sciendo

REM - Research on Education and Media
Vol. 11, N. 1, Year 2019
ISSN: 2037-0830 – DOI: 10.2478/rem-2019-0013

A "psychopathic" Artificial Intelligence: the possible risks of a deviating AI in Education

Margot Zanetti [a], Giulia Iseppi [b], Francesco Peluso Cassese [c1]

[a] *Università Niccolò Cusano, Italy, margot.zanetti@unicusano.it, ORCID: 0000-0002-9702-9840*
[b] *Independent Researcher, United Kingdom, giseppi.research@gmail.com*
[c] *Università Niccolò Cusano, Italy, francesco.peluso@unicusano.it*

Abstract

This work analyses the use of artificial intelligence in education from an interdisciplinary point of view. New studies demonstrated that an AI can "deviate" and become potentially malicious, due to programmers' biases, corrupted feeds or purposeful actions. Knowing the pervasive use of artificial intelligence systems, including in the educational environment, it seemed necessary to investigate when and how an AI in education could deviate. We started with an investigation of AI and the risks it poses, wondering if they could be applied also to educative AI. We then reviewed the increasing literature that deals with the use of technology in the classroom, and the criticism about it, referring to specific use cases. Finally, as a result, the authors formulate questions and suggestions for further research, to bridge conceptual gaps underlined by lack of research.

Keywords: *technology; artificial intelligence; education; risks*

Introduction

Artificial intelligence (AI) has invaded our life – as demonstrated by over increasing academic, business and public interest on the topic (for a layered approach on the topic, see Shoham et al., 2018), allowing us to achieve impressive results such as monitor complex systems, predict weather changes and improve knowledge of illnesses. Nevertheless, AI has also been demonstrated to be faulty, discriminatory and limited. Our interest was caught by the announcement that in 2018 the Massachusetts Institute of Technology (MIT), researchers such as Yanardag, Cebrian and Rahwan created the first psychopathic AI, Norman. To do so, Yanardag and colleagues fed the AI with truculent data obtained from Reddit, a famous American social news aggregation. Then, Norman underwent a Rorschach test and scored as a "human psychopath".

The aim of MIT scientists' work was to underline the fact that an AI could "deviate" not only due to programmers' biases expressed in the algorithms but also because of the feed given to the machine.

Technology's role in education is ever increasing (Almohammadi et al., 2017), as is human–computer interaction which introduces new methods to stimulate students' attention, leading to improved lessons, for example through the use of multimodal, affective, attentive and perceptual user interfaces (Bevilacqua et al., 2009). Thanks to innovations based on neural networks, machine learning, deep learning and data mining, artificial intelligence has offered the hope of overcoming the majority of obstacles in education, such as one-to-one tutoring, lifelong learning and support for students with special needs.

But what if these technologies could deviate, in a way similar to the one that MIT researchers induced? Have current risks of these tools, both technically and psychologically, been analysed?

Following this idea, basic and general information about AI and some related risks are given, as well as a review of the most common devices and tools used in education.

1. Artificial Intelligence

[1] Introduction and conclusion by the authors; paragraphs 1; 4 by Giulia Iseppi and paragraphs 2; 3; 5 by Margot Zanetti. Supervision by Prof. Francesco Peluso Cassese.

The concept of an intelligence able to replicate human reasoning and thinking initially tantalised many philosophers (e.g. Descartes but also Hobbes and Leibniz). They saw the possibility of an "intelligent machine" as a neutral, positive, achievement. When – with the Industrial Revolution – the mechanisation of previously exclusively human activities pervaded everyday life, the earlier attempts to demonstrate the innocence of perceived "artificial intelligence" failed. The first act of hacking performed by the Luddites (revolutionary group of English textile workers protesting violently against machines) demonstrates the resistance to new technology that could take their place (Baggaley, 2010).

With the twentieth century and the rise of computers, engineering met programming and studies on AI flourished. In 1936, Alan Turing proved that it was possible to design a machine that, with a limited set of operations, was theoretically capable of carrying out any calculation (Turing, 1937). This meant, in practice, that it was then possible to instruct a machine to emulate (and so create) another computer. The idea opened the way to early attempts to create intelligent programs with the evolving symbol manipulation languages (Rajan and Saffiotti, 2017). In 1943, McCulloch and Pitts proved that it was possible to create artificial neurons, modelling the "intelligent" machine on brains. The turning point for AI, however, remains the paper entitled "Computer Machinery and Intelligence", published by Turing not in a science journal, but in *Mind*, a philosophy journal in 1950. The article presented the first version of the so-called "Imitation Game", the rules for an assessment of an intelligent computer conversing like a human.

AI is a field of study which indicates an intelligent agent (software) based on different technologies. Amongst others, McDowell et al. (2018) identify:

- voice recognition
- natural language processing
- machine learning
- analytical and predictive statistics,
- deep learning
- neural networking,
- cognitive computing.

All software is based on algorithms: lists of instructions used to solve a problem. The instructions to solve the simple problem "how do you make a cup of tea?" is an algorithm, presented in natural language (a language used naturally by humans), as compared to programming languages, used by computers. Since AI is an intelligent software, it is based on algorithms.

Many categorisation attempts describe the technologies differently and enumerate "most used" types of AI. These are in reality strictly interrelated – machine learning can be seen as a subfield of AI, and sometimes they sustain each other (see, among others, Schmidhuber, 2015). IBM Watson, the famous "Cognitive Computing" system, is based on machine and deep learning (IBM, 2019).

The concept of "a program as a recipe for creating a new computer" (p. 39) used by Urwin (2017) is a powerful image that translates the complexity of AI into simplest reality: AI is a program, a recipe to emulate human thinking. We can also make use of the working definition offered by Kaplan and Haenlein (2018), "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation".

AI in education is, then, a system's ability to interpret data inputted by the teacher or the student, learn from such data and use those learnings to achieve specific teaching goals and tasks through flexible adaptation to the environment.

We noticed that in non-technical publications and in newspapers, we frequently found reference to the "most used" techniques for AI: machine learning, deep learning and neural networks. We tried to identify which, if any, was the main technology for AI in the classroom and if we could find data to demonstrate such statements. We performed a simple content analysis (Neuendorf, 2016) of documents, through online and offline research (using journals, market research publications, newspapers and books) of technology labelled as "artificial intelligence" and available to the public. We excluded professional consultancy services for education and created a database of resources, prototypes and theoretical designs. At this stage, we studied the documents to identify the nature of the technology which powers the service.

We discovered that producers of AI solutions for education focus their attention on the more familiar technique of machine learning while they seem to ignore neural networks or deep learning. More interestingly, the majority of them avoid references to specific techniques, preferring instead to declare that their service is "powered by AI and/or data science". Open-source technologies are few and far between. Considering the enthusiasm for AI and the benefits for teacher and students, we find the lack of clarity on algorithms and impact evaluation disconcerting.

In line with the current results, we define machine learning as a group of techniques that discover relations in raw data to make predictions, recognise patterns and more specifically and apply new relations to new situations.

2. The use of technology in education

Technology has overcome unexpected limits and helped to reach new goals. But it has to be admitted that if we think about a classroom today or 50 years ago, few things have changed. Something is different in superior education where new methods, which provide for students' and workers' difficulties, have been improved. The era of the Internet has brought many innovations: e-learning, u-learning, b-learning, and m-learning, to cite some examples (Kostecka &

Sczewz, 2017). But more was asked: human–computer interaction studies have shown that for a good learning experience, it was not enough to be able to study from a screen whenever the student had the time to do it. It is now highlighted that a more personalised method is needed and that the teacher's presence is desirable for better, more active learning. Spitzer, for example, states that e-learning methods have already failed because "learning by doing" is very difficult, professors have almost no control over what students really learn and do, and the emotional part, fundamental for good and lasting learning, is missing (Spitzer, 2013).

The already cited studies of human–computer interaction (HCI) tried to develop solutions to the problems written above, such as the control over the students, by creating new interfaces:

- Perceptual User Interface (PUI): this allows the observation of users' explicit and implicit behaviour, so as to deduce their intentions;
- Attentive User Interface (AUI): to deduce and manage users' attention;
- Affective User Interface (AFUI): to analyse the emotional status of the user and adapt to it.

With these new interfaces, it is possible to investigate the emotional reactions of learners, through posture or facial expression analysis. If the system notices signs of confusion, boredom or loss of attention, it provides an appropriate incentive (Bevilacqua, Capuano et al., 2009). Of course, this is not easily applied to all the online learners, but thanks to these resources and studies, new online courses are structured to stimulate attention and emotions in the learner.

3. Artificial Intelligence in Education (AIEd)

AI has spread rapidly in recent years and started covering lots of aspects of our daily life. But has something really changed in school? Mostly, the answer is negative. AI, above all in the United States, has interesting proposals to help face new teaching and learning challenges. Several studies showed, for example, that one of the most effective teaching methods is one-on-one tutoring (Bloom, 1984), but this fact did not imply that it had been possible to implement this aspect in schools. AIEd found different ways to overcome this problem by proposing tools that are flexible, personalised, inclusive and effective. It is hoped not only to make the AI do some of the tasks expected of teachers but also to face all the limits of current education, integrating a pedagogical model (represented by the effective approaches to teaching). But where is AIEd now, not including educational data mining?

The most cited and probably used AI instruments are provided by private companies or universities and are (Luckin, 2016) as follows:

- Personal tutors for every learner;
- Intelligent support for collaborative learning;
- Intelligent virtual reality.

Intelligent Tutoring System (ITS) simulates one-to-one human tutoring which respects the cognitive needs of the learner: it gives appropriate activities and feedback and in some cases leaves the control of the learning to the learner himself, to help them develop self-regulation skills. The majority of ITS use machine learning mechanisms and neural networks, but they do not require the teacher to have specialist training. However, it would be better that the person who uses them has a knowledge of the tool they are using (Dermeval et al., 2018). Some of the most used ITS products have been developed by the Carnegie Learning Company, whose major aim is to improve math education. It offers a Middle School Math Solution (MATHia, grades 6–12) and a High School Math Solution (MATHiaU). The company affirms that the proposed tools continually adjust to each student, making sophisticated pedagogical decisions and delivering a personalised learning path with ongoing formative assessments. The solution should deliver differentiated learning experiences that support students who encounter difficulties while incite those who are ready for more. The company also provides a software platform designed for the study of computer science: Zulama. Another, different, product to pursue improvements of math learning is Thinkster Math that offers the support of expert tutors and coaches along with a math tutor app.

Intelligent support for collaborative learning helps to overcome the usual difficulties that students have with interacting and working together. The support can be of different types: it can be a tool to form groups thanks to analyses made on individuals, or a facilitator that can understand when students are having trouble in understanding some concepts. There is also the possibility to provide intelligent virtual agents (e.g. McLaren, 2010). One of the most used platforms of this kind is, for example, Brainly, which makes students from around the world interact.

With an intelligent virtual reality, it is possible to recreate different kinds of environments to simulate some aspects of the real world and let people immerse themselves in it. This is an "intelligent" virtual reality: one in which is possible to interact with the elements in specific ways. This can be helpful in the study or in-depth analysis of many subjects. Another interesting use is to simulate everyday life situations to guide the learner in understanding different ways to behave (think of bullying situations, e.g. Vannini, 2011).

Finally, there is one last tool that deserves mention: content creation. These applications, for example those offered by the Content Technology Inc., may break down and disseminate textbooks' content in smaller and smarter study guides that include mini-tests, summaries and flashcards. Teachers might also compose the curricula they prefer, inserting videos, self-assessments, etc.

It is obvious that all these tools are filled and fed with a large amount of data. They are fundamental to the machine itself and possess lots of details of how the students' minds work (Almohammadi el al., 2017). This information must be protected and probably continuously supervised to maintain the correctness of the contents.

4. General Risks

This paragraph does not aim to be a comprehensive list of risks for and created by AI. A more appropriate location for an in-depth analysis of the topic would be a risk assessment paper.

We define risk as the impact that the exploitation of a vulnerability of an asset by a threat can create (Gritzalis et al., 2018). An asset is something of value, whereas a vulnerability is a flaw, weakness or exposure point of the asset. Finally, a threat is anything that can alter, delete or in any way diminish the value of the asset.

Initially, research focused on the risk of malfunction, or the achievement of erroneous results. Errors were made by programmers in the design of new solutions, due to the infancy of the discipline and sometimes to lack of funding. One famous case is that of the Logic Theory Machine (Newell & Shaw, 1957) that excluded the very algorithms useful to solve the problem at hand, or the results presented in the ALPAC report (Automatic Language Processing Advisory Committee, 1966) that took natural language research off the table for years.

Thankfully, AI researchers can be light-hearted and learn continuously from their mistakes, as an article on natural stupidity by McDermott demonstrates (1976). The author lists some failures of AI research (and his thinking) in his time because "if we can't criticize ourselves, someone else will save us the trouble" (p. 4). The show of crippled ideas starts with the use of "wishful mnemonics", a legacy of traditional programming where functions refer to their purpose rather than the correct AI sectorial question – or better a property – that loops attempt to answer. So a function that tries to operate a "FETCH & TRY-NEX" acquires a bigger and bolder meaning of "GOAL". A dog's goal is not to fetch a stick and try next the stick. Labels enriched of meanings are repeated in the program, causing mistakes and inheriting the lack of clarity of human language. The Author underlines the risk of creating an internally consistent system known only to initiated, which produces inexact results.

Among others, another inconsistency of AI programming is the use of "Unnatural Language": language enriched by additional meaning derived by the role the chosen concept plays in the context. In a family with children, PARENT-1 and PARENT-2 will convey all the information about them that the programmer identifies (also that they are "father" and "mother" if they are of different sexes). But this is not pre-programmed in the meaning of the PARENT-1, it is the programmer interpretation of it. Similar to this fallacy is the so-called "natural language fallacy", where we tend to connect those extrinsic and not defined information against logic: "the left arm of my chair" is one object, but can be represented like that "the arm", "that arm", it, and so on. It remains, nonetheless, one object and not many. If language is used improperly – meaning simplistically – in databases, then it will form fallacious structures that transfer the wrong message. Human language is, in fact, focused on brevity, leaving to the hearer to reorganise concepts and think more to compensate the concise message. Computers work in a completely different way, and the lack of understanding of subliminal or enriched vocabulary can constitute a serious problem. The Author continues explaining the risks of using human language structures indiscriminately, suggesting a savvy perspective to reduce the risks inherent to naming (and calling out even Linguists).

Another general risk of AI solutions is that of biased results. A bias is a prejudiced result which is caused by erroneous assumptions. The history of AI is dotted with cases of programs who transfer their "imperfections" to algorithms, unconsciously or consciously according to different authors (Friedman & Nissenbaum, 1996; Koliska & Diakopoulos, 2018; Osoba & Welser, 2017 and the previous examples offered by McDermott).

Recent famous cases include Amazon Alexa's discrimination and research on public tools to evaluate recidivism (Larson et al., 2016). The risk of biased decisions originating from biased databases has initiated a movement to guarantee the cleanliness of datasets "fed" to AI. This is the case of Partnership for AI (created by Apple, Amazon, DeepMind, Google, Facebook, IBM and Microsoft) and the use for research purposes of "clean" datasets like MSCOCO.

Biased results can also be caused by the user. Google's pre-compilation mechanism in the search bar can link innocent individuals with crimes, as users kept googling the wrong name associated with the event (Cheung, 2015). AI reproduces language and bias that users and developers take with them, so if this contains racist or discriminatory slurs or concepts, the algorithms will repeat or make them more evident (Caliskan et al., 2017). Hopefully, the research on limitation of effects of biases will continue according to new interpretations (Amini et al., 2019).

Literature on existential risk is also famous (see Bostrom amongst others, 2014). The risk that AI will take over humans is considered real when referred to Artificial General Intelligence (AGI). AGI attempts to create thinking machines, generalistic systems comparable to humans (Goertzel, 2014). Research agrees that we are not there yet. However, literature investigates the possibility of laying ethical or legal rules to reduce specific risks of AI Armageddon (Michie, 1973; Sherer, 2015) even if we are still working on Weak Artificial Intelligence, AI technology that can execute one task only, or solve specific problems (Lu et al., 2018). We already have robots and technologies that recognise emotions (McStay, 2018), and the World Economic Forum considers the possibility of "affective computing" that can create radicalisation of various types.

What has been discussed so far falls into the definition of "artificial intelligence hazard" created by Bostrom (2011), or "*computer-related risks in which the threat would derive primarily from the cognitive sophistication of the program rather than the specific properties of any actuators to which the system initially has access*".

If the possible impact of risks created unintentionally is relevant, possible intentional production of malevolent AI could have devastating effects. In the interesting paper by Pistono and Yampolskiy (2016), the authors list possible risks that a faulty AI can create, and the impact and possible security (and societal measures) to be implemented.

AI is under attack, too. Not only by scholars but also research has demonstrated that AI can be hacked.

Adversarial attacks, for example, manipulate the content of the feed of data on which the algorithm is trained to deceive the AI into doing something wrong. One example is that of Sharif and his research team, that in 2016 managed to confound intelligent CCTVs for face recognition thanks to special glasses, or the case of Google's trained algorithm that could not distinguish between a tortoise and a gun (Athalye et. al., 2018).

We think the identified risks could be extended to AI used in education: AI could influence the student and incite a deviant behaviour, or be a vehicle for bullying and questionable content or cause radicalisation.

5. Problems posed by the use of AIEd

As it is possible to read from all the recent and enthusiastic papers about AIEd, the hopes for the future are really high, above all those applied to STEM (science, technology, engineering, mathematics) education (Chaudhri et al., 2013). Researchers are really excited and convinced of the possibility to have one-to-one tutoring, and to be able to overcome all the learning difficulties of people mostly excluded from a normal class. One example is the European project MaTHiSIS (http://mathisis-project.eu/), now at the end of the third experimentation. One of its aims is to overcome the difficulties of people with serious learning problems, due to pathologies or mental disabilities. This platform, with its tools, can analyse facial expressions and students' reaction, similar to some technologies cited above. Moreover, as AI is born as something interdisciplinary, it is integrating studies and models from neuroscience, psychology and pedagogy to become more effective (Luckin et al., 2016). So, it is really surprising that very little literature can be found about the possible risks of its use. The most cited risks deal with the problems linked to privacy factors and not to the possible questions about the effects on a person's mental and behavioural development. The enthusiasm is amazing even when reading the ambiguous results given by general meta-analysis on the effects of the use of technology in education (Vivanet, 2014).

Far from wanting to neglect and not appreciate these innovations, it is believed necessary to pose some of the problems that the use of AI in education could trigger, from different points of view.

- How can the use of AI in education influence brain development? Millennia of evolution have shaped the brain to learn, to acquire and process information in specific ways and through precise steps. In using technology, some of these steps are jumped over to arrive directly to a more efficient and lasting learning. But how can the shortening of these passages affect the shaping of the mind, brain and behaviour?
- If the use of technology influences development, thought and behaviour, what could happen if an AI "deviates"? And what if AI is used by already deviated, even unconsciously, people?
- Can learning algorithms and content created by industries be considered always reliable? Should they not be controlled and supervised by an ethics committee? Considering that there is the concrete risk of a diffusion of fake news or of conditioning political opinions, a major supervision of the technology used to transmit culture, and that enables a person to shape their beliefs, should require more and more attention.

These might seem not scientifically required questions but there are already experts that point out issues near to the ones cited (Riek & Howard, 2014). It should also be highlighted that the Global Risk Report 2019 deals extensively with problems that will be brought by technology: from the pervasive control that surveillance systems will have, to the emotional disruption that an AI which answers to human emotion could bring: *But the adverse consequences, either accidental or intentional, of emotionally "intelligent" code could be profound. [...] To help mitigate these risks, research into potential direct and indirect impacts of these technologies could be encouraged. Mandatory standards could be introduced, placing ethical limits on research and development. Developers could be required to provide individuals with "opt-out" and greater education about potential risks—both for people working in this field and for the general population—would also help* (World Economic Forum, Global Risk Report, 2019).

Conclusion

This article faced the issue of AIEd from an interdisciplinary point of view. The starting point was the work of MIT, the creation of a "psychopath" AI. The possible application of researchers' discoveries was more extended than previously examined and technological progress is going fastest in private industries, making it almost impossible for academic research to follow. AI has pervaded our everyday life and every possible consequence of its use should be inspected. If it is also used in our education systems, especially by children, it would be opportune to ask whether these technologies can "deviate" and whether this could influence the mental and behavioural development of students. The hope is that researchers can deepen the studies into AIEd and the possible influences that the use of them can have, deviating or not, on pupils.

In the future, we would like to investigate the effective use of AI by teachers in class, and create metrics to identify improvements and critical issues. We believe that different types of AI in education should be classified and characteristics and metrics should be identified for each of them. This will allow the researchers to clarify the phenomenon even further, and perform a risk assessment of the technology.

References

Almohammadi, K., Hagras, H., Alghazzawi, D., Aldabbagh, G. (2017). A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms. *Journal of Artificial Intelligence and Soft-Computing Research*, 7(1), 47-64.

Amini, A., Soleimany, A., Schwarting, W., Bhatia, S., Rus, D., (2019). Uncovering and Mitigating Algorithmic Bias through Learned Latent Structure, *Conference on Artificial Intelligence, Ethics and Society.*

Athalye, A, Engstrom, L., Ilyas, A., Kwok, K., (2018). Synthesizing Robust Adversarial Examples, *Proceedings of the 35 th International Conference on Machine Learning*, Stockholm, Sweden, PMLR 80.

Automatic Language Processing Advisory Committee. (1966). *Languages and machines: computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. Washington, D.C.: National Academy of Sciences, National Research Council, 1966. (Publication 1416.)

Baggaley, J. (2010). The Luddite Revolt continues. *Distance Education,* 31(3), 337–343.

Bevilacqua, L., Capuano, N., Ceccarini, F., Corvino, F. (2009). Interfacce Utente Avanzate per l'e-learning, *Journal of E-Learning and Knowledge Society*, 5(3), 95-104.

Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-To-One Tutoring, *Educational Researcher*, 13(6), 4-16.

Bostrom, N. (2011). Information Hazards: A Typology of Potential Harms From Knowledge. *Review of Contemporary Philosophy*, 10, 44-79.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press.

Caliskan, A., Bryson, J.J., Narayanan, A.(2017). Semantics derived automatically from language corpora contain human-like biases, *Science*, 183-186.

Chaudhri, V. K., Gunning, D., H. Lane, H. C., Roschelle, J. (2013). Intelligent Learning Technologies: Applications of Artificial Intelligence to Contemporary and Emerging Educational Challenges, (Introduction to the Special Articles in the Fall and Winter Issues), *AI Magazine*, 10-12.

Cheung, Anne S. Y. (2015). Defaming by Suggestion: Searching for Search Engine Liability in the Autocomplete Era. in *Comparative perspectives on the fundamental freedom of expression*, (ed. András Koltay).

Dermeval, D., Paiva, R., Bittencourt, I., Vassileva, J., Borges, D. (2018). Authoring Tools for Designing Intelligent Tutoring Systems: a Systematic Review of the Literature. *International Journal of Artificial Intelligence in Education*, 28(3), 336- 384.

Friedman, B., Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems* (TOIS), 14 (3), 330- 347.

Goertzel, B. (2014). Artificial General Intelligence: Concept, State of the Art, and Future Prospects, *Journal of Artificial General Intelligence,* 5(1), 1-46.

Gritzalis, D., Iseppi, G., Mylonas, A., & Stavrou, V. (2018). Exiting the Risk Assessment Maze: A Meta-Survey. *ACM Computing Surveys (CSUR)*, *51*(1), 11.

IBM. (2019). *IBM Watson*, About. [Retrieved 10/04/2019] https://www.ibm.com/watson/about/index.html.

Kaplan, A., Haenlein, M. (2018). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence, *Business Horizons*, (62)1, 15-25.

Koliska, M., Diakopoulos, N. (2018). Disclose, Decode and Demystify: An Empirical Guide to Algorithmic Transparency. *The Routledge Handbook of Developments in Digital Journalism Studies*. Eds. Scott Eldridge II and Bob Franklin.

Kostecka- Szewc, A. (2017). Nuove tecnologie-nuove sfide alla didattica. *Annales Universitatis Paedagogicae Cracoviensis*, 9(3), 158- 166.

Larson, J., Mattu,S., Kirchner, L., Angwin, J. (2016). *How We Analyzed the COMPAS Recidivism Algorithm*, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

Luckin, Rose; Holmes, Wayne; Griffiths, Mark and Forcier, Laurie B. (2016). *Intelligence Unleashed: An argument for AI in Education*. Pearson Education, London.

Lu, H., Li, Y., Chen, M., Kim, H., & Serikawa, S., (2018). Brain intelligence: go beyond artificial intelligence. *Mobile Networks and Applications*, *23*(2), 368-375.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115-133.

McDermott, D. (1976). Artificial intelligence meets natural stupidity. *ACM SIGART Bulletin*, (57), 4-9.

McDowell Marinchak, C. L., Forrest, E., & Hoanca, B. (2018). The Impact of Artificial Intelligence and Virtual Personal Assistants on Marketing. In M. Khosrow-Pour, D.B.A. (Ed.), *Encyclopedia of Information Science and Technology*, Fourth Edition (pp. 5748-5756). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-2255-3.ch499.

McLaren, B. M., Scheuer, O., & Mikšátko, J. (2010). Supporting Collaborative Learning and e-Discussions Using Artificial Intelligence Techniques. *International Journal of Artificial Intelligence in Education*. 20(1), 1–46.

McStay, A. (2018). *Emotional AI: The Rise of Emphatic Media*. London: Sage Publication.

Michie, D. (1973). Machines and the theory of intelligence, *Nature*, 241(23.02.1973), 507-512.

Minsky, M.L., Papert,S. (1969). *Perceptrons: an introduction to computational geometry*, Cambridge, Mass.: MIT Press.

Neuendorf, K. A. (2016). *The content analysis guidebook*. Sage. Thousand Oaks.

Newell, A & C., Shaw, J. (1957). Programming the logic theory machine. *Western Computing Proceedings*, 128.

Osoba, O. A., Welser, W. (2017). *The Risks of Artificial Intelligence to Security and the Future of Work.* Santa Monica, CA: RAND Corporation. [Retrieved 10/04/2019] https://www.rand.org/pubs/perspectives/PE237.html.

Pistono, F., & Yampolskiy, R. V. (2016). *Unethical research: how to create a malevolent artificial intelligence*. arXiv preprint arXiv:1605.02817.

Rajan, K., Saffiotti, A. (2017). Towards a science of integrated AI and Robotics. *Artificial Intelligence*. 1-9.

Riek, L.D., Howard, D. (2014). A Code of Ethics for the Human-Robot Interaction Profession. *We Robot*. 1-10.

Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harv. JL & Tech.*, *29*, 353.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 5-117.

Shoham,Y., Perrault, R., Brynjolfsson, E., Clark, J., Manyika, J., Niebles, J., C., Lyons, T., Etchemendy, J., Grosz, B., Bauer, Z. (2018). *The AI Index 2018 Annual Report*. AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA.

Spitzer, M. (2013). *Demenza digitale. Come la nuova tecnologia ci rende stupidi*. Milano: Corbaccio.

Sharif, M., Bhagavatula, S., Reiter, M.K., Bauer, L., (2016). *Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition*, CCS'16 October 24-28, 2016, Vienna, Austria.

Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London mathematical society*, 2(1), 230-265.

Urwin, R. (2017). *Artificial Intelligence - The quest for the ultimate thinking machine*. Arcturus Publishing Limited, London.

Vannini, N., Enz, S., Sapouna, M., Wolke, D., Watson, S., Woods, S., Aylett, R. (2011). "FearNot!": A Computer-Based Anti-Bullying-Programme Designed to Foster Peer Intervention. *European Journal of Psychology of Education*.
26(1), 21-44.

Vivanet, G. (2014). Sull'efficacia delle tecnologie nella scuola: analisi critica delle evidenze empiriche. *TD Tecnologie Didattiche*, 22(2), 95-100.

Turing, A. M. 1950. Computer Machinery and Intelligence. *Mind: A quarterly Review of Psychology and Philosophy*, 433-460.

World Economic Forum, *Global Risk Report*, 2019. [Retrieved 10/04/2019] https://www.weforum.org/reports/the-global-risks-report-2019.