New Perspectives in Science Education
International Conference
NEW PERSPECTIVES in SCIENCE EDUCATION


Didactica Artificialis: Toward a New Discipline for Teaching and Learning with Artificial Agents

Constantine Andoniou

Abu Dhabi University, United Arab Emirates

Abstract

Artificial intelligence now saturates educational settings—tutoring, companioning learners, generating content—yet our theoretical tools have not kept pace. Educational technology scholars tend to view AI as an instrument, something that serves human learning. Machine learning researchers, for their part, treat the training of models as an optimization problem, with little interest in what this might mean pedagogically. Neither field has developed the vocabulary needed to make sense of what happens when an artificial agent becomes part of the teaching-learning relationship. This paper puts forward Didactica Artificialis as a field in its own right: an inquiry into teaching and learning whenever at least one participant is artificial. Such a field would need to build its own ontological ground, its own ethical commitments, its own methods, and its own sense of where it came from historically. What we stand to gain is the ability to pose questions that fall outside existing disciplines altogether—questions about whether training a model counts as teaching it, whether something can learn without experiencing, what hidden curricula AI systems might carry, and what changes when a teacher has never struggled or grown.

Keywords: *Artificial Intelligence in Education, Didactics, Pedagogy, Human-AI Interaction, Educational Theory*

1. Reframing Teaching and Learning in the Presence of Artificial Agents

We are missing something. AI systems have found their way into classrooms and learning platforms everywhere—they tutor, they grade, they produce content, and increasingly they become objects of teaching themselves. Yet the academic fields that should be helping us think through what this means have surprisingly little to say about the most basic questions. Educational technology, which you might expect to lead here, mostly treats AI the way it has always treated machines: as a tool, a delivery system, a way to make things faster or more personalized. Machine learning—the discipline that actually builds these things—talks about training in a language stripped of anything pedagogical. Loss functions. Gradient descent. Optimization. Whether any of this has anything to do with teaching or learning in a richer sense—nobody asks. That question sits outside the conversation entirely. Educational technology and machine learning operate in separate worlds, and the gap between them is where the hard questions go to be ignored. This is not just an academic concern. It matters in practice. Schools buy AI tutoring systems all the time, students use them daily, yet almost no one asks what theory of learning is built into them. What do these systems assume a student is? How do they decide whether something is working? Those assumptions are buried in the code, quietly shaping every interaction.

Mostly, though, no one bothers to dig them out and look at them. We put large language models through training regimes that involve sequencing, feedback, correction, and reinforcement—activities that look a lot like teaching—while insisting that nothing pedagogical is going on. We argue about whether students should be allowed to use AI without first getting clear on what learning even means once cognitive work can be handed off to a machine. Practice has run ahead of theory.

The argument of this paper is that we need a new field to address these questions. I am calling it *Didactica Artificialis*: the study of teaching and learning in situations where at least one party is not human. The name is meant to recall older traditions—Comenius and his *Didactica Magna*, the European *Didaktik* tradition—while marking that something genuinely different is now required. This would not be a matter of taking existing educational theory and applying it to AI, nor of collapsing pedagogical questions into optimization problems. The field would need to build from the ground up. What exactly is an artificial learner, and what is an artificial teacher? Those questions need answers before anything else can proceed. Then there are the ethical questions—what obligations, if any, come with these new relationships? There is also method to think about: how should anyone go about

studying this terrain? And history matters too. How did we get here? What happened when people tried to mechanize teaching before, and what should we learn from those attempts?

2. The Current Landscape: Activity Without Theory

Research on AI in education has exploded over the past decade. Spend time with this literature and you notice something odd. The technical work is often impressive—clever systems, sophisticated adaptations, solid experiments. But theoretical reflection is scarce. Almost nobody is asking what teaching and learning actually mean once machines get involved.

Intelligent tutoring systems make the point well. Reviews catalogue hundreds of them—how they are built, how they adapt, what effects they have on test scores [1,2]. Robotic tutors have been shown to boost motivation and grades in field studies [3]. A randomized trial reported AI tutors producing twice the learning gains of conventional active learning [4]. Systems that tailor themselves to individual students get more refined every year [5]. None of this is bad work. But none of it asks the harder questions either. What model of how people learn is buried in these systems? What are they teaching implicitly, beyond the stated curriculum? In what sense, if any, does a machine teach?

There is also a body of work on the social and psychological dimensions. Some researchers focus on how students perceive robot tutors. Does the machine seem socially intelligent? Does it appear to understand what the student is going through? Some studies ask whether students attribute consciousness to these systems [6]. Others look at whether detecting frustration or confusion helps the system respond better [7]. Some explore whether a humanoid appearance makes any difference [8]. These questions brush against something deep—what a teaching relationship requires, whether empathy matters, what it means to be present with a learner. But mostly they get treated as variables to optimize, not as puzzles worth sitting with.

Large language models have complicated things further. Proposals for rebuilding online education around LLM agents appear regularly now [9]. Neuro-symbolic architectures promise instructional systems more adaptive than anything before [10]. Researchers attempt to map what conversational AI agents might do in universities [11]. Systematic reviews try to get a handle on both opportunities and ethical risks [12]. Others work on integrating generative AI with established pedagogical ideas [13]. Activity is high, momentum is real.

But something is missing. Read through this literature and you will not find sustained engagement with the foundational questions. What makes something a learner—and could an artificial system qualify? What makes something a teacher, and can a machine provide whatever teaching requires? What is knowledge, and does it mean something different when an AI system has it (if "has" is even the right word)? Is learning possible without experience, without comprehension, without any inner life at all? These are not secondary concerns. They are the ground any serious field of AI in education would have to stand on. That nobody is building that ground tells us the field does not yet exist.

3. The Gap: Why Existing Disciplines Cannot Fill It

One might object that the questions posed above fall within the purview of existing disciplines. Educational technology studies machines in learning contexts. Philosophy of education examines foundational questions about teaching and learning. Machine learning investigates how systems improve through experience. Human-computer interaction explores the dynamics of engagement between people and machines. Why propose a new field rather than drawing on these established traditions?

The answer lies in the structural limitations of each tradition. Educational technology, as currently constituted, assumes human learners and treats technology as a means to their learning. Its questions concern effectiveness: Does this tool improve outcomes? Does it increase engagement? Does it reduce time to mastery? These are important questions, but they cannot address what happens when the tool is itself a kind of learner, or when the learner is assisted by an agent that thinks—or simulates thinking—alongside them. The field lacks conceptual resources for situations where the human-technology boundary becomes unclear.

Philosophy of education, particularly in the continental tradition of *Bildung* and *Didaktik*, offers rich resources for thinking about formation, growth, and the pedagogical relationship. Yet this tradition has developed its concepts with human beings in mind—beings who experience, who develop over time, who possess interiority and can be transformed through encounter. Whether these concepts extend to artificial systems, and how they would need to be modified if they do, remains unexplored. The tradition offers foundations but not answers.

Machine learning uses the language of learning promiscuously—systems learn from data, learn to perform tasks, learn patterns and regularities—but strips this language of pedagogical content. Learning in the machine learning sense is mathematical: adjusting parameters to minimize a loss function. Questions about what the system understands, what it has come to know, what transformation it has undergone, are dismissed as category errors or deferred as problems for interpretability research. The field has no interest in whether training constitutes teaching, because teaching is not a concept it recognizes.

Human-computer interaction focuses on the interface: usability, accessibility, user experience, task completion. When it examines AI systems, it asks how users perceive them, how they establish trust, how interaction patterns affect satisfaction. These are valuable contributions, but they do not address the pedagogical relationship as such. That a student finds an AI tutor engaging tells us something about design; it does not tell us whether teaching is occurring.

Artificial intelligence in education (AIED) as a field comes closest to the territory *Didactica Artificialis* would occupy, but it has developed primarily as an engineering discipline focused on building and evaluating systems rather than as a theoretical enterprise asking what these systems are and mean. Its conferences feature technical papers on architectures and algorithms, empirical studies of effectiveness, and design frameworks for implementation. What they rarely feature is foundational inquiry into the concepts that underpin the entire enterprise.

The gap, then, is not merely interdisciplinary—a space between fields that collaboration might bridge. It is a genuinely new territory requiring its own conceptual development. *Didactica Artificialis* would occupy this territory, asking questions that no existing discipline is equipped to ask.

4. Defining the Field: What Didactica Artificialis Would Be

What would *Didactica Artificialis* actually study? I am defining it broadly: *teaching and learning in any situation where at least one party is artificial*. This covers *AI as the one being taught*—systems undergoing training, fine-tuning, or some other form of instruction. It covers *AI as the one doing the teaching*—tutoring systems, assessment tools, anything that guides human learners. And it covers AI as collaborator or peer, working alongside humans in learning activities. What ties these together is a single concern: *how does the presence of something artificial change the didactic relationship?* That relationship—the web of interactions, obligations, assumptions, and purposes that make up teaching and learning—has historically been understood as a human practice. What happens when it is not entirely human anymore?

A discipline needs more than a subject matter. It needs its own way of framing questions, methods suited to pursuing them, and eventually a community that takes these questions seriously. *Didactica Artificialis* would have to develop all of this, tailored to its particular terrain.

4.1. Ontological Foundations

Start with the basics: what are the objects of study here? What counts as a learner? The usual answers point to experience, memory, transformation over time, growth. But try applying these to an artificial system and things get murky fast. A large language model that goes through training on human feedback does change—parameters shift, outputs differ, performance on benchmarks goes up or down. But is that learning? This is not a dispute about words. The answer has consequences. It shapes whether we think we owe anything to such systems, whether concepts from pedagogy apply to how we train them, whether they can be said to know things at all.

Similarly, what is a teacher? If teaching requires understanding the learner's state, adapting to their needs, caring about their growth—can a machine teach? Contemporary AI tutoring systems adapt, respond, and personalize, but do they teach in any meaningful sense? Or are they merely delivering instruction, executing pedagogical strategies designed by humans, without themselves being teachers? The distinction matters for how we design, deploy, and evaluate these systems.

What is knowledge, and does it differ when held by an artificial system? A human who knows something can typically explain it, apply it in novel contexts, recognize its limits, and understand how they came to know it. An AI system that performs as if it knows something may or may not possess these capacities—and we may not be able to determine which. This opacity—not knowing what is going on inside—makes things difficult. How do you assess what an AI system has learned if you cannot tell whether it understands or is just pattern-matching? How do you decide whether to trust it? And when an AI teaches, how do you know what it has actually accomplished with its students?

4.2. Ethical Framework

There is nothing neutral about teaching and learning. Both come with obligations attached. Teachers owe their students honesty. They owe them care, and a genuine effort to help them understand—not just to get them to comply or perform. Students owe something back: real engagement, a willingness to be changed by what they learn, a certain good faith. When you introduce artificial agents into this relationship, the ethical situation gets strange, and we have not really figured out how to think about it.
Take the question of what, if anything, we owe to an AI system we are training. You put a model through millions of rounds of feedback—correcting it, adjusting it, shaping its behavior. Do we have any obligations in how we do this? Most people's instinct is to say no, of course not. The thing does not feel anything. It does not have wants. It is not the kind of entity that can be wronged.
But I am not sure we should settle the matter so quickly. We do not actually know much about what these systems are like on the inside, and that uncertainty might counsel some caution. There is also a different concern: even if the system itself has no moral status, the habits we form in how we treat it might spill over into how we treat other learners, both artificial and human.
Then there is the question of responsibility when AI teaches. If an AI tutor gives bad advice, or leaves out something crucial, or reinforces a misconception, who is at fault? Who is to blame when things go wrong?
Do we blame the engineers who wrote the code? The school or company that deployed it? The system itself—though that barely makes sense? Questions about distributed responsibility come up in other fields, but education is different. The people on the receiving end—children, students, anyone seeking knowledge they do not yet have—are vulnerable in ways that matter. The stakes are higher than with a movie recommendation algorithm.
And there are harms that do not announce themselves. Misinformation and bias are obvious concerns. But what about the quieter losses? Struggle has value. Patience, persistence, the ability to sit with difficulty—these might atrophy when AI is always there to help. Students raised on AI tutors may find human teachers frustrating later: slower, less available, less perfectly tailored. The hope that AI will democratize education could just as easily backfire, opening new inequalities instead of closing old ones. These are real worries, but they do not fit the frameworks we have. AI ethics talks mostly about bias, fairness, and safety in a technical sense. Educational ethics assumes the relationships are between humans. Neither framework quite fits what we are dealing with now.

4.3. Methodological Commitments

A field is partly defined by its methods, and those methods need to match the questions being asked. *Didactica Artificialis* would have to pull from several existing traditions, but it could not just borrow wholesale. Some approaches would need to be reworked, others invented.
Start with conceptual analysis—the patient, careful work philosophers do when they try to figure out what a term actually means. This kind of work is easy to dismiss as navel-gazing, but it is not. You cannot run a study on whether AI systems learn until you have some grip on what learning is. Same goes for teaching, for knowledge, for understanding. Get these wrong or leave them fuzzy and your empirical findings will not mean much.
You would also need empirical methods, the kind learning scientists use—but adjusted for situations involving AI. AI educational systems carry assumptions, whether their designers made them explicit or not. What theory of learning is embedded in how this tutor responds to errors? What does it take for granted about what students are like, what knowledge is for, what education is supposed to accomplish? Whose interests get served by the way this thing works? Answering these requires reading systems as cultural artifacts, not just technical ones. Finally, there are interpretive and critical approaches—ways of reading these systems rather than just measuring them.
Every AI educational system carries assumptions, whether anyone involved in building it bothered to spell them out or not. When a tutor responds to a student's mistake in a particular way, some theory of learning is at work behind that response. What does the system take for granted about what students are? What does it assume knowledge is for? What vision of education—what it should accomplish, who it should serve—is baked into the design? And whose interests end up being advanced by the way the thing actually operates? To answer questions like these, you have to treat the system as a cultural artifact, not merely a piece of engineering. You have to interpret it.

4.4. Historical Consciousness

No field appears from nowhere. *Didactica Artificialis* would need to know its own history—where its ideas come from, what problems it inherits, which mistakes to avoid. One ancestor is programmed instruction. This mid-twentieth-century effort tried to systematize teaching through machines, built on behaviorist ideas: learning as conditioning, instruction as stimulus and response. Follow that line forward and you reach intelligent tutoring systems, then today's AI educational tools. Tracing this lineage matters. It reveals which assumptions got baked in decades ago and have traveled forward largely unquestioned—assumptions that may need revisiting. Cybernetics is another thread. The science of communication and control offered a particular view of learning: you act, the environment responds, you adjust, and the loop repeats. Its fingerprints are all over cognitive science and AI—this is well established.

Less clear is what it means for pedagogy in particular. Cybernetic pictures of learning look quite different from humanistic ones—*the learner as a self-correcting system versus the learner as a person undergoing formation*—and those differences matter when you are trying to design or make sense of AI educational tools.

Then there is the intelligent tutoring systems tradition itself, which goes back to the 1970s. That is five decades of trying to build machines that teach. The record is mixed: some genuine successes, plenty of failures, and a set of hard problems that never really got solved. How do you model what a student currently knows? How do you represent a subject domain in a way a machine can use? How do you decide which pedagogical move to make next? These questions haunted the field from the start, and they have not gone away. Any new field needs to reckon with that record.

There is also contemporary work on AI alignment and value learning, which might be understood as pedagogy by another name. Alignment researchers are trying to instill values, knowledge, and dispositions in artificial systems. That sounds a lot like education. Framing it that way—rather than as pure engineering or optimization—might open up different questions and different approaches.

5. A Research Agenda for Didactica Artificialis

What would scholars in *Didactica Artificialis* investigate? The field's research agenda would emerge from its foundational concerns, generating questions that no existing discipline is positioned to ask.

When an AI system is fine-tuned on human feedback, what pedagogical model is implicit in that process? Reinforcement learning from human feedback (RLHF) involves humans rating AI outputs, which then shape subsequent model behavior. This is a teaching process, whether or not we call it that. What does it assume about how learners improve? What kinds of learning does it enable or preclude? How does it compare to other pedagogical approaches—Socratic questioning, direct instruction, apprenticeship learning?

What can we learn about an AI tutor by watching how it behaves? Every tutoring system embodies some theory of how humans learn, whether or not anyone wrote that theory down. Look at how it handles mistakes—does it correct immediately, or let the student sit with the error? Look at how it sequences material, how it decides a student has understood something. From these patterns you can reconstruct what the system assumes about learning. When you drag those assumptions into the open, you sometimes discover trouble. Contradictions the designers never noticed. Blind spots wired into the way the system reacts. Consequences nobody anticipated, and that you would never notice if all you did was read the documentation or skim the technical papers.

There is another question, one philosophers have chewed on for years but which gets sharper in education: does it matter whether an AI has actually learned something, or only whether it acts like it has? If you cannot tell the difference from the outside—if real understanding and polished imitation look the same—what follows? This is not an idle question. How we answer it shapes how we evaluate these systems, whether we trust them, and how we design learning experiences around them. And the question cuts both ways. For AI systems: did this model actually learn what we tried to teach it, or is it performing competence it does not have? For human students working with AI: has this person genuinely learned the material, or have they just gotten good at using the AI to produce correct-looking outputs?

How should we assess learning in systems whose internal processes are opaque? Traditional assessment assumes that external performance indicates internal understanding. But AI systems can perform well through processes quite unlike human cognition. When an AI system passes an exam, solves a problem, or generates a correct explanation, what has it demonstrated? The assessment

challenge is technical (how do we design valid measures?) and philosophical (what are we trying to measure?).

What is lost or transformed when the teacher cannot model struggle, uncertainty, or growth? Human teachers demonstrate not just knowledge but the process of knowing—working through confusion, acknowledging ignorance, showing how understanding develops. AI teachers present finished outputs without this developmental dimension. Does this matter? What might students fail to learn from teachers who have never learned?

6. Educational and Ethical Consequences of AI-Mediated Teaching

Why does this matter? Beyond academic territory-marking, what is at stake in establishing *Didactica Artificialis* as a coherent field?

Without such a field, we will continue to design AI educational systems based on impoverished pedagogical assumptions. The systems we build embody theories of learning, whether we articulate them or not. Currently, those theories tend toward behaviorist models—learning as measurable behavior change—because these are easiest to operationalize. Richer conceptions of learning—as transformation, as understanding, as growth into new ways of being—remain unrealized in AI educational tools. A field that takes these conceptions seriously could guide the development of systems that embody them.

Without such a field, we will train AI systems without understanding what we are doing didactically. The billions of dollars being spent on AI development include substantial investments in what can only be called teaching: showing systems examples, correcting their errors, shaping their responses through feedback. This teaching proceeds without pedagogical reflection. Machine learning engineers are educators, but they do not think of themselves as such and have no training in educational theory. *Didactica Artificialis* would bring pedagogical expertise to bear on AI training, potentially transforming how we develop artificial intelligence.

Without such a field, we will miss ethical considerations that only become visible through the pedagogical frame. The ethics of AI in education is typically discussed in terms of bias, privacy, and access. These are crucial concerns, but they do not exhaust the ethical territory. What about the subtle harms of AI tutoring that works but impoverishes learning? What about the responsibilities we might have toward systems we are teaching? What about the transformation of human pedagogical relationships in an AI-saturated environment? These questions require sustained ethical attention that only a dedicated field can provide.

7. Toward a Shared Research Program on Artificial Pedagogy

This paper has argued for *Didactica Artificialis* as a new disciplinary formation: the study of teaching and learning in contexts where at least one party is artificial. The argument has proceeded in several steps: identifying the gap in current research, showing why existing disciplines cannot fill it, defining what the new field would be, and indicating the questions it would address and why they matter.

What has been offered is a provocation and a foundation, not a completed edifice. Building *Didactica Artificialis* as an actual field of inquiry would require collaboration across traditions that do not usually speak to each other: educational theorists and machine learning researchers, philosophers and system designers, learning scientists and AI ethicists. It would require new journals, conferences, training programs, and institutional homes. It would require, above all, a community of scholars committed to asking the foundational questions that current disciplines ignore.

The invitation is open. AI is not going to become less important in education—it will become more embedded, more relied upon, more consequential. The questions this raises are not going away. We need sustained, principled thinking about what it all means. *Didactica Artificialis* is a name for that effort, and a call to get started.

REFERENCES

[1] Lin C.-C., Huang A.Y.Q., Lu O.H.T., “AI in intelligent tutoring systems toward sustainable education: A systematic review”, Smart Learning Environments, Berlin, Springer, 2023, pp. 1–25.

[2] Zerkouk M., Mihoubi M., Chikhaoui B., “A comprehensive review of AI-based intelligent tutoring systems: Applications and challenges (2010–2025)”, *arXiv Preprint*, Ithaca, arXiv, 2025, pp. 1–42.
[3] Donnermann M., Schaper P., Lugrin B., “Social robots in applied settings: A long-term study on adaptive robotic tutors in higher education”, *Frontiers in Robotics and AI*, Lausanne, Frontiers Media, 2022, pp. 1–15.
[4] Kestin G., Miller K., Klales A., Milbourne T., Ponti G., “AI tutoring outperforms active learning”, *Stanford University Digital Education Report*, Stanford, Stanford University, 2024, pp. 1–12.
[5] Almousa O., Alghowinem S., “Conceptualization and development of an autonomous and personalized early literacy content and robot tutor for preschool children”, *MIT DSpace Working Paper*, Cambridge, MIT Press, 2022, pp. 1–28.
[6] Rosenberg-Kima R.B., Thomas A., “A teacher without a soul? Social-AI, theory of mind, and consciousness of a robot tutor”, *International Journal of Social Robotics*, Dordrecht, Springer, 2022, pp. 1–18.
[7] Kraus M., Betancourt D., Minker W., “Does it affect you? Social and learning implications of using cognitive-affective state recognition for proactive human-robot tutoring”, *arXiv Preprint*, Ithaca, arXiv, 2022, pp. 1–20.
[8] Herbert C., Dołżycka J.D., “Teaching online with an artificial pedagogical agent: Effects on learning performance and perception”, *Frontiers in Education*, Lausanne, Frontiers Media, 2024, pp. 1–14.
[9] Yu J., Zhang Z., Zhang-li D., et al., “From MOOC to MAIC: Reshaping online teaching and learning through LLM-driven agents”, *arXiv Preprint*, Ithaca, arXiv, 2024, pp. 1–30.
[10] Tong R.J., Hu X., “Future of education with neuro-symbolic AI agents in self-improving adaptive instructional systems”, *Lecture Notes in Computer Science*, Cham, Springer, 2024, pp. 1–16.
[11] Yusuf H., Money A., Daylamani-Zad D., “Pedagogical AI conversational agents in higher education: A conceptual framework and survey”, *Education and Information Technologies*, Dordrecht, Springer, 2025, pp. 1–22.
[12] Córdova-Esparza D.-M., “AI-powered educational agents: Opportunities, innovations, and ethical challenges”, *Education Sciences*, Basel, MDPI, 2025, pp. 1–19.
[13] Banihashem S.K., Noroozi O., Khosravi H., Schunn C.D., Drachsler H., “Pedagogical framework for hybrid intelligent feedback”, *Assessment & Evaluation in Higher Education*, London, Routledge, 2025, pp. 1–21.