13

Artificial Intelligence and Education

Different Perceptions and Ethical Directions

Inge Molenaar, Duuk Baten, Imre Bárd, and Marthe Stevens*

13.1 INTRODUCTION

Recently there has been much discussion about AI in all application domains, especially in the field of education.[1] Since the introduction of ChatGPT, a storm has swept through the educational landscape.[2] The awareness that AI will impact education has now reached the general public. For instance, teachers are confronted with AI in their daily practices when students, from late primary education to university, find their way to generative AI as an easy help to support homework, write essays, and make assessments.[3] In this way, generative AI comes into schools through the backdoor, and educational professionals struggle to respond meaningfully. This stands in stark contrast with the instructional design approach and responsible research and innovation trajectories, in which applications of technology and AI are carefully designed for use in education, relevant stakeholders are included in the development process, and diverse societal and ethical implications are assessed.[4] In this chapter, we argue that these recent developments further increased the need for ethical approaches that stimulate the responsible use of AI in education.

Although AI in education has been a scientific field for over 35 years,[5] policy-oriented developments and ethical approaches directly focused on AI and education are more recent. Following the development of general guidelines for developing

[*] At the time of writing this chapter, Imre Bárd was also a part-time Trust and Safety contractor at OpenAI.

[1] Wayne Holmes and Ilkka Tuomi, "State of the art and practice in AI in education" (2022) *European Journal of Education*, 57: 542; Inge Molenaar, "Towards hybrid human-AI learning technologies" (2022) *European Journal of Education*, 57: 632.

[2] Enkelejda Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education" (2023) *Learning and Individual Differences*, 103: 102274.

[3] Cindy Gordon, "How are educators reacting to Chat GPT?" (*Forbes*), www.forbes.com/sites/cindygordon/2023/04/30/how-are-educators-reacting-to-chat-gpt/, accessed August 4, 2023.

[4] Jack Stilgoe, Richard Owen, and Phil Macnaghten, "Developing a framework for responsible innovation" (2013) *Research Policy*, 42: 1568; Molenaar, "Towards hybrid human-AI learning technologies" (n 1).

[5] "International AIED Society," https://iaied.org/about, accessed August 4, 2023.

and using AI,[6] the first international event on AI in education with a policy and ethics perspective was organized by UNESCO in 2019.[7] The resulting statement, the Beijing consensus,[8] was followed up by numerous NGO initiatives to support governments toward policy for responsible use of AI in education. Examples are the *OECD Digital Education Outlook 2021: Pushing the frontiers with AI, Blockchain and robots*[9] and the European Commission's *Ethical guidelines on using artificial intelligence (AI) and data in teaching and learning for educators*.[10]

In this chapter, we discuss why AI in education is a special application domain and outline different perspectives on AI in education. We will provide examples of various specific-purpose AI applications used in the educational sector and generic-purpose AI solutions moving into schools (Section 13.2). Next, we will outline ethical guidelines and discuss the social impact of AI in education (Section 13.3), elaborating on initial steps taken in the Beijing consensus and ethical guidelines for AI and data use in education from the European Union. Finally, we describe concrete examples from the Netherlands, where the *Dutch value compass for digital transformation* and the *National Education Lab AI (NOLAI)* serve as an illustration of how a collaborative research-practice center can facilitate proactive ethical discussions and support the responsible use of AI in education (Section 13.4), and conclude (Section 13.5).

13.2 AI IN EDUCATION: A SPECIAL APPLICATION DOMAIN OF AI

It has been argued that AI in education is a special application area of AI.[11] To explain why the use of AI in education is unique, we build on the distinction between the *replacement* and *augmentation* perspectives on the role of AI in education.[12] In many application areas of AI, the replacement perspective is most dominant

[6] Anna Jobin, Marcello Ienca, and Effy Vayena, "The global landscape of AI ethics guidelines" (2019) *Nature Machine Intelligence*, 1: 389.

[7] Fengchun Miao and Wayne Holmes, "International forum on AI and the futures of education, developing competencies for the AI era, December 7–8, 2020: Synthesis Report" (UNESCO, 2021), https://unesdoc.unesco.org/ark:/48223/pf0000377251, accessed August 3, 2023.

[8] UNESCO, "Beijing consensus on artificial intelligence and education – UNESCO Digital Library," https://unesdoc.unesco.org/ark:/48223/pf0000368303, accessed August 4, 2023.

[9] "OECD digital education outlook 2021 – Pushing the frontiers with AI, blockchain, and robots," https://digital-education-outlook.oecd.org/, accessed August 4, 2023.

[10] European Union, "Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators – Publications Office of the European Union," https://op.europa.eu/en/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1/language-en, accessed August 4, 2023.

[11] Ryan S. Baker, "Artificial intelligence in education: Bringing it all together" (OECD, 2021), www.oecd-ilibrary.org/education/oecd-digital-education-outlook-2021_f54ea644-en, accessed August 4, 2023; Inge Molenaar, "Personalisation of learning: Towards hybrid human-AI learning technologies" (OECD, 2021), www.oecd-ilibrary.org/education/oecd-digital-education-outlook-2021_2cc25e37-en, accessed August 4, 2023.

[12] R. Luckin and W. Holmes, "Intelligence unleashed: An argument for AI in education" (UCL Knowledge Lab, 2016) Report www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/innovation/Intelligence-Unleashed-Publication.pdf, accessed August 4, 2023.

and considered appropriate. This means that the focus is on replacing human behavior with AI systems. For example, the application of AI in the self-driving car explicitly aims to offload driving from humans to AI. In contrast, AI in education aims to optimize human learning and teaching.[13] It is important to note that humans and artificial intelligence have different strengths.[14] While AI systems are good at quickly analyzing and interpreting large amounts of data, humans excel at social interaction, creativity, and problem-solving. The augmentation perspective strives for a meaningful combination of human and artificial intelligence.

Current AI systems cannot make broad judgments and considerations as humans do: they merely recognize patterns and use those to optimize learning outcomes or mirror human behavior. In addition, the function of education is broader than the development of knowledge and skills; general development, socialization, and human functioning are critical aspects.[15] With a restricted focus on optimizing learning outcomes, there is a considerable risk that these broader education functions will be lost out of sight.[16] Consequently, it is important to ensure that critical processes for human learning and teaching are not offloaded to AI. For example, adaptive learning technologies (ALTs) can take over human regulation, that is, control and monitoring of learning, in optimizing the allocation of problems to learners.[17] Similarly, automated forms of feedback may reduce social interaction between learners and teachers.[18] Hence, it is important to understand how the application of AI in education offloads elements from human learning and teaching.[19]

This notion of offloading can also help us understand the storm that the introduction of ChatGPT has created in educational institutions around the world. Students bypass the intended learning process when they use generative AI for homework. Homework is designed to help students engage in cognitive processing activities to integrate new knowledge into their mental models and develop a more elaborate understanding of the world.[20] Hence, students using generative AI for homework brings considerable risks of offloading and reduced learning. At the same time,

[13] Inge Molenaar, "The concept of hybrid human-AI regulation: exemplifying how to support young learners' self-regulated learning" (2022) *Computers and Education: Artificial Intelligence*, 3: 100070.

[14] Zeynep Akata et al., "A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence" (2020) *Computer*, 53: 18.

[15] Gert Biesta, "Risking ourselves in education: Qualification, socialization, and subjectification revisited" (2020) *Educational Theory*, 70: 89.

[16] Neil Selwyn, "Should robots replace teachers?: AI and the future of education," 145.

[17] Inge Molenaar, Anne Horvers, and Ryan S. Baker, "What can moment-by-moment learning curves tell about students' self-regulated learning?" (2021) *Learning and Instruction*, 72: 101206.

[18] Cultuur en Wetenschap Ministerie van Onderwijs, "Inzet van intelligente technologie – Advies – Onderwijsraad" (September 28, 2022), www.onderwijsraad.nl/publicaties/adviezen/2022/09/28/inzet-van-intelligente-technologie, accessed August 4, 2023.

[19] Molenaar, "Towards hybrid human-AI learning technologies" (n 1).

[20] Jeroen J. G. Van Merriënboer, and Paul A. Kirschner, *Ten Steps to Complex Learning: A Systematic Approach to Four-Component Instructional Design* (Routledge, 2017).

combining generative AI with effective pedagogics may provide new education opportunities.[21] For example, dynamic assessment in combination with collaborative writing, where the students write a paragraph and generative AI writes the next paragraph, can help students develop new writing skills while still ensuring students' conscious processing and engagement with the instructional materials offered and challenging them to make a cognitive effort to learn. Despite these good examples, many questions about implementing AI that augments human learning and teachers remain. Therefore, it is important to understand how AI offloads human learning and teaching. A careful analysis of the pedagogical and didactical arrangements can ensure that we do not offload critical processes for learning or teaching.

13.2.1 *Understanding Offloading in Education*

In order to better analyze how AI is offloading human learning and teaching, two different models can be used.[22] First, the Detect-Diagnose-Act Framework distinguishes three mechanisms underlying the functioning of AI in education (see Figure 13.1). In *detect*, the data that AI uses to understand student learning or teacher teaching are made explicit. The constructs AI analyses to understand the learning or teaching process are outlined in the *diagnosis*. Finally, *act* describes how the diagnostic information is translated into didactic pedagogical action. For example, an ALT for mathematical learning uses the learners' answers to questions as input to diagnose a learner's knowledge of a specific mathematical topic.[23] This insight is used to adjust the difficulty level of problems provided to the learner and to determine how a learner should continue to practice this topic. Below, we provide an example of how this can look like in practice under "Case 1."

From the teaching perspective, the task of adjusting problems to students' individual needs is offloaded to ALT. The technology and the teachers share the task of determining when a learner has reached sufficient mastery. Although these technologies support teachers,[24] it is important to ensure that teachers stay in control. From the learner's perspective, the need to monitor and control learning is reduced as the technology supports learning by adjusting the problem's difficulty, which decreases

21 Mike Sharples, "Towards social generative AI for education: Theory, practices and ethics" (2023) *Learning: Research and Practice*, 9(2): 159–167.

22 Molenaar, "Personalisation of learning" (n 11).

23 Inge Molenaar and Annemarie van Schaik, "A methodology to investigate the usage of educational technologies on tablets in schools," (2017) *Tablets in Schule und Unterricht*.

24 Carolien A. N. Knoop-van Campen, Alyssa Wise, and Inge Molenaar, "The equalizing effect of teacher dashboards on feedback in K-12 classrooms" (2021) *Interactive Learning Environments*, 31(6): 3447–3463; Anouschka van Leeuwen et al., "How teacher characteristics relate to how teachers use dashboards: Results from two case studies in K-12" (2021) *Journal of Learning Analytics*, 8(2): 6–21.

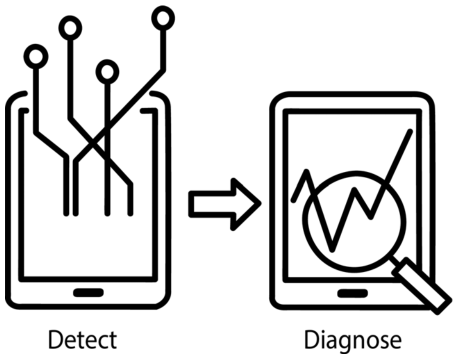

FIGURE 13.1 Detect-Diagnose-Act Framework[25]

the need for learners to self-regulate their learning and may affect the development of these skills.[26]

In this way, the Detect-Diagnose-Act Framework helps analyze offloading by AI, illustrating how particular AI solutions function in educational arrangements. At the same time, this model only describes the AI's roles and largely ignores the roles of learners and teachers. Here *the six levels of the automation model* can be used to understand the division of control between AI, learners, and teachers in education. This model distinguishes six levels of automation in which the degree of control gradually transfers from the teacher to the AI system. The model starts with full teacher control and ends with full automation or AI control (see Figure 13.2). Hence the model goes from no offloading to AI to full offloading.

This model includes elements from the detect-diagnose-act framework. The input lines at the bottom represent detection and data collection in intelligent technologies. The data forms the basis for the AI system to diagnose meaningful constructs for learning and teaching, as described earlier. Hence, more data and different data streams are required for further automation. The diagnostic information is consequently transformed into different pedagogical didactical actions that can be taken in response. The main focus of this model is to make explicit which actors, that is, teachers, learners, or the AI system, perform those actions. This largely determines the position of an educational arrangement with AI on the model.

This model has distinct levels of automation at which AI can execute actions.

First, the AI system can provide information and *describe* student behavior without taking over any control (*teacher assistance level*). The information provided is known

[25] Molenaar, "Personalisation of learning" (n 11).
[26] Molenaar, "The concept of hybrid human-AI regulation" (n 13).

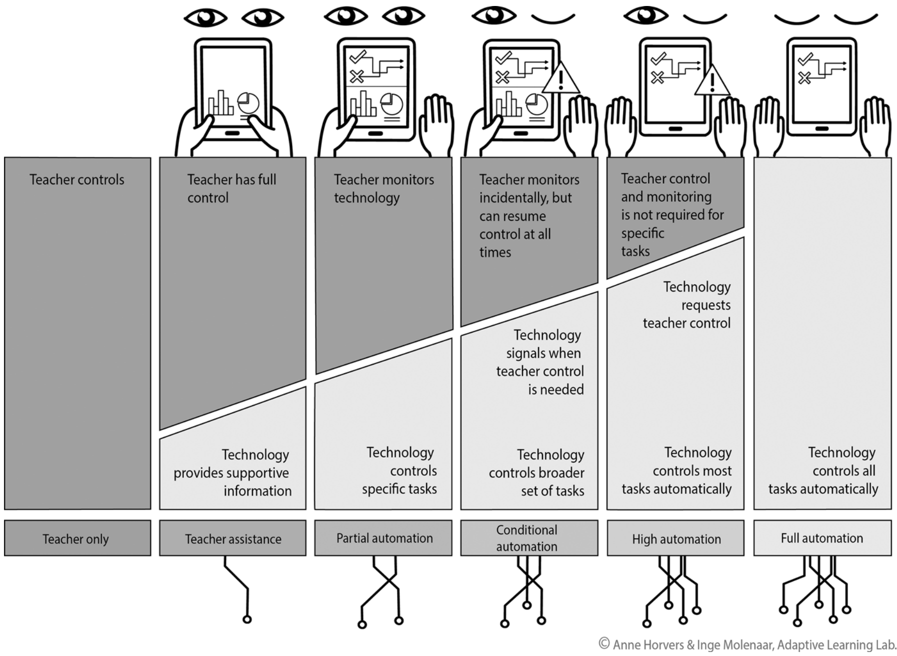

FIGURE 13.2 Six Levels of Automation Model[27]

to impact teacher behavior.[28] It can be communicated in different forms describing, guiding, and even recommending particular actions.[29] Second, the AI system can *enact* simple actions during learning. These actions typically are at three levels: the *step* level providing feedback within a problem, the *task* level adjusting the task to the student's needs, or the *curriculum* level optimizing the order of learning topics. In this *partial automation* level, AI only takes over tasks at one particular level, either enacting step, task, or curriculum adaptivity in interaction with the learner. In Case 1, an example of task adaption is outlined. In *conditional automation*, multiple tasks are taken over by AI, which can be a combination of different levels of adaptivity. With the transition of tasks to the AI system, the importance of the interface between the system and the teacher increases. For teachers to orchestrate the learning scenarios in the classroom, AI must inform the teacher adequately about the actions taken. Hence coordination between AI and humans becomes more critical. In *high automation*, control transfers primarily to AI and teachers step in only for specific tasks. Teacher actions are needed in case AI does not oversee the learning context. Here AI steers learning to a large extent. Finally, in *full automation*, the system autonomously supports learning and teaching without human control.

[27] Molenaar, "Personalisation of learning" (n 11).
[28] Carolien Knoop-Van Campen and Inge Molenaar, "How teachers integrate dashboards into their feedback practices" (2020) *Frontline Learning Research*, 8: 37.
[29] Van Leeuwen et al. (n 24).

This model is functional for describing the augmentation perspective of AI in education, positioning the current role of AI in education, and discussing the future development of the role of AI in education. It can also help foster the discussion about the envisioned role of AI in education, in which it should be made explicit that the goal is not to reach full automation. Successful augmentation requires an ongoing interaction between humans and AI, and the interface between humans and AI is critical.[30] The *Detect-Diagnose-Act Framework* and the *Six Levels of Automation Model* help to understand offloading by AI in specific educational arrangements and analyze the implications of AI in education more broadly. These insights can help teachers and educational professionals understand different applications of AI in the educational domain, allow scientists from different disciplines to compare use cases and discuss implications, and enable companies to position their products in the EdTech market.

CASE 1 Adaptive Learning Technologies

Adaptive Learning Technologies (ALT) and Intelligent Tutoring Systems (ITS) have become increasingly prevalent in European primary and secondary schools. These technologies personalize learning for students in foundational mathematics, grammar, and spelling skills. Using tablets and computers allows rich data on student performance to be captured during practice sessions. For instance, the Snappet technology,[31] widely used in the Netherlands, is typically employed in combination with the pedagogical direct instruction model. In this approach, the teacher activates prior knowledge through examples and explains today's learning goal to the students. Smartboard materials support this direct instruction phase, and students work on adaptively selected problems during the individual practice phase. This practice is tailored to the needs of each student, with the technology providing direct feedback during the process. Teacher-facing dashboards give educators the information they need to make informed decisions about providing feedback and additional instruction. They can also optimize the balance between digital and face-to-face lesson components.

The current generation of ALTs uses data on student performance to adapt problems to learners' predicted knowledge levels and to provide additional information on their progress in teachers' dashboards. These technologies enable more efficient teaching of foundational skills, and free time to focus on more complex problem-solving, self-regulation, and creativity. Adaptive learning technologies offer advantages, including advanced personalization of practice materials tailored to each

30 Akata et al. (n 14).
31 "Homepage" (*Snappet*), https://snappet.nl/, accessed August 4, 2023.

student's needs and the ability for teachers to devote more time to tasks beyond the reach of technology, such as providing elaborate feedback or helping students who need additional instruction. This case represents an example of partial automation, in which the teacher and the ALT work closely together. The functions of the ALT are to describe, diagnose, and advise the teacher through the dashboards based on ongoing student activities and, in specific cases, to select student problems. Teachers continue to control most organizational tasks in this learning scenario and remain responsible for monitoring the functioning of the technology, in which teacher dashboards play an important role. ALTs are one example of AI in education, below we provide an overview applications.

13.2.2 *Applications of AI in Education*

Generally, applications of AI in education can be divided into student-faced, teacher-faced, and administrative AI solutions, depending on the actor/stakeholder they support.[32] Below the most commonly used AI systems of each type are shortly outlined.

13.2.2.1 Student-Facing AI in Education

AI for learners is directed at human learners to support learning and make it more efficient, effective, or enjoyable. A large range of ALTs and intelligent tutor systems (ITS) adjusts to the needs of individual learners.[33] These technologies mostly show three levels of adaptivity: step, task, and curriculum adaptivity. In step adaptivity, the learner receives feedback or support within a particular learning task, for example, elaborative feedback on a mistake made in solving math equations providing automatic formative assessment. Task adaptivity aims to give students a task that fits their progress or interest. For example, when a learner is making progress, the next problem selected can be more difficult than when a learner is not making progress. Finally, curriculum adaptivity is directed at supporting learners' trajectories and selecting fitting next learning goals or topics to address. Intelligent Tutoring Systems often combine multiple levels of adaptivity[34] and have been shown to improve students' learning.[35] Most adaptive technologies focus on analyzing students' knowledge; these systems often do not measure other important learning constructs such as self-regulated learning, motivation, and emotion. Case 2 provides an example of how to develop systems that also consider these broader learning characteristics. New developments are chatbots for learning with a more dialogic character, dialogue-based

[32] Holmes and Tuomi (n 1).

[33] Vincent Aleven et al., "Instruction based on adaptive learning technologies" (2016) *Handbook of Research on Learning and Instruction*.

[34] Kurt VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems" (2011) *Educational Psychologist*, 46: 197.

[35] James A. Kulik and J. D. Fletcher, "Effectiveness of intelligent tutoring systems: A meta-analytic review" (2016) *Review of Educational Research*, 86: 42.

tutoring systems, exploratory learning environments with games, learning network orchestrators, simulations, and virtual reality.[36]

13.2.2.2 Teacher-Facing AI in Education

Teacher-facing AI applications are mostly systems that help teachers to optimize their instruction methods. The best-known solutions are teacher dashboards that have been shown to impact teacher feedback practices during lessons. Teachers provide different feedback,[37] allocate it to different students,[38] and reduce inequality.[39] Classroom orchestration can also help teachers when teaching classes to make changes based on how students respond to their teaching.[40] Automatic summative assessment systems directly assess students' work. More recently, double-teaching solutions and teaching assistants have provided teachers with instructional support in their classrooms.[41] Finally, classroom monitoring systems[42] and plagiarism detection are helping teachers ensure academic integrity and maintain a fair learning environment in their classrooms.

13.2.2.3 Administrative AI in Education

Administrative AI solutions are directed at helping schools to enact education in an efficient matter. Here, AI is used for administrative purposes such as financial planning, course planning, and making schedules.[43] Quality control is another application that uses predictive analytics of how students develop, both for admission and to identify at-risk students.[44] Finally, e-proctoring monitors students during exams.[45]

36 Holmes and Tuomi (n 1).

37 I. Molenaar and C. Knoop-van Campen, "How teachers make dashboard information actionable" (2018) *IEEE Transactions on Learning Technologies*.

38 Knoop-Van Campen and Molenaar (n 28).

39 Kenneth Holstein et al., "The classroom as a dashboard: Co-designing wearable cognitive augmentation for K-12 teachers," *ACM International Conference Proceeding Series* (2018), https://dl.acm.org/doi/abs/10.1145/3170358.3170377?casa_token=dn-UmbWKvosAAAAA:mfruhjvvLGSfKtF5fZUUd5km5WypTmZAPsLE2vXLt4CXTWtMyYMl-TvebU-POtCQsJe_x1Vjh8c, accessed March 3, 2020.

40 Pierre Dillenbourg, "Classroom analytics: Zooming out from a pupil to a classroom" (OECD, 2021), www.oecd-ilibrary.org/education/oecd-digital-education-outlook-2021_336f4ebf-en, accessed August 4, 2023.

41 Alex Guilherme, "AI and education: The importance of teacher and student relations" (2019) *AI and Society*, 34: 47.

42 Qui X. Lieu, Dieu T. T. Do, and Jaehong Lee, "An adaptive hybrid evolutionary firefly algorithm for shape and size optimization of truss structures with frequency constraints" (2018) *Computers & Structures*, 195: 99.

43 Kirsty Kitto et al., "Towards skills-based curriculum analytics: Can we automate the recognition of prior learning?," *ACM International Conference Proceeding Series* (Association for Computing Machinery, 2020).

44 Alex J. Bowers, "Early warning systems and indicators of dropping out of upper secondary school: The emerging role of digital technologies" (OECD, 2021), www.oecd-ilibrary.org/education/oecd-digital-education-outlook-2021_c8e57e15-en, accessed August 4, 2023.

45 Aditya Nigam et al., "A systematic review on AI-based proctoring systems: Past, present and future" (2021) *Education and Information Technologies*, 26: 6421.

CASE 2 Student-Facing Dashboards for Self-Regulated Learning

Recent advancements in learning technologies have expanded the focus of personalized education beyond learner knowledge and skills to include self-regulated learning, metacognitive skills, monitoring and controlling learning activities, motivation, and emotion. Research shows that self-regulated learning, motivation, and emotion play a vital role in learning. Incorporating self-regulated learning in personalized education can improve current and future learning outcomes.

The Learning Path App[46] is an example of this development. The app uses ALT's log data to detect self-regulated learning processes during learning. The moment-by-moment learning algorithm was developed to visualize the probability that a learner has learned a specific skill at a specific time. The algorithm provides insight to learners on how accurately they worked (monitoring) and when they need to adjust their approach (control). Personalized dashboards were developed for students to provide feedback, changing the role of learner-facing dashboards from discussing *what* learners learned to also incorporating *how* learners learned.

Results indicate that learners with access to dashboards improved control and monitoring of learning and achieved higher learning outcomes and monitoring accuracy. Widening the indicators that are tracked and increasing the scope of diagnosis can further personalize learning and advance our ability to accurately understand a learner's current state and improve the prediction of future development. This supports better approaches toward the personalization of learning that incorporate more diverse learner characteristics and a better understanding of the learner's environment.[47]

The above-illustrated perspectives on the use of AI in education offers insights into how AI can offload human learning, how that affects the roles of teachers and learners and which different AI solutions exist. Still, many challenges and questions remain, and many initiatives have been taken to steer the development of AI in education in a desirable direction. The next section will reflect on those policy, governance, and ethical initiatives, starting with a cursory view of the AI ethics discourse developed over the past decade. We then concentrate on the specific realm of education, discussing major ethical frameworks chronologically. The section concludes with a closer look at the Netherlands' pioneering role in addressing the ethical dimensions of digital applications in education.

[46] "Leerpaden – Apps op Google Play," https://play.google.com/store/apps/details?id=com.leerpaden.rickdijkstra.iprogress20&hl=nl, accessed August 4, 2023.

[47] S. H. E. Dijkstra, M. Hinne, E. Segers, & I. Molenaar. "Clustering children's learning behaviour to identify self-regulated learning support needs" (2023) *Computers in Human Behavior*, 145, 107754.

13.3 TOWARD THE DEVELOPMENT OF RESPONSIBLE AI FOR EDUCATION

13.3.1 *Overview of AI Ethics Frameworks*

The mid-2010s saw a surge in AI ethics discussions, spurred by rapid advances in deep learning and growing controversies surrounding the technology's implications. More specifically, the years between 2016 and 2019 have seen the proliferation of AI ethics guidelines issued by technology companies, NGOs, think tanks, international organizations, and research institutions.[48] Jobin et al.[49] analyzed 84 published sets of ethical principles for AI, which they concluded converged on five areas: transparency, justice and fairness, non-maleficence, responsibility, and privacy. Similarly, a comparative analysis by Fjeld et al.[50] identified an emerging normative core comprised of 8 key themes: privacy, accountability, safety and security, transparency and explainability, fairness and nondiscrimination, human control of technology, professional responsibility, and the promotion of human values. While this convergence may be seen as a sign of maturation and a key step for the development of binding rules and laws, a review by Blair Attard-Frost et al.[51] revealed a disproportionate emphasis on principles intended for the governance of algorithms and technologies and little attention to the ethics of business practices and the political economies within which AI technologies are embedded. These latter aspects are of key importance in the context of education, given that the adoption of AI in schools can accelerate the commodification of education and further embed large private tech companies into the provision of public goods.[52]

In recent years the AI ethics discussion gradually moved from the enumeration of key values toward efforts to translate abstract principles into real-world practices. However, this is wrought with several difficulties, and the field is currently exploring various approaches.[53] For example, at the time of writing, the OECD's Policy Observatory catalogues[54] over 500 procedural, educational, and technical tools intended to support trustworthy and responsible AI development. However, there is currently little evidence about this uptake and impact. A 2021 review of AI

48 "AI ethics guidelines global inventory by AlgorithmWatch" (*AI Ethics Guidelines Global Inventory*), https://inventory.algorithmwatch.org, accessed August 4, 2023.

49 Jobin, Ienca, and Vayena (n 6).

50 Jessica Fjeld et al., "Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI" (2020) *SSRN Electronic Journal*, https://papers.ssrn.com/abstract=3518482, accessed August 3, 2023.

51 Blair Attard-Frost et al., "The ethics of AI business practices: a review of 47 AI ethics guidelines" (2023) *AI and Ethics*, 3: 2.

52 Niels Kerssens and José van Dijck, "The platformization of primary education in the Netherlands" (2021) *Learning, Media and Technology*, 46: 250.

53 Jessica Morley et al., "From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices" (2020) *Science and Engineering Ethics*, 26: 2141.

54 "OECD AI policy observatory," https://oecd.ai/fr/, accessed August 4, 2023.

impact assessments and audits concluded that most approaches suffered from a lack of stakeholder participation, failed to utilize the full range of possible techniques and that internal self-assessment methods exhibited scarce external verification or transparency mechanisms.[55]

Finally, in addition to developments in AI ethics, there has been increasing regulatory attention in several jurisdictions, including the EU,[56] the UK,[57] the United States,[58] and China, along with calls for international harmonization. The European Union adopted the world's first comprehensive regulation, the AI Act, in July 2024, which enshrines several previously voluntary ethical principles into law. As a result, schools will need to implement a comprehensive AI governance strategy to adequately deal with transparency, data protection and risk assessment requirements. The law also classifies certain uses of AI in education as high risk, including systems that determine access to educational institutions, determine the appropriate education level for students, evaluate learning outcomes, or monitor students for prohibited behaviour during tests. These use-cases are subject to additional regulatory requirements.[59]

Still, AI represents a uniquely difficult technology for lawmakers to regulate.[60] Given the pace, potential scale, and complexity of AI's societal impacts, ethical frameworks, guidelines, and tools for responsible technology development will likely continue to evolve alongside regulatory efforts.

13.3.2 *Ethics of AI in Education*

When AI is applied in the domain of education, it may substitute, augment, modify, or redefine existing educational practices.[61] Consequently, the ethics of AI in education should not just be based on an ethics of AI, but also based on an ethics of

55 Jacqui Ayling and Adriane Chapman, "Putting AI ethics to work: Are the tools fit for purpose?" (2021) *AI and Ethics*, 2(3): 405.

56 Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts 2021.

57 "A pro-innovation approach to AI regulation" (*GOV.UK*), www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper, accessed August 4, 2023.

58 "Oversight of A.I.: Rules for artificial intelligence" (2023) U.S. Senate Committee on the Judiciary, www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-rules-for-artificial-intelligence, accessed August 4, 2023.

59 Clara Hawking, "The EU AI Act, for Schools" (2024) LinkedIn https://media.licdn.com/dms/document/media/D4D1FAQHbIET4k7CRKA/feedshare-document-pdf-analyzed/0/1721386685072?e=1723680000&v=beta&t=mMr1FocvwoqiptNP-rjrOm5888BbHeZ8fvUAfOVaXBQ, accessed August 2, 2024.

60 Richard Wheeler and Fiona Carroll, "An explainable AI solution: Exploring extended reality as a way to make artificial intelligence more transparent and trustworthy" (2023) *Springer Proceedings in Complexity* 255.

61 Erica R. Hamilton, Joshua M. Rosenberg and Mete Akcaoglu, "The Substitution Augmentation Modification Redefinition (SAMR) Model: A critical review and suggestions for its use" (2016) *TechTrends* 60.

education.[62] Aiken and Epstein set the stage for ethics in AI education back in 2000. They advocated for focusing on human needs rather than letting technology dictate decisions.[63] They considered a multidimensional view of humans, looking at ethical, social, intellectual, and other aspects. This laid the groundwork for today's human-centered AI ethos.[64] Aiken and Epstein's guidelines emphasized positive, adaptive AI that supported diverse learning approaches and cultures, respected teachers and underscored the human role in education. However, principles we commonly see in AI ethics guidelines today, such as transparency, explainability, and avoiding bias, are notably absent, as these emerged later due to the rise of data-driven deep learning systems. Despite the changes in AI technologies, Aiken and Epstein's principles still remind us to prioritize human-centered educational values in AI development and align well with the augmentation perspective on AI in education.

13.3.2.1 Framework of the Institute for Ethical AI in Education

Discussions about the ethics of AI in education were further galvanized in the late 2010s as part of the broader engagement with the risks and opportunities of machine learning systems. The Institute for Ethical AI in Education in the UK developed a framework[65] iteratively based on extensive consultations with stakeholders, including policymakers, academics, philosophers and ethicists, industry experts, educators, and young people. The framework acknowledges the necessity of wider educational reform to ensure that AI can benefit all learners while expressing the hope that AI might help "*combat many of the deep-rooted problems facing education systems and learners themselves: from a narrow and shallow curriculum to entrenched social immobility. AI could allow societies to move away from an outdated assessment system, and it could also enable high-quality, affordable lifelong learning to become universally available.*" (The Institute for Ethical AI in Education, 2021: 4)[66]

The framework recognized the power of public institutions in setting high-quality standards for product development and was therefore intended for those making procurement and application decisions related to AI systems in education. The framework advanced nine overall objectives that AI systems must adhere to: AI in education should focus on achieving well-defined educational goals beneficial to learners, support the assessment of a broader range of talents, and increase organizational capacity while respecting human relationships. It should promote equity and learner autonomy and

[62] Holmes and Tuomi (n 1).

[63] Robert M. Aiken and Richard G. Epstein, "Ethical guidelines for AI in education: Starting a conversation" (2000) *International Journal of Artificial Intelligence in Education*, 11: 163.

[64] Ben Shneiderman, "Human-centered AI" (2022) Human-centered AI 1; Raphael Koster et al., "Human-centred mechanism design with democratic AI" (2022) *Nature Human Behaviour*, 6(10):1398.

[65] "The institute for ethical AI in education" (*University of Buckingham*), www.buckingham.ac.uk/research/research-in-applied-computing/the-institute-for-ethical-ai-in-education/, accessed August 4, 2023.

[66] Ibid.

uphold privacy. Human oversight and accountability must be maintained, educators and learners should understand AI implications, and ethical design principles should be followed by those creating AI resources. Many of these principles track broader values articulated in AI ethics guidelines – such as autonomy, privacy, and transparency – but we also find education-specific values. These high-level objectives were further specified into a list of criteria and a checklist of concrete questions to guide pre-procurement, procurement, implementation, and monitoring and evaluation phases.[67]

13.3.2.2 The UNESCO 2019 Beijing Consensus

At a global level, UNESCO's 2019 Beijing Consensus[68,69] was a major accomplishment toward defining the requirements for the sustainable development of educational AI technologies. The drafting committee consisted of selected members from the electoral world districts who were invited to focus on the 2030 agenda for sustainable development with specific attention to Sustainable Development Goal 4 to ensure high-quality education for all learners. In different sessions, a broad range of topics around AI and education were discussed: from AI for learning to learning in an AI era, as well as societal consequences and labor market impacts. It was also explicitly recognized that demands differ depending on the broader socioeconomic characteristics of member countries.

The Beijing Consensus emphasized the utilization and scaling of intelligent ALTs for foundational skills, such as math and language learning, while highlighting the need for developing unique human competencies such as problem-solving, creativity, and the regulation of learning processes in an AI era. Ensuring teacher professional development and using formative assessment was crucial for effective AI implementation, and governments were encouraged to harness AI for optimizing educational policies and understanding system effectiveness. The consensus accentuates the importance of lifelong learning, AI literacy skills, and inclusivity for all demographics. Ethical considerations were emphasized as important and included equitable and inclusive use of AI, addressing the needs of vulnerable groups such as minorities and students with learning impairments or disabilities. The consensus also highlighted the importance of ensuring gender equity and maintaining auditable transparency in data use. Finally, attention was also placed on evidence-based AI applications and establishing novel regulatory frameworks.

Overall, there was a strong agreement on the human-centred approach to AI in education, whereby teachers were considered the central focus, and AI

67 Ibid.

68 "Beijing consensus on artificial intelligence and education – UNESCO digital library" (n 8); F. Miao and W. Holmes, "Artificial intelligence and education. Guidance for policy-makers" (United Nations Educational, Scientific and Cultural Organization (UNESCO) 2021) Report, https://unesdoc.unesco.org/ark:/48223/pf0000376709, accessed August 4, 2023.

69 Author Inge Molenaar was a member of the expert panel that produced the Beijing Consensus.

in education should always be human-controlled. Following up on the Beijing Consensus, UNESCO issued guidance on AI and education intended for policymakers to support the achievement of SDG4 to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all".[70] The document reaffirmed the principles of the Beijing Consensus. It emphasized that for AI to be best exploited for the common good, it should be used to reimagine teaching and learning rather than just automating often outmoded existing practices. This involves adopting a system-wide vision for AI in education that puts teachers and learners in the center. Importantly, the document recognized that getting AI right in the context of education requires an integrated approach that involves interdisciplinary planning, fostering ethical and inclusive AI use, developing a comprehensive plan for AI in educational management and teaching, conducting pilot testing and evaluations, and encouraging local AI innovations in the field of education.[71]

13.3.2.3 European Commission's Ethical Guidelines on the Use of AI and Data in Teaching and Learning for Educators

The European Union is one of the main actors in the discussion around AI ethics, governance, and regulation. This started with the European Commission's establishment of a High-Level Expert Group on AI in 2018, which drafted a set of Ethics Guidelines with seven key requirements for trustworthy AI: human agency and oversight, transparency, diversity, nondiscrimination and fairness, societal and environmental well-being, privacy and data governance, technical robustness and safety, and accountability.[72] The European Commission's Digital Education Action Plan[73] specifically describes the development of "ethical guidelines on the use of AI and data in teaching and learning for educators" (in priority 1, action 6). To achieve this action, it set up a European Expert Group for this specific purpose, resulting in guidelines on the use of AI and data in teaching and learning for educators, published in 2022.[74,75]

[70] F. Miao and W. Holmes, "Artificial intelligence and education. Guidance for policy-makers" (United Nations Educational, Scientific and Cultural Organization (UNESCO) 2021) Report, https://unesdoc.unesco.org/ark:/48223/pf0000376709, accessed August 4, 2023, page 5. See also Beijing consensus on artificial intelligence and education – UNESCO digital library (n 8).

[71] Miao and Holmes (n 70).

[72] "Ethics guidelines for trustworthy AI | Shaping Europe's digital future" (April 8, 2019), https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, accessed August 4, 2023; Nathalie A. Smuha, "The EU approach to ethics guidelines for trustworthy artificial intelligence" (2019) *Computer Law Review International*, 20: 97.

[73] "Digital education action plan (2021–2027) | European education area," https://education.ec.europa.eu/focus-topics/digital-education/action-plan, accessed August 3, 2023.

[74] "Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators – Publications office of the EU" (n 10).

[75] Authors Inge Molenaar and Duuk Baten were involved as invited members of the European Expert Group.

The guidelines consist of four main elements: a description of the EU policy and regulatory context, examples of AI and data use in education, ethical considerations and requirements, and guidance for teachers and school leaders. The background report specifically mentions how the ethics of AI and the ethics of education are deeply related and highlights the importance of interpreting the ethical dilemmas and challenges of AI in education in the context of educational practices.[76] The guidelines are based on Biesta's three key objectives of education: qualification, socialization, and subjectification.[77] From an ethical perspective, the guidelines focus on four interrelated dimensions of ethics: *human agency*, *fairness*, *humanity*, and *justified choice*. These are seen as guiding the choices around using AI systems in education.[78]

The guidelines also have a strong basis in the requirements set by the European Commission's High-Level Expert Group on AI, and rearticulate the abovementioned seven key requirements for trustworthy AI in the context of education. Using these requirements as a scaffolding, the document offers guiding questions for schools and educators as a starting point for reflection and constructive dialogue among various stakeholders about using AI systems in educational practices. In line with this, the guidelines describe the competencies necessary to successfully implement and use AI systems in education. The existing European Framework for the Digital Competence of Educators (DigCompEdu)[79] provides a basis for developing the integral skills and capacities necessary within the educational system.

A critical note here would be to wonder whether educators and schools are equipped to ask and answer these questions, some of which require extensive technical understanding as well as access to elaborate system documentation; think of "*Are the appropriate oversight mechanisms in place for data collection, storage, processing, minimisation and use?*" and "*Are there monitoring systems in place to prevent overconfidence in or overreliance on the AI system?*"[80] Dealing with these questions around the application of AI in education is not an easy feat and requires extensive collaboration among educators, schools, and public institutions. This is why the guidelines proposed that schools adequately prepare for the effective use of AI and recommended raising awareness around the challenges. Such reflective action will require additional resources to be committed to supporting schools or new organizations to address these issues together with teachers and schools. Hence, NGOs have raised awareness of the need to implement ethical guidelines and have requested national governments to act accordingly. In the next section, we turn to the example of the Netherlands to show how this can be organized at a national level.

[76] Ibid., p.7.
[77] Gert J. J. Biesta, "Good education in an age of measurement: Ethics, politics, democracy." 159.
[78] "Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators – publications office of the EU" (n 10) 18.
[79] Ibid., p. 18
[80] Ibid., pp. 19–21.

13.4 THE DUTCH EXPERIENCE: A NATIONAL AMBITION TOWARD VALUE-DRIVEN EDUCATIONAL INNOVATION

13.4.1 *The Value Compass for the Digital Transformation of Education*

Within the Netherlands, there has been an increasing discussion of the impact of digital technologies on education. This has been symbolized by a call for action in a national newspaper by the rector magnifici of all Dutch public universities,[81] warning about the influence of large tech corporations on the public educational system and calling for the higher education sector to take responsibility for its public values. This can be seen as the start of a national discussion on prioritizing educational values in public education, which was reflected in a subsequent advisory report on public values in education by the Dutch Association of Universities (UNL).[82] Public values are the values that ground and give meaning to our interactions, societal institutions and political structures.[83] Public values are not fixed, but are the result of continuous societal and political processes.[84] And the public interests they represent are of such importance that they need to be safeguarded within the public sector.[85] The underlying thought is that the digitalization of public services needs to be guided by "fundamental public values such as privacy, autonomy, equity and equality."[86]

As technologies become more pervasive in educational institutions, we see, on the one hand, how these technologies start to shape educational practices and, on the other hand, how the dependency on existing software providers can become problematic. AI applications in education only compound those challenges, raising questions about how the roles of students and teachers change and which new responsibilities emerge for educational institutions.[87] To facilitate navigating the influence of digital technologies in education, Kennisnet,[88] SURF,[89] and the Rathenau Institute[90] developed a Value Compass as a reference framework for public values in education (see Figure 13.3).[91]

[81] "Digitalisering bedreigt onze universiteit. Het is tijd om een grens te trekken" (*de Volkskrant*, December 22, 2019), www.volkskrant.nl/columns-opinie/digitalisering-bedreigt-onze-universiteit-het-is-tijd-om-een-grens-te-trekken~bff87dc9/, accessed August 6, 2023.

[82] "Advisory report on public values in education," www.universiteitenvannederland.nl/files/documenten/Advisory%20report%20on%20public%20values%20in%20education_EN_vnov22.pdf, accessed August 4, 2023.

[83] "Value lines" (iHub Radboud University) https://ihub.ru.nl/valuelines/, accessed August 2, 2024.

[84] José Van Dijck, Thomas Poell, and Martijn De Waal, *De Platformsamenleving: Strijd Om Publieke Waarden in Een Online Wereld* (Amsterdam University Press, 2016).

[85] Wat dan weer komt van: WRR (2000). Het borgen van publiek belang. Den Haag: Sdu Uitgevers.

[86] Rinie van Est et al., "Valuable digitalisation: How local government can play the 'technology game' in the public's interest."

[87] John Walker and Duuk Baten, "Promises of AI in education" (Zenodo, 2022), https://zenodo.org/record/6874315, accessed August 4, 2023.

[88] "Laat ict werken voor het onderwijs" (*Kennisnet*, July 4, 2023), www.kennisnet.nl/, accessed August 4, 2023.

[89] "SURF is de ict-coöperatie van onderwijs en onderzoek | SURF.nl," www.surf.nl/, accessed August 4, 2023.

[90] "Onderzoek en debat over de impact van wetenschap, technologie en innovatie op ons leven | Rathenau Instituut," www.rathenau.nl/nl, accessed August 4, 2023.

[91] "Value compass for digital transformation of education," www.surf.nl/files/2022-01/surf-value-compass-english.pdf, accessed August 4, 2023.

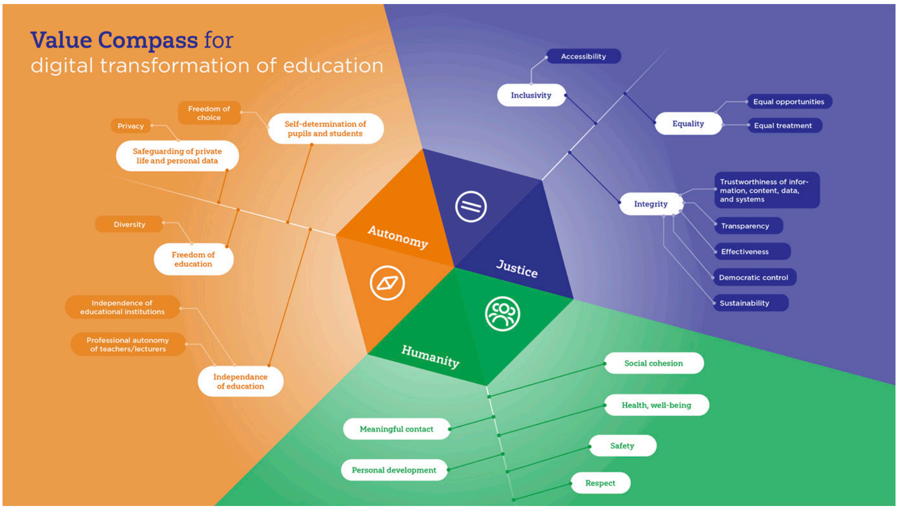

FIGURE 13.3 The value compass[92]

This common language aims to elicit a discussion that transcends functionalities, costs and benefits toward formulating shared ambitions for a future of digital education.

The value compass emphasizes three core values as central to educational practices: autonomy, justice, and humanity. These values are loosely defined as constellations of other values, such as privacy (autonomy), inclusivity (justice), and well-being (humanity). Here autonomy is seen as the ability "to live under your own laws" for diverse educational stakeholders, the students and teachers, and the institutions themselves.[93] Justice is defined as collected values that mainly describe the importance of treating others in terms of themselves and treating them equally in an inclusive manner.[94] The human aspect of education is central to the value of humanity, consisting of the meaningful contact, respect, safety, and well-being necessary to value the unique aspects of each student.[95] The framework was developed through deliberative engagement with sector stakeholders, including a published beta version for public consultation, as can be read in the Rathenau report[96] which describes in more detail how this compass of values was developed. Working from a previous list of seven key themes and questions around the digital transformation of the public sector,[97] the value compass conceptualizes themes such as *power*

92 Ibid., p. 7.
93 Ibid., p. 5.
94 Ibid., p. 6.
95 Ibid., pp. 142–147.
96 "Naar hoogwaardig digitaal onderwijs | Rathenau Instituut," www.rathenau.nl/nl/digitalisering/naar-hoogwaardig-digitaal-onderwijs, accessed August 4, 2023.
97 L. Kool et al., *Urgent Upgrade: Protect Public Values in Our Digitized Society* (Rathenau Instituut, 2017), www.rathenau.nl/en/file/33578/download?token=FA8OpJuY, accessed August 4, 2023.

relations and *control over technology* into values such as autonomy. The three main values of autonomy, justice, and humanity can be seen as equally important, with underlying instrumental values that allow the operationalization of these values.

The Value Compass is not a normative framework for digital transformation but a basis for considering educational values in decision-making.[98] The value compass is used for deliberative workshops, for example, a workshop of SURF with the national student bodies ISO and LSVB on the ethics of using online proctoring in examinations.[99] Here the values in the value compass helped guide the discussion through all relevant perspectives. A more normative approach can be achieved through value hierarchies, where one conceptualizes values into norms and design requirements to guide choices in digital transformation.[100] For example, Kennisnet conceptualized the value of inclusivity for a digital learning system into the norm "accessible for all students," which led to the design requirement of "meets web accessibility requirements." These requirements can then be taken into account within the development or procurement processes. Through a lens of public values, educational institutions can proactively structure the digital transformation by "weighing values," using them as a guide in shaping considerations and priorities.[101] By looking at the design, procurement, and use of new technologies through the lens of these values, the education sector can become an active participant in the digitalization of education.[102]

13.4.2 *The NOLAI as an Approach to Drive Responsible Use of AI in Education*

To finalize this chapter, we introduce the example of the National Education Lab AI (NOLAI)[103] in the Netherlands, which pursues an embedded ethics approach to iteratively develop, prototype, implement, evaluate and scale responsible AI technologies for primary and secondary education.

NOLAI is an innovative research initiative at the Faculty of Social Sciences of Radboud University in the Netherlands in collaboration with several strategic partners[104] and the Dutch Growth Fund.[105] NOLAI's main goal is to develop intelligent

98 "Value compass for digital transformation of education" (n 91). P.41.

99 "Publieke waarden in de praktijk: met het Ethiekkompas in gesprek over online proctoring | SURF Communities" (April 4, 2022), https://communities.surf.nl/publieke-waarden/artikel/publieke-waarden-in-de-praktijk-met-het-ethiekkompas-in-gesprek-over, accessed August 4, 2023.

100 Ibo Van de Poel, "Translating values into design requirements" (2013) Philosophy and engineering: Reflections on practice, principles and process 253.

101 Kennisnet. (2020). "Weighing values: An ethical perspective on digitalisation in education." Kennisnet (Authors: Pijpers, R., Bomas, E., Dondorp, L., and Ligthart, J.) – www.kennisnet.nl/app/uploads/Kennisnet-Waardenwegen-ENG.pdf, accessed August 4, 2023.

102 "Value compass for digital transformation of education" (n 91). P.3.

103 "NOLAI | National Education Lab AI," www.ru.nl/en/nolai, accessed August 4, 2023.

104 www.nolai.nl

105 Ministerie van Economische Zaken en Klimaat, "Home – Nationaal Groeifonds" (June 30, 2023), www.nationaalgroeifonds.nl/, accessed August 4, 2023.

technologies that improve the quality of primary and secondary education. The institute aims to achieve this goal by developing innovative prototypes that use AI and promoting the responsible use of AI in education. This is done in two programs: the co-creation program and the scientific program. The co-creation program develops innovative prototypes and applications of AI in co-creation with schools, scientists, and businesses. The scientific program develops knowledge in five focus areas: teacher professionalization, technology for AI in education, sustainable data, pedagogy & didactics and embedded ethics.

NOLAI's main activities are developing state-of-the-art applications of AI in education and investigating their use NOLAI's activities start with dialogues with schools to explore their needs for using AI to improve education and literature reviews that, map current knowledge. After, NOLAI brings scientists and educational practitioners together to develop AI prototypes, explore current applications of AI technologies and ambitions for the future with businesses. An example of a co-creation project by NOLAI is the visualization of student data collected across different learning management, ALTs, and summative assessment systems. This project is a collaboration between three schools, an ALT company, an assessment company, and pedagogical and AI scientists. The collaborative and interdisciplinary approach ensures the connection between educational practice, science, and business development.

NOLAI stimulates the responsible development of AI in education. NOLAI has a dedicated Data and Privacy Officer who helps the co-creation projects comply with relevant privacy and data protection regulations. Also, all projects need approval from institutional ethical committees that monitor the ethical conduct of the research being conducted. In addition, as there are many open questions about the ethical issues that will emerge throughout the development and implementation of AI in education, NOLAI strives to further the discussions around the responsible use of AI in education in the Netherlands with its embedded ethics approach.

Embedded ethics approaches have most famously been developed in computer science education, where it is referred to as an approach for teaching ethics in computer science curricula and aims to incorporate ethics into the entire engineering process in an integrated, interdisciplinary, and collaborative way.[106] As such, embedded ethics is aptly seen as an ongoing process of anticipating, identifying, and addressing ethical features of technological innovations by helping developers to integrate ethical awareness and critical reasoning in their technical projects, thereby benefitting individuals and society at large.[107]

The embedded ethics approach developed within NOLAI complements such existing approaches with for example, insights from the "ethics parallel research"

[106] Hannah Bleher and Matthias Braun, "Reflections on putting AI ethics into practice: How three AI ethics approaches conceptualize theory and practice" (2023) *Science and Engineering Ethics*, 29.

[107] Daniel W. Tigard and others, "Toward best practices in embedded ethics: Suggestions for interdisciplinary technology development" (2023) *Robotics and Autonomous Systems*, 167: 104467.

approach that has been developed to provide ethical guidance parallel to the development process of emerging biomedical innovations.[108] This last approach has many similarities with the embedded ethics approaches as it can also be characterized by focusing on bottom-up, inductive ethical dilemmas and stimulating ethical reflexivity and awareness. An important difference is that the ethics parallel research approach argues for the inclusion of a wider variety of stakeholders in the deliberation process (beyond engineers and ethicists) and is less focused on the design of technology but also its broader sociopolitical implications.

This means that for the embedded ethics approach within NOLAI, ethicists will closely collaborate with various stakeholders in co-creation projects, including technologists, company representatives, scientists, and educational professionals. As the co-creation projects develop and mature, the ethicists aim to provide ethical support and develop sustainable processes to advance responsible innovation, navigating the "messy" reality of the co-creation projects and the ethical questions, complex dilemmas, and practices that emerge. This means they will advise stakeholders and support them with anticipating, identifying, and addressing moral dilemmas iteratively and continuously.

In addition to ethical literature and theory, ethicists within NOLAI will conduct empirical research within the co-creation projects to inform and advance their ethical support. Through various qualitative research studies, including participant observations, focus groups, and surveys, the ethicists will study the effects and implications of introducing AI systems in education. For example, they will study students' and teachers' moral beliefs, intuitions, and reasoning using AI systems within NOLAI. These findings can help align the AI systems with students' and teachers' needs and wishes. In another study, ethicists will use qualitative research methods to explore value conflicts that emerge when AI systems are introduced in classrooms and how these conflicting values are balanced in practice. The findings help formulate "best practices" regarding implementing AI systems in education.

As a last step, the ethicists within NOLAI will use the insights gained from their participation in the co-creation projects and the results of their empirical studies to inform more abstract ethical debates about AI in education. The diversity and large amounts of co-creation within NOLAI provide an exceptional opportunity to help answer complex ethics questions outlined in this chapter.

13.5 CONCLUSION

In this chapter, we explained that education is a special application domain of AI that optimizes human learning and teaching. The replacement and augmentation perspectives were contrasted, and we emphasized the importance of human

[108] Karin R. Jongsma and Annelien L. Bredenoord, "Ethics Parallel Research: An Approach for (Early) Ethical Guidance of Biomedical Innovation" (2020) *BMC Medical Ethics*, 21.

learners and teachers staying in control over AI. We outlined a variety of AI applications used in education, covering student-faced, teacher-faced, and administrative-oriented systems. AI in education is about carefully designing learning and teaching in a way that technologies augment human learning. As we have recently witnessed the increasing presence of generative AI, developments outside educators' control raise questions and impact the educational system.[109]

Subsequently, we discussed the ethical and social impacts of AI in education. We outlined how ethics in AI and education developed, describing general AI ethics developments, the Beijing consensus based on UNESCO's conference on AI in Education in 2019, and the recent European Commission's ethical guidelines on the use of AI and data in teaching and learning for educators. Finally, we outlined the example of the Netherlands with the Dutch value compass and the embedded ethics approach of NOLAI, as concrete illustrations of how AI ethics can be embedded in the educational context.

One of the central distinguishing features of ethical frameworks for AI in education has been to prioritize decision-making aligned with ethical values and sound pedagogical objectives. This call has been echoed in numerous frameworks ever since Aiken and Epstein[110] first put AI and education on the agenda, and has been reaffirmed by UNESCO's and the EU's guidelines. Efforts to combine pedagogical and didactical values with generic ethical values in a way that ensures a sound approach to ethics in education are still in their infancy. This also requires the understanding and navigation of potential misalignments in interests between stakeholders, including students, parents, teachers, schools, companies, and policymakers.[111] As Selwyn[112] notes, the ethics of AI is not a clear-cut case of solving technological challenges or doing the right thing intuitively but requires an ongoing, morally reflective process.[113]

[109] Walker and Baten (n 87).
[110] Aiken and Epstein (n 63).
[111] Miao and Holmes (n 70); Wayne Holmes and Kaska Porayska-Pomsta, *The Ethics of Artificial Intelligence in Education : Practices, Challenges, and Debates*, www.routledge.com/The-Ethics-of-Artificial-Intelligence-in-Education-Practices-Challenges/Holmes-Porayska-Pomsta/p/book/9780367349721, accessed August 3, 2023.
[112] Neil Selwyn, "AI, education and ethics – starting a conversation."
[113] Holmes and Porayska-Pomsta (n 111).