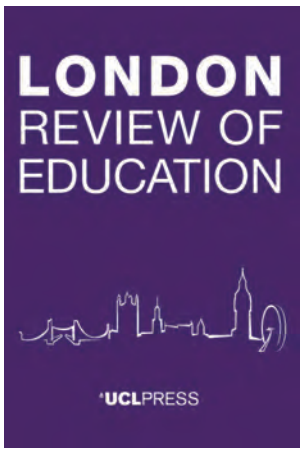

UCLPRESS

LONDON REVIEW OF EDUCATION

e-ISSN: 1474-8479

Journal homepage:
https://www.uclpress.co.uk/pages/london-review-of-education

The role and challenges of education for responsible AI

Virginia Dignum

How to cite this article
Dignum, V. (2021) 'The role and challenges of education for responsible AI'. *London Review of Education*, 19 (1), 1, 1–11. https://doi.org/10.14324/LRE.19.1.01

Submission date: 4 March 2020
Acceptance date: 24 August 2020
Publication date: 13 January 2021

Peer review
This article has been peer-reviewed through the journal's standard double-blind peer review, where both the reviewers and authors are anonymized during review.

LONDON REVIEW OF EDUCATION

Dignum, V. (2021) 'The role and challenges of education for responsible AI'. *London Review of Education*, 19 (1), 1, 1–11. https://doi.org/10.14324/LRE.19.1.01

The role and challenges of education for responsible AI

Virginia Dignum* – *Umeå University, Sweden*

Abstract

Artificial intelligence (AI) is impacting education in many different ways. From virtual assistants for personalized education, to student or teacher tracking systems, the potential benefits of AI for education often come with a discussion of its impact on privacy and well-being. At the same time, the social transformation brought about by AI requires reform of traditional education systems. This article discusses what a responsible, trustworthy vision for AI is and how this relates to and affects education.

Keywords: artificial intelligence, responsible AI, ethics, trustworthy AI

Introduction

Digital transformation is currently seen as the driving force of global, innovative, inclusive and sustainable growth. Education is necessary to prepare people for the opportunities and challenges of globalization and the digital revolution and to ensure that everyone can fully participate in, benefit from and adapt to new occupations and skills needs. In many cases, artificial intelligence (AI) is considered to provide an essential contribution to ensure the inclusive and cohesive societies and resilient economies that are expected to arise from this digital transformation. However, given the wide range of views on the role and possibilities of AI in society, it is not always clear how to bring AI and education together towards these aims. Even at the beginning of the twenty-first century, Aiken and Epstein (2000: 164) stated: 'Unless we seriously discuss our philosophical premises before AI moves in any significant way into the classroom, we will limit the scope, effectiveness and positive contributions that AI can bring to learning.' Understanding the possibilities and limits of AI and ensuring the necessary human skills need to be the focus of education and training programmes at different levels. Nevertheless, and even though the quantity and quality of robust evidence that shows that well-designed AI does work in education is increasing (Luckin and Cukurova, 2019), the field of education often questions the educational value of technology (Selwyn, 2015; Slay *et al.*, 2008).

There is a general feeling that the current educational system is not generating enough experts, and everyday users are increasingly not able to understand the systems with which they are faced. These issues have been extensively studied by others. This article is not about all the potential impacts and uses of AI in education, but it focuses on what a responsible, trustworthy vision for AI is and how this relates to and affects education and learning studies. Even though science, technology, engineering and mathematics (STEM) education is necessary, responsible AI renders the need for education in arts and humanities even more necessary. In a world where machines

*Email: virginia@cs.umu.se

can find (all) answers, it becomes imperative that all people are well trained in asking questions and evaluating answers.

The many faces of AI

Understanding the role of education in the realization of the digital transformation fuelled by AI requires that we first are able to understand what AI is and what makes AI different from other technologies, so specific educational changes are needed. The recent success of AI, and the hype around it, have created a plethora of definitions, ranging from ultra-simplified ones, such as that recently put forward in a White Paper by the European Commission (2020: 2) – 'AI is a collection of technologies that combine data, algorithms and computing power' – to purely magical ones. Depending on the focus and the context, Theodorou and Dignum (2020) have identified the following ways in which AI has been viewed:

1. A (computational) technology that is able to infer patterns and possibly draw conclusions from data. (Currently AI technologies are often based on machine learning and/or neural networking based paradigms.)
2. A next step in the digital transformation. This view brings under the general denominator of AI many different technologies, from robotics to the Internet of Things, and from data analytics to cybersecurity, the result of which is that everything is considered to be AI.
3. A field of scientific research. This is the original reference, and it is still predominant in academia. The field of AI includes the study of theories and methods for adaptability, interaction and autonomy of machines (virtual or embedded).
4. An (autonomous) entity (for example, when one refers to 'an' AI). This is the most usual reference in media and science fiction, endowing AI with all-knowing, all-powerful qualities and bringing with it the (dystopic) view of magical powers and the feeling that AI happens to us without us having any power to control it.

A more informed view describes AI as a software system (possibly embedded in hardware) designed by humans that, given a complex goal, is able to take a decision based on a process of perception, interpretation and reasoning based on data collected about the environment and that meets the properties of:

- *autonomy*, meaning that the system is able to deliberate and act with the intent of reaching some task-specific goal without external control
- *adaptability*, meaning that the system is able to sense its environment and update its behaviour to changes in the environment
- *interactivity*, meaning that the system acts in a physical or digital dimension where people and other systems coexist.

Even though many AI systems currently only exhibit some of these properties, it is their combination that is at the basis of the current interest in, and results of, AI and that fuels the public's fears and expectations (Dignum, 2019). The scientific study of AI includes several approaches and techniques, including not only machine learning (deep learning and reinforcement learning are specific examples), which is currently the approach to AI that is most visible and successful, but also reasoning (for example, planning, scheduling, knowledge representation, search and optimization), multi-agent systems (for example, distributed problem solving and communication) and intelligent robotics (for example, control, perception, sensors and actuators, and the integration of physical/hardware components).

Responsible AI

To understand the societal impact of AI, one needs to realize that AI systems are more than just the sum of their software components. AI systems are fundamentally socio-technical, including the social context in which they are developed, used and acted upon, with its variety of stakeholders, institutions, cultures, norms and spaces. That is, it is fundamental to recognize that, when considering the effects and the governance of AI technology, or the artefact that embeds that technology, the technical component cannot be separated from the socio-technical system (Hendler and Mulvehill, 2016; Dignum, 2019). This system includes people and organizations in many different roles (developer, manufacturer, user, bystander, policymaker and so on), their interactions, and the procedures and processes that organize those interactions. Guidelines, principles and strategies must focus on this socio-technical view of AI. In fact, it is not the AI artefact or application that is ethical, trustworthy or responsible. Rather, it is the people and organizations that create, develop or use these systems that should take responsibility and act in consideration of human values and ethical principles, such that the overall system and its results can be trusted by society. The ethics of AI is not, as some may claim, a way to give machines some kind of 'responsibility' for their actions and decisions and, in the process, discharge people and organizations of their responsibility. On the contrary, AI ethics requires more responsibility and accountability from the people and organizations involved: for the decisions and actions of the AI applications and for their own decision to use AI in a given application context.

The processes by which systems are developed entail a long list of decisions by designers, developers and other stakeholders, many of them of a societal, legal or ethical nature. Typically, many different options and decisions are taken during the design process, and in many cases there is not one clear 'right' choice. These decisions cannot just be left to be made by those who engineer the systems, nor to those who manufacture or use them; they require societal awareness and informed discussion. Determining which decisions an AI system can take, and deciding how to develop such systems, are at the core of a responsible approach to AI. At the very least, responsible AI means that these choices and decisions must be explicitly reported and open to inspection.

The ART (accountability, responsibility, transparency) principles for responsible and trustworthy AI (Dignum, 2019), depicted in Figure 1, are therefore meant for the

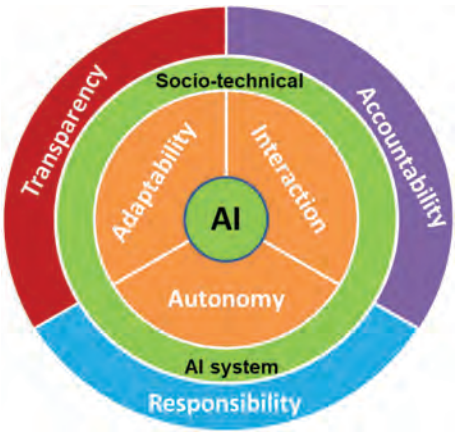

Figure 1: The ART of AI (Source: Dignum, 2019)

whole AI socio-technical system, rather than being focused on the software. In other words, addressing ART will require a socio-technical approach to design, deployment and use of systems, interweaving software solutions with governance and regulation.

The ART principles for responsible AI can be summarized as the following requirements for the socio-technical AI system:

- *Accountability* refers to the necessity for the system to provide account, that is, to explain and justify its decisions to users and other relevant actors. To ensure accountability, decisions should be derivable from, and explained by, the decision-making mechanisms used. It also requires that the moral values and societal norms that inform the purpose of the system, as well as their operational interpretations, have been elicited in an open way involving all stakeholders.
- *Responsibility* refers to the role of people themselves in their relation to AI systems. As the chain of responsibility grows, means are needed to link an AI system's decisions to its input data and to the actions of stakeholders involved in the system's decision. Responsibility is not just about making rules to govern intelligent machines. but also about the role of people and institutions in the effects of developing and using the system.
- *Transparency* indicates the capability to describe, inspect and reproduce the mechanisms through which AI systems make decisions and learn to adapt to their environment and the provenance and dynamics of the data that is used and created by the system. Moreover, trust in the system will improve if we can ensure openness of affairs in all that is related to the system. As such, transparency is also about being explicit and open about choices and decisions concerning data sources, development processes and stakeholders. Stakeholders should also be involved in decisions on all models that use human data, affect human beings or have other morally significant impacts.

Responsible AI and education

Current awareness of ethical issues relating to AI and education are mostly restricted to privacy, security and the appropriate uses of personal data. In addition to these, in settings where there is an increasing availability of virtual assistants supporting students' learning processes or teachers' performance, concerns are growing around the impact of such assistants on the learner (Tuomi, 2018; Popenici and Kerr, 2017). Students may also find it difficult to experience or believe in success when using a learning companion that remembers and reminds the student of their past failures (Luckin *et al.*, 2016). A study reported in Hastie *et al.* (2016) shows worse results with the empathic version of an agent that reminded students of what they had learnt. Wearables and smart assistants keep track of many of our activities and those of our children. A visit to the gym, therapist or supermarket, or a quiet afternoon at home, can produce psychometric, physiological, financial, emotional and social information that could be used to build an affective user model and, as such, potentially improve personalized and appropriate responses (Richards, 2017). However, the sharing of user models that capture an individual's inner thoughts and feelings could potentially impact that individual if revealed to their employer, family, friends or to the wider public. When we consider the context of learning and education, such monitoring could be seen as a way to identify bullying and supporting children to cope (Hall *et al.*, 2009).

Hudlicka (2016) identifies several ethical issues that go beyond the general concerns of data privacy and that are specific to virtual agents. These concerns include affective privacy (the right to keep your thoughts and emotions to yourself), emotion

induction (changing how someone feels) and virtual relationships (where the human enters a relationship with the agent). Concerning student data and student privacy, potential conflicts can arise from the proliferation of systems for continuous evaluation of students. In fact, even if parents want to protect the personal data of their children, they also want access to that data themselves, while children and young people may want privacy from their parents. For example, in the Netherlands, the Magister tracking system used by most secondary schools provides parents with direct access to their children's results and activities at school. Often, the parent will know before the child whether they passed or failed a test. In a study, 70 per cent of parents indicated that they thought that their child was not bothered by the fact that parents receive school information about their child through the tracking system. The parents also thought that their involvement increased by having direct access to information on grades, homework and absence (Boon, 2018). However, concerns over breaching children's right to privacy have been raised in parliament and by national ombudsmen (Kok, 2018). In fact, adolescents need a certain degree of freedom to learn to take responsibility. Schools and parents must allow that space. Dealing with these conflicts of interest, and the protection of the rights of vulnerable people such as children and the mentally disabled, will need carefully designed standards and legislation.

Potential social and ethical issues are raised as educational technology is increasingly endowed with more smart functionalities (Richards and Dignum, 2019). A literature review focusing on the use of (humanoid) robots in the classroom has analysed their ethical impact along four themes: (1) privacy; (2) replacing humans; (3) effects on children; and (4) responsibility (Serholt *et al.*, 2017). At the same time, many fundamental questions about AI – on the nature of intelligence, how to balance individual and collective interests, how to solve ethical dilemmas and how automation will impact the labour market – cannot be answered by technology alone. These questions require interdisciplinary approaches and, as such, a change in the role and content of educational programmes (Dwivedi *et al.*, 2019). AI offers the potential for augmentation and potential replacement of human tasks and activities within a wide range of applications. The current pace of change for AI innovation is high, requiring societal, institutional and technological adjustments, and new opportunities for continued innovation across different domains, including business and management, government, public sector and science and technology. In order to navigate this potential, explore opportunities and mediate challenges, it is essential to integrate humanities and social sciences into the conversation on law, economics, ethics and the impact of AI and digital technology. Only together can we chart a way forward into a beneficial and trustworthy future with our increasingly algorithmic societal systems.

The rapidly growing capabilities and increasing presence of AI-based systems in our lives raise pressing questions about the impact, governance, ethics and accountability of these technologies around the world (Dwivedi *et al.*, 2019). How can decisions be made on when, what for and how AI should be applied? How can the variety of views and requirements of people who use, interact with and are impacted by these technologies be integrated? How do we harness the potential of AI systems, while ensuring that they do not exacerbate existing inequalities and biases, or even create new ones? These questions cannot be answered from a computer science or engineering perspective alone. In fact, we can say that AI is no longer an engineering discipline, but requires a broad involvement of different disciplines and participants. Here is where education and learning studies have an important role to play. The learning sciences field is interdisciplinary and encompasses psychology,

sociology, computer science, education and cognitive science. Bringing learning studies together with AI research and development will increase the understanding of teaching and learning within those who develop AI, which can contribute to better machine-learning techniques and applications. At the same time, such collaborations contribute to the ability of education specialists, teachers and learners to understand and be confident using AI (Luckin and Cukurova, 2019). However, most current AI and robotics curricula worldwide provide engineers with a too-narrow task view. The wide impact of AI on society, as argued in Dignum (2020), requires a broadening of engineering education to include:

1. analysis of the distributed nature of AI applications as these integrate socio-technical systems and the complexity of human–agent interaction
2. reflection on the meaning and global effect of the autonomous, emergent, decentralized, self-organizing character of distributed learning entities and how they operate
3. incremental design and development frameworks, and the unforeseen positive and negative influence of individual decisions at a system level, as well as how these impact human rights, democracy and education
4. the consequences of inclusion and diversity in design, and how these inform processes and results
5. understanding of governance and normative issues, not only in terms of competences and responsibilities, but also in the case of views on health, safety, risks, explanations and accountability
6. the underlying societal, legal and economic models of socio-technical systems.

Broadening AI and engineering curricula is possibly also a way to attract a more diverse student population. When AI curricula are known to be transdisciplinary, it can be expected that female students, who traditionally choose humanities and social sciences subjects over engineering ones, may be motivated to choose AI. In parallel, humanities and social sciences curricula also need to include subjects on the theory and practice of AI. For example, law curricula need to prepare law experts on how to address legal and regulatory issues around AI.

Guidelines and regulatory frameworks for AI

There is increasing recognition that AI should be developed with a focus on the human consequences as well as the economic benefits. As such, most guidelines and recommendations explicitly consider the need for education and training, in particular to ensure the technical skills needed to drive the role of AI in the digital transformation, even though they are particularly short on providing concrete approaches for educational transformation.

Governance is necessary for the reduction of incidents, to ensure trust, and for the long-term stability of society through the use of well-established tools and design practices. Well-designed regulations do not eliminate innovation, as some claim; instead, they can enhance innovation through the development and promotion of both socio-legal and technical means to enforce compliance (Monett and Lewis, 2017). Moreover, policy is needed to ensure human responsibility for the development and deployment of intelligent systems, filling the gap that emerges from the increased automation of decision making. The ultimate aim of regulation is to ensure our – and our societies' – well-being in a sustainable world. That is, research, development and

use of AI should always be done in a responsible way – what is often referred to as *responsible AI*. When developing intelligent systems, besides the obvious necessity to meet legal obligations, societal values and moral considerations must also be taken into account, weighing the respective priorities of values held by different stakeholders in various multicultural contexts. Human responsibility to ensure flourishing and well-being of our societies should always be at the core of any technological development (Floridi *et al.*, 2018).

According to the Organisation for Economic Co-operation and Development (OECD, 2019: n.p.): 'Governments should take steps, including through social dialogue, to ensure a fair transition for workers as AI is deployed, such as through training programmes along the working life, support for those affected by displacement, and access to new opportunities in the labour market.'

The European High Level Expert Group on AI devotes a whole chapter of its recommendations document to skills and education, encouraging member states to increase digital literacy across Europe, to provide data literacy education to government agencies, to ensure a leading role for Europe in fundamental AI research programmes, retaining and attracting world-class researchers and to ensure the upskilling and reskilling of the current workforce (European Commission, n.d.). This means that proficiency in education in STEM subjects can position students to be competitive in the workforce. In the view of the High Level Expert Group:

> Conscious and well-informed children and other individuals will create a solid foundation for responsible and positive uses of AI systems and digital technologies more generally, and strengthen their personal skills on cognitive, social and cultural levels. This will not only increase the available talent pool, but also foster the relevance and quality of research and innovation of AI systems for society as a whole. (European Commission, n.d.)

At the same time, as suggested by the G20, ongoing training in the workplace should be reinforced to help workers adapt (Twomey, 2018). It is not only that a human impact review should be part of the AI development process and that a workplace plan for managing disruption and transitions should be part of the deployment process, but governments should also plan for transition support as jobs disappear or are significantly changed.

Besides the need for increased attention to AI technology skills, several groups recognize the need for introducing ethics into STEM education (Villani, 2018, Taebi *et al.*, 2019). This is aligned with the overall view that ensuring students are prepared for the changing labour market will be the main challenge for education curricula. Curricula should focus on the development of those skills that are likely to remain in demand (sometimes referred to as 'twenty-first-century skills') and thus prioritize teaching critical thinking, problem solving and teamwork across subject areas and at all education levels, from kindergarten to university. Teaching students to become analytical thinkers, problem solvers and good team members will allow them to remain competitive in the job market, even as the nature of work changes.

The role of education is also recognized as a way to build and sustain trust in AI systems, alongside reliability, accountability, processes to monitor and evaluate the integrity of AI systems over time, and the tools and techniques ensuring compliance with norms and standards (Jobin, 2019).

Challenges for education

At all levels and in all domains, businesses and governments are, or will soon be, applying AI solutions to a myriad of products and services. It is fundamental that the general public moves from passively adopting or rejecting technology to being in the forefront of the innovation process, demanding and reflecting on the potential results and reach of AI. The success of AI is therefore no longer a matter of financial profit alone, but of how it connects directly to human well-being. Putting human well-being at the core of development provides not only a sure recipe for innovation, but also both a realistic goal and a concrete means to measure the impact of AI.

The extent to which various technologies are adopted and supported in educational institutions is often politically driven or at least constrained (Selwyn, 2020; Jandrić, 2017). While governments may launch grand-sounding initiatives such as the 'Digital Education Revolution' (Buchanan, 2011), often the aims are to provide all students with computers and schools with access to networks and digital resources and to equip teachers, students and parents to use digital technologies. While laudable, these aims are not sufficient, and they are certainly not concerned with the use of state-of-the-art technology or new pedagogies. Moreover, teachers often do not have permission, interest, capacity or energy to participate in trials and experiments with advanced technology applications in the classroom, even when the materials are carefully aligned to the national curriculum and designed with teachers to cover concepts and topics that students find challenging (Jacobson *et al.*, 2016).

Before committing to a future where AI pervades learning, educationalists and technologists need to guide society and governments to understand the potential social and ethical implications of this technology. Rather than worrying about AI taking over the world (or at least the classroom), as in the fear of the technological singularity warned of by some (Bostrom, 2016), the main concern should be people's readiness to blindly accept the technology as a given, which leads to practical problems, including misuse, overreliance, ignorance or underutilization, abuse and use without concern for the consequences (Parasuraman and Riley, 1997). It is also striking to see that most guidelines and principles that have sprouted in the last couple of years, from governments, policy organizations, social agencies or tech companies, are mostly lacking in concrete proposals for education, even though most recognize that education will play an important role in ensuring trustworthy and responsible AI.

From an education perspective, a pressing question is how to ensure the knowledge and skills to develop and deploy AI systems that align with fundamental human principles and values, and with our legal system, and that serve the common good. As industry, research, the public sector and society in general are increasingly experimenting with, and applying, AI across many different domains, governments and policymakers are looking at AI governance, that is, the means to shape the process of decision making in ways that ensure public safety, social stability and continued innovation. For policy to deliver that goal, it needs to be supported by a solid knowledge not only of the technical aspects of AI, but also of its implications for law, ethics, economy and society. This requires a multidisciplinary approach to the study and development of AI.

The digital transformation of society is possibly the main challenge of this century. By the end of 2013, those who have grown up in a digital world started to outnumber those who had to adapt to it. However, capacity building to ensure that everybody is

able to contribute to the digital ecosystem, and to fully participate in the workforce, is lagging behind, and current education curricula are perhaps not the most suitable to meet the demands of future work.

Considering that, as the media theorist Marshall McLuhan is claimed to have said, 'the tools that we shape will thereafter shape us', the digital ecosystem will bring about a redefinition of fundamental human values, including our current understanding of work and wealth. In order to ensure the skills needed for resilient and sustainable capacity building for the digital ecosystem, the following aspects must be central in education curricula across the world:

- *Collaborate:* The digital ecosystem makes possible, and assumes, collaboration across distance, time, cultures and contexts. The world is indeed a village, and all of us are the inhabitants of this village. Skills are needed to interact, build relationships and show the self-awareness needed to work effectively with others in person and virtually, across cultures.
- *Question:* AI systems are great at finding answers and will do this increasingly well. It is up to us to ask the right questions and to critically evaluate results to contribute to responsible implementation of solutions.
- *Imagine:* Skills to approach problem-solving creatively, using empathy, logic and novel thinking, are needed. For this, it is imperative to nurture humanities education and to include humanities subjects in all technology curricula.
- *Learn to learn:* The ability to adapt and pick up new skills quickly is vital for success, requiring us to continuously learn and grow and to adapt to change. Being able to understand what it is necessary to know, and knowing when to apply a particular concept, as well as knowing how to do it, are key to continuous success.

The digital age is a time for reinvention and creativity. Capacity building must embrace these skills alongside technological expertise. This shows that the traditional separation between humanities, arts, social sciences and STEM is not suitable for the needs of the digital age. More than multidisciplinary, future students need to be transdisciplinary – to be proficient in a variety of intellectual frameworks beyond the disciplinary perspectives. In fact, we have stated that artificial intelligence is not a pure STEM discipline (Dignum, 2019, 2020); it is in essence transdisciplinary and requires a spectrum of capabilities not covered by current education curricula. It is urgent we redesign studies. This will provide a unique opportunity to truly achieve inclusion and diversity across academic fields.

Acknowledgements

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

Notes on the contributor

Virginia Dignum is Professor of Responsible Artificial Intelligence at Umeå University, Sweden and Scientific Director of WASP-HS, the Wallenberg Program on Humanities and Society for AI, Autonomous Systems and Software. She is a Fellow of the European Artificial Intelligence Association, member of several policy advice groups including the European High Level Expert Group on AI, and author of *Responsible Artificial Intelligence: Developing and using AI in a responsible way* (Springer, 2019).

Declarations and conflict of interests

The author declares no conflict of interest with this work.

References

Aiken, R.M. and Epstein, R.G. (2000) 'Ethical guidelines for AI in education: Starting a conversation'. *International Journal of Artificial Intelligence in Education*, 11, 163–76. Accessed 25 November 2020. www.researchgate.net/publication/228600407_Ethical_guidelines_for_AI_in_education_Starting_a_conversation.

Boon, M. (2018) 'Online volgsystemen handig maar niet zaligmakend'. *Ouders en Onderwijs*, 5 May. Accessed 25 November 2020. www.oudersonderwijs.nl/nieuws/online-volgsystemen-handig-maar-niet-zaligmakend/.

Bostrom, N. (2016) *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press.

Buchanan, R. (2011) 'Paradox, promise and public pedagogy: Implications of the federal government's digital education revolution'. *Australian Journal of Teacher Education*, 36 (2), 67–78. https://doi.org/10.14221/ajte.2011v36n2.6.

Dignum, V. (2019) *Responsible Artificial Intelligence: How to develop and use AI in a responsible way*. Cham: Springer.

Dignum, V. (2020) 'AI is multidisciplinary'. *AI Matters*, 5 (4), 18–21. https://doi.org/10.1145/3375637.3375644.

Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... and Williams, M.D. (2019) 'Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy'. *International Journal of Information Management*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

European Commission (n.d.) 'Ethics guidelines for trustworthy AI'. Accessed 25 November 2020. https://ec.europa.eu/futurium/en/node/6945.

European Commission (2020) *White Paper on Artificial Intelligence – a European approach to excellence and trust*. Accessed 25 November 2020. https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-european-approach-excellence-and-trust.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R. ... and Vayena, E. (2018) 'AI4People – an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations'. *Minds and Machines*, 28 (4), 689–707. https://doi.org/10.1007/s11023-018-9482-5.

Hall, L., Jones, S., Paiva, A. and Aylett, R. (2009) 'FearNot! Providing children with strategies to cope with bullying'. *Proceedings of the 8th International Conference on Interaction Design and Children (IDC '09)*. New York: Association for Computing Machinery, 276–7. https://doi.org/10.1145/1551788.1551854.

Hastie, H., Lim, M., Janarthanam, S., Deshmukh, A., Aylett, R., Foster, M. and Hall, L. (2016) 'I remember you! Interaction with memory for an empathic virtual robotic tutor'. In J. Thangarajah, K. Tuyls, S. Marsella and C. Jonker (eds), *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016)*. Singapore, 9–13 May, 931–9.

Hendler, J. and Mulvehill, A.M. (2016) *Social Machines: The coming collision of artificial intelligence, social networking, and humanity*. New York: Apress.

Hudlicka, E. (2016) 'Virtual affective agents and therapeutic games'. In D.D. Luxton (ed.), *Artificial Intelligence in Behavioral and Mental Health Care*. London: Elsevier, 81–115.

Jacobson, M.J., Taylor, C.E. and Richards, D. (2016) 'Computational scientific inquiry with virtual worlds and agent-based models: New ways of doing science to learn science'. *Interactive Learning Environments*, 24 (8), 2080–108. https://doi.org/10.1080/10494820.2015.1079723.

Jandrić, P. (2017) *Learning in the Age of Digital Reason*. Rotterdam: Sense.

Jobin, A., Ienca, M. and Vayena, E. (2019) 'The global landscape of AI ethics guidelines'. *Nature Machine Intelligence*, 1, 389–99. https://doi.org/10.1038/s42256-019-0088-2.

Kok, L. (2018) 'D66 wil schoolkinderen meer privacy geven'. *Algemeen Dagblad*, 12 February. Accessed 25 November 2020. www.ad.nl/politiek/d66-wil-schoolkinderen-meer-privacy-geven~aa794f5b/.

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), 2824–38. https://doi.org/10.1111/bjet.12861.

Luckin, R., Holmes, W., Griffiths, M. and Forcier, L.B. (2016) *Intelligence Unleashed: An argument for AI in education*. London: Pearson. https://doi.org/10.13140/RG.2.1.3616.8562.

Monett, D. and Lewis, C.W.P. (2017) 'Getting clarity by defining artificial intelligence: A survey'. In V.C. Müller (ed.), *Philosophy and Theory of Artificial Intelligence*. Berlin: Springer, 212–14.

OECD (Organisation for Economic Co-operation and Development). (2019) *Recommendation of the Council on Artificial Intelligence*, 22 May. Accessed 25 November 2020. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

Parasuraman, R. and Riley, V. (1997) 'Humans and automation: Use, misuse, disuse, abuse'. *Human Factors*, 39 (2), 230–53. https://doi.org/10.1518/001872097778543886.

Popenici, S.A. and Kerr, S. (2017) 'Exploring the impact of artificial intelligence on teaching and learning in higher education'. *Research and Practice in Technology Enhanced Learning*, 12 (1), 22. https://doi.org/10.1186/s41039-017-0062-8.

Richards, D. (2017) 'Intimately intelligent virtual agents: Knowing the human beyond sensory input'. *Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents*, 39–40. https://doi.org/10.1145/3139491.3139505.

Richards, D. and Dignum, V. (2019) 'Supporting and challenging learners through pedagogical agents: Addressing ethical issues through designing for values'. *British Journal of Educational Technology*, 50 (6), 2885–901. https://doi.org/10.1111/bjet.12863.

Selwyn, N. (2015) 'Technology and education – Why it's crucial to be critical'. In S. Bulfin, N.F. Johnson and C. Bigum (eds), *Critical Perspectives on Technology and Education.* New York: Palgrave Macmillan, 245–55.

Selwyn, N. (2020) *Telling Tales on Technology: Qualitative studies of technology and education.* London: Routledge.

Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S. and Paiva, A. (2017) 'The case of classroom robots: Teachers' deliberations on the ethical tensions'. *AI & Society*, 32 (4), 613–31. https://doi.org/10.1007/s00146-016-0667-2.

Slay, H., Siebørger, I. and Hodgkinson-Williams, C. (2008) 'Interactive whiteboards: Real beauty or just lipstick?'. *Computers & Education*, 51 (3), 1321–41. https://doi.org/10.1016/j.compedu.2007.12.006.

Taebi, B., van den Hoven, J. and Bird, S.J. (2019) 'The importance of ethics in modern universities of technology'. *Science and Engineering Ethics*, 25, 1625–32. https://doi.org/10.1007/s11948-019-00164-6.

Theodorou, A. and Dignum, V. (2020) 'Towards ethical and socio-legal governance in AI'. *Nature Machine Intelligence*, 2, 10–12. https://doi.org/10.1038/s42256-019-0136-y.

Tuomi, I. (2018) *The Impact of Artificial Intelligence on Learning, Teaching, and Education*. JRC Science for Policy Report, European Joint Research Centre. Accessed 25 November 2020. https://publications.jrc.ec.europa.eu/repository/bitstream/JRC113226/jrc113226_jrcb4_the_impact_of_artificial_intelligence_on_learning_final_2.pdf.

Twomey, P. (2018) *The Future of Work and Education for the Digital Age*. Centre for International Governance Innovation (CIGI). G20 Argentina 2018. Accessed 25 November 2020. www.g20-insights.org/wp-content/uploads/2018/07/TF1-1-11-Policy-Briefs_T20ARG_Towards-a-G20-Framework-For-Artificial-Intelligence-in-the-Workplace.pdf.

Villani, C. (2018) *For a Meaningful Artificial Intelligence: Towards a French and European strategy.* AI for Humanity. Accessed 25 November 2020. www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf.