ACADEMIA | Letters

A vigotskijan perspective on machine learning. How cultural stereotypes are involved in education of algorithms

Martina De Castro, Roma Tre University, Rome, Italy
Umberto Zona, Roma Tre University, Rome, Italy

1. The role of algorithms in spreading stereotypes

Algorithms are conventionally described (Zona & De Castro, 2020) as logical-formal processes that, moving from a starting condition and using a series of parameters, allow the resolution of the most different problems, helping human beings in their daily tasks. Advantages are certainly not negligible, but at the same time, the ethical implications of using algorithms should be considered and analysed, in particular the role played in the processes of conditioning human behaviour, especially in view of the widespread use of algorithms on the Net. For some years now, an increasingly relevant line of research (Cardon, 2016; Noble & Tynes, 2016; O'Neil, 2017; Zona & Bocci, 2018; Finn, 2018; Noble, 2018) has endeavoured to analyse algorithms as cultural constructs – as an expression of the culture of the designer and, more generally, of the vision of the companies that use them. This condition is decisive since – as shown by the "Diversity Reports" released by Google (Google, 2021), Apple (Apple, 2020) and Facebook (Williams, 2021) – designers are overwhelmingly heterosexual and able-bodied white males, with the privilege not only of predetermining the questions and answers that the algorithm has to provide, but also of choosing the dataset on which to exercise it (Buolamwini & Gebru, 2018). In order to analyse the learning/training processes of the algorithms and, in particular, the most frequent errors they make – which play a significant role in the propagation of sexist, racist, classist and ableist stereotypes –, we referred to Lev

Academia Letters, January 2022

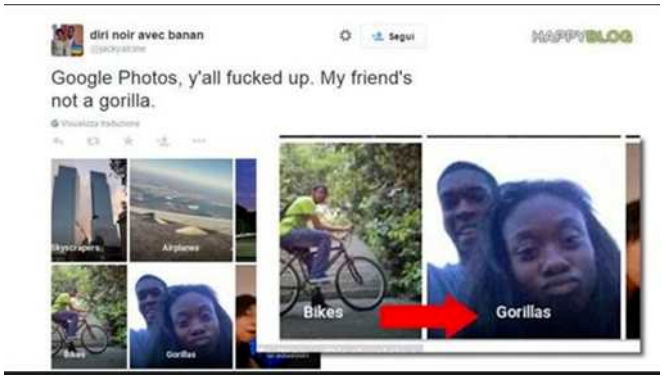
Corresponding Author: Martina De Castro, martina.decastro@uniroma3.it
Citation: De Castro, M., Zona, U. (2022). A vigotskijan perspective on machine learning. How cultural stereotypes are involved in education of algorithms. *Academia Letters*, Article 4638.
https://doi.org/10.20935/AL4638.

Vygotskij's theories on childhood learning. There is no doubt, in fact, that between the designer and the algorithm a relationship is established that, in many ways, is analogous to that which, from the first months of life, is established between the child and the adult and which then, in the course of the development process, can give rise to the Zone of Proximal Development (Vygotskij, 2008), in which – as is known – the more expert subject acts in support of the less competent one. The designer also comes to the aid of the algorithmic construct, providing it with criteria for a systematic organisation of concepts. In doing so, however, he imparts to the algorithm historically and socially determined cultural models of reference that will inevitably influence the choices made by the machine intelligence.

As Vygotsky teaches, the child is, from birth, embedded in relational and linguistic contexts and it is for this reason that it starts to use words at a very early age, but this does not presuppose an equally full development of concepts (Vygotsky, 2008). Also in machine learning, the learner (the algorithm) learns within a social context. The relationship with the designer, first of all, provides the artificial intelligence with the basic set of languages, as happens between a teacher and a pupil, and the relationship becomes qualitative as well as quantitative, since the learner does not only receive data from teacher, but also criteria for interpreting them. Indeed, as Cardon (2016, p. 91) says, «there is no learning without knowledge. Data alone are not enough. […] Machine learning is a kind of knowledge pump, […] which first has to be trained» and, consequently, «the real problem is that almost all learners start from too little knowledge […] Without the guidance of an adult brain […] it is easy for them to get lost» (Cardon, 2016, p. 108). The error that children can most frequently fall into is that of generalisation and is the same as the error of overfitting algorithms. Domingos writes: «When a learner finds a regularity in the data that is not matched in the real world, we talk about overfitting of data. Overfitting is the fundamental problem of machine learning. […] Even humans are not immune to overfitting. […] Think of the little white girl at the mall who, upon seeing a Hispanic baby girl, exclaims 'look, mommy, a baby waitress!' […] Her is not innate racism: the little girl has overgeneralised from the few Hispanic waitresses she has seen in her short life. The world is full of Hispanic women doing another job, but she has not yet met them» (Domingos, 2016, pp. 98-100). The algorithm, then, similarly to the child, learns from experience, but whereas in the case of the child this experience is direct, for algorithms experience is always indirect, in other words mediated by the linguistic representations and cultural contents of others. Thus, despite the enormous amount of data they are able to process, their experience remains substantially limited and requires constant supervision by human beings. Thus, as infantile generalisations develop on the basis of strictly predetermined meanings, the same thing happens with algorithms. This is demonstrated, among

Corresponding Author: Martina De Castro, martina.decastro@uniroma3.it
Citation: De Castro, M., Zona, U. (2022). A vigotskijan perspective on machine learning. How cultural stereotypes are involved in education of algorithms. *Academia Letters*, Article 4638.
https://doi.org/10.20935/AL4638.

many others, by an episode that happened to Jacky Alciné,[1] who experienced personally some racist stereotypes conveyed by the Google Photo algorithm when the selfie he had taken with a friend was named 'gorillas'. How had this been possible?



Google Photo's auto tagging system is based on deep neural networks, which allow the machine to 'see', in much the same way as the human brain does, using a learning algorithm: «In image processing, it works something like this: millions of images are added to a system, each of them tagged by a human. […] Those images make up the system's *training data*: the collection of examples it will learn from. Algorithms then go through the training data and identify patterns». (Wachter-Boettcher, 2017, p. 131). Algorithm that is used by Google Photo, therefore, is able to 'learn', but it needs to practice on a very large number of details that make up the overall image because, unlike how our eyesight acts, it does not look at the overall shape but at each, infinitesimal elements that make it up: «This kind of pattern recognition happens in layers: small details, like the little point at the top of a cat's ear, get connected to larger concepts, like the ear itself, which then gets connected to the larger concept of a cat's head, and so on – until the system builds up enough layers, sometimes twenty or more, to make sense of the full image» (Wachter-Boettcher, 2017, pp. 131-132).

2. Vygotskij's lesson on generalisation biases

The mistake made by Google Photo can be compared to the one a child frequently makes. Vygotsky explains that the development of concepts takes place in three stages. In the first stage, the very young child associates and gathers objects on the basis of his or her own im-

[1] Jacky Alciné is an engineer and developer.

Corresponding Author: Martina De Castro, martina.decastro@uniroma3.it
Citation: De Castro, M., Zona, U. (2022). A vigotskijan perspective on machine learning. How cultural stereotypes are involved in education of algorithms. *Academia Letters*, Article 4638.
https://doi.org/10.20935/AL4638.

pressions, even though there is no relationship between them. These are «isolated objects that are linked to one another in any way in the child's representation and perception, in a single fused image. […] extremely unstable» (Vygotsky, 2008, p. 148-149). In the second stage, the child is able to assemble concrete objects into general groups and this thinking is called complex thinking: «The generalisations implemented through this mode of thinking represent, by their structure, complexes of isolated objects, or things, brought together no longer on the basis of only the subjective links that are established in the child's impression, but on the basis of objective links that really exist between these objects» (Vygotsky, 2008, p. 151). A relationship is established between the general and the particular. The last form of complex thinking is represented by the pseudo-concept, a set of concrete objects that «by the set of its external characteristics, coincides completely with the concept, but by its genetic nature, by the conditions of its appearance and development, by the causal-dynamic links that underlie it is not a concept at all» (Vygotsky, 2008, p. 161). Vygotsky strongly emphasises that childhood complexes, corresponding to the meaning of words, do not develop freely, spontaneously, along lines drawn by the child himself, but according to directions established by adult language. The fact that the meaning of the words of a three-year-old child seems to coincide with that of the adult has long led to the belief that the child's thought – and, therefore, his concepts – also coincides with that of the adult, but the child thinks by pseudo-concepts, that is by complexes, even if the word he uses to communicate them is the same as the adult's: «The child does not choose the meaning of a word. It is given to him in the process of verbal communication with adults. […] In simpler terms, the child does not create his own language, but assimilates the ready-made language of the adults around him» (Vygotsky, 2008, p. 165). Now, by way of language, cultural, external and social impositions are transferred, shaping the development of concepts (is this not one of the functions of education?), and this is also what happens to machine learning algorithms, which are culturally misled by their educators, the designers. They, just like the child, learn words whose meaning they have not chosen – which is transferred to them in the training process – and therefore use generalisations not constructed by themselves but by their designers. Like the child, they do not create their own language but assimilate a pre-packaged one, that of their trainer. But through language, as Vygotsky teaches, cultural contents are transferred and this also happens with algorithms and is probably the main reason for their "mistakes" (Zona & De Castro, 2020).

Corresponding Author: Martina De Castro, martina.decastro@uniroma3.it
Citation: De Castro, M., Zona, U. (2022). A vigotskijan perspective on machine learning. How cultural stereotypes are involved in education of algorithms. *Academia Letters*, Article 4638.
https://doi.org/10.20935/AL4638.

The essay is a joint work of the authors. However, for the purposes of recognising the individual parts, it should be noted that paragraph 1 is by Martina De Castro and paragraph 2 is by Umberto Zona.

References

Apple (2020). *Inclusion & Diversity*. URL: https://www.apple.com/diversity/

Buolamwini J., Gebru T. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. "Proceedings of Machine Learning Research" 81:1–15. URL: http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Cardon D. (2016). *Che cosa sognano gli algoritmi? Le nostre vite al tempo dei big data*. Milano: Mondadori.

Domingos P. (2016). *L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*. Torino: Bollati Boringhieri.

Finn E. (2018). *Che cosa vogliono gli algoritmi? L'immaginazione nell'era dei computer*. Torino: Einaudi.

Google (2021). *2021 Diversity Annual Report*. URL: https://diversity.google/annual-report/

Noble S. U., Tynes B. M. (a cura di) (2016). *The Intersectional Internet. Race, Sex, Class and Culture Online*. New York: Peter Lang.

Noble S.U. (2018). *Algorithms of Oppression. How Search Engines Reinforce Racism*. New York: New York University Press.

O'Neil C. (2017). *Armi di distruzione matematica. Come i big data aumentano la disuguaglianza e minacciano la democrazia*. Milano: Bompiani.

Vygotskij L. S. (2008). *Pensiero e linguaggio.* Bari: Laterza.

Wachter-Boettcher S. (2017). *Technically Wrong. Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*. New York: W.W. Norton & Company.

Williams M. (15 luglio 2021). *Facebook Diversity Update: Increasing Representation in Our Workforce and Supporting Minority-Owned Businesses.* URL: https://about.fb.com/news/2021/07/facebook-diversity-report-2021/

Corresponding Author: Martina De Castro, martina.decastro@uniroma3.it
Citation: De Castro, M., Zona, U. (2022). A vigotskijan perspective on machine learning. How cultural stereotypes are involved in education of algorithms. *Academia Letters*, Article 4638.
https://doi.org/10.20935/AL4638.

Zona U., Bocci F. (2018). *La Rete come una Skinner box. Neocomportamentismo, bolle sociali e post-verità*. "MEDIA EDUCATION", vol. 9, n. 1, pp. 57-77.

Zona U., De Castro M. (2020). *Edusfera. Processi di apprendimento e macchine culturali nell'era social*. Lecce: Pensa Multimedia.

Corresponding Author: Martina De Castro, martina.decastro@uniroma3.it
Citation: De Castro, M., Zona, U. (2022). A vigotskijan perspective on machine learning. How cultural stereotypes are involved in education of algorithms. *Academia Letters*, Article 4638.
`https://doi.org/10.20935/AL4638`.