Educating For Artificial Intelligence: Critical Thinking, Responsibility And Resistance

Giulia Mauti [1]*

[1] University of Cassino and Southern Lazio, Italy
* giulia.mauti@unicas.it

Abstract: Within the epistemological framework of the symposium dedicated to the theme of Artificial Intelligence (AI), this contribution reflects on human thought as the true subject and protagonist of the ongoing revolution (Malavasi), capable of achieving a conscious and reflective use of the resources generated by technological progress. In the face of the digital revolution, which has modified the perception of presence and position through the re-ontologisation of modernism (Floridi), educating for A.I. implies an in-depth reflection on meditating thought (Heidegger), capable of discernment and thoughtful choices in the use and management of technological devices. Such critical thinking (Dewey) is closely connected to the exercise of responsibility and judgement (Arendt; Jonas; Spina), as well as the education to resist (Houdé) cognitive automatisms (bias). The time of onlife requires educational work to promote critical thinking, capable of dealing with the invisibility of data and algorithms, in order to achieve their correct interpretation, through the processes of awareness, trustworthiness, and explainability (Rivoltella, Panciroli). Within the domain of a pedagogy of human development (Malavasi), the alliance between man and A.I. implements the need to appeal to one's conscience in a critical manner and to question it constantly, in order to address issues concerning ethics and responsibility, such as privacy, security and fairness (Di Tore), in order to ensure an ecologically sustainable future for mankind (Malavasi).

Keywords: A.I. education, critical thinking, responsibility, resilience, integral ecology.

1. Human thought and the digital revolution: between presence and position

An increasing number of people are living onlife and in the infosphere, both digitally and analogically. The ongoing digital revolution has also affected the way reality is conceived and understood. As Floridi (2022, p. 27) points out, there is an unprecedented disconnection between presence and position: today a person can be physically in a café and interactively present on a Facebook page (ibid.). Position and presence seem to have changed their constitutive essence as attributes of the same human situation. What has just been described is just one of the many changes taking place with the advent of the digital era, which is also redefining, from an epistemological point of view, modern mentality, and therefore many established ideas and conceptions. This process, defined by Floridi as the re-ontologisation of modernism (ibid., p. 31), is well conveyed by this analogy, which the Author himself proposes:

"Let us imagine two people, A and H. They are married and really want to make their relationship work. A, who is always doing more at home, is inflexible, stubborn, intolerant of mistakes and reluctant to change. H, on the other hand, is exactly the opposite, but is also becoming progressively lazier and more dependent on A. The

result is an unbalanced situation in which A ends up shaping the relationship and distorting H's behaviour, practically, if not intentionally. If the relationship works, it is because it is carefully tailored around A. The relationship becomes interpretable in terms of the Hegelian slave-master dialectic. Now, smart technologies play the role of A in the above analogy, while their human users are clearly H" (ibid., pp. 61-62).

This suggests that, faced with the future of artificial intelligence, pedagogical reflection plays a crucial role in protecting the human. As Malavasi observes, it has a twofold mission: on the one hand, to promote awareness of the technological potential, and the benefits and dangers related to the use of tools and devices; on the other hand, to focus attention on the protagonist of today's rapid changes, namely human thought (Malavasi, 2019a, p. 82). "The authentic ethical-educational challenge underlying artificial intelligence is a reform of human thought, which makes the most of the resources generated by technological progress, in order to increase social and economic cohesion by defeating absolute poverty" (Ibid.). Also according to Agrusti, "in order to make the potential of new automation technologies, such as A.I., useful to society as a whole and not just circumscribe them to purely self-referential purposes for a specific elite, it is [...] necessary to return to philosophical thinking, placing the transformation of the world we are living under ontological and ontogenetic investigation" (Agrusti, 2023, p. 26). From this perspective, educating for A.I. therefore implies an in-depth reflection on the meditating thought (Heidegger, 1983, pp. 30-31) of Heideggerian memory, capable of discernment and thoughtful choices in the use and management of technological devices.

2. Critical thinking

Man seems to have acquired a "new" position in the world and with respect to the (physical and digital) world: this seems to confirm the Heideggerian hypothesis of a possible absence of thought, as a "disturbing guest that creeps in everywhere" (ibid., p. 29). However, in the same Author's reflection, it should be pointed out that we can by no means relinquish the ability to think, since it constitutes the essence of human beings (ibid.). Just as a highway on which nothing grows can never become uncultivated and, similarly, man can only become deaf because he has the ability to hear, in the same way, we can only surrender to poverty of thought because we possess the ability to think; the human being is determined to think (ibid.).

By following Heidegger's speculation (and, in this regard, Malavasi's [2019a, pp. 46-49]), we note that the products of technology have shaped man without his knowledge and so strongly that he is now dependent on them (Heidegger, 1983, p. 37). Nevertheless, Heidegger says, we can "say yes to the unavoidable use of technological devices and at the same time [...] say no to them, insofar as we do not permit them to gain the upper hand [...], to distort, embroil and devastate [...] [the] being" (ibid., p. 38). In this way, the relationship with the world of technology will become simple and safe. This attitude, which says both yes and no to the world of technology, is defined by Heidegger as "abandonment in the face of things" (ibid.). The latter, combined with openness to the mystery, allows us to dwell in the world in a different way, promising a new ground upon which we can stand and endure without being imperiled (ibid., p. 39). Abandonment in the face of things and openness to the mystery "can only arise from incessant and passionate thought" (ibid., p. 40).

This means that the production and use of technology require another way of relating to them, which passes through the "path of meditating thought". This per-

tains to the kind of thinking that does not follow a single path unilaterally, but opens up to multiple points of view.

Even Dewey identifies the best way of thinking in ongoing reflective thought (Dewey, 2019, p. 3). With reference to education for artificial intelligence, starting from these premises, "it becomes - as Malavasi says - increasingly important to be aware of what [pertains to the digital universe], and educate to critical thinking on technosciences, which use data on a massive scale" (Malavasi, 2019a, p. XIV). In *How We Think* (2019, p. 13), Dewey gives the example of a man who, travelling along an unfamiliar road, comes to a crossroads. In the absence of sure knowledge, he has two alternatives: continue arbitrarily, trusting in the fortunate outcome of his attempt, or experiment to assess the best path to follow. Among these trials is that of climbing a tree, in order to get a different perspective on the path to follow. With reference to the reflection on A.I., placed within a horizon imbued with educational intentionality, we certainly cannot opt for the first choice. Therefore, we metaphorically climb a tree, "that is, strive to find a standpoint from which we may survey additional facts and, getting a more commanding view of the situation, may decide how the facts stand related to one another" (ibid.). This is what Rivoltella suggests when, while reflecting on the arrival of robots and spaces and the possibilities that remain for critical thinking in such a scenario, he insists on the need for a metacritical work of thought, capable of "addressing the suggestions that devices [now able to learn from user searches, to know their habits, to suggest choices to be made and to predict the behaviour of things and others] [...] leave with regard to data" (Rivoltella, 2019, p. 53). This appears to be possible through educational itineraries aimed at educating to critical thinking: "it is thoughts that inhabit us and that, once formulated, shape the forms of our very being in the same way as steps that, treading on a piece of land, come to form a path, to the point of defining its entire course" (Tota, 2023, p. 231). Hence the need for a commitment to thinking well, which is achieved, according to Morin, by "practising thinking that relentlessly strives to contextualise and globalise its information and knowledge, that relentlessly struggles against error and self-deception" (Morin, 2000, p. 62).

3. Responsibility

Such critical and meditative thinking is closely connected to the exercise of responsibility and judgement (Arendt, 2004; Jonas, 2009; Spina, 2023). According to Panciroli and Rivoltella, "the form of digital and social media calls for critical sense to be accompanied by responsibility, because the spectator also becomes an author and when placing his contents in public space must be able to take responsibility for the consequences that may ensue" (Panciroli, Rivoltella, 2023, p. 99). Malavasi also emphasises that "human intentions, choices and actions are responsible for regulating the functioning and scope of technological tools and devices" (Malavasi, 2019a, p. XIV). In questioning the human faculty of judgement, when faced with events for which there are no precedents (such as the digital revolution, for example), Arendt (2004, p. 23) insists on the need to judge rationally, without being conditioned by emotions, prejudices and personal interests. "Thinking always means examining and questioning" (ibid., p. 88). The philosopher also warns of the widespread danger of not wanting to judge at all (ibid., p. 126) and maintains that the capacity for judgement is the manifestation of the wind of thought (ibid., p. 163). It is then necessary for this capacity for judgement to be imbued with responsibility, to be understood, as Spina

(2023, p. 140) warns, in its qualification of acting (rather than action), since responsibility always calls into question the human being as subject, who participates actively and personally in decision-making processes. According to Jonas, it goes without saying that "the humanum has similarities with the sphere of responsibility" (2009, p. 124): in his reflection on the ethics for the technological civilisation, he recalls the need to learn respect in order to protect man from the vagaries of his own power. The constant commitment to respect for the human is, according to the author, the prerequisite for the future integrity of human identity. In this regard, promoting and supporting the development of a responsible conscience in those who interface in various capacities with intelligent machines and humanoid robots is an unavoidable pedagogical imperative. It is a question of "educating for individual and community intra- and inter-generational discernment in the complexity of social needs and emerging technologies" (Malavasi, 2019a, p. XVII), encouraging a culture of responsibility and care for the human being. "A pedagogy of artificial intelligence - Malavasi observes - requires the responsibility of acting as care for the person and the common home, which is the planet, and takes the form of design interpretation, discernment and thoughtful choice" (ibid., p. 32).

4. Resistance

Closely related to critical thinking and responsibility is what Houdé calls education to resist cognitive automatisms, otherwise known as bias. [This term] "refers to the tendency of individuals to allow themselves to be conditioned by habitual patterns of understanding or prejudices, in the evaluation of their experiences, in solving problems or in making decisions" (Houdé, 2023, p. 20 [N.d.T.]). As Rivoltella points out, "the automatisms that are produced in the brain in response to the stimuli coming from digital screens [represent] one of the most pervasive and effective sources of the development of those simplifications and stereotypes that also function as biases in the child's relationship with the world and with others" (Rivoltella, 2023, p. 13). According to Houdé, teaching to resist, that is, implementing educational processes based on the inhibition of these biases, becomes fundamental in order to facilitate a conscious use of technologies, and create the conditions in which respect may animate reciprocal relationships in online environments. Antonietti also recognises the need to train the capacity for inhibition for the development of critical thinking, which is, in his opinion, closely related to the ability to carefully consider contextual information, to examine its reliability, evaluate the various options and, subsequently, verify the accuracy of the reasoning (Antonietti, 2024, pp. II-V). By providing for the creation of awareness, education for inhibition allows us to guard our noological postures and prevent stereotypes and false beliefs, which are nurtured in digital environments and social media. Rivoltella argues that learning to resist our instinctive responses means controlling ourselves and inspiring our behaviour with reflexiveness. Learning to resist our stereotypes and beliefs implies refraining from classifying others, while valuing their diversity, and making ourselves available to listen and change (Rivoltella, 2023, p. 15). With the acquisition of the ability to resist understood in this way, critical consciousness is achieved.

Morin likewise emphasises the need for an apprenticeship to life capable of developing an awareness of the need for lucidity and understanding (as well as all human aptitudes) in order to live a "real life" (Morin, 2000, p. 53). According to the author, the initiation to lucidity, which incorporates the apprenticeship to self-observation, "is

in itself inseparable from an initiation to the omnipresence of the problem of error" (ibid., p. 50). In this regard, educating to resist implies promoting the individual's capacity to become aware of reality, to observe himself and observe the interlocutor with whom he relates, in order to analyse their characteristics critically (and in a "lucid" manner) and make consciously matured and thoughtful choices. This appears to be of crucial importance with reference to the reflection concerning the relationship with A.I. and data literacy processes, insofar as both A.I. and data qualify as privileged interlocutors of the man as "spectator-author" (Panciroli, Rivoltella, 2023, p. 13), called upon to act in the platform society (Colombo, 2020).

5. Educating to Artificial Intelligence for an ecologically sustainable future

By being aware of the principles linked to the ecology of action (when an action is taken, actions and retroactions come into play, which may divert it from its ends and generate a result contrary to the one initially expected; the ultimate consequences of an action are unpredictable [Morin, 2000, p. 62]), it is necessary to acknowledge that the time of onlife (Rivoltella, Panciroli, 2023, p. 71) requires the work of education to promote critical thinking, in addition to the instances of responsibility and resistance. These elements appear to be able to cope with the invisibility of the data and the algorithm, in order to achieve their correct interpretation, through the processes of awareness, trustworthiness, explainability (ibid., p. 81). A critical discussion of the data (awareness) is needed to test its trustworthiness (trustworthy) and explanation (explainability). As D'Ignazio and Bhargava (2015) argue, this recognises the crucial role of a critical and proactive data literacy that takes into account the fact that users are often unaware of the data being collected, they are unaware of the complexity of the collection techniques and they do not play an active role in this regard.

It is crucial that everyone who interacts with A.I. have knowledge of the ethical issues and responsibility associated with its use, such as privacy, security and fairness. In general, according to Di Tore (2023, p. 477), we need to have a combination of technical and transversal skills to be able to interact effectively with A.I., especially in learning processes. It is a matter of promoting the development of the skills, the outcome of a process of the ideal combination of elements/resources sifted through metacognitive reflection, which according to Birbes (2020) are necessary "in order to make innovation work" (ibid., p. 227) (including technological innovation). According to the author, in view of the human capacity for transformative action, it is possible to inhabit the world "in a better way, by creating new value, reconciling tensions, collaborating with imagination, intellectual curiosity, perseverance and self-discipline" (ibid., p. 232). In order to address the complexity of reality (including the digital revolution) and achieve a conscious exercise of mental attitudes, the conditio sine qua non is, also according to Morin (2000, p. 109), a reform of thought. In this perspective, within the territory of a pedagogy of human development (Riva, 2019, p. XI), the alliance between man and A.I. once again implements the need to appeal to our conscience in a critical manner and to question it constantly, in order to ensure an ecologically sustainable future for mankind. Such a future is to be understood as part of a humanism of life (Malavasi, 2019b, p. 81), capable of cultivating the resources of A.I., without overlooking the risks and uncertainties associated with its use. According to Nida-Rümelin and Weidenfeld, it is a question of moving towards a digital humanism, capable of taking the form of an ethics for the age of A.I. (2019, p. 21): this digital humanism "recommends a coherent and well thought-out use of the potential

of digital technologies in order to improve the protection of the life and health of all parties involved [...] but at the same time it warns against the inhuman consequences of an optimising calculation that compensates for human lives with human lives, human lives with health, the health of some with the health of others, individual rights with individual rights" (ibid., p. 91).

Educating to A.I. implies considering the world of onlife through the critical lens of the complexity that characterises it, by recognising its constituent elements, relationships and underlying logics, in order to realise the care of the human (Malavasi, 2019b, p. 79). In the light of the issues that have emerged, the educational task appears to be, among others, that of promoting a habitus of thought, capable of inspiring ethical action while interacting with technological devices: an action characteristic of a critical, responsible and resistant conscience.

References

AA.VV. (2019). *Arrivano i robot. Riflessioni sull'intelligenza artificiale*. Milano: EDUCatt.

Agrusti, F. (2023). L'*AI literacy* per un'educazione attenta agli algoritmi. In F. Agrusti (Ed.). *Educazione e Intelligenza artificiale* (pp. 19-32). Roma: RomaTre-Press.

Antonietti, A. (2024). Pensiero critico e capacità di inibizione. *La Vita Scolastica*, 35 (79), II-V.

Arendt, H. (2003). *Responsibility and Judgment*. New York: Schocken Books (trad. it. Responsabilità e giudizio, Einaudi, Torino, 2004).

Birbes, C. (2020). La competenza: valore e risorsa della persona. *Annali online della Didattica e della Formazione Docente*, 20 (12), 222-233.

Colombo, F. (2020). *Ecologia dei media. Manifesto per una comunicazione gentile*. Milano: Vita e Pensiero.

D'Ignazio, C., Bhargava, R. (2015). Approaches to Building Big Data Literacy. *Bloomberg Data for Good Exchange Conference*, 28-Sep-2015, New York City, NY, USA (https://www.media.mit.edu/publications/approaches-to-building-big-data-literacy/).

Dewey, J. (1986). *How we think*. Southern Illinois University: Board of Trustees (trad. it. Come pensiamo, Raffaello Cortina, Milano, 2019).

Di Tore, A. (2023). Intelligenza Artificiale e processi educativi secondo l'Intelligenza Artificiale. *Q-Times*, 1 (2), 464-481.

Floridi, L. (2022). *Etica dell'intelligenza artificiale sviluppi opportunità e sfide*. Milano: Raffaello Cortina.

Heidegger, M. (1959) *Gelassenheit*. Pfullingen: Verlag Günther Neske (trad. it. *L'abbandono*, il melangolo, Genova, 1983).

Houdé, O. (2019). *Apprendre à résister*. Le Pommier/Humensis: Paris (trad. it. *Imparare a resistere. Educazione al pensiero* [P.C. Rivoltella Ed.], Scholé, Brescia 2023).

Jonas, H. (1979). *Das Prinzip Verantwortung*. Frankfurt am Main: Insel Verlag (trad. it. Il principio responsabilità. Un'etica per la civiltà tecnologica, Einaudi, Torino, 2009).

EU R *EDIZIONI UNIVERSITARIE ROMANE*
editoria scientifica dal 1952

López López, M.C., Pizzi, F., Russo, P. (Eds.). *The frontiers of inclusive education, Theory and practice in Italy and Spain*. Lecce: Pensa Multimedia.

Malavasi, P. (2019a). *Educare robot? Pedagogia dell'intelligenza artificiale*. Milano: Vita e Pensiero.

Malavasi, P. (2019b). Educare robot per un umanesimo della vita. In AA.VV., *Arrivano i robot. Riflessioni sull'intelligenza artificiale* (pp. 79-82). Milano: EDUCatt.

Morin, E. (1999). *La tête bien faite*. Paris: Seuil (trad. it. La testa ben fatta. Riforma dell'insegnamento e riforma del pensiero, Raffaello Cortina, Milano, 2000).

Nida-Rümelin, J., Weidenfeld, N. (2018). Digitaler Humanismus. Eine Ethik für das Zeitalter der Künstlichen Intelligenz. München/Berlin: Piper Verlag GmbH (trad. it. *Umanesimo digitale. Un'etica per l'epoca dell'intelligenza artificiale*, FrancoAngeli, Milano, 2019).

Panciroli, C., Rivoltella, P.C. (2023). *Pedagogia algoritmica. Per una riflessione educativa sull'intelligenza artificiale*. Brescia: Scholè.

Riva, M.G. (2019). Prefazione. In P. Malavasi (2019). *Educare robot? Pedagogia dell'intelligenza artificiale* (pp. IX-XI). Milano: Vita e Pensiero.

Rivoltella, P.C. (2019). L'educazione al tempo degli algoritmi. In AA.VV., *Arrivano i robot. Riflessioni sull'intelligenza artificiale* (pp. 51-54). Milano: EDUCatt.

Rivoltella, P.C. (2023). Pensare contro se stessi. In O. Houdé. *Imparare a resistere. Educazione al pensiero* (P.C. Rivoltella Ed.) (pp. 6-15). Brescia: Scholé.

Spina, C. (2023). "Acting like Humans". For a Pedagogy of Ethical and Inclusive Responsibility. In M.C. López López, F. Pizzi, P. Russo (Eds.), *The frontiers of inclusive education, Theory and practice in Italy and Spain* (pp. 137-156). Lecce: Pensa Multimedia.

Tota, A.L. (2023). *Ecologia del pensiero. Conversazioni con una mente inquinata*. Torino: Einaudi.