Why Are We Not Teaching Machine Learning at High School? A Proposal

Ignacio Evangelista, Germán Blesio, Emanuel Benatti
Department of Physics – Polytechnic Institute of Rosario
National University of Rosario
Rosario, Argentina
nachoeva@fceia.unr.edu.ar

***Abstract*—Although there are many online resources for youngsters to start learning about machine learning on their own, a majority of them require solid mathematical or programming background. In addition, there is a lack of material on how to effectively teach machine learning. This paper is motivated by a survey conducted among young students in a technical high-school to inquire on their interest in learning how computers learn. Given the fact that almost every high-school student has heard about artificial intelligence and is curious about it, this article proposes a way to give a friendly introduction to machine learning in the context of a short workshop. Through a series of problem-based activities, students are expected to understand the foundations of what does 'learning' mean for a computer. In addition, through analogies as well as toy and real problems, this short workshop will tackle students' preconceptions, give them an insight of what tools are important in order to deal with this popular topic and address the ethical issues that arise. Moreover, it is desirable that this stimulates students' interest for STEM degrees. Finally, the activities proposed can be easily adapted for an introductory engineering course.**

Keywords—machine learning; artificial intelligence; problem-based learning

I. Introduction

Artificial Intelligence (AI) is a discipline that has the potential of helping to improve people's life quality. It is an exciting topic that has changed the way we live. Rich data sources and powerful data analysis are shaping our everyday lives [1] and transforming the way in which we do science [2]. Notwithstanding, lacking proper background, the artificial intelligence that surrounds us is seen as a kind of black-box magic.

Machine Learning (ML), in some of its flavors, be it Data Science, Data Mining or Statistical Learning is part of every degree in Computer Science, Information Theory, and Electrical Engineering, among others.

Even for a university student coming from another branch different than Computer Science, it is still possible to learn ML at their own pace thanks to a wide range of resources available on the Internet. MOOCs on ML can be easily found on the major platforms: Coursera, EDX or Udacity.

But think about a freshman or a high-school student, they will certainly struggle if they picked up a ML book. This can be discouraging. ML is a complex world and algorithms are just the tip of the iceberg; there is also a solid mathematical and statistical background that is often let aside.

Unfortunately, there is a deficiency of literature addressing the proper way to teach ML [3], perhaps trusting excessively on online resources for self-paced learning. Whereas there is a large number of literature in computational thinking and computer science education, there is a lack of research on how we educate in ML. Indeed, ML articulates with different disciplines and should be tackled as a whole.

We argue that ML is an exciting topic that can be delivered at an early age by using an active learning methodology [4]. In this paper, we present a proposition for a short workshop meant for high-school students that aims at teaching the intuition between some ML aspects. This workshop is intended to demystify ML and to provide sensible insight into the topic.

The proposition is based on a survey conducted among young students to inquire on their perceptions on ML. The results of the survey motivated the following question: How can we address the machine learning topic in order to arouse interest on the topic but without relying excessively on mathematical jargon?

II. Motivation

We conducted a short survey among high-school students with the purpose of exploring their conceptions on AI and ML. The survey was answered by 98 students from 14 to 17 years old. The survey contained Likert-scale questions (1-totally disagree, 5-totally agree) and open questions.

First of all, students were asked whether they had heard about AI and ML. Results are shown in Figure 1. It is remarkable that, whereas the majority of them has heard about ML, fewer know the concept of ML. Indeed, half of the students have scarcely heard about ML.

When inquired about the words that they associated to AI among a set of given concepts, most of them (90%) chose robotics, programming and computers. Around a half of the students marked mathematics. Furthermore, between 15% and 20% of the students selected one of the suggested possible fields of applications of ML: cars, business, psychology, feelings. Only 18% of the students chose ethics as a concept linked to AI.

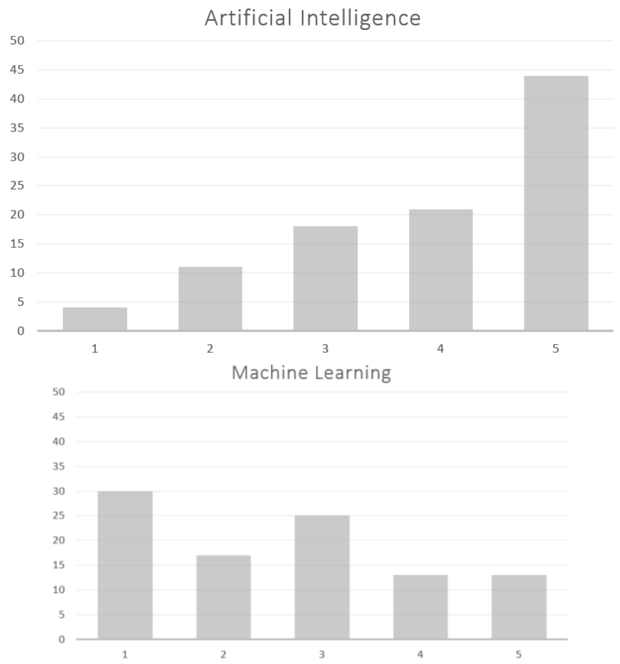

Fig. 1. Answers to the questions "Have you ever heard about AI?" and "Have you ever heard about ML?". Answers range from totally disagree (1) to totally agree (5).

We gathered information on what does learning mean for a computer. Interestingly, 30% of the students related "learning" to "data" and "information", which are essential concepts for ML. However, around 16% of the students explicitly stated that ML was about memorizing information. Around 10% said that the ability to learn implied "doing things automatically". Another 10% of the students compared ML to human learning. 15% mentioned that learning (for computers) is related to "improving" and "being able to do more things". Probably based on their experience with recommender systems, 12% of the surveyed teenagers stated that ML implied learning human behavior.

The surveyed students express a great interest on learning ML as shown in Figure 2. This large level of agreement is the main reason for envisioning an extracurricular course.

We observed that students focus on particular applications on ML: robotics or recommender systems. While it is true that these fields of application benefit from ML, we would like to deliver a broader picture of what ML is, in order to lay the foundations for further self-guided exploration.

III. Workshop Proposal

We provide a thorough description of the activities involved in the workshop so that the reader can understand the topics that are aimed for and the resources that are used to deliver them in a gentle manner.

The main objectives of this workshop are:

- Prove that ML aims at providing a model for understanding nature based on data;

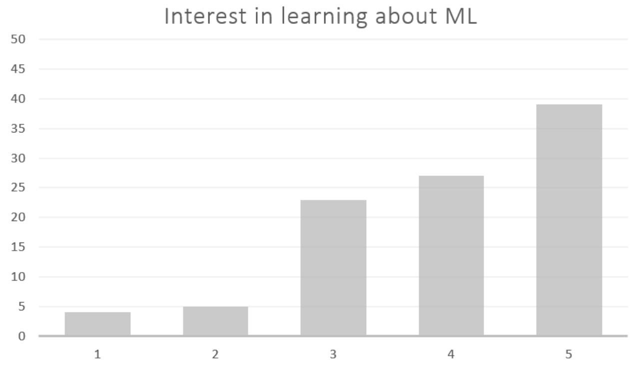

Fig. 2. Answers to the question "Would you be interested in learning about ML?". Answers range from totally disagree (1) to totally agree (5).

- Make students understand that ML is about pattern recognition, mathematics, algorithms and statistics;
- Provide insight on the most important concepts in ML;
- Show that there are similarities and differences in the way computers and people "learn".

The workshop is suited for a group of around 15 students regardless their background knowledge. The approach for most of the activities is an active one, intended to boost students' participation. There are some side benefits to this approach, namely: encouraging teamwork and sharpen problem-solving skills. The activities are grouped in four sessions, each of them targeting different concepts.

The first session is based on a toy problem: "Sophie's dad loves making cookies; she has found out she likes most of the cookies but there are some she would rather give to her brother. She keeps a diary where she writes down the size of the cookie and the number of chocolate chips and tag it as like or dislike". Students are delivered a set of toy cookies with their tags (called "observations"). In groups of 5, they are asked to find any regularity in Sophie's taste. During this task, instructors just provide minimum guidelines but students are encouraged to propose a solution by themselves.

The only hint that is given is to "plot" the size of the cookies and the number of chocolate chips (these are the "variables") and to indicate the label. Given the plot (shown in Figure 3) it is easy to observe a pattern and students are likely to describe the behavior: "Sophie does not like cookies that are too big and have few chocolate chips". How can this be done using a computer? Assuming that students know that computer can make comparisons, this exercise is followed by a description of the key algorithm for the workshop: decision trees.

We believe that decision trees are well-suited for this workshop since they are simple to build, they involve simple rules and they can be represented in a plot.

Indeed, we have a set of data elements for which we have measured two "variables" or "features" and labeled them (i.e. we have given them a "class"). Decision trees are based on

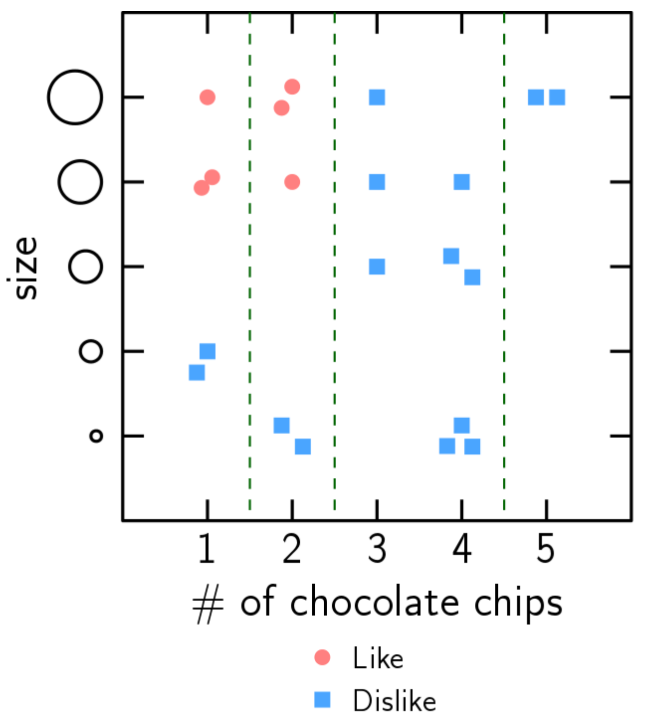

Fig. 3. Plot of the size of the cookies and the number of chocolate chips of the first dataset, as could be obtained by the students. The blue squares represents cookies that Sophie likes, and the red circles the one she does not like.

comparisons of variables with a threshold: therefore, given three possible cut points for the number of chocolate chips which one is the most appropriate? This questions allows for intuitively building up the concept of entropy as a measure of the "class disorder" in the two regions determined by the threshold. After choosing the best threshold for the number of chocolate chips, the procedure is repeated with the size of the cookie and the tree is done. It is easy to see that this set of steps (i.e. an "algorithm") can be easily programmed in the computer so that it can infer the pattern in the data through a series of comparison rules: at each step, the computer seeks the comparison that enables the largest "information gain". In fact, students with some programming experience are encouraged to code the algorithm; nonetheless, the choice of the best threshold can even be done using a spreadsheet.

The main take-away idea is that it is possible to build a model that can "learn" a set of rules from data to describe a certain behavior. What is this useful for? We expect that upcoming cookies will fall into the behavior modeled by the decision tree so that we can make predictions on future data: that is what supervised learning is. It is implicitly hypothesized that there is a black-box process [5] or a phenomenon that labels the data that is collected and therefore it is expected that the model is appropriate for unseen data [6]. The extent to which the model emulates nature's box is a measure of how well the model can reproduce the natural phenomenon producing the data [5].

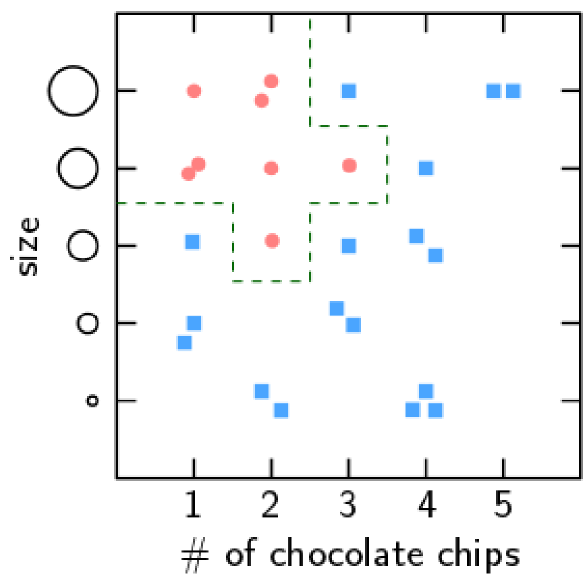

Fig. 4. Plot of the size of the cookies and the number of chocolate chips of the second set of data, that allows to show the risk of overfitting.

It is obvious that this dataset lacks something that is characteristic of experimental measurements: uncertainty (or "noise"). Thus, students are afterwards presented with a more realistic dataset that will enable the introduction of new concepts.

Given the cookies plotted in Figure 4 students are once again asked to find the pattern. Whereas it is possible to come up with a conclusion like "Sophie does not like cookies that are too big and have few chocolate chips unless they have size 6 and 5 chocolate chips or size 4 and 5 chocolate chips". This is an overcomplicated statement: there is "overfitting". Overfitting is like preparing for an exam by memorizing all the examples and thus being unable to generalize to unseen problems. It is possible to prevent overfitting by "pruning" a decision tree.

Complex models, while good to "fit" the data, are usually not good enough to generalize to new samples since they are too stiff. Occam's Razor encourages simple models even if that implies some errors. That allows to define the "accuracy" of the model as well as to build the "confusion matrix" and to define other metrics in terms of "false positives" and "false negatives": "precision and recall".

Likewise, there are some cases in which it is impossible to fit all the data, no matter how complex the model that is used. Students' groups are presented with different samples from a new dataset (Figure 4) each and are asked to find a model that they consider good. They are not asked to follow the algorithm so that they can possibly propose different options. How can they assess whether their model is good or not? They are asked to tell their colleagues "how good" is their model.

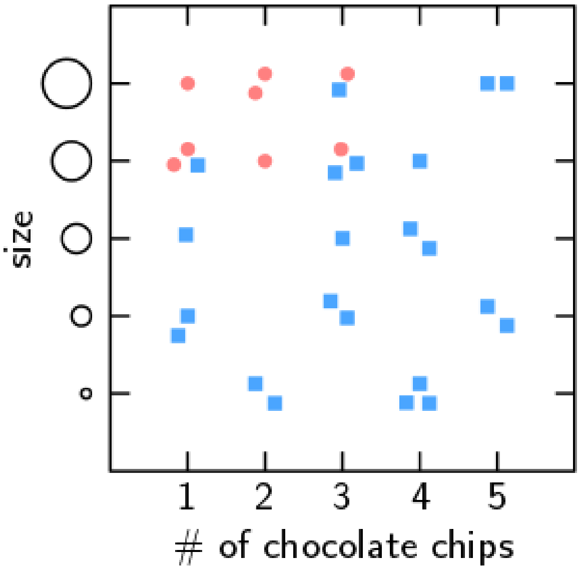

Fig. 5. Dataset in which perfect class separation is not achievable.

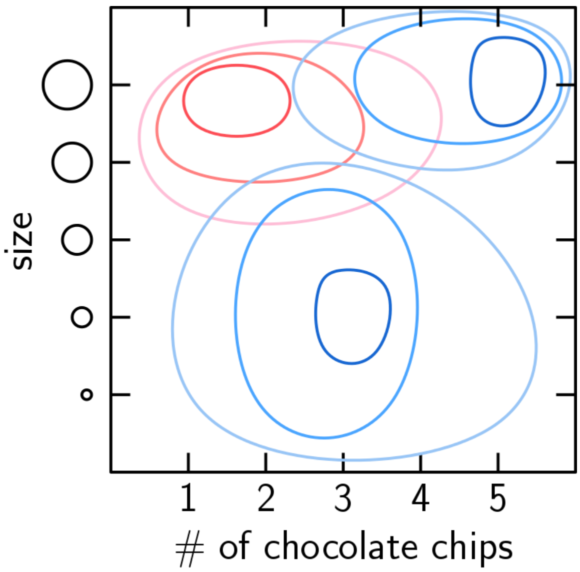

Fig. 6. Continuous representation of the probability distribution that generates the cookies. The results should be discretized to generate the dataset.

It is evident that the accuracy of the model is an indicator of the performance of the classifier. However, the model was fit to that data and we cannot say if it will show a comparable behavior on unseen data. Indeed, we assume that the sample that we have is representative of the true distribution of the labels. That is to say that we expect that unseen data will behave similarly and thus, that the model will behave comparably well. In some cases, it is not possible to simply collect more data so an estimate of the performance must be provided. That leads to the concepts of “train and test splits”, students are asked to randomly split their sample in two sets (it is possible to work on how to do that, maybe through a coin or a dice) to use one to fit the model and to use the other one to evaluate their model. They are finally given a new set of cookies and they are asked to evaluate their model: was the accuracy predicted from the test set a good approximation of the true behavior on unseen data?

In order to clarify on the last dataset: samples are generated from a probability distribution (Figure 6) and then the set is split into four parts, one for each of the groups and the last one for the last evaluation.

To sum up, the first session, which is the most concept-intensive one, focuses on the concept of a learning algorithm and of how it is fitted to data. In addition, following a path made of datasets that ranges from a very simple one to a one that can be separated accurately, different concepts are introduced: overfitting, accuracy and train-test splitting. There are also some concepts like sampling and probability distributions that are skimmed through.

The second session is based on a very powerful concept in machine learning: bagging. This idea postulates that complex models can be combined in order to improve performance. Consequentially, it is possible to benefit from the decision trees from session one to build random forests.

The dataset used in this session will be the last one from the first session. Each student will be presented with samples from the dataset and will be asked to train a tree on that model, encouraging overfitting. While this may seem confusing, bagging support this procedure and tends to attenuate the effect by averaging the trees. The forest is built by voting of each of the trees and it will be evaluated using a new sample from the same dataset (actually, the same group of samples in which the models were evaluated by the end of the first session).

This exercise demonstrates how complex models can be added up together to make a better model and possibly prevent overfitting. Additionally, it permits talking about averaging personal estimations to draw a group conclusion, i.e. the “wisdom of the crowds” [7] and even the “regression towards the mean” (which is a marvelous opportunity for lightly introducing the concept of sampling distributions).

With the objective of deepen the understanding of the random forest model, we encourage working with a more complex set of data. To achieve that, we will add two features to the cookies problem: the time that they were kept in the oven and the amount of butter. By doing this, the original problem becomes much more complicated and additionally it is not possible to solve it graphically. In groups of three, students will be assigned two out of the four features and a set of samples, and will be asked to fit a tree on that data. Having five trees, we can provide a test set and make each sample go through each of the trees so that we finish with five outputs. The number of trees that predicted a certain class is an indicator of the certainty of the prediction made by the forest. The threshold depends on the tolerance for false negatives and false positives. It is interesting to discuss what applications have higher costs for finding false positives or false negatives.

By the end of this session, students will have gained valuable insight into an extremely strong concept in ML: bagging. Bagging improves existing models (in terms of accuracy) at the cost of losing interpretability [8].

The third session is about clustering: the problem of identifying groups or clusters of data points [9]. In the clustering scenario, we do not assume that data are labeled and we would like grouping elements that are similar and making dissimilar items fall into different categories. The applications of clustering are almost endless, during this session we encourage discussion on the possible uses of clustering. We emphasize that not only is it about enabling the computer to “learn” the different categories or clusters but also about providing humans with valuable knowledge about the world.

The scenario of clustering is also that of “unsupervised learning”: we try to identify the underlying structure of data so as to increase our understanding.

The dataset used for this session consists on measures of two dimensions of oranges and lemons [10], we want to identify groups of fruits and would expect at the end of the day to recover the oranges on one hand and the lemons on the other one (although we assume that we do not know the classes). The clustering algorithm par excellence is k-means.

The k-means algorithm consists in iteratively identifying, for one center, which of the samples are closer to it than to any other one; and for each group of samples belonging to the same group, computing the new center. This procedure can be easily illustrated in a 2D plot. There are some discussions that should be encouraged, namely: “how are the initial centers chosen?” and “when does the iterative process stop?”. At this point, by focusing on the algorithmic aspect of ML, students should understand that the procedures are flexible and allow some degree of tuning that will have an impact on the final results. Students are encouraged to “try” the algorithm graphically.

For the last session, students have to search for applications of machine learning in science or industry and give a presentation to their colleagues. Rather than properly understanding the methods and techniques used, we prefer to focus on how powerful and heterogeneous the applications are. Then, we would like to address the ethical issues involved in ML, such as: data collection and privacy, transparency, manipulation or failures [11]

Finally, we consider it important to discuss similarities and differences between ML and human learning. ML certainly focuses on the ability of generalizing rather than merely learning by memorization [6]. As well as computers, humans learn from data collected with experience. ML is useful for tasks that are beyond human capabilities (very large and complex datasets) or for those that are performed routinely but with people not having the capacity to elaborate a well-defined model [6]. In some sense, as well as human learning, ML is characterized but adaptivity, i.e. algorithms adapt to input data. Nevertheless, ML fails to replicate human ability to learn a new concept from very few examples and to build rich causal models of the world that support explanation and understanding [12], [13], [14].

Ultimately, students should understand that ML is both a discipline *per se* and a conjunction of programming, statistics, probabilities and science in general. Deep understanding of the topic requires further deepening on these building blocks: learning a programming language, notions on probability and statistics, and the scientific discipline in general.

IV. Conclusions

We believe that, being ML an extremely popular topic, there is no reason for not teaching it from an early age. We proposed a set of problem-based activities [15] to make students get in contact with the concepts in ML that we consider essential. Through several simple albeit theoretically rigorous activities, we expect to raise awareness on the fact that ML is a blend of disciplines and thus to provide the tools for further exploring the topic. Indeed, after this workshop, students will very likely have a better understanding of what does it mean for a computer to learn and what are the tools required to make it happen.

The activities presented above might be regarded as too simple to apprehend the potential of ML. Nevertheless, we find simple examples the most adequate in order to lay the foundations for further learning. It is expected that students will end the workshop having a toolbox to confront a more advanced course (such as an online one). It is evident that this short workshop is not enough

There are some simple concepts that were not included in this workshop, namely: k-nearest neighbors or the perceptron (as a prelude to neural nets). Ideas related to model evaluation such as cross-validation could have been included too. The selection of topics is a trade-off between simplicity and usefulness. It is important to keep in mind that we prefer a discover activity rather than an exposition. Reinforcement learning and recommender systems might be included in the workshop provided that the content is adapted to the audience. For these two topics, it might be beneficial to have some programming knowledge.

While not disruptive, this paper tries to develop a strategy to tackle ML education and hopes to spur discussions on what is the best way to teach how computers learn.

The lack of experiences at high school make this paper a unique contribution. This innovative short workshop is expected to be carried out during the beginning of 2019.

The choice of an academic degree could be powerfully influenced by the experiences and learning outcomes of secondary school. Generating enthusiasm for learning scientific subjects could possibly influence the decision of choosing a scientific degree. In addition, addressing technological trends at high school awakens curiosity and can be regarded as an appropriate means of stimulating interest for STEM subjects. Should it not be for activities delivered at an early stage, some students may never get in contact with some topics and therefore will not be able to consider a related field of study.

Furthermore, adapting this course for an introductory engineering course is simple. The aforementioned exercises can be enhanced by virtue of some prior background knowledge of a programming language.

REFERENCES

[1] E. Brynjolfsson, A. McAffee. The second machine age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & Company, 2014.

[2] R. Kitchin. Big Data, new epistemologies and paradigm shifts. Big Data & Society, 2014, vol. 1, no 1.

[3] A. J. Ko. We need to learn how to teach machine learning. Medium, 2017.

[4] M. Prince. Does active learning work? A review of the research. Journal of engineering education, 2004, vol. 93, no 3, p. 223-231.

[5] L. Breiman et al. Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science, 2001, vol. 16, no 3, p. 199-231.

[6] S. Shalev-Shwartz, S. Ben-David. Understanding machine learning: From theory to algorithms. Cambridge University Press, 2014.

[7] J. Ugander, R. Drapeau, C. Guestrin. The wisdom of multiple guesses. En Proceedings of the Sixteenth ACM Conference on Economics and Computation. ACM, 2015. p. 643-660.

[8] L. Breiman. Bagging predictors. Machine learning, 1996, vol. 24, no 2, p. 123-140.

[9] C. M. Bishop. Pattern recognition and machine learning. Springer, 2006.

[10] https://homepages.inf.ed.ac.uk/imurray2/teaching/oranges_and_lemons/oranges_and_lemons.pdf

[11] N. Bostrom, E. Yudkowsky. The ethics of artificial intelligence. The Cambridge handbook of artificial intelligence, 2014, vol. 316, p. 334.

[12] B. M. Lake et al. Building machines that learn and think like people. Behavioral and Brain Sciences, 2017, vol. 40.

[13] J. B. Tenembaum et al. How to grow a mind: Statistics, structure, and abstraction. Science, 2011, vol. 331, no 6022, p. 1279-1285.

[14] B. M. Lake; R. Salakhutdinov, J. B. Tenembaunm. Human-level concept learning through probabilistic program induction. Science, 2015, vol. 350, no 6266, p. 1332-1338.

[15] C. E. Hmelo-Silver. Problem-based learning: What and how do students learn?. Educational psychology review, 2004, vol. 16, no 3, p. 235-2.