University Press, 286–312.

Domínguez D (2017) Heuristics and web skills acquisition in open learning environments. *Journal of Educational Technology & Society* 20(4): 102–111. Available at: <https://drive.google.com/file/d/14rxFbbWU-VFDpVun5kB2hTXbvsX8NShB/view> (accessed 2 March 2023).

European Commission (2020) *White paper on artificial intelligence: A European approach to excellence and trust* (COM(2020) 65 final). Available at: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (accessed 2 March 2023).

Eynon R (2021) Becoming digitally literate: Reinstating an educational lens to digital skills policies for adults. *British Educational Research Journal* 47(1): 146–162.

Gal A, Elkin-Koren N, Montagnani ML, et al. (2023). Responsible AI, law, ethics & society. Available at: <https://learn.responsibly.ai/> (accessed 2 March 2023).

Gervais J (2016) The operational definition of competency-based education. *The Journal of Competency-Based Education* 1(2): 98–106.

Getoor L (2019) Responsible data science. In: 2019 IEEE International Conference on Big Data (Big Data). IEEE, pp.1–1.

Ghalla M (2019) Responsible AI: Requirements and challenges. *AI Perspectives* 1(1): 1–7.

Government of Canada (2021) *Directive on automated decision-making*. Available at: <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592> (accessed 2 March 2023).

Gupta S, Kamboj S and Bag S (2021) Role of risks in the development of responsible artificial intelligence in the digital healthcare domain. *Information Systems Frontiers* 25: 1–18.

Hagendorff T (2020) The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines* 30(1): 99–120.

Hew KF, Lan M, Tang Y, et al. (2019) Where is the “theory” within the field of educational technology research? *British Journal of Educational Technology* 50(3): 956–971.

Hod S, Chagal-Feferkorn K, Elkin-Koren N, et al. (2022) Data science meets law. *Communications of the ACM* 65(2): 35–39.

Hoidn S and Reusser K (2020) Foundations of student-centered learning and teaching. In: Hoidn S and Klemenčič M (eds) *The Routledge International Handbook of Student-Centered Learning and Teaching in Higher Education*. London: Routledge, pp. 17–46.

Hundhausen CD, Douglas SA and Stasko JT (2002) A meta-study of algorithm visualization effectiveness. *Journal of Visual Languages & Computing* 13(3): 259–290.

Jagadish HV (2023) Data science ethics. Available at: <https://www.coursera.org/learn/data-science-ethics> (accessed 2 March 2023).

Jobin A, Ienca M and Vayena E (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1(9): 389–399.

Jones M and Wiggins C (2023) Data: Past, present, and future. Available at: <https://data-pf.github.io/> (accessed 2 March 2023).

Karpati A (2011) *Digital literacy in education*. UNESCO Institute for information technologies in Education. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000214485> (accessed 2 March 2023).

Kateryna A, Oleksandr R, Mariia T, et al. (2020) Digital literacy development trends in the professional environment. *International Journal of Learning, Teaching and Educational Research* 19(7): 55–79.

Kerr A, Barry M and Kelleher JD (2020) Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance. *Big Data & Society* 7(1): 2053951720915939.

Kim S, Jang Y, Kim W, et al. (2021) Why and what to teach: AI curriculum for elementary school. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp.15569–15576.

Koltay T (2015) Data literacy: In search of a name and identity. *Journal of Documentation* 71(2): 401–415.

Koltay T (2017) Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science* 49(1): 3–14.

Laupichler MC, Aster A, Schirch J, et al. (2022) Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence* 100101: 1–15.

Law N, Woo D, de la Torre J, et al. (2018) A global framework of reference on digital literacy skills for indicator 4.4. 2. UNESCO. Available at: <http://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf> (accessed 2 March 2023).

Leaton Gray S (2020) Artificial intelligence in schools: Towards a democratic future. *London Review of Education* 18(2): 163–177.

Lee I, Ali S, Zhang H, et al. (2021) Developing middle school students' AI literacy. In: Proceedings of the 52nd ACM technical symposium on computer science education. pp.191–197.

Lee I, Zhang H, Moore K, et al. (2022) AI Book club: An innovative professional development model for AI education. In: Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1. pp.202–208.

Lewis A and Stoyanovich J (2022) Teaching responsible data science: Charting new pedagogical territory. *International Journal of Artificial Intelligence in Education* 32: 783–807.

Long D, Blunt T and Magerko B (2021) Co-designing AI literacy exhibits for informal learning spaces. *Proceedings of the ACM on Human-Computer Interaction* 5(CSCW2): 1–35.

Long D and Magerko B (2020) What is AI literacy? Competencies and design considerations. In: Proceedings of the 2020 CHI conference on human factors in computing systems. pp.1–16.

Lu Q, Zhu L, Xu X, et al. (2022) Responsible AI Pattern Catalogue: a Multivocal Literature Review. *arXiv preprint arXiv:2209.04963*.

Luckin R and Cukurova M (2019) Designing educational technologies in the age of AI: A learning sciences-driven approach. *British Journal of Educational Technology* 50(6): 2824–2838.

Mandinach EB and Gummer ES (2013) A systemic view of implementing data literacy in educator preparation. *Educational Researcher* 42(1): 30–37.

Mertala P, Fagerlund J and Calderon O (2022) Finnish 5th And 6th grade students' pre-instructional conceptions of artificial intelligence (AI) and their implications for AI literacy education. *Computers and Education: Artificial Intelligence* 3: 100095.

Montreal Declaration Responsible AI. (2018) Montreal declaration for a responsible development of artificial intelligence. Available at: https://www.montrealdeclaration-responsibleai.com/_files/ugd/ebc3a3_5c89e007e0de440097cef36dcd69c7b0.pdf (accessed 2 March 2023).

Murad M (2022). Back to basics: Revisiting the responsible AI framework. Towards Data Science. Available at: <https://towardsdatascience.com/back-to-basics-revisiting-the-responsible-ai-framework-847fd3ec860b> (accessed 2 March 2023).

Ng DTK, Leung JKL, Chu SKW, et al. (2021) Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence* 2: 100041.

Ng DTK, Leung JKL, Su J, et al. (2023) Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development* 71(1): 137–161.

Ng DTK, Luo W, Chan HMY, et al. (2022) Using digital story writing as a pedagogy to develop AI literacy among primary students. *Computers and Education: Artificial Intelligence* 3: 100054.

OECD (2019) *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449). OECD. Available at: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> (accessed 2 March 2023).

OECD (2021) *21st-century readers: Developing literacy skills in a digital world*. OECD Publishing.

Papert SA (1980) *Mindstorms: Children, computers, and powerful ideas*. New York: Basic books.

Pegrum M, Dudeney G and Hockly N (2018) Digital literacies revisited. *European Journal of Applied Linguistics and TEFL* 7(2): 3–24.

Peters D, Vold K, Robinson D, et al. (2020) Responsible AI—two frameworks for ethical design practice. *IEEE Transactions on Technology and Society* 1(1): 34–47.

Potyrała K and Tomczyk Ł (2021) Teachers in the lifelong learning process: Examples of digital literacy. *Journal of Education for Teaching* 47(2): 255–273.

Rasi P, Vuojärvi H and Rivinen S (2021) Promoting media literacy among older people: A systematic review. *Adult Education Quarterly* 71(1): 37–54.

Rayendra AI (2020) Assessment of digital literacy competence for non-formal education study program students. In: Proceedings of the 4th International Conference on Learning Innovation and Quality Education. pp.1–4.

Reddy P, Sharma B and Chaudhary K (2020) Digital literacy: A review of literature. *International Journal of Technoethics* 11(2): 65–94.

Richterich A (2018) *The big data agenda: Data ethics and critical data studies*. London: University of Westminster Press.

Sangrà A, Raffaghelli JE and Veletsianos G (2019) Lifelong learning ecologies: Linking formal and informal contexts of learning in the digital era. *Wiley Online Library* 50(4): 1615–1618.

Schiff D, Biddle J, Borenstein J, et al. (2020) What's next for AI ethics, policy, and governance? A global overview. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. pp.153–158.

Schiff D, Rakova B, Ayesh A, et al. (2020) Principles to practices for responsible AI: Closing the gap. *arXiv preprint arXiv:2006.04707*.

Smythe S and Breshears S (2017) Complicating access: Digital inequality and adult learning in a public access computing space. *Canadian Journal for the Study of Adult Education* 29(1): 67–81. Available at: <https://cjsae.library.dal.ca/index.php/cjsae/article/view/5362> (accessed 1 June 2023).

Steinbauer G, Kandhofer M, Chklovski T, et al. (2021) A differentiated discussion about AI education K-12. *KI-Künstliche Intelligenz* 35(2): 131–137.

Stoyanovich J (2023) Responsible Data Science. Available at: <https://dataresponsibly.github.io/rds23/> (accessed 2 March 2023).

Su J, Zhong Y and Ng DTK (2022) A meta-review of literature on educational approaches for teaching AI at the K-12 levels in the Asia-Pacific region. *Computers and Education: Artificial Intelligence* 100065: 1–18.

Touretzky D, Gardner-McCune C, Martin F, et al. (2019) Envisioning AI for K-12: What should every child know about AI? In: Proceedings of the AAAI conference on artificial intelligence. pp.9795–9799.

UNESCO (2004) *The Plurality of literacy and its implications for policies and programmes: position paper*. Paris: UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000136246> (accessed 2 March 2023).

UNESCO (2021). Recommendation on the ethics of artificial intelligence (SHS/BIO/PI/2021/1). Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000136246>

unesco.org/ark:/48223/pf0000381137.locale=en (accessed 2 March 2023).

United States Congress (2022) Algorithmic Accountability Act of 2022 Text - H.R.6580. Available at: <https://www.congress.gov/bill/117th-congress/house-bill/6580/text> (accessed 2 March 2023).

Voorhees AB (2001) Creating and implementing competency-based learning models. *New Directions for Institutional Research* 110: 83–95.

Vuorikari R, Kluzer S and Punie Y. (2022) *DigComp 2.2: The Digital Competence Framework for Citizens*. Luxembourg: Publications Office of the European Union.

Wang Y, Xiong M and Olya H (2020) Toward an understanding of responsible artificial intelligence practices. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences (HICSS), pp.4962–4971.

Wang Z (2020) Exploring different notions of literacy: A literature review analysis of literacy research related to artificial intelligence and big data application. In: IOP Conference Series: Materials Science and Engineering. IOP Publishing, p.012023.

Williams R (2021) How to train your robot: Project-based AI and ethics education for middle school classrooms. In: Proceedings of the 52nd ACM Technical Symposium on Computer Science Education. pp.1382–1382.

Williams R, Park HW, Oh L, et al. (2019) Popbots: Designing an artificial intelligence curriculum for early childhood education. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp.9729–9736.

Wilton L, Ip S, Sharma M, et al. (2022) Where is the AI? AI literacy for educators. In: Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium: 23rd International Conference, AIED 2022, Durham, UK, July 27–31, 2022, *Proceedings, Part II*. Springer, pp.180–188.

Wolff A, Gooch D, Montaner JJC, et al. (2016) Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics* 12(3): 9–26.

Yang W (2022) Artificial intelligence education for young children: Why, what, and how in curriculum design and implementation. *Computers and Education: Artificial Intelligence* 3: 100061.

Yannier N, Hudson SE and Koedinger KR (2020) Active learning is about more than hands-on: A mixed-reality AI system to support STEM education. *International Journal of Artificial Intelligence in Education* 30: 74–96.

Yi Y (2021) Establishing the concept of AI literacy. *Jahr-European Journal of Bioethics* 12(2): 353–368.

Zawacki-Richter O, Marín VI, Bond M, et al. (2019) Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education* 16(1): 1–27.

Zhang H, Lee I, Ali S, et al. (2022) Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: An exploratory study. *International Journal of Artificial Intelligence in Education* 33: 1–35.

Zwitter A (2014) Big data ethics. *Big Data & Society* 1(2): 2053951714559253.