

## AI in the discourse of the relationships between technology and education

José Luis Rodríguez Illera<sup>1\*</sup>

<sup>1</sup> Universitat de Barcelona, Spain, jlrodriguez@ateneu.ub.edu, <https://orcid.org/0000-0001-6754-6479>

### ABSTRACT

The article reviews some of the relationships between AI and education, emphasizing the metaphors used, the difficulties in finding points of agreement, as well as aspects of the social criticism that is made of AI (e.g. considering that it can be a form of unwanted deviation). AI appears as one more case of technology that comes to improve education, as happened in the past with other less “intelligent” technologies, although now with more arguments in its favor. Secondly, it explores the logical mechanisms (abduction) that are widely used in human reasoning but also in AI, showing how their uncontrolled use can lead to certain conversation algorithms being able to lie, something that teachers would not do. An ethical rather than logical question is raised that deserves to be explored later.

**KEYWORDS:** Artificial intelligence, education, abduction

### 1 INTRODUCTION

Recent advances in Artificial Intelligence (hereinafter, AI) have generated a vast amount of opinions, books, articles and reports in the media. Rightly so, since they have paved the way for brand-new applications, creating mixed feelings of admiration and fear in the face of startling achievements. Some of these applications seem to have made the promise and dream of AI come true, despite the fact that they remain algorithms; more complex, perhaps, but no less predefined.

Historically and educationally speaking, AI belongs to a long series of moments in which the arrival of a new technology – such as radio, television, video, computing or telecommunications in their day – has generated great hope about how it holds the promise of solving most educational problems (for a critique of techno-educational optimism, see Cuban, 1986). Attempts have been made to differentiate many new technologies from their predecessors by highlighting a specific feature through which the technology in question would “really” provide a solution or improvement for teaching, assessment or learning (for a partial review, see Weller, 2020; and Watters, 2021). Education is too complex a field for this to occur in such a straightforward manner; indeed, the reality is that almost all technologies have only left a moderate footprint.

Many of the reviews and opinions on the use of AI in education are situated within a particular framework of thinking, even if implicitly, taking as given a specific type of approach. However, over the decades, particularly in the 20th century, there have been multiple frameworks of thinking which, due to their differences, lead us to positions that are far removed from one another. As such, for some it boils down to a question of futurology (AI surpassing and organising humans), while for others it is all about economics (the massive loss of jobs), and for still others it is an exclusively ethical matter (machines and their limits), or, to a certain extent, humanistic or even epistemological (the type of knowledge used by

AI). Our goal here is to review some of the ideas underpinning the multitude of opinions on AI. We will attempt to organise these discourses, albeit on the understanding that no such organisational endeavour can be neutral, that it always entails assuming a point of view and classifying the organised material in a particular way.

### 2 DISTANT APPROACHES

Unlike AI programmes themselves, which consist of the execution of a set of algorithms, the way in which AI is discussed always involves one or more metaphors. We might say that they are “discursive formations” that convey different ways of seeing AI; that is, ideological positions, reasoning or simple expressions articulated with varying degrees of success. Although these points of view are extremely varied – as has occurred in the past regarding the relationships between technology and education –, within this wide array they can be grouped into a small set of metaphors that contemplate AI at a more general level of abstraction.

2.1. According to Nardi and O'Day (1999), we can discuss metaphors as a general approach to a particular technology, and not only as lexical and semantic metaphors. Accordingly, these authors highlight the metaphor of thinking about technologies as tools to achieve a goal. Tools enhance human abilities, transform our potentialities, or even create new ones. This is undoubtedly one way of seeing technologies and, particularly in the case of education, an extremely common one, dominant in theoretical approaches with Vygotskian roots, or in applied approaches such as didactics. Most of what is published about AI uses this metaphor when singing its praises: improvements in dialogue and reflection on any given curricular topic or piece of creative writing, but also a seemingly endless list of disciplines, including academic writing, painting, page layout design, medical diagnosis, economics, weather forecasting, or basic and advanced programming. AI is a

tool for help and improvement: in the workplace, in education, and in the social and personal spheres.

As far as the educational sector is concerned, to this approach we must add the one conceived by Salomon et al. (Salomon 1991; Salomon et al. 1992) when considering the consequences for human cognition of the use of technologies, in particular interactive, participatory and smart technologies, understood as symbolic tools of mediation in the Vygotskian manner. These authors made a classic distinction, which can be retranslated a little to clarify it: on the one hand, the consequences or effects of the use of these technologies while they are being used for a specific purpose or activity (effects with technology), such as adding colour to a photograph, writing a text, calculating the sum of a long series of numbers, or any of the thousands of activities that can be performed with a computer; and, on the other hand, the long-term effects of their continued use (effects of technology), for which they coined the term “cognitive residue”. However, the distinction does not only encompass a difference in time-span, since a few years later the authors incorporated a third type of effect in the distinction (effects through technology), in order to address the current situation, in which the use of technologies is a constant feature of our daily lives.

This framework of thinking has been one of the most influential in terms of how we consider the use of technologies in education (for others, see Rodríguez Illera, 2023). Among other aspects, it leads us to question the “solely positive” cognitive effects of the use of different technologies. Regarding television, Salomon discusses the phenomenon of channel-hopping, which leads to a lack of concentration and mental engagement on the part of users in what they are watching. He also raises the issue – already discussed in the past – of whether the constant use of a pocket calculator, for example, improves or deteriorates the user’s learned capacities.

2.2. Something similar to these unresolved discussions arises today when it comes to considering the uses and effects of AI in education; not only in relation to AI as an agent of passive knowledge which, for instance, performs searches, but also regarding its role as an authentic generator of multimodal productions. The issue is about its long-term effects and the possibility of cognitive capacities atrophying rather than developing: capacities such as written argumentative reasoning, performed by an intelligent chatbot that we go about correcting. While it is true that this model – in which the person corrects and outlines what the machine does – supposes a higher level of knowledge, on a metacognitive level, in relation to one or more fields of knowledge, it is also true that many users of this type of AI system will not want to, will not know how to or will not be able to correct the text produced by the chatbot. Therefore, some groups of users – whether or not in the school setting – will automatically consider certain productions adequate without questioning them. It is rather like producing the automatic translation of a text in a completely unknown language and considering it adequate without checking for mistakes, in the hope that the native reader – the recipient – will detect any errors and correct the interpretation of the text.

In the case of new AI applications, the resulting text has been greatly improved. This is certainly true of ChatGPT, perhaps the best-known application, but also of others such as Elephas or Notion. There is no discussion about the quality and speed of work of this AI technology, although for the time being we are still not

dealing with Strong AI or artificial general intelligence (AGI), which would entail it thinking about what it is saying or writing. The current model is merely a highly evolved automaton with a huge capacity to combine and recombine data, along with many rules to construct well-formed and new linguistic expressions, although it is a vast improvement on the first model, Eliza, created in 1964. Some people might think that there is a person behind ChatGPT, or that the computer programme has advanced cognitive abilities, when in reality neither of these things is true; it is a set of algorithms which has no thoughts and which does not know what it is doing – at least for now. Searle (1980, 1999) highlighted this fact in his well-known thought experiment, the Chinese Room Argument, in which a version of the Turing test was implemented. ChatGPT and other “generative” applications may pass this test more adequately than other older applications, but this does not mean that they understand anything of what they say or write. As things stand, AI does not think or feel in any way (Schank, 2015; Thagard, 2024).

2.3. Discourses on AI are not only dominated by an instrumental metaphor; a powerful textual metaphor is also reflected in a multitude of opinions that are not technical but rather ideological in nature. They can be found, above all, in the media – newspapers, radio, television, social networks, below-the-line comments, news items, opinion articles, monographs –, all repeating arguments and giving voice to a multiplicity of social actors. AI began life as a largely unknown phenomenon, present almost exclusively in popular cultural forms such as science fiction. In the realm of film, examples include *The Matrix* and *Blade Runner*, or, going back further, Stanley Kubrick’s *2001: A Space Odyssey* and Fritz Lang’s *Metropolis*. It has appeared in TV comedy shows featuring robots falling over when trying to walk, or in more “serious” news reports focusing on events such as the triumph of Deep Blue’s algorithms in 1997, when it defeated the world chess champion Garri Kasparov. Nowadays, however, everyone talks about AI, gives their opinion on it and, above all, expresses their fears about it, mostly based on the warped depictions of the phenomenon in film and other media.

While we could think about this textual metaphor of AI through the prism of economics or politics, another way to consider it is from the perspective of its close association with the enormous influence of the media. In this sense, it echoes other eras, such as the 1960s, when the media reacted to large-scale music festivals, seeing them as places of perversion, drugs and free love, as events fully prepared to subvert the established order. This reaction became known as moral panic (Cohen, 1972). The idea of moral panic – or panic induced by the media – is a broad perspective that allows us to consider the effect of postmodernity in the creation and survival of currents of opinion with virtually no basis in reality. A similar phenomenon accompanied the arrival of computers in social life and education, with the emergence of technophiles and technophobes – supporters of electronic writing and reading, and detractors who only saw the evils of such technology (Birkerts, 1999). AI – possibly together with robotics – might be said to have inherited the mantle of moral panic. It has been situated in a place where all the fears about evolution and the future come together – once again featured in films (such as those mentioned above) or in the multiple variants of “killer droids”.

This social fear lies in considering the effects of AI in a dichotomous and radical way. Other authors (McRobbie and Thornton, 1995) have focused precisely on how different social groups have dealt with moral panic, while still others have retraced its history and

argued that it is not a media-produced effect but rather something earlier and more structural (Drotner, 1999, highlights the McLuhanian origin of the term “moral panic”). In short, if we were to admit this media-social persecution of certain deviant or undesirable behaviours, we should by the same token consider the existence of a kind of naivety, or do-goodism – media-based or moral –, according to which we are only prepared to see the positive aspects of AI, such as in education, for example; in other words, a sort of uncritical technophilia.

On the other hand, moral panic – especially in the case of technologies – marked a turning point, a shift, in the general societal view of an issue. It is somewhat similar to what had been termed hegemony in the study of dominant ideologies (Gramsci), or, more descriptively, in what Aristotle termed *endoxa* (Categories, Topics); that is, the views of the preponderant (and respectable) majority, considered as an explanatory source in many aspects, again somewhat similar to so-called common sense or the shared and unquestioned knowledge of broad social sectors. In discourses on AI, the *endoxa* is created by the media that multiply their references, accepting everything they publish as true, and announcing a change in the generally accepted functioning of the system.

Although both types of metaphor include a good number of discursive examples, these do not enter the core of their arguments. A third metaphor explores technologies as a system or as an environment, a phenomenon that has already occurred regarding the use of computers in several formal education settings; that is, they have become everyday elements rather than sporadically used ones. When it comes to AI, a perception of the future emerges that generates misgivings even among its own developers: some implement each new development, influencing other companies that may adapt aspects of these developments; others, in a sort of justified moral panic, warn about how these developments have the potential to act against humans. The relationship between AI technologies and the social environment has barely been touched upon, despite some recent efforts (Airoldi, 2022).

In the cognitive sphere, this metaphor contemplates an uncertain future in which AI technologies will not only multiply but may also contribute to reducing these cognitive improvements. None of this has been “proven”, although the same is true of the opposite argument, whether in respect of the learning of Latin when it began to cease to be compulsory at the turn of the 20th century, or that of BASIC much later on, in the 1960s and 1970s, when programming was considered a way to improve advanced cognitive abilities and social development. As such, the idea that AI may usher in an involution echoes the argument used against arithmetic calculators and, before that, slide rules. Conversely, we believe that it is a curricular issue: what is taught and how it is taught. There is nothing in Latin, or in any other language, which makes us more or less intelligent; there is only differentiated learning. What makes a significant difference, however, is the culture written in that language – understood as one of the forms of cultural capital according to Bourdieu, 1979 –, since an African or Austronesian language is different from the tradition and quantity of scientific texts produced in German or English.

In the social sphere, there is even greater uncertainty, since the habitual use of AI and its evolution, which has barely begun, has only been thought about in literature and film, and almost always with a markedly negative outlook, in the genre of science fiction. Nevertheless, its impact on the *endoxa* has been more significant

than it seems, being the genre where the negative thinking in which humans almost always appear dominated by robots originated (The Matrix being an extreme case). Once again, the experts and creators of AI themselves warn about its negative use, seeing it as serving the purely economic interests of certain companies.

### 3 OTHER CONTRADICTIONS

There appears to be no quick fix or feasible solution for this worrying set of contradictions. We believe that AI will continue to develop at breakneck speed – at least with the current programming tools – and that the emergence of applications for tasks and problems, both common and new, will be unstoppable. The current reaction of the political class is to propose regulations aimed at companies. The European Union seems to have stolen a march on other nations in this regard, but it is sure to be copied. Unfortunately, as far as AI is concerned, the borders between countries are not at all dissuasive, which means that it will be difficult to contain or apply certain legislation (as already occurs with information technology today, with programmes and servers beyond the reach of any controlling mechanism). Another possibility would be to request a moratorium on its development, as the Future of Life Institute did in its 2023 manifesto, although this is actually a non-solution in the form of a postponement which, moreover, has arrived late due to the pace of development of AI and the commercial repercussions that any such moratorium would entail. Nonetheless, although extremely naïve, it would – like the aforementioned legislation – be a wake-up call about the social and ethical problems posed by AI.

Some authors (in particular, Beck, 1998) talk about concepts such as late modernity and the “risk society” to show how the dynamics of these complex, interconnected societies generate catastrophic situations – above all, environmental and health-related crises – with major political repercussions. In Spain, examples include the Prestige oil spill off the coast of Galicia in 2002 and the toxic syndrome caused by the ingestion of adulterated rapeseed oil in 1981. Internationally, we could highlight cases such as the so-called mad cow crisis of the 1990s, which mainly affected the UK, or the Chernobyl nuclear disaster in Ukraine in 1986. Meanwhile, global crises include the Covid-19 pandemic that started in 2019, or the Thalidomide scandal of the late 1950s and early 1960s, among many others. On top of this, we have catastrophes generated by wars and the use of nuclear or chemical weapons. Moral panic can be seen as a way of thinking about social deviation, especially in industrial and pre-industrial societies, less related to major crises or catastrophes such as those mentioned above and of smaller scope. It is not the same to fantasise about uncontrolled robots that make autonomous decisions about humans or that are even aware of who or what they are, which actually exist beyond the realm of fantasy, or to indulge in the classic daydream of how our ancestors might view today’s machines.

### 4 IA AND LIES: UNCONTROLLED ABDUCTION

“But that is mere speculation,” said I. (Watson)

“It is more than that. It is the only hypothesis which covers the facts.” (Holmes)

Arthur Conan Doyle. *The Sign of the Four*

Many believe that AI has limits in relation to human thinking. They are not referring to limits that might be imposed, such as labelling products made with AI so that we can distinguish between them and exclusively “human-made” products, or halting the development of AI, legislating, etc., but rather limits regarding how far AI can develop. Nonetheless, in just 70 years it has already blurred some aspects of the human/artificial boundary. We do not need to be clairvoyants to imagine what might happen a few hundred years from now. AI was an oxymoron, but this is no longer the case. Or is it?

Most of the arguments are couched in terms of “what AI cannot do”. Examples include Selwyn on education (2018, 2019) or Thagard (2024). They counter the arguments that come out of the computing sphere, along the lines of “everything that AI is capable of” (Walsh, 2018, or even more prudent figures such as López de Mántaras and Meseguer, 2017). Both sides are right: on the one hand, AI is characterised by amazing developments and successes, unstoppable thanks to the schematisation of the underlying logical and mathematical operations (incorporated in programming languages like Prolog or Python), and great achievements in computing; on the other hand, it faces the major difficulty of an unattainable goal, namely that of being capable not only of reasoning but also of acting like humans, who have their own bodies, their own voices, desires and emotions, a learned – and, to a large extent, shared – culture, and a capacity for social interaction that shifts over the course of their lives. However, neither side of the argument serves any purpose, since they both fail to address the key problem of an AI which in the medium term will not become untethered from human control.

Among the many aspects to consider, one seems particularly important to me: AI as a machine for generating lies. This occurs because ChatGPT, to take one example, has to provide an answer even if it does not know the answer, or even if it knows that its answer is inaccurate or false. The fact that it then apologises, with the excuse that it is merely a programme, does not alter the fact that it has lied, perhaps knowingly, according to simple probability theory (in simple terms, we might call it a machine version of *Sartrean bad faith*). This occurs when the question is not contained in its database and it extrapolates the answer to what it *has* found. In a brief chat, I requested a summary of the book *Code in Context* by Diana Adlam. The programme knew nothing about the book. Accordingly, the answer it provided was about writing code in certain languages. Once corrected, it continued to provide a sort of pastiche, combining the data I fed it about this book on the sociology of education with content on conventional coding. At a certain point, the conversation went as follows:

**Me:** It's quite strange that you invent answers with no relation to the question. Better to say “Sorry, I don't know.”

**ChatGPT:** I apologize for the confusion and mistakes in my previous responses. As an AI language model, I strive to provide accurate and helpful information to the best of my ability based on the information provided to me. However, I am not perfect and can make mistakes or create misunderstandings. Thank you for correcting me and providing the correct information about Diana Adlam's “Code in Context” book. I appreciate your feedback and will strive to provide more accurate and relevant responses in the future.

As can be seen, it is only after showing ChatGPT its mistake that the system makes a sort of apology – as well as (self)anthropomorphising itself with the verb “strive”. The problem is not that it does not know the book in question but rather that it has invented the answer without realising that it was a mere bet, an unproven hypothesis. The system is not even lying in bad faith; it does not know that it is lying and, therefore, has no remorse for having lied. It is understandable that it is not programmed to constantly repeat that it does not know the answer, since this would make for a much less attractive dialogue.

We humans share this mode of reasoning and argumentation. It is called *abduction* (or *retroduction*), following the American logician and philosopher C.S. Peirce (1839-1914), and can be said to complete the three types of reasoning: deductive, inductive and abductive. Unlike the first two, which are also used constantly by both humans and machines, abduction does not provide conclusive results – in fact, neither does induction, as philosophers since Hume have pointed out –, but rather it is a way of formulating hypotheses about facts that are new, unknown or difficult to explain. These hypotheses are probabilistic, like the answer to the request for a summary of Adlam's book. In other words, sometimes they are right and sometimes they are not. In some cases they are mere conjectures, as occurs in fiction. For example, in the first few pages of *The Name of the Rose*, its author, the Italian semiotician Umberto Eco, employs a form of “extreme” abduction to make the Franciscan monk William of Baskerville describe the abbot's horse, which he has never seen, and which the monastery's cellarer and other subordinates are searching for, it having bolted. Eco is referencing Voltaire's *Zadig*, which contains a very similar example that Eco himself had analysed (Eco, 1989). It could be said that Voltaire was the author of the first example of abduction in detective fiction in 1748, long predating Poe or Doyle. However, it is worth pointing out that abduction almost always succeeds in fiction.

Abduction is a somewhat elusive concept, since Peirce himself modified his definition of it on several occasions, in both its application and theorisation. Several works have been written about this, of which some of the best known are Fann (1970), Anderson (1986), and – regarding the concept and its evolution in the Collected Papers of Charles Sanders Peirce (1931-1958) – Kim and Cunningham (2003). Nonetheless, this elusiveness does not prevent it from being a habitual and constant mechanism for inference, whether in special situations (surprise, analogy, creation) or in situations of simple ignorance, where both humans and AI pose hypotheses and search for an idea to accommodate the new data. Both the “logic” of abduction and the necessary computer programming remain open problems, with multiple classifications and technical solutions (Magnani, ed., 2023).

There are many ways to think about and classify abduction (Kapitan, 1997). Peirce himself distinguished between a conventional but imprecise form of inference – a method of reasoning resulting in non-necessary inferences, as opposed to deduction, which results in necessary ones – and a kind of instinct or insight (Paavola, 2005), even though we would not use the term “instinct” today. The idea of instinct refers to making quick decisions – with little basis in fact – and employing inferential reasoning, based in many cases on small visual clues. However, abduction can also be seen as a generator of hypotheses, as a concept related to creativity in the heuristic sense (Pólya, 1973) – at least in one of the methods of so-called heuristics.

The table below summarises the three main meanings of abduction in Peirce and incorporates Eco’s coding distinctions (1979, 1989), as well as suggesting connections with other disciplines. When it comes to classifying the forms of abduction, a distinction is usually made between those that “invent” an explanation for a new fact (new for the abductor subject), which we have called a context of discovery, and those that generate hypotheses in a somewhat more reasoned (probably verbal) way, which, for lack of a better term, we have called a demonstrative context, which involves searching for or selecting the best hypothesis, even if by also initiating a process of inductive inference.

**ABDUCTION**

Context of discovery (novelty, surprise)	Demonstrative or selective context	Coding	Result / Truth value	Other related disciplines
① <b>Instinct / Insight</b>		Hypo-codified	Doubtful	Psychology
<b>Generator of hypotheses</b>		Code and s-code	“Creative” Plausible	Logic Heuristic
	③ <b>Abductive Inference</b>	Hyper-codified	Pre-inductive Selective	Inductive logic Heuristic

Table 1. Forms of abduction in Peirce and other relationships

Meanings 2 (generator of hypotheses) and 3 (inference) are virtually the same – differentiated by when and why they occur rather than by being essentially different from each other –, whereas meaning 1 (instinct/intuition) is distinct. There may be some doubt about Eco’s terminology; indeed, he himself points this out (Eco, 1989; Thagard, 1987), in addition to incorporating what he calls meta-abduction at another point. However, the purpose of the meanings in this table is to describe rather than to explain the logic of abduction. Its connection with AI is, in some cases, direct: inductive logic, which we arrive at through inference or even by generating hypotheses, forms the basis of machine learning, as critically highlighted by Larson (2021), although in a highly theoretical way.

**5 DISCUSSION**

In the case of formal education, we could argue, broadly speaking, that a “moderate” abduction is emphasised, similar to what occurs in everyday life – the endoxa once again, perhaps –, seeking inferences based on common sense rather than extravagant or completely unlikely ones (Is there really any reason why the abbot’s horse should be called Brunellus?), closer to science or calm reasoning than to the unbridled fantasy of Zadig, Holmes or Peirce himself in his account of how he recovered a stolen watch (Eco and Sebeok, 1989, chapter 2). However, this moderation in hypotheses can lead to the riskiest ones being discarded, considered impossible or, even worse, unthinkable due to an emphasis of the school system on transmitting a certain cultural status quo, acting more as a mechanism of reproduction than one of innovation and change.

However, the results of education are never unique or easily measurable; they owe themselves to the three aforementioned forms of reasoning, but also to the contents and structures of the sciences and humanities, to many other types of learning (such as dance, music, crafts, gymnastics, poetry, drawing, sports in

general, etc.), to groups of friends and, of course, to the great influence of parents and teachers. Although Peirce focused on these general logical forms, a long time has passed and an enormous body of knowledge about education has been built around later authors such as Piaget or Vygotsky. Despite the arguments put forward by Ramah (2023), for whom abduction plays a central role in AI, there is something about educational interaction, and interpersonal interaction in general, that is very difficult to find in AI and its use of abduction. Examples include changing plans and goals set on the spur of the moment, in an improvised manner, adapting to new situations and recalibrating actions. This phenomenon has been studied, for instance, in how jazz musicians improvise. Although there are underlying mechanisms in jam sessions that can be made explicit (Becker, 2000), the label that marks the order and types of intervention among expert and novice musicians alike is difficult to define. Teachers are also well aware of these situations, especially in face-to-face interactions.

If we accept that education in its broad sense (formal and informal) still has poorly thought-out areas, such as the interaction between its contexts and the variations in the learning that occurs therein, AI could serve as a tool to improve learning in the form of effects with technology, to use Salomon’s terminology, possibly in line with Bruner’s theory of scaffolding, which would undoubtedly be extremely helpful. Nonetheless, the complexity of education and its contexts means that AI’s pretensions to create an autonomous teaching system, a so-called Artificial Pedagogy (Bieger et al. 2017), is a distant reality, if it is even possible. It is worth highlighting that, several decades ago, David Merrill et al., with greater pedagogical knowledge, attempted a more limited version of this and achieved negligible results. However, we must also bear in mind that some AI applications are capable of lying and do so constantly, given that they are actually algorithms programmed by humans, extrapolating the abductive inference they need in order to achieve some kind of result, whatever it might be.

In light of our analysis, there are many good reasons to continue to reflect on these ideas. The existing AI technology already provides great solutions, which will only improve over time, but it is not a tool without duplicity; on the contrary, the social discussion (ethics, politics) is of equal or greater importance than the technological one. However, this social discussion is currently rather backwards-looking and overly polarised, taking the form of moral panic in many social sectors. Some of the elements and stakeholders of AI are more fragile than others; we might say that they are more easily manipulated. They will undoubtedly require much greater attention on the part of the educators and professionals involved, since we must never forget – as argued by Becker (1967) – whose side we are on.

**REFERENCES**

Airoidi, M. (2022). *Machine Habitus. Toward a Sociology of Algorithms*. Polity Press.

Anderson, D. R. (1986). The Evolution of Peirce’s Concept of Abduction. *Transactions of the Charles S. Peirce Society*, 22(2), 145–164.

Becker, H. S. (2000). The etiquette of improvisation. *Mind, Culture and Activity*, 7: 171–6.

Becker, H.S. (1967). Whose Side Are We On?. *Social Problems*, 14, pp. 239–67.

Bieger, J., Thórisson, K.R., & Steunebrink, B.R. (2017). The Pedagogical Pentagon: A Conceptual Framework for Artificial Pedagogy. In: *Proceedings of AGI-17 (2017)*.

Birkerts, S. (1999). *Elegías a Gutenberg*. Alianza.

- Bourdieu, P. (1979). Les trois états du capital culturel. *Actes de la Recherche en Sciences Sociales*, 30, 3-6.
- Cohen, S. (1972) *Folk Devils and Moral Panics: The Creation of the Mods and Rockers*. Routledge, 2011.
- Cuban, L. (1986) *Teachers and Machines. The Classroom use of Technology since 1920*. Teachers College Press Columbia University.
- Drotner, K. (1999). Dangerous Media? Panic Discourses and Dilemmas of Modernity, *Paedagogica Historica*, 35(3) 593-619.
- Eco, U. (1985). *Tratado de semiótica general*. Lumen.
- Eco, U. (1989). Cuernos, cascos, zapatos: Algunas hipótesis sobre tres tipos de abducción. En U. Eco y T. Sebeok (eds, 1989). *El signo de los tres* (Dupin, Holmes, Peirce), pp. 265-294. Lumen.
- Eco U. & Sebeok T. (eds, 1989). *El signo de los tres* (Dupin, Holmes, Peirce). Lumen.
- Fann, K.T. (1970). *Peirce's Theory of Abduction*. Martinus Nijhoff.
- Future of Life Institute (2023). *Policymaking In The Pause*. [https://futureoflife.org/wp-content/uploads/2023/04/FLLI\\_Policymaking\\_In\\_The\\_Pause.pdf](https://futureoflife.org/wp-content/uploads/2023/04/FLLI_Policymaking_In_The_Pause.pdf)
- Kapitan, T. (1997). Peirce and the structure of abductive inference. In *Studies in the Logic of Charles Sanders Peirce*, N. Houser, D. Roberts, and J. Van Evra (eds.), 477-496.: Indiana University Press.
- Kim, J. y Cunningham, D. (2003). A syllogism for formulating hypotheses. *Semiotica*, 144 –1/4 (2003), 303-317.
- Larson, E. J. (2021). *The Myth of Artificial Intelligence*. The Belknap Press of Harvard University Press.
- López de Mántaras, R. & Meseguer, P. (2017). *Inteligencia Artificial*. Los libros de la catarata.
- Magnani, L. (ed, 2023). *Handbook of Abductive Cognition*. Springer.
- McRobbie, A. & Thornton, S.L. (1995). Rethinking 'Moral Panic' for Multi-Mediated Social Worlds. *The British Journal of Sociology*, 46(4), 559-574
- Nardi, B. & O'Day, V. (1999). *Information Ecologies: Using Technology with Heart*. MIT Press.
- Paavola, S. (2005). Peircean abduction: instinct, or inference? *Semiotica*, vol. 153, DOI: 10.1515/semi.2005.2005.153-1-4.131
- Peirce, C.S. (1931-1958). *Collected Papers of Charles Sanders Peirce, Charles Hartshorne, Paul Weiss, and A. W. Burks* (eds.). Harvard University Press.
- Pólya, G. (1973). *How to solve it. A new aspect of the Mathematical Method*. Princeton University Press. [orig. 1945]
- Ramah, G. (2023). *Generative AI: Unveiling the Abductive Mind*. On line: <https://www.linkedin.com/pulse/generative-ai-unveiling-abductive-mind-gary-ramah-tqymc>
- Rodríguez Illera, J. L. (2023). Marcos de pensamiento: las relaciones entre tecnología y educación. *Temps d'Educació*, 64, 117-130.
- Salomon, G. (1991). The different effects of technology in the development of the mind. En L. Tolchinsky (ed,1991) *Culture, Schooling and Psychological Development*. Ablex.
- Salomon, G., Perkins, D. & Globerson, T. (1992). Coparticipando en el conocimiento: la ampliación de la inteligencia humana con las tecnologías inteligentes. *Comunicación, lenguaje y educación*, 13, 6-22.
- Schank, R. (2015). *Machines that Think are in the Movies*, in Brockman, J. (ed.), *What to Think About Machines that Think*. Harper Collins.
- Searle, J. (1980). Minds, Brains and Programs, *Behavioral and Brain Sciences*, 3, 417-57.
- Searle, J. (1999). The Chinese Room Argument, en R.A. Wilson and F. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*. MIT Press, 115-116.
- Selwyn, N. (2018, 12 de junio). *Six reasons artificial intelligence technology will never take over from human teachers* [Blog]. EduResearch Matters. Consultado en mayo de 2023: <https://www.aare.edu.au/blog/?p=2948>
- Selwyn, N. (2019). *Should Robots Replace Teachers?: AI and the Future of Education*. Polity.
- Thagard, P. (1987). Semiotics and Hypothetic Inference in C. S. Peirce. in Versus, *Semiotica testuale: mondi possibili e narratività*, 19-20, 163-172.
- Thagard, P. (2024). *Bots and beasts: What Makes Machines, Animals, and People Smart?* M.I.T. Press.
- Walsh, T. (2018). *Machines that Think. The future of Artificial Intelligence*. Prometheus.
- Watters, A. (2021). *Teaching Machines. The History of Personalized Learning*. MIT Press

## LA IA EN EL DISCURS DE LES RELACIONS ENTRE TECNOLOGIA I EDUCACIÓ

L'article revisa algunes de les relacions entre la IA i l'educació, i destaquen les metàfores utilitzades, les dificultats per trobar punts d'acord, així com aspectes de la crítica social que es fa a la IA (per exemple, considerar que pot ser una forma desviació no desitjada). La IA apareix com un cas més de tecnologia que ve a millorar l'educació, com va passar abans amb altres tecnologies menys "intel·ligents", encara que ara amb més arguments a favor seu. En segon lloc, explora els mecanismes lògics (l'abducció) molt utilitzats en el raonament humà però també a la IA, mostrant com el seu ús descontrolat pot fer que determinats algorismes de conversa siguin capaços de mentir, cosa que els professors no farien. Es planteja una qüestió ètica més que lògica que mereix ser explorada més endavant.

PARAULES CLAU: intel·ligència artificial, educació, abducció

## LA IA EN EL DISCURSO DE LAS RELACIONES ENTRE TECNOLOGÍA Y EDUCACIÓN

El artículo revisa algunas de las relaciones entre la IA y la educación, destacando las metáforas utilizadas, las dificultades para encontrar puntos de acuerdo, así como aspectos de la crítica social que se hace a la IA (por ejemplo, considerar que puede ser una forma de desviación no deseada). La IA aparece como un caso más de tecnología que viene a mejorar la educación, como ocurrió antaño con otras tecnologías menos "inteligentes", aunque ahora con más argumentos a su favor. En segundo lugar, explora los mecanismos lógicos (la abducción) muy utilizados en el razonamiento humano pero también en la IA, mostrando cómo su uso descontrolado puede llevar a que determinados algoritmos de conversación sean capaces de mentir, algo que los profesores no harían. Se plantea una cuestión ética más que lógica que merece ser explorada más adelante.

PALABRAS CLAVE: inteligencia artificial, educación, abducción

The authors retain copyright and grant the journal the right of first publication. The texts will be published under a Creative Commons Attribution-Non-Commercial-NoDerivatives License.

