Project Convergência do Saber: Exploring Artificial Intelligence Fundamentals

Juan Marques Jordão[1], Eduardo do Valle Simões[1]

[1]Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo (USP)
Avenida Trabalhador São-carlense, 400 - Centro CEP: 13566-590 - São Carlos - SP

juanmarquesjordao@usp.br, simoes@icmc.usp.br

Abstract. *The integration of AI in education raises ethical and cognitive challenges not yet addressed by the Brazilian curriculum (BNCC). To bridge this gap, Project Convergência do Saber introduces AI fundamentals to public school students, blending Paulo Freire's pedagogy with unplugged activities and practical tools like Scratch and Teachable Machine. By minimizing programming complexity, the curriculum prioritizes critical AI literacy and conceptual understanding of neural networks and supervised learning. Error is embraced as an epistemic learning tool, while ethical reflection permeates discussions on algorithmic bias and opacity. Ultimately, this framework provides an accessible and scalable model for democratizing AI education.*

1. Introduction

The rapid integration of Artificial Intelligence (AI) in education has transcended simple tool usage, sparking urgent debates on cognitive offloading and the erosion of critical thinking [Fischer et al. 2023, Luckin et al. 2016]. While the Brazilian National Common Curricular Base (BNCC) emphasizes digital culture, it lacks specific guidelines for AI literacy, a gap that exacerbates educational inequalities in public schools [Brasil 2018, CGI.br 2022].

This paper presents a pedagogical curriculum design titled *Projeto Convergência do Saber*, that means integration of technical, social, and experiential knowledge, to enable deeper and more critical learning, aimed at high school students with ages from 16–18 in São Carlos, Brazil. Grounded in Paulo Freire's critical pedagogy [Freire 1996] and Constructivist Project-Based Learning, the framework shifts the focus from AI as a "black box" to AI as a human-constructed, fallible system. Unlike traditional literacy programs, our contribution lies in three core pillars. First, it provides a contextualized scaffolding through a sequence of unplugged and block-based activities explicitly designed for low-resource environments. Second, it emphasizes the epistemic role of error, using machine learning's iterative optimization as a metaphor for human knowledge construction. Third, it promotes sociopolitical literacy by moving beyond functional tool usage to address AI as a sociotechnical infrastructure involving labor, bias, and power dynamics [Holmes et al. 2022, Selwyn 2019].

The ethical dimension is transversal to all classes, addressing topics such as bias in datasets, misinformation, limitations of generative models, and the need for critical validation of responses produced by AI systems. This approach is aligned with the Grand Challenges in Computing defined by the Brazilian Computer Society and with international discussions on AI Education [SBC 2026, Holmes et al. 2022].

2. Methodology

The project proposes the introduction of fundamental concepts of AI and Machine Learning through a practical pedagogical approach that combines unplugged and computational activities. This strategy is aligned with studies in computing education, indicating that alternating between concrete and digital activities supports the understanding of abstract concepts, especially in basic education contexts [Bell et al. 2009, Long e Magerko 2020]. The methodological focus is on the conceptual understanding of model training processes, emphasizing the internal functioning of these technologies (such as data, parameters, and adjustment cycles) rather than mathematical or technical complexity. This approach makes the content accessible to the target audience, as recommended by international guidelines for AI education in K–12 settings, which is a policy document that establishes guidelines and standards for computing education [Touretzky et al. 2019].

The methodology adopts error as a structuring element of the learning process, particularly in the context of generative models and systems based on iterative learning, in which the continuous cycle of attempt, error, and adjustment is an essential part of the model operation. This perspective is connected to constructivist theories and to the literature on learning from failure, which recognizes error as a fundamental component for the consolidation of more robust and transferable mental models [Kapur 2016]. In the AI context, this process is explored pedagogically to demonstrate how data, labels, and feedback influence system behavior, bringing human reasoning closer to the functioning of machine learning algorithms [Mitchell 1997].

This iterative learning cycle is articulated transversally with the ethical discussion about the development, training, and use of AI systems. Recent research in AI applied to education indicates that understanding the limits, uncertainties, and potential failures of models is a determining factor in preventing the uncritical use of these technologies by students [Selwyn 2019]. By making explicit the role of error and system limitations, the course seeks to demystify the notion of algorithmic neutrality and to foster an investigative, critical, and reflective stance toward the results produced by computational models.

The ethical dimension is integrated throughout the course, with the objective of promoting the responsible use of technology, preventing discriminatory or exclusionary practices, and aligning system development and use with current legislation and international guidelines. As a normative reference, the principles of the ACM Code of Ethics and Professional Conduct are adopted, widely recognized in the computing and AI education literature, addressing aspects such as professional responsibility, fairness, transparency, privacy, and social impact [ACM 2018]. This approach is consistent with international recommendations that advocate the explicit incorporation of ethics into introductory AI curricula, especially in basic education [SBC 2026].

2.1. Class Structure

2.1.1. Methodological Design and Progression

The framework follows a qualitative-applied approach, blending Project-Based Learning with Freirean critical pedagogy [Freire 1996]. Content is organized in a concrete-to-abstract sequence: each module starts with an unplugged, analog activity to reduce cognitive load, followed by computational practice to formalize AI concepts (Table 1).

This iterative design treats error as an epistemic element, bridging human learning with machine optimization [Kapur 2016].

Table 1. Course schedule and module organization, detailing the progression from unplugged activities to digital practices.

Class	Main Topics	Unplugged Activity	Computational Practice
1	Intro to AI & Supervised Learning	Act 1: The Learning Model	Pract 1: The Intelligent Character
2	Classification & Societal Bias	Act 2: Classification of Digital Content	Pract 2: Implementing the Classifier
3	Unsupervised Learning (Clustering)	Act 3: Entertainment Clusters	Pract 3: Algorithmic Playlist Organizer
4	Artificial Neural Networks	Act 4: Human Neural Network	Pract 4: Mini Neural Network

2.1.2. Tools and Environments

Computational activities utilize Scratch[1] and Teachable Machine[2], chosen for their visual, block-based interfaces that minimize syntactic barriers. This setup allows students in low-resource public schools to focus on high-level decision logic and model behavior.

2.2. Class 1 – Introduction to Artificial Intelligence and Supervised Learning

2.2.1. Activity 1 – The Learning Model

A student assumes the role of the ”learning algorithm”, like a AI model, while the class provides examples, simulating training data and immediate ”yes/no” corrections for a supervised feedback. The student has a limited number of attempts to infer a secret binary rule based on peer inputs.

Crucially, the student’s inevitable initial failures are not treated as mistakes, but as structural epistemic elements required for optimization [Kapur 2016]. Once the rule is inferred, the ethical dimension is explicitly introduced: the teacher problematizes rules based on stereotypes, allowing students to directly experience how historical or social biases encoded in training data automatically propagate into discriminatory models.

2.2.2. Practice 1 – The Intelligent Character

The computational practice consists of developing a game in Scratch in which a character interacts with the user through binary questions. The responses are processed by a model

[1] `https://playground.raise.mit.edu/create/`
[2] `https://teachablemachine.withgoogle.com/`

previously trained on the Teachable Machine platform, and performance is recorded by counting correct and incorrect predictions. This activity allows students to observe, in a concrete way, the influence of the quality and diversity of training data on model behavior, reinforcing the ethical discussion initiated in the unplugged activity.

This practice materializes the abstract discussion of data quality. By intentionally testing the digital model with diverse or ”edge-case” inputs and observing its failures, students concretely validate the unplugged activity's core lesson: a model's worldview is strictly bounded by its training data. This reinforces the sociotechnical reality that algorithmic accuracy are fundamentally human design responsibilities, setting the ethical foundation for the rest of the curriculum.

2.3. Class 2 – How Models Classify

2.3.1. Activity 2 – Classification of Digital Content

This activity demonstrates how AI models use multiple pieces of information to categorize data, intuitively introducing the functioning of decision trees. The class is divided into groups, each receiving cards representing digital content from the daily lives of the students. Methodologically, the proposal is grounded in principles of critical pedagogy inspired by Paulo Freire [Freire e Macedo 1987], emphasizing contextualized learning. Examples drawn from social media, music, games, and familiar scenarios are intended to enhance engagement and meaning-making. Additionally, unplugged computing activities support experiential learning of AI concepts, including in settings with limited technological resources [Bell et al. 2009]. Based on the analysis of the cards, the groups identify patterns and propose classification criteria, which are later formalized collectively by the teacher as chained binary decisions, making explicit the “if–then” logic characteristic of decision trees.

Next, the groups repeat the classification using an explicitly constructed decision tree developed by themselves, allowing a comparison between intuitive classifications and rule-based classifications. This stage highlights the advantages and limitations of structured approaches in comparison to informal decisionmaking.

The ethical dimension is addressed by discussing how historical and social biases present in data may result in unfair or discriminatory classifications, emphasizing the importance of data curation, diversity, and proper treatment of training data to mitigate these effects.

2.3.2. Practice 2 – Implementing the Classifier

The computational practice consists of developing, in the Scratch environment, a classifier based on a decision tree. Students receive a base code containing sprites representing different types of content and implement the decision logic using conditional “if/else” structures until reaching the final classification. The activity reinforces the understanding that explicit rules also constitute valid ways of implementing classification models.

After implementation, new sprites that were not part of the original dataset are used to test the generalization ability of the classifiers. The analysis of the results allows discussion of how the model makes approximations based on the training it received and how it should behave when facing unknown cases, introducing the concept of responsible responses and the importance of recognizing limitations to avoid arbitrary or incorrect classifications.

2.4. Class 3 – Clustering

2.4.1. Activity 3 – Entertainment Clusters

Students receive cards representing cultural consumption habits (e.g., movies, music) and self-organize into groups without explicit teacher instruction, simulating unsupervised learning. By observing how patterns emerge purely from data proximity, students contrast this with the rule-based supervised learning explored in previous classes. To deepen the Freirean sociopolitical dimension [Freire e Macedo 1987], the activity critically examines how real-world recommendation algorithms cluster *users*. We discuss the Brazilian General Data Protection Law and the risks of using indirect proxies for sensitive data. Students reflect on how unsupervised grouping is not neutral; it can inadvertently create cultural echo chambers, dictate consumption patterns, and reinforce social marginalization. This highlights that mathematical ”distance” in feature space often translates to real-world social consequences.

2.4.2. Practice 3 – Algorithmic Playlist Organizer

The computational practice consists of implementing, a clustering system that simulates a playlist organizer, in the Scratch environment. Students use sprites representing songs and movies from different genres and receive a pre-trained model for cluster identification. They must implement the grouping logic using variables, loops, and conditional structures. Initially, a fixed number of clusters (k = 5) is defined, covering music genres, film genres, and one group for exceptions or intermediate cases

After implementation, new data outside the original pattern are introduced to evaluate the behavior of the system. The analysis of the results allows discussion of how different coding decisions produce different groupings for the same dataset, as well as the phenomenon of hallucinations in models, showing that in the absence of proper exception handling, systems tend to generate incorrect responses with high confidence. It is emphasized that an ethically appropriate response may be the explicit recognition of the inability to classify certain data.

2.5. Class 4 – Artificial Neural Networks

2.5.1. Activity 4 – Human Neural Network

Students physically enact a feedforward neural network for video recommendation. Individuals act as nodes (three inputs, a two-neuron hidden layer, and one output), using strings to represent weights. By manually passing signals and applying activation thresholds, students experience how simple operations combine into complex decisions.(Figure 1). Crucially, as they struggle to reverse-engineer the final recommendation path, the concept of algorithmic opacity, often cited as a major ethical risk in

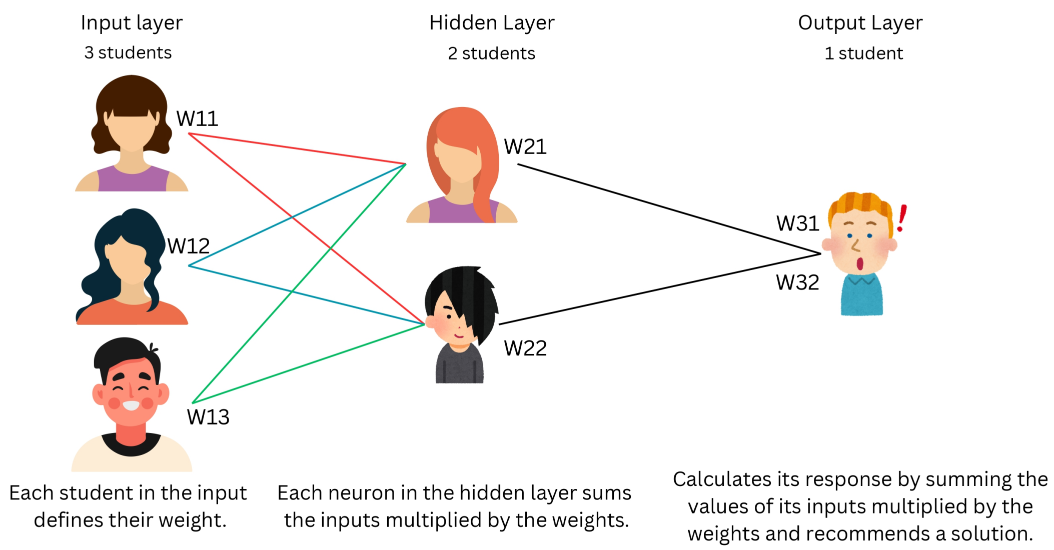

Figure 1. Example of the dynamics of a physical neural network, where the line represents the weight state, and the pulling force will define the weight.

AI [Holmes et al. 2022], becomes a concrete, physical reality. This embodied experience anchors critical reflections on deploying opaque models in high-stakes domains (e.g. healthcare, justice), reinforcing that AI must remain a human-supervised support tool.

2.5.2. Practice 4 – Mini Neural Network

The Scratch practice consists of a simulation in the Scratch environment of a mini neural network for content recommendation, composed of an input layer, a hidden layer, and an output layer. Students manually adjust the input weights through sliders, observing how variations in these values change the behavior of the model, based on the calculation of the weighted sum and the application of the activation threshold.

An output labeled "I do not know" is explicitly included, representing situations in which the network does not have sufficient information to classify the data. This mechanism is used to discuss the phenomenon of hallucination in AI models and the importance of systems capable of expressing uncertainty. At the end, students test the network with data outside the initial set and compare different weight configurations, discussing bias, generalization, and algorithmic responsibility, highlighting that learning in neural networks fundamentally consists of weight adjustment and that network design decisions directly influence the results produced.

3. Expected Results

Previous studies, such as Ojeda-Ramírez et al. [Ojeda-Ramírez et al. 2024], have demonstrated that pairing unplugged logical analogies with Scratch applications effectively fosters AI literacy, particularly among multilingual learners. Project Convergência do Saber builds upon this solid methodological baseline but significantly expands its theoretical scope. While notable contemporary frameworks, such as the John Masla [Masla, J. et al. 2025] , have made crucial strides by embedding applied ethics and stakeholder analysis into foundational AI education, they frequently operate within a design-centric paradigm focused on harm mitigation. In contrast, our curriculum distinctly integrates a Freirean sociopolitical lens [Freire e Macedo 1987]. We leverage the unplugged-to-digital progression not merely to teach computational thinking or evaluate functional ethical impacts, but to emancipate students to interrogate the underlying power matrices, cultural biases, and the epistemic nature of model failure directly within their lived realities.

Consequently, the intentional inclusion of applied AI ethics transitions from a peripheral topic to the central pedagogical driver. The design mandates ongoing reflection on training data curation, the phenomenon of hallucination, and the societal impacts of model opacity [Selwyn 2019].

Finally, the proposal addresses a critical gap in democratizing AI education within resource-constrained environments. By relying exclusively on low-cost analog materials and open-access platforms, and by tightly aligning with the BNCC, this framework offers a highly contextualized and scalable blueprint for public schools. To operationalize this scalability, we propose a decentralized ”train-the-trainer” methodology. Specifically, university extension groups, such as the PET-Computação (Programa de Educação Tutorial em Computação) at the universities of the state of São Paulo or Brazil, will serve as regional deployment hubs. By initially training these university students to act as facilitators, they can subsequently instruct and empower local public school educators, creating a self-sustaining network of teachers with the autonomy to reproduce the curriculum. Future work will focus on this empirical implementation to collect data, validate these theoretical projections, and systematically assess the framework's impact on students' critical AI literacy.

4. Final Considerations

The proposed initiative seeks to support the development of a critical and contextualized understanding of how AI models are built and how they operate in everyday contexts. The methodology integrates technical foundations of AI with the development of ethical and critical competencies, emphasizing that these systems are created by humans and are therefore subject to limitations, biases, and social impacts.

The course is structured progressively, beginning with unplugged activities, moving to visual programming environments, and culminating in the creation of student developed projects. This organization was designed to lower entry barriers and encourage interest in computing as a potential academic and professional field. At the same time, it aims to strengthen conceptual understanding of AI while fostering intellectual autonomy and active participation in learning.

By introducing AI concepts in the early stages of formal education, the proposal aims to support the development of essential twenty first century competencies, including computational thinking, problem solving, and informed decision making. Ethical reflection, awareness of bias, responsible technology use, and recognition of the limits of automated systems are addressed throughout the course, reinforcing the importance of a thoughtful and balanced relationship with digital technologies.

References

Association for Computing Machinery (ACM) (2018). ACM Code of Ethics and Professional Conduct. `https://www.acm.org/code-of-ethics`.

Bell, T., Alexander, J., Freeman, I., and Grimley, M. (2009). Computer Science Unplugged: School Students Doing Real Computing Without Computers. *The New Zealand Journal of Applied Computing and Information Technology*, 13(1):20–29.

Brasil. Ministry of Education (2018). National Common Curricular Base. `http://basenacionalcomum.mec.gov.br`.

Brazilian Internet Steering Committee (CGI.br) (2022). TIC Kids Online Brasil.

Fischer, C., et al. (2023). AI Tools, Cognitive Offloading, and Critical Thinking: Implications for Education. *Societies*, 13(2):45.

Freire, P. and Macedo, D. (1987). *Literacy: Reading the Word and the World*. Routledge, London.

Freire, P. (1996). *Pedagogia da Autonomia*. Paz e Terra, São Paulo.

Holmes, W., Bialik, M., and Fadel, C. (2022). *Ethics of Artificial Intelligence in Education*. Routledge, London.

Kapur, M. (2016). Examining Productive Failure, Productive Success, Unproductive Failure, and Unproductive Success in Learning. *Educational Psychologist*, 51(2):289–299.

Long, D. and Magerko, B. (2020). Foundations of Artificial Intelligence for K–12 Education. In *Proc. Int. Conf. on Computing Education Research*, pages 1–10.

Luckin, R., Holmes, W., Griffiths, M., and Forcier, L. (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson, London.

Mitchell, T.M. (1997). *Machine Learning*. McGraw-Hill, New York.

Ojeda-Ramírez, M., et al. (2024). AI Literacy for Multilingual Learners: Combining Unplugged and Block-Based Programming. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED)*.

Brazilian Computer Society (SBC) (2026). Grand Challenges for Computer Science Research in Brazil. `https://www.sbc.org.br`. Last accessed 2026/01/15.

Selwyn, N. (2019). *Should Robots Replace Teachers? AI and the Future of Education*. Polity Press, Cambridge.

Touretzky, D., Gardner-McCune, C., Martin, F., and Seehorn, D. (2019). Envisioning AI for K–12: What Should Every Child Know about AI? In *Proc. AAAI Conf. on Artificial Intelligence*, volume 33, pages 9795–9799.

UNESCO (2023). Guidance for Generative AI in Education and Research. `https://www.unesco.org`.

Masla, J. et al. 2025. Supporting AI Literacy Teaching Through the Development of Assessments for Classroom Use. Proceedings of the AAAI Conference on Artificial Intelligence. 39, 28 (Apr. 2025), 29178–29185. `https://doi.org/10.1609/aaai.v39i28.35191.`