




Ethical competencies in machine learning from a communicational perspective in the educational process¹

<https://doi.org/10.34766/fetr.v58i2.1277>

Justyna Horbowska^a 

^a Justyna Horbowska, MA, <https://orcid.org/0000-0002-0723-0939>,

KUL Doctoral School, Faculty of Philosophy, The John Paul II Catholic University of Lublin, Poland

 Corresponding author: jhorbowska@kul.lublin.pl

Abstract: The term „artificial intelligence” (AI) refers to computer programs equipped with numerous competencies, such as making calculations, grouping and categorizing data, or communicating with the user in ethnic languages. On the other hand, artificial intelligence systems do not have certain properties, and among them, apart from the lack of „creative abilities,” is indifference to the moral aspect of actions when searching and compiling data or cataloging phenomena. This study aims to discuss selected reasons for this state of affairs in the context of machine learning (ML) methodology, including the issues and applications of artificial intelligence from the perspective of scientific communication that occurs in the educational process. Given this goal, a research problem was formulated in the form of a question: How are ethical competencies developed in the machine learning process in the context of communication occurring in the educational process? In order to answer this research question, the text analysis method and the synthesis method were used. As a result of the research, it was determined that conducting machine learning with human participation, as well as using artificial intelligence systems previously learned with human participation, may enable the transmission of moral content in the normative sense to a cybernetic machine. Since human participation allows supervised learning of cybernetic machines, this type of learning, used as the sole method or in combination with another method, offers the opportunity to provide applications with the desired information about socio-cultural rules. Fully independent training of cybernetic machines does not ensure they collect information on ethical aspects desirable in communication during the educational process because open data sets on which machine learning takes place may contain harmful content, amplifying negative social phenomena.

Keywords: artificial intelligence, machine learning, ethical competencies, communication

Introduction

Education, in a broad sense, involves activities related to people, and these activities are to be consistent with the professed system of values (Okoń, 1998). Also, applications based on artificial intelligence (AI), according to McCarthy’s original definition, are understood as the behavior of a cybernetic machine that would be considered intelligent in the case of a human (McCarthy, Minsky et al., 1955), the ability to evaluate is expected, treating it as crucial for numerous AI applications. First, the results obtained by artificial intelligence must be adequate for a given discipline and fit into its paradigm. Their correctness is determined by compliance with the assumed formal and content expectations. Language learning software

is supposed to enable proper use of the language, and the translation device is supposed to translate the text as best as possible (Massey, Ehrensberger, 2017).

Therefore, if we consider purposefully obtaining a result corresponding to the query as a condition for effective communication (understood here as the transmission and reception of communicates that constitute information)², the basic axiological assumption seems to be fulfilled here: the good as a value obtained as a result of the application’s work will determine the correctness of the result.

However, it is essential to note that such a result does not refer to moral good resulting from the social, cultural, religious, or ideological system

1 Article in polish language: https://www.stowarzyszeniefidesetratio.pl/fer/58P_Horb.pdf

2 The initiator of cybernetics, N. Wiener, wrote about the method of transmitting information and communicating between man and mechanism and between machine and machine. (See: N. Wiener, 1961, p. 7).

or moral obligation. As Hoes puts it, “artificial intelligence lacks a moral compass”; he adds that any other one also (Hoes, 2019). This is a significant limitation of AI systems, as they are not equipped with self-awareness, nor have they been found in them (Bishop, 2017). They are also not entities of social communication called proper communication, shaped by co-intentionality as the ability to share and co-shape goals and intentions specific to human cooperation (Tomasek, 2022). The status of AI may approximate the Chinese room argument introduced into the philosophical debate by Searle and intended to show in an experiment that a cybernetic machine, even if it processes data, does not have to understand it (Searle, 1980).

Since AI systems do not have a nature, they are not governed by natural law per se, nor do any rules other than those of logic apply unless they are imposed on them – unless laws are considered as self-evident as the principles of theoretical rationality applicable in science (Finnis, 2001).

Among the many divisions of AI³, those based on the type of machine learning used are noteworthy. The source of acquired knowledge plays a unique role in the learning process. Despite the undoubted differences, the process of acquiring knowledge through digital applications, like in the case of students, proceeds by obtaining it through communication with a subjectively treated teacher or the educational environment. It seems that it is possible to point out the relationship between the algorithm’s learning model and the possibility of embedding the work of AI in a moral context.

1. Machine learning methods

Machine learning is a set of activities included in artificial intelligence. Opinions about AI’s creative possibilities are divided, but it is assumed that even if independent, in the case of programs, teaching does not contain an element of creativity.

Although many machine learning methods and their divisions have been developed, they can be divided into the following three categories:

1. supervised learning,
2. unsupervised learning,
3. reinforcement learning (RL).

It is worth adding that a different learning model characterizes each of the above types of machine learning. Moreover, each is dedicated to specific problems that can be solved with its help (Flasiński, 2020).

1.1. Supervised learning

The first type involves teaching cybernetic machines by providing them with „labeled” data – with ready-made answers to the problem attached. This is done so that the system receives sample information and its classification as correct or incorrect (e.g., A or B). After collecting and analyzing the appropriate amount of such data, it can independently sort new information into one of the groups, with correctness depending on the number of previous examples and the homogeneity of their categories. In this case, we can talk about „reasoning” by analogy, when the machine refers to a similar or comparable situation and assigns new information to one of the states based on previously acquired knowledge.

Supervised learning aims to equip the machine to predict an object’s value or class. The machine achieves this by first learning many examples provided with labels for values or classes. While value prediction is used in solving a regression problem, predicting an object’s class enables its appropriate classification (Domingos, 2015).

Examples of supervised learning algorithms solving the regression problem include:

- linear regression,
- polynomial regression,
- regression tree,
- neural networks.

³ In addition to the divisions according to the learning mechanism, which is discussed in more detail here, there are other divisions, e.g., according to the data delivery method, application, or classification of machine learning systems. (Author’s note)

In turn, examples of supervised learning algorithms solving the classification problem include:

- decision trees,
- k-nearest neighbors method,
- support vector machines (SVM – support vector machine),
- naive Bayes classifier,
- random forest,
- logistic regression,
- neural networks.

The supervised learning method allows a person to follow learning processes because it assumes control over the acquired content and its interpretation in the sense of sequencing based on imposed labels. If labeling is based on moral norms, the AI will learn to imitate the recognition of data as good or bad, just as students do by reading folk tales full of morally polarized characters or in an arranged educational situation involving meetings with characters who they tell of noble deeds performed.

However, this method is rarely used because it is time-consuming and high-cost. It is worth emphasizing that although neural networks as artificial intelligence applications are supposed to enable solving both of the problems mentioned above, their work also turns out to be time-consuming; the costs of computer equipment with high computing power are also not low. Higher generations of AI are less effective than humans in „teaching” machines to recognize good and evil unless they are specially trained.

1.2. Unsupervised learning

This type of learning means that data is sent to the IT system without any suggestions regarding their classification. The computer collects and analyzes the data, then finds common elements between the data and, based on them, combines the data into groups. Man appears here only at the stage of interpreting the divisions made. This model may imply the need to limit the number of groups the system generates as a prerequisite.

When it comes to unsupervised learning (so-called learning without a teacher), its distinguishing feature is the lack of labels or classes assigned to the examples learned by the cybernetic machine. Finding connections between data is a task for the application itself. Such learning prepares the machine to group data, as in the case of clustering algorithms. In addition to clustering, algorithms that visualize unlabeled data in two or three dimensions will be used to group it with the assumption of its cause in dependencies (Sala, 2017).

The most important algorithms used in unsupervised learning – depending on the problems they deal with – include algorithms for cluster analysis, such as:

- k-means method,
- hierarchical cluster analysis,
- density-based data grouping (clustering) algorithm (DBSCAN – density-based spatial clustering of applications with noise) (Starczewski, Goetzen, 2020).

Examples of algorithms for extracting associations through visualization and dimensionality reduction are:

- principal component analysis (PCA),
- nuclear principal component analysis.

If the problem is reversed, the unsupervised learning method, given the task of distinguishing items that deviate from clusters, will eliminate groups and select only those dispersed items that do not belong to any of them. A one-class support vector machine can be used to solve anomaly and novelty detection problems.

Unsupervised learning results in a data set grouped according to some common feature or features. When interpreting the results of unsupervised learning, also in the ethical aspect, generative forms of artificial intelligence may not prove to be as effective as humans because the human, without limiting the free learning of AI by any preconditions for selection (including information about morality), will want to make the final choice on its own with selected opportunities. In this model, there is no question of

the „correctness” or „incorrectness” of the solution – evaluation is replaced by distinguishing equivalent sets or elements. In this case, lacking a teacher in the face of AI without insight into moral norms means that unsupervised learning models are morally insensitive. However, the effects of their work do not have to be morally indifferent.

1.3. Semi-supervised learning

Supervised and unsupervised learning are the main methods traditionally used in machine learning, not only separately but also as a combination of both approaches.

Semi-supervised learning algorithms are constructed using data marked with labels or classes to eliminate the inconveniences associated with supervised learning, such as its time-consuming and expensive nature, and to provide the highest possible quality to unsupervised learning algorithms. However, such labeling usually concerns a small group of data. Semi-supervised algorithms are combinations of supervised and unsupervised learning algorithms and are primarily implemented into neural networks.

A type of machine learning different from those described above is reinforcement learning.

1.4. Reinforcement learning

This type of teaching occurs when the system does not receive sample data with their classification (as correct/incorrect or good/inadequate) as in learning by analogy but also does not learn by ranking the supplied data. Its task is to provide an answer to a query based on searching an available database (often World Wide Web resources) without training, as if ad hoc and only the human reaction to the answers proposed by the machine is a source of knowledge for the system, which learns when the human selects one of them – remembering the association of the question with the answer chosen as a reinforcement, i.e., a positive signal. Machine learning of this type is carried out by interacting with the environment

based on information received from it on an ongoing basis – without previously implementing training data. Data is acquired automatically, and its submission in response to a query triggers a reaction confirming the accuracy of the choice (reward) or denying it. Reinforcement learning has vital elements: environment, agent, and buffer.

The environment is understood here as the task (real or simulation) with which the algorithm, referred to as the agent or player, interacts. The goal of reinforcement learning is to maximize an agent’s reward from the environment so the agent learns to achieve the highest possible score in a given environment.

The agent is an element that interacts with the environment to perform the task of learning to achieve the highest possible result and thus maximize the reward. A function returning an action, called a policy, is responsible for the agent’s behavior. Most often, the policy is implemented using a neural network⁴.

In turn, the buffer is a database that stores information collected by the agent during training, which is then used to train it.

As seen from the above, this learning model may, over time, assimilate the cultural, religious, or legal norms to which it receives access. However, it should be remembered that creating a sufficiently comprehensive database of controlled data may exceed the capabilities of software developers, and applications are most often given the broadest possible – uncontrolled – access to the World Wide Web so that they can use as much information as possible to answer questions effectively. This can be compared with the broadly understood educational environment traditionally described in pedagogy instead of the educational situation. While the academic environment surrounds the student in an uncontrolled manner, the problem „draws” him in, and its elements are intended to serve a didactic effect, reserving space for ethical interpretations. A narrowly understood educational environment assumes the autonomy of the student drawing from it because „it is no longer the teacher who teaches using demonstrative means, but the environment of the school classroom (not

⁴ *Reinforcement Learning and the Importance.* (From:) <https://datascience.eu/machine-learning/machine-learning-for-humans-part-5-reinforcement-learning/> (Access: January 20, 2024).

only the classroom) is a direct source of important educational impulses and cognitive conflict” (Kruk, 2009, p. 494).

Comparing the machine learning process to the developed methods of teaching students shows that, just like in the case of the classic education model where there are a student and a teacher, the machine can have a mentor. However, it will not always be human because artificial intelligence may play this role. Unsupervised learning brings to mind anti-pedagogy – perhaps it would be justified to treat this model as a distant echo of Rousseau’s concept of education – and reinforcement learning may resemble training animals so that, without understanding the reason or purpose, they perform specific actions in response to a stimulus, becoming more and more effective at recognizing it.

2. Human in the machine learning process

Initially, in each of the types of machine learning described here, human participation was considered necessary. However, current artificial intelligence models based on neural networks that have been developed in recent years are programmed in such a way as to gain the power to replace humans in the machine learning process. This means that AI can participate in both supervised and unsupervised learning and reinforcement learning. However, it seems it cannot replace humans when assigning data to their labels. Nor can it be applied to teaching methods postulated for students when the teacher presents specific contexts from which the student draws only to the extent he deems sufficient (Thomas, Brown, 2011).

The issue of machine learning, especially in the context of language models, seems to raise questions about the relationship of the categories used here to the communicability criteria commonly adopted in education and science (Kulczycki, 2017). Already in supervised learning systems, the correctness of an answer means its compliance with the collected selection of examples, and incorrectness implies the lack of such compliance. This means that correctness does not have to be identical to truthfulness and

any law or principle – it is a statistical probability. Unsupervised learning does not even contain an element of evaluation regarding the degree of correctness because the system divides the elements of the set into groups based on the standard features it searches for and interprets them as equivalent in value. Compared to training, reinforcement learning brings to mind a system of punishments and rewards that seems somewhat derived from behaviorism.

The use of artificial intelligence to improve the processes described above deepens the outlined tendencies to prefer majority judgments or conclusions based on the number of phenomena. Making the way AI processes information similar to the mechanisms of the nervous system when determining such priorities means that the correctness of the effect of reasoning – which cannot be traced in such a situation – may constitute an unsolvable puzzle for humans.

Therefore, the expectation of potential future usefulness in the case of machine learning remains outside the context of morality, which is an essential element in communication and education. This is probably because applications are trained to serve as tools. However, approaching them this way implies further doubts, making the issue more complex.

If a machine, after „training,” is to become a tool, who is to be its creator and user? Humans have traditionally made tools to aid humans; their criterion was utility. Therefore, machines taught by providing information selected by humans could, after training, be a valuable tool for interpersonal communication. However, its usefulness for humans may be questionable when AI trains a tool. After training, such a tool will probably be able to serve the AI algorithm, while AI is assumed to be used by humans. However, the lack of knowledge about how artificial intelligence obtains results may raise doubts about the possibility of using them even if it recognizes them as correct (Kasperska, 2017).

This situation is related to the black box problem, widely described in the literature, when the input and output data are known, but there is no information about the process connecting them that takes place inside (Chojnowski, 2019). However, since a person cannot determine how the result resulted from the cause, he cannot be sure of its value.

Another issue arising from the above is algorithmic decision-making (ADM). AI can make them and use the acquired knowledge. However, suppose it collects it based on rules set by itself. In that case, the decision it makes regarding a human seems to carry a danger related to its consequences for the human – the so-called risk of algorithmic discrimination (Tolan, 2018).

Although research based on modern knowledge proves the lack of awareness and the ability to think abstractly in a classical machine undergoing training and in generative AI, the nature of AI reasoning processes based on neural networks is no more evident than human thought processes. The worldview structure underlying a human decision and the system of values an individual realizes determines the clarity of their intentions, intentions, and actions. In the case of AI, the decision may depend on specific assumptions formulated by a human. Still, the inability to trace the process results in ignorance as to other assumptions made by the AI, for example, based on the criterion of the most excellent repeatability and the resulting higher effectiveness of the solution. This also shows the issue of responsibility (co-responsibility) for a decision of this type (Sierocka, 2016) when it is made by a human based on an AI recommendation and concerns other people.

3. Ethical aspect of using Big Data systems

The specificity of machine learning raises problems proportional to the advancement of IT systems, which are not observed in modern school education. One of them is the issue of liability related to algorithmic decision-making mentioned above. To be effective, neural networks learn by analyzing giant data sets (Big Data). Although such learning ensures higher effectiveness of solutions, it loses the opportunity to determine precisely what factors influenced the choice. It is, therefore, impossible to decide on the reason for a possible wrong decision.

Another issue is the strengthening of prejudices and the repetition of harmful content. When applications learn from the wide range of data they access, they also absorb the biases and other negative social phenomena contained therein, without interpreting them but reproducing and strengthening them. In the absence of guidelines regarding taking into account at least the system of values applicable in a given culture, which is part of the cultural code, this may lead to the spread of discrimination and violence. Therefore, compliance with the assumptions of pedagogical axiology seems to be particularly important in the machine learning of systems trained for use in education (Maj, 2016).

Data collection by machines during training should not violate the privacy of the people whose data the application uses. However, private data helps make solutions as effective as possible. As a result, machines that need a considerable amount of data can use all the information they have access to for training; they can collect, analyze, and process them in various ways without the consent or even the knowledge of the people who introduced them. It also risks revealing private information about users, including their data, when the cyber machine responds.

To avoid the problems mentioned above, the European Union adopted the AI Act, which specifies prohibited practices in using AI and the method of approving high-risk software. Prohibited practices include:

- artificial intelligence systems using subliminal, manipulative techniques to shape the behavior of individual people or communities, making conscious decision-making difficult;
- biometric categorization systems based on race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation (except for network filtering in cases of law violations);
- AI systems using information about age, disability, social problems, or economic situation – showing significant social harm;

- artificial intelligence systems that rate or classify individuals or groups based on social behavior or personal characteristics, which results in harmful or disproportionate treatment in unrelated contexts or is unjustified or disproportionate to their behavior.⁵

Finally, it is worth emphasizing that a cybernetic machine may also be susceptible to attacks and manipulations during learning; it then collects incorrect training data. Even if the application software is protected against attacks, the environment from which it draws data may not meet the security requirements. Another possibility is the possible inconsistency of the environment with the programmer's expectations, resulting, for example, from the protection of some potentially needed data by their owners and, as a result, a numerical advantage in the advertising content searched by the application. In such a situation, a properly constructed application will be „unconsciously” trained to return incorrect solutions from the point of view of its creators, becoming a source of disturbances in the communication process.

Bibliography

Artificial Intelligence Act, European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). (From:) https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf (access: 15.03.2024).

Bishop, J.M. (2018). Is Anyone Home? A Way to Find Out If AI Has Become Self-Aware, *Frontiers in Robotics and AI*, 5. <https://doi.org/10.3389/frobt.2018.00017>

Chojnowski, M. (2019). *Zrozumieć decyzje podejmowane przez maszyny.* (From:) <https://www.sztucznainteligencja.org.pl/badacze-z-google-brain-opracowali-system-pozwalajacy-wydobyc-z-modeli-si-informacje-o-stosowanych-kryteriach-oceny/> (access: 15.02.2024).

Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world.* New York: Basic Books.

Finnis, J. (2001). *Prawa naturalne i uprawnienia naturalne.* Warszawa: Dom Wydawniczy ABC.

Fłasiński, M. (2020). *Wstęp do sztucznej inteligencji.* Warszawa: Wydawnictwo Naukowe PWN.

Summary

Artificial intelligence can be constructed considering its ethical aspect, which determines its use with respect for man as a person in the social and cultural dimension. It is essential to conduct machine learning and use artificial intelligence responsibly and according to moral standards.

AI is a valuable tool in education and can bring numerous benefits in terms of communication for students, teachers, and other participants in educational processes. Still, at the same time, it raises many ethical and legal dilemmas. Compliance with appropriate regulations and shaping social awareness are necessary to ensure its use under the system of values applicable to culture. It is essential to provide users with comprehensive information about how their data is collected, processed, and used by applications. The prohibition on using AI for purposes harmful to people and society contained in the AI Act and the conditions for approving high-risk artificial intelligence systems must be respected.

Hoes, F. (2019). *The Importance of Ethics in Artificial Intelligence.* (From:) <https://towardsdatascience.com/the-importance-of-ethics-in-artificial-intelligence-16af073dedf8> (access: 2.02.2024).

Kamiński, E., *Uczenie maszynowe: z nadzorem i bez nadzoru.* (From:) <https://analitik.edu.pl/uczenie-maszynowe-z-nadzorem-vs-bez-nadzoru/> (dostęp:10.03.2024).

Kasperska, A. (2017). Problemy zastosowania sztucznych sieci neuronalnych w praktyce prawniczej. *Przegląd Prawa Publicznego*, 11.

Kulczycki, E. (2017). *Komunikacja naukowa w humanistyce.* Poznań: Wyd. IF UAM.

Maj, A. (2016). Aksjologia pedagogiczna. (In:) K. Chałas, A. Maj (eds.), *Encyklopedia aksjologii pedagogicznej*, Radom: Polskie Wydawnictwo Encyklopedyczne.

Massey, G., Ehrensberger-Dow, M. (2017). Machine learning: Implications for translator education. *Lebende Sprachen*, 62(2).

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C.E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, 27(4), 12. <https://doi.org/10.1609/aimag.v27i4.1904>

Reinforcement Learning and the Importance. (From:) <https://datascience.eu/machine-learning/machine-learning-for->

5 Artificial Intelligence Act, European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). (From:) https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf (access: 15.03.2024).

- [humans-part-5-reinforcement-learning/](#) (Accessed: January 20, 2024).
- Okoń, W. (1998). *Nowy słownik pedagogiczny*. Warszawa: Wydawnictwo Akademickie „Żak”.
- Sala, K. (2017). Przegląd technik grupowania danych i obszary zastosowań, *Spoleczeństwo i Edukacja. Międzynarodowe Studia Humanistyczne*, 2(25).
- Sierocka, B. (2016). Etyka współodpowiedzialności czyli moralność wywiedziona z międzyludzkiej komunikacji. *Rocznik Bezpieczeństwa Międzynarodowego*, 10(1), 186–196.
- Starczewski, A., Goetzen, P., Er, M.J. (2020). A New Method for Automatic Determining of the DBSCAN. *Parameters, Journal of Artificial Intelligence and Soft Computing Research*, 10(3).
- Thomas, D., Brown, S.J. (2011). *A New Culture of Learning: Cultivating the Imagination for a World of Constant Change*. Lexington, KY: Creative Space.
- Tomasello, M. (2002). *Kulturowe źródła ludzkiego poznawania*. Warszawa: PWN.
- Wiener, N. (1961). *Cybernetyka i społeczeństwo*. Warszawa: Wyd. Książka i Wiedza.