иуащкуhttps://doi.org/10.48417/technolang.2026.03.02
Research article

Learning Machines: From the Automatic Teacher to Large Language Models

Lorenzo De Stefano (✉)
University of Naples “Federico II”, Corso Umberto I, 40, 80138, Naples, Italy
lorenzo.destefano@unina.it

Abstract

A reconstruction is offered of the genealogy of educational automation, personalisation, and epistemic agency. The contemporary debate on large language models in education often treats generative AI as an unprecedented rupture. Yet its promises of personalisation, immediate feedback, and relief from repetitive teaching belong to a longer history. This article reconstructs educational automation from Pressey and Skinner through cybernetics, computer-assisted instruction (CAI), and intelligent tutoring systems (ITS) to large language models. It advances three claims. Educational automation redistributes epistemic labour rather than expressing a linear increase in machine intelligence. Personalisation repeatedly depends on prior standardisation of objectives, knowledge, and evidence of learning. Generative AI introduces a discontinuity by producing the explanatory discourse through which understanding is ordinarily recognised. Bringing these claims together, the genealogy reveals how successful performance can become detached from the formation of judgement. Its critical force lies in examining this tension against educational commitments to developing learners’ capacities, while making the normative priority of epistemic agency explicit. The decisive question concerns how relations among learners, teachers, and machines enable subjects to direct inquiry, evaluate claims, and regulate their dependencies. This criterion remains relevant even if models exhibit limited forms of understanding: the competence of an assisting system does not establish the educational development of its user.

Keywords: Teaching machines; Programmed instruction; Computer-assisted instruction; Adaptive learning; Large language models; Epistemic agency; Personalisation

Citation: De Stefano, L. (2026). Learning Machines: From the Automatic Teacher to Large Language Models. Models *Technology and Language, 7*(3), 10-34. https://doi.org/10.48417/technolang.2026.03.02

© De Stefano, L.This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

УДК 084:37
https://doi.org/10.48417/technolang.2026.03.02
Научная статья

Обучающие машины: От автоматического учителя к большим языковым моделям

Лоренцо Де Стефано (✉)
Неаполитанский университет имени Фридриха II, Корсо Умберто I, 40, 80138, Неаполь, Италия
lorenzo.destefano@unina.it

Аннотация
Современные дебаты о больших языковых моделях в образовании часто рассматривают генеративный ИИ как радикальный разрыв с прошлым. Однако его обещания персонализации, немедленной обратной связи и избавления от рутинного преподавания имеют долгую историю. В данной статье реконструируется автоматизация образования от Пресси и Скиннера через кибернетику, компьютерное обучение (CAI) и интеллектуальные обучающие системы (ITS) к большим языковым моделям. Выдвигаются три утверждения. Автоматизация образования перераспределяет эпистемический труд, а не выражает линейный рост машинного интеллекта. Персонализация постоянно зависит от предварительной стандартизации целей, знаний и свидетельств обучения. Генеративный ИИ вносит разрыв, порождая объяснительный дискурс, посредством которого обычно распознаётся понимание. Соединяя эти утверждения, генеалогия показывает, как успешное выполнение может отрываться от формирования суждения. Её критическая сила заключается в рассмотрении этого напряжения в свете образовательных обязательств развивать способности учащихся, при одновременном прояснении нормативного приоритета эпистемической агентности. Решающий вопрос состоит в том, как отношения между учащимися, преподавателями и машинами позволяют субъектам направлять исследование, оценивать утверждения и регулировать свои зависимости. Этот критерий остаётся значимым, даже если модели демонстрируют ограниченные формы понимания: компетентность вспомогательной системы не устанавливает образовательного развития её пользователя.

Ключевые слова: Обучающие машины; Программированное обучение; Компьютерное обучение; Адаптивное обучение; Большие языковые модели; Эпистемическая активность; Персонализация

Для цитирования: De Stefano, L. Learning Machines: From the Automatic Teacher to Large Language Models // Technology and Language. 2026. № 7(3). P. 10-34. https://doi.org/10.48417/technolang.2026.03.02

© Де Стефано, Л. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

INTRODUCTION: THE DREAM OF THE AUTOMATIC TUTOR

The public release of ChatGPT in November 2022 brought large language models (LLMs) into everyday educational practice and revived old promises in a new form: continuous access to a patient, responsive, and individually adaptable form of "intelligence" that can serve as a personal tutor. Among other functions, generative systems can reformulate explanations, produce examples, ask questions, comment on exercises, and adjust vocabulary and complexity to conversational cues. These affordances appear to address a basic limitation of mass education—the scarcity of sustained individual attention—while also raising questions about reliability, assessment, dependency, and governance (Kasneci et al., 2023; OECD & European Commission, 2026; UNESCO, 2023).

However, the promise of individualised learning is much older. In the 1920s, Sidney Pressey built mechanical devices that administered tests, recorded errors, provided immediate knowledge of results, and required the correct answer before progression. In 1958, Burrhus Frederic Skinner argued that rising educational demand could not be met simply by constructing schools and training teachers; education also required labour-saving equipment. In 1966, Patrick Suppes predicted that computers could eventually provide millions of pupils with something formerly reserved for Alexander the Great: a personal Aristotle, an informed and responsive private tutor (Pressey, 1926; Skinner, 1958; Suppes, 1966). The history of learning machines thus lends itself to a genealogy: every historically designed machine embodied pedagogical models, engineering solutions, and pre-understandings of what the student and the teacher were supposed to be.

Following Matteo Pasquinelli, automation can be analysed as a process in which socially organised labour and knowledge are decomposed, formalised, and made technically executable (Pasquinelli, 2023; Pasquinelli & Joler, 2021). Applied to education, the question becomes: which components of teaching—assessment, correction, reinforcement, sequencing, diagnosis, explanation—must first be isolated and operationalised before they can be delegated to a machine? The answer is never merely technical, because each delegation selects what counts as an educationally relevant operation.

Educational automation is also a problem of cognitive labour. For Pasquinelli, rendering socially organised knowledge technically executable reorganises labour as well as technique (Pasquinelli, 2023; Pasquinelli & Joler, 2021). Once assessment, diagnosis, sequencing, or explanation become machine-executable, parts of teachers' intellectual work become substitutable. In institutions governed by capitalist imperatives of scalability and labour-cost reduction, technical relief can therefore become a means of workforce substitution. In the forms of automated education, the political question is who captures efficiency gains and who retains authority over educational judgement.

In this paper, I adopt a genealogical and hermeneutical standpoint, rather than a derivational one. Pressey's apparatus, Skinnerian programmes, computer-assisted instruction (CAI), intelligent tutoring systems (ITS), and transformer-based LLMs are technically and epistemologically heterogeneous. LLMs were developed as general-purpose statistical systems for generating language, not primarily as teaching machines;

transformer architectures use self-attention to estimate distributions over tokens rather than to execute explicitly authored instructional frames (LeCun et al., 2015; Russell & Norvig, 2021; Vaswani et al., 2017). The point is therefore to identify recurrent educational problems and the redistribution of functions while preserving technical discontinuities.

This genealogy operates on two levels. Historically, it makes different machines, philosophies, and engineering criteria available for comparison: scoring, reinforcement, diagnosis, and generation embody different decisions about what learning is and how to implement it. Critically, it exposes a recurrent tension between the promise to develop learners' capacities and the a priori decision of what education should be and mean in a particular sociotechnical environment. In this sense, a teaching machine embodies political choices. This is a point of entry, not a deduction of educational values from historical facts. The further claim that education should cultivate the capacity to assess reasons and participate in inquiry is a normative commitment of the present argument, developed through epistemic agency below. This approach, therefore, rejects technological neutrality. In different ways, scholars such as Günther Anders (1956) and Langdon Winner insist that technical means are not just means; they cannot be understood as indifferent carriers of independently chosen ends. Winner's argument is especially precise: technologies are "ways of building order in our world" (1980, p. 127), embodying decisions about authority, access, scale, dependency, and competence. Artefacts are also embedded in the institutional and economic arrangements captured by Andrew Feenberg's concept of the technosystem, while their emergent effects can exceed the intentions of both designers and users through what Roberto Redaelli calls preter-intentional effects (Feenberg, 2017; Redaelli, 2025). Educational technologies must therefore be analysed both as products of social organisation and as mediations capable of reorganising it. Nicola Russo's distinction, which reframes the categories of Pseudo-Aristotle's *Mechanica* in terms of the difference between a machine and its mechanism, sharpens the point: a machine cannot be reduced to how its parts operate, because its organisation is intelligible only in relation to the end for which it has been designed (Russo, 2023). Educational machines are consequently materialised theories of learning by design: their architectures embody assumptions about what learning is, which traces count as evidence of it, and which educational functions may legitimately be delegated. However, properties not initially foreseen by the design or by the prior pedagogical theory can also emerge from the human-machine interaction, and it is precisely this domain that a genealogical hermeneutics of technologies can bring to light.

Across the genealogy, the principal object of automation shifts without forming a linear progression. Pressey automated scoring and corrective progression; Skinner organised reinforcement and behavioural shaping; cybernetic systems made feedback adaptive; CAI made performance histories computationally available; ITS constructed explicit learner models and diagnostic inferences; later adaptive systems expanded profiling and selection; LLMs automate the generation of explanations, arguments, counterarguments, calculations, instructional dialogue, and even coding. Correspondingly, the learner appears as an examinee, a behaviourally modifiable organism, a participant in a feedback loop, an information-processing subject, a

computational profile, and, finally, an interlocutor within a hybrid human-machine epistemic process. This is not an incidental effect: we do not merely shape technologies; they also shape us.

The article advances three connected claims. Educational automation redistributes functions rather than following a linear increase in machine intelligence; personalisation repeatedly depends upon prior standardisation; generative AI makes the relation between epistemic performance and epistemic formation newly problematic. Historical studies already challenge narratives of technological rupture and document the politics of personalised learning (Cuban, 1986; Watters, 2021), while contemporary educational research examines LLM opportunities, risks, and instructional design (Kasneci et al., 2023). The contribution proposed here is to connect these perspectives through a comparative analysis of what each configuration delegates, what it recognises as evidence of learning, and who retains authority to evaluate that evidence. This makes visible a problem that evaluations centred on output quality cannot resolve: a system may improve the recognised product of learning while taking over the activity through which judgement was to be formed. The decisive question is therefore how educational arrangements redistribute epistemic labour and enable learners to acquire, exercise, and revise the capacities for which those arrangements are responsible.

To this end, we draw on the concept of epistemic agency: the situated capacity to direct inquiry, evaluate reasons, revise claims, and regulate reliance on others and on technical supports — a capacity that includes the opportunity to exercise these abilities, not merely their individual possession (Stroupe, 2014). Section 9 develops this concept and its consequences for human–machine interaction.

BEFORE THE AUTOMATIC TEACHER: TESTING, EFFICIENCY, AND THE TECHNICISATION OF EDUCATION

Educational technology precedes educational machinery. Examination, curricular sequencing, standardised testing, and the organisation of school time are already techniques for making learning observable and governable (Foucault, 1977; Saettler, 2004). Drawing on Foucault's account, educational practices can also be interpreted as technologies of the self through which subjects participate in their own formation (Foucault, 1988). Modern teaching machines emerged when these practices converged with mass schooling, experimental psychology, and managerial ideals of efficiency. Raymond E. Callahan's "cult of efficiency" documents the entry of business categories into American educational administration, while the expansion of standardised assessment made performance increasingly recordable, comparable, and administratively actionable (Callahan, 1962). In the nineteenth century, Friedrich Nietzsche's critique of education subordinated to state and economic utility already exposed the tension between mass formation and the cultivation of free individuals (Nietzsche, 2016). The mechanical apparatus was therefore preceded by an epistemic transformation whose values could subsequently be sedimented in the machine: before instruction could be automated, learning had to appear as observable, recordable, manageable, and comparable performance.

Edward Lee Thorndike's 1912 image of a book in which each page became visible only after the preceding task had already been completed contained several elements later associated with teaching machines: segmentation, active response, conditional progression, and control of sequence (Watters, 2021). Ludy T. Benjamin likewise shows that nineteenth-century patents included devices for spelling, testing, and self-instruction (Benjamin, 1988). Pressey was therefore not the origin of educational machinery as such. What distinguishes the modern teaching machine is the integration of eliciting a response, evaluating it, and making the next instructional step conditional upon that evaluation.

A further conceptual shift accompanied this technicisation: educational problems were increasingly framed in terms of operations that could be specified, observed, and revised. In this limited sense, the relevant precursor to automation is not any particular machine but the algorithmic form of organisation: a procedure identifies states, conditions, and permissible transitions, joining logic to control (Goffey, 2008). This does not mean that education became literally algorithmic in every respect. It means that those dimensions selected for automation had first to become sufficiently explicit to be represented as executable procedures.

This integration ties technical control to epistemic normativity. In early machines, the relevant norm is explicitly encoded: an answer key, programmed frame, branching rule, or domain model establishes in advance what counts as correct or relevant. This point will matter later. In LLMs, a next-token probability distribution models statistical likelihood but does not by itself constitute a norm of truth, justification, or pedagogical relevance. The genealogy is therefore also a history of changing relations between technical regulation and epistemic criteria.

SIDNEY PRESSEY'S AUTOMATIC TEACHER: INDIVIDUALISATION THROUGH STANDARDISATION

Pressey demonstrated an automatic testing apparatus in 1924 and described it in his 1926 article, *A Simple Apparatus Which Gives Tests and Scores—and Teaches*. A question appeared in a window, and the learner selected one of four answers by pressing a key. In testing mode, the device recorded responses and advanced through the test; in teaching mode, an incorrect response was counted but blocked progression until the correct option was chosen. Immediate correction and conditional progression thus converted automatic scoring into what Pressey called an Automatic Teacher (Benjamin, 1988; Pressey, 1926; Watters, 2021).

The apparatus emerged from Pressey's work on standardised testing and from the labour of administering and scoring it. Automating routine correction, he argued, could free teachers for guidance, motivation, and intellectual stimulation. The promise already combines two recurrent justifications for educational automation: efficiency and the restoration of a supposedly more human pedagogical relation (Watters, 2021, pp. 44–46).

Yet Pressey's individualisation depended upon standardisation. Learners could proceed at different speeds and repeat material according to need, but questions, possible answers, and criteria of correctness were fixed in advance. Technical constraints could even reshape assessment formats: because answer keys and sheets had to align precisely,

standardisation of the materials became part of making the machine workable (Petrina, 2004, p. 323). Personalisation concerned the learner's route and pace through a common normative structure, not the co-construction of the epistemic problem itself.

Pressey linked the apparatus to Thorndike's laws of recency, exercise, and effect: immediate confirmation, repetition, and obligatory correct completion were materially built into the device (Petrina, 2004, p. 316). He later described the wider programme as an "industrial revolution" in education, explicitly connecting labour-saving machinery with the expansion of instruction (Pressey, 1932; Watters, 2015). The rhetoric is important because standardisation, repeatability, and individualisation were not presented as opposites: standardised content and automated routines were precisely what made individual pacing economically imaginable.

The commercial project failed for reasons that included cost, reliability, classroom routines, and the difficulty of aligning curricular practices with the device (Petrina, 2004). Its conceptual legacy was nevertheless durable. Stephen Petrina describes Pressey's ambition as simultaneously to normalise, socialise, and liberate, condensed in the problem of how to "individualize the masses" (2004, pp. 307–308). The learner became technically legible as score, pace, error frequency, and master, in a manner similar to the process by which the worker is incorporated into the industrial assembly line. From a critical standpoint, we are therefore within the ideal of mechanisation typical of the era, and of the automation of cognitive labour and of learning. The machine registered selections and progression; it did not register the reasons by which an answer was reached. In other words, it registered the output, not the epistemic process.

SKINNER: FROM AUTOMATIC TESTING TO THE ENGINEERING OF BEHAVIOUR

Skinner shifted the emphasis from testing acquired knowledge to shaping new behaviour. After observing an arithmetic lesson in 1953, he criticised uniform pacing and delayed correction; although he acknowledged Pressey's priority, he treated the earlier apparatus largely as a testing device. Skinnerian programmes instead presented new material in carefully ordered sequences and required the learner to construct rather than merely recognise responses (Benjamin, 1988; Skinner, 1958).

This design followed operant conditioning. Multiple-choice formats were suspect because they exposed learners to plausible errors and encouraged recognition rather than production. A programme should elicit frequent active responses, provide immediate knowledge of results, arrange material in small steps, and keep error rates low enough for reinforcement to remain effective. Arthur A. Lumsdaine accordingly identified continuous responding, rapid confirmation or correction, and self-pacing as characteristic features of teaching machines (Lumsdaine, 1959). Instruction was operationalised as presentation, response, evaluation, correction, and progression.

The decisive "machine" was therefore the programme rather than the hardware. Knowledge had to be decomposed into ordered discriminations, prerequisites, and terminal performances; sequences were revised when learners repeatedly failed. Teaching labour was not simply removed but displaced upstream into programme design. Jason K.

McDonald's reconstruction shows how programmed instruction linked behavioural assumptions to ideals of social efficiency and to the conviction that educational problems could be reformulated as experimentally tractable design problems (McDonald, 2003, pp. 22–31). Highly specific rules for frame length, error reduction, and sequence revision made the programme empirically corrigible, while also narrowing what could be represented as an instructional objective.

Skinner's epistemology is explicitly articulated in his theoretical works. He argued that not every correct response is a symptom of understanding. In *About Behaviorism*, in the chapter on knowing, for example, he distinguishes behaviour shaped directly by contingencies from the ability to state rules describing those contingencies, noting that neither form implies the other. He also distinguishes repetition or appropriate responding from a deeper sense of understanding in which one can identify relevant controlling variables or reasons (Skinner, 1974). His methodological move is therefore not to abolish distinctions among forms of knowing, but to reconstruct them without appealing to an autonomous inner possession that subsequently causes behaviour. Observable and controllable relations among behaviour, rules, and contingencies nevertheless remain epistemically central.

This matters for the politics and phenomenology of the teaching machine. The learner must act continuously, but activity is not equivalent to epistemic self-direction: the field of possible responses, the sequence, and the reinforcements are organised in advance. Skinner himself denied that the apparatus possessed independent pedagogical agency—"The machine itself, of course, does not teach" (1958, p. 971)—because its operation placed the learner in contact with a programme authored by others. The machine condenses prior decisions about what should be learned and how successful behaviour will be recognised. Its *telos* is, so to speak, external.

Criticism of programmed instruction therefore concerned both scientific validity and authority. Behavioural engineering promised systematic and reproducible control of the learning environment, provoking objections about freedom and manipulation (Watters, 2021). Its later decline also revealed the limits of formal personalisation. Learners could proceed at different speeds and, in branching systems, along different routes, but the relevant distinctions and terminal behaviours remained preselected; empirical research did not uniformly confirm the superiority of the most rigid programming rules (McDonald, 2003, pp. 31–42). The lesson is not that structure is educationally illegitimate, but that adaptivity within a programmed space should not be confused with intellectual openness.

The epistemic difference from Pressey is therefore clear. Pressey's machine primarily evaluated whether the learner could identify a predetermined answer. Skinner sought to organise the behavioural sequence through which the response would be formed. The object of automation shifted from scoring to the programmed production of performance.

THE CYBERNETIC TURN: BRANCHING, ADAPTATION, AND CONVERSATION

Programmed instruction was not theoretically homogeneous. Norman A. Crowder’s branching or “intrinsic” programming differed from Skinner’s linear approach, mainly in the importance attributed to errors. Some of them were treated as diagnostically informative. An incorrect response could redirect the learner to remedial material before returning to the main sequence (Crowder, 1960, 1963). Feedback thereby acquired a stronger regulatory role: the learner’s response did not merely receive confirmation but altered the subsequent behaviour of the instructional system.

This is the logic of cybernetics. In Norbert Wiener’s formulation, feedback enables the consequences of prior action to modify subsequent action; applied to education, this makes teaching describable as a dynamically regulated relation rather than a one-way delivery sequence (Wiener, 1961). The shift should not be exaggerated: goals and regulatory criteria still have to be specified. What changes is the temporal organisation of control, which can now respond to the state produced by previous interaction.

Gordon Pask is especially important because he moves from the regulation of responses towards the regulation of a cognitive and conversational relation. In his conversation theory, human educational interaction cannot be adequately reduced to stimulus and response. The learner participates in an exchange in which conceptual relations must be reconstructed and explained, and adaptive teaching concerns the changing relation between learner, task, and instructional strategy (Lewis & Pask, 1965; Pask, 1975). His steady-state approach sought to regulate difficulty so that tasks were neither incomprehensible nor insufficiently demanding.

This makes adaptation theoretically ambivalent. Pask’s experiments indicate that adaptive regulation can improve performance, yet the best performers may also be those predisposed to adopt the learning strategy embodied in the machine (Pask, 1975). The system therefore does not simply adapt to a pre-given learner; it can also shape the learner towards the distinctions and strategies through which adaptation is organised. Personalisation is relational and potentially bidirectional.

Pask’s use of teachback – the learner’s explanation of a conceptual relation back to an interlocutor – also complicates any equation of learning with correct response. After instruction, learners were required to produce a self-constructed account of the conceptual domain, whose content and order were analysed for errors, deductions, inventions, redundancies, and departures from the sequence imposed by the teaching programme. Teachback therefore provided a way of examining how acquired conceptual relations were reorganised by the learner rather than merely whether an expected response could be reproduced (Pask, 1975, pp. 62–67). His notion of a redundant conceptual repertoire is significant: multiple routes to the same relation may appear inefficient, yet can support reconstruction when contexts change or ambiguities arise.

The cybernetic turn therefore changes the technical image of the learner. The system increasingly treats behaviour as information about an evolving relation and uses that information to modify recursively the instruction. But adaptation still presupposes a model of what counts as relevant performance and a criterion by which regulation is

judged. The question becomes not whether control disappears, but how it is redistributed through feedback.

COMPUTER-ASSISTED INSTRUCTION: FROM THE PROGRAMMED SEQUENCE TO THE LEARNER'S HISTORY

Early computer-assisted instruction did not abolish programmed instruction; rather, it expanded its computational capacity. The computer could administer exercises and, crucially, store and process information about previous responses. The learner's history became available as a basis for future instructional decisions. Suppes in 1966 argued that computers could follow each student's successes and failures and use past performance to select the problems and concepts presented next (Suppes, 1966). Individualisation was no longer only self-pacing; it increasingly depended on computational memory.

The same period broadened the repertoire of educational computing beyond drill-and-practice to tutorial systems, problem solving, simulations, dialogue, databases, and games. Yet technical possibility did not automatically produce pedagogical transformation.

Two major projects make this infrastructural transformation particularly visible. PLATO (Programmed Logic for Automatic Teaching Operations), begun at the University of Illinois in 1960, was one of the earliest large-scale university systems for computer-assisted instruction. Rather than a single-purpose teaching apparatus, it developed into an integrated hardware-and-software environment through which different instructional programmes could be delivered. PLATO IV, operational from 1972, connected learners to a central computer through terminals equipped with keyboards and touch-sensitive displays, while the TUTOR authoring language allowed instructors and course designers to construct instructional sequences across different disciplines. The teaching machine was thus becoming less a discrete mechanical device than a programmable instructional infrastructure. Yet much TUTOR-authored material retained conventional drill-and-practice or tutorial forms, illustrating that increased computational capacity did not by itself transform the underlying educational logic (Saettler, 2004).

TICCIT (Time-Shared, Interactive, Computer-Controlled Information Television), conceived by the MITRE Corporation in 1969 and later developed for mathematics and English instruction in community colleges, explored a different dimension of computer-assisted education: learner control. Students interacted with lessons through a television terminal and a keyboard equipped with dedicated "rule", "example", "practice", and "help" commands. They could consult a map of the lesson, choose topics, alter the order of presentation, or skip material. This apparent increase in autonomy was nevertheless bounded by a tutorial structure that remained predetermined by the programmer. TICCIT therefore offers an early instance of a distinction that becomes increasingly important in adaptive educational technologies: control over the path through instructional material does not necessarily entail control over its epistemic structure or aims. Significantly, the project was also intended to demonstrate that computer-assisted education could provide more effective teaching at a lower cost than conventional instruction (Saettler, 2004).

Educational computing also intersected with the broader cognitive turn in psychology. Paul Saettler describes a movement away from treating the learner primarily as a reactive organism towards models concerned with information processing, prior knowledge, strategies, and individual differences. Computers were well suited to this change because programmes could become "response sensitive": instructional decisions could be conditioned by stored traces of the learner's earlier performance (Saettler, 2004). The technical object of personalisation thus shifted from pace alone towards an increasingly articulated representation of the learner.

Institutional and social history further limits technologically deterministic explanations. Larry Cuban shows that classroom adoption depends on schedules, curricular obligations, maintenance, teacher knowledge, organisational incentives, and whether a technology solves problems teachers recognise (Cuban, 1986). Technical capacity alone does not explain patterns of adoption, since advanced technologies may still be used in conventional ways. CAI nevertheless marks a durable epistemic change: past performance becomes a machine-readable resource. The learner is no longer represented only by the current response but by a stored history from which future interventions can be selected. This creates the conditions for a stronger form of personalisation in which the system claims not only to present material differently but to infer something about the learner from accumulated traces.

INTELLIGENT TUTORS, ADAPTIVE LEARNING, AND THE DATAFIED LEARNER

Intelligent tutoring systems developed this point by separating functions that earlier programmes had largely combined. Classical architectures typically include domain expertise, a model of the student, pedagogical strategy, and the interface through which communication takes place (Wenger, 1987, pp. 5–7, 14–21). Saettler describes the learner model as a dynamic representation of the learner's emerging knowledge and skill, while the tutorial component uses that representation to regulate instructional dialogue (Saettler, 2004). Systems such as SCHOLAR, GUIDON, SOPHIE, BUGGY, and John R. Anderson's tutors made diagnosis and knowledge representation central to computational instruction (Anderson et al., 1985; Wenger, 1987).

These systems did not perform the same pedagogical function. BUGGY encoded recurrent arithmetic errors as runnable faulty procedures, treating mistakes as evidence of systematic, internally represented "bugs" rather than as isolated incorrect responses. DEBUGGY subsequently extended this approach into a diagnostic system capable of inferring a student's procedure from observed performance (Wenger, 1987, pp. 153–161, 182). SOPHIE, by contrast, was conceived less as an autonomous tutor than as a reactive learning environment: students could manipulate a simulated electronic circuit, formulate and test hypotheses, request explanations, and receive context-sensitive critiques. SOPHIE-I had neither a student model nor an explicit representation of tutoring strategies; later versions added demonstrations of qualitative and causal troubleshooting (Wenger, 1987, pp. 55, 59–69). BUGGY therefore foregrounded the computational reconstruction of erroneous procedure, whereas SOPHIE organised an environment in

which inquiry could be supported without fully transferring initiative to the system, encouraging learners' autonomy.

Research on aptitude-treatment interactions provides another important stage in this history. Richard E. Clark and Richard E. Snow argued that instructional technologies should be analysed by decomposing programmes into components and examining how their effects interact with learner aptitudes, tasks, and conditions rather than asking only whether a medium works "in general" (Clark & Snow, 1975). Individual differences thereby became an experimentally tractable variable to which instruction might be matched. Personalisation increasingly depended on making both instructional treatment and learner analytically decomposable.

This is an epistemic shift. Earlier machines reacted to manifest answers; these systems attempt to infer an underlying state that explains them. A wrong answer may be classified as evidence of a misconception, missing prerequisite, or faulty procedure, and the learner model is updated accordingly. Adaptivity therefore presupposes a technical epistemology of the learner: in order to decide what should happen next, the system must produce a selective computational account of what the learner is presumed to know.

This cognitive turn does not dissolve formalisation. Domain knowledge, permissible problem states, diagnostic categories, and tutoring strategies must still be represented in machine-actionable form. The learner model is not the learner; it is a task-relative representation constructed from observable traces and data. Personalisation consequently becomes inseparable from datafication: responsiveness depends upon decisions about which behaviours count as evidence and how uncertainty about the learner is represented (Saettler, 2004; Williamson, 2017).

Knowledge Space Theory makes this structure particularly explicit. It represents domains through sets of knowledge states and uses assessment to estimate which state best fits a learner before selecting material compatible with that representation (Doignon & Falmagne, 1985). Such systems are especially straightforward to formalise where competencies can be decomposed, and relations among prerequisites can be explicitly represented. This is a structural design condition, not a general empirical claim that adaptive systems are most effective in such domains.

The distinction between adaptation and autonomy therefore remains essential. A system may offer different paths to different learners while preserving fixed objectives and criteria. Saettler's account of TICCIT already showed that learner-controlled sequencing could remain embedded within a programmer-defined tutorial; ITS deepen the representation of the learner without automatically granting the learner authority over the aims or standards of the process. Personalisation describes a relation between a model and a designed space of possibilities, not necessarily an increase in epistemic self-determination.

At the same time, the history of ITS cautions against reducing educational value to automation. Kurt VanLehn's review shows that tutoring effectiveness varies significantly with interaction and design, while Saettler notes both the promise and the unresolved limits of intelligent tutoring (Saettler, 2004; VanLehn, 2011). The genealogy thus reaches a point at which the system claims to know the learner computationally. LLMs inherit

this ambition but alter the underlying configuration, since they can generate the linguistic content of tutoring even in the absence of comparably explicit learner and domain models.

LARGE LANGUAGE MODELS: FROM PROGRAMMED INSTRUCTION TO GENERATIVE INTERACTION

LLMs introduce a substantial discontinuity without stepping outside the preceding genealogy. A classical teaching machine contains an explicit instructional structure: prompts, expected responses, criteria of correctness, and rules of progression. An LLM instead generates text from parameters learned during large-scale training. Autoregressive transformer-based LLMs use self-attention to model dependencies in sequences and generate each token conditionally on the preceding context (Vaswani et al., 2017). Fine-tuning and preference-based optimisation can shape model behaviour to follow requests, while retrieval-augmented generation and tool use can connect generation to external sources or operations (Lewis et al., 2020; Ouyang et al., 2022; Schick et al., 2023). None of these additions converts the model into a traditional authored programme.

Generative responsiveness is not ideologically neutral in the relevant sense. This does not mean that an LLM is an intentional subject that possesses an ideology. Training corpora, filtering, annotation, and post-training criteria are socially situated, and models can reproduce stereotypes embedded in their data (Blodgett et al., 2020; Coeckelbergh & Gunkel, 2025; Jeoung et al., 2023). Mark Coeckelbergh and David J. Gunkel reject the "mirror" defence that models are neutral because they merely reflect language: bias mitigation through curation, guardrails, or other normative fixes itself requires decisions about which values should be operationalised and who may decide them. In education, generated discourse can therefore reproduce social hierarchies while the systems that constrain it also encode institutional choices about relevance, ethics, and acceptability.

In programmed instruction, pedagogical logic is principally specified before interaction; in an LLM, much of the response is produced at runtime from a learned statistical model conditioned by the current prompt and context. This moves educational automation from the regulation of a predefined path towards the generation of instructional discourse. Many earlier systems selected among authored exercises, hints, branches, or explanations, although simulation-based and dialogue-based tutors could also construct responses from explicit domain representations. LLMs can instead formulate novel explanations, analogies, examples, questions, summaries, and counterarguments without each item having been explicitly written into the system beforehand. This greatly widens the range of domains addressable without pre-authoring each instructional item, thereby increasing the plasticity of interaction and the range of possible uses.

Generation also changes the form of interaction. Learners can ask open-ended questions in ordinary language, challenge an explanation, request reformulation, or redirect the discussion. Yet conversational openness is not equivalent to pedagogical competence. A general-purpose assistant is usually optimised for responsiveness, whereas good tutoring may require withholding an answer, eliciting a reason, preserving productive uncertainty, or maintaining a long-term instructional strategy. Research on

LLM tutoring therefore often reintroduces structures familiar from ITS—domain grounding, pedagogical policies, learner modelling, and guardrails—to constrain generation (Chowdhury et al., 2024; Puech et al., 2025).

The same distinction applies to personalisation. A base conversational model can adapt its vocabulary, tone, and apparent level of explanation to the current exchange without possessing a validated, persistent student model comparable to a traditional ITS. Specialised systems may add profiles, memory, retrieval, or learner models, but these layers should not be confused either with the foundation model itself or with the user "training" it. Prompting can influence interaction, yet it is an unevenly distributed and fragile means of pedagogical control (Zamfirescu-Pereira et al., 2023). Linguistic responsiveness can therefore create the illusion of being understood without establishing that the system has diagnosed the learner's conceptual structure.

This distinction should not be made to settle the broader question of machine understanding. Bender and Koller (2020) argue that training on linguistic form alone cannot supply meaning in the sense of a relation to communicative intentions and the world. Their argument concerns the conditions of semantic grounding; describing an optimisation objective is not, by itself, an inventory of everything a trained model can represent. Reports of abilities appearing at larger scales challenge simple accounts of models as repositories of memorised phrases (Wei et al., 2022). Li and colleagues (2023), for example, found internal representations of board states in a sequence model trained on Othello moves, and interventions on those representations changed its predictions. This is evidence from a restricted synthetic domain, not a demonstration of general linguistic understanding, but it gives the representational counterargument genuine force.

This gap between linguistic responsiveness and epistemic diagnosis is central to the educational problem. Edoardo Loru and colleagues introduced the term Epistemia for a condition in which the simulation of judgement by LLMs can displace the practices through which claims are ordinarily evaluated (Loru et al., 2025). Building on this diagnosis, Walter Quattrociocchi, Valerio Capraro, and Matjaž Perc characterise Epistemia as a situation in which linguistic plausibility substitutes for epistemic evaluation, producing "the feeling of knowing without the labor of judgment" (2025, p. 1). On their account, fault lines between human and artificial epistemic processes concern grounding, parsing, experience, motivation, causal reasoning, metacognition, and value. Whatever one's broader position on machine understanding, the educational point is narrower: a surface or an output that resembles the product of judgement does not by itself establish the conditions under which that judgement is warranted.

In a recent article, Giulia Bini and Walter Quattrociocchi call the recurrent attribution of epistemic capacities on the basis of formally aligned outputs "epistemic pareidolia" (2026). Their historical example is Stanley H. Erlwanger's study of the student Benny in *Individually Prescribed Instruction*. Benny repeatedly satisfied the mathematics programme's mastery criteria and was considered a strong pupil, yet interviews revealed internally consistent but mathematically ungrounded rules. Correct output had functioned as operational evidence of understanding while obscuring the reasoning that produced it (Bini & Quattrociocchi, 2026; Erlwanger, 1973). The example demonstrates that the evidential problem predates AI: surface correspondence and

epistemic opacity can already separate institutional recognition of mastery from conceptual understanding.

LLMs intensify this problem because they can generate not only answers but the discursive signs traditionally used to probe understanding: explanations, justifications, objections, reformulations, and even apparent metacognitive commentary. Pask's teachback criterion therefore acquires a new ambiguity. If the capacity to reconstruct and explain a relation once provided richer evidence than selecting the correct option, generative systems can now produce that explanatory form on behalf of the learner. The issue is no longer only whether an answer is correct, but whose epistemic work the answer evidences.

Recent empirical results reinforce the need to distinguish assisted performance from learning. Hamsa Bastani and colleagues found that access to GPT Base, a general-purpose GPT-4 interface without pedagogical guardrails, improved performance during mathematics practice but was associated with lower subsequent unaided performance than in the control condition; a constrained GPT Tutor designed to preserve learning substantially mitigated this effect (Bastani et al., 2025). The result does not show that generative AI necessarily harms learning. It shows that successful task completion in the presence of AI is not sufficient evidence of acquired competence.

Conversely, Greg Kestin and colleagues reported larger immediate learning gains in less time with a carefully structured AI tutor than with an in-class active-learning condition in a crossover study of undergraduate physics. The intervention, however, used expert-crafted prompts, question-specific solutions, sequential scaffolding, and a purpose-built platform, and the authors caution against generalising beyond the studied conditions (Kestin et al., 2025). Taken together, these studies indicate that "using an LLM" is not an adequate description *per se* of an educational intervention. The relevant unit of analysis is the sociotechnical arrangement—the model, prompts, constraints, content, assessment regime, learner, and teacher—within which generation is organised.

Evidence outside these two settings further differentiates the forms of assistance. In two randomised field experiments in undergraduate computer science, Kumar and colleagues (2024) examined LLM-guided reflection after assignments. In the first experiment, the reflection group performed better on an examination two weeks later; in the second, LLM-guided and questionnaire-guided reflection performed similarly and both exceeded slide review. The comparison supports attention to the (meta)reflective activity organised around a tool, rather than attributing every gain to generation itself. In writing, a qualitative multiple-case study of CyberScholar involved 143 students in grades 7–11 across five US schools. Students described benefits from rubric-based feedback, while teachers and students also identified inconsistent ratings and misalignment with assignment expectations (Zheldibayeva et al., 2026). This preprint paper supplies contextual evidence about use and perceived value; it does not establish a causal effect on durable writing competence.

Philosophy, as a subject, provides an interesting disciplinary comparison because its central activities—interpreting texts, reconstructing arguments, formulating objections, and revising judgements—can be supported or displaced by the same generative system. The emerging literature remains mainly conceptual, design-oriented,

or descriptive. Mouser (2024) defends dialogical writing with ChatGPT as a practice that need not be epistemically inferior to writing alone, whereas Landers (2024) proposes the CUPID (Catch, Utilize, Prevent, Ignore, and Disincentivize) model and recommends low-stakes journals and reflections that reduce incentives for substitutive use of the machine. Lemasters and Hurshman (2025) respond to generated essays by giving renewed weight to oral assessment in university philosophy. In a master's thesis addressing secondary education, Frais (2024) develops ten competence-oriented applications ranging from thought experiments to school-leaving examination questions. A small reflective case report from a first-year philosophy teacher-education course describes students using ChatGPT and Gemini for exploratory questioning, role assignment, group criticism, and the organisation of philosophical essays (Ortiz et al., 2025). These sources document disciplinary designs and an applied classroom experience; they do not yet provide controlled evidence of durable gains in philosophical reasoning. They nevertheless show why substitution must be analysed at the level of the operation: producing a finished essay can displace argument formation, while criticising a generated objection or explaining a concept back to a model can make the learner's reasoning explicit and available for scrutiny.

These examples are illustrative and do not constitute a systematic review, and they do not warrant conclusions across all ages, disciplines, or educational systems. In writing, feedback on a learner's draft delegates a different operation from generating its argument; in language learning, conversational practice differs from outsourcing the production one is learning to perform. These are task-level distinctions proposed for empirical investigation, not established rankings of cognitive risk. The same caution applies to critical thinking: more articulate output is not a validated measure of its development. The claim defended here is conditional: substitution becomes educationally problematic when it displaces a targeted formative activity without providing an alternative way of developing and assessing the relevant capacity or epistemic virtue.

The discontinuity therefore concerns how epistemic criteria are related to technical production, rather than the mere presence or absence of standards. Pressey's key encoded a determinate answer; Skinnerian programmes specified desired responses; classical ITS used explicit domain and learner models. Such standards could themselves be mistaken or educationally inadequate. Conversely, an LLM's training and deployment can incorporate truth-oriented feedback, retrieval, tools, and ethical constraints. Its next-token probabilities nevertheless do not, by themselves, constitute criteria of truth, justification, or pedagogical relevance. The resulting output must still be assessed through appropriate evidence and procedures. Moreover, on the account developed by Quattrociocchi and colleagues, linguistic fit can displace this epistemic evaluation (Quattrociocchi et al., 2025). The educational implication is that neither an authored answer key nor a generated explanation supplies sufficient evidence of the learner's understanding merely by producing a recognisably successful result.

Earlier teaching machines primarily automated the administration, evaluation, and regulation of responses whose relevant norms had been authored in advance. LLMs automate the production of discourse that can itself help define what appears to count as an answer. They can formulate questions, select distinctions, organise arguments, propose

evidence, anticipate objections, and write conclusions. This capacity expands assistance, but it also places the machine within the activity whose educational function is often precisely to form judgement. The genealogy therefore culminates not in an autonomous artificial teacher, but in a more difficult problem concerning the distribution of epistemic labour.

EPISTEMOLOGICAL AND ANTHROPOLOGICAL SYNTHESIS: DELEGATION, FORMATION, AND EPISTEMIC AGENCY

The present genealogy should not be read as a narrative of progressive automation of education leading to dehumanisation. Human cognition and education have always depended on external supports. André Leroi-Gourhan traces the exteriorisation of human capacities, including memory, through technical supports. Andy Clark and David J. Chalmers argue that reliably integrated external resources can become constituents of cognitive processes, allowing cognition to extend beyond the individual organism. Lambros Malafouris likewise treats cognition as enacted through relations among brains, bodies, and things (Clark & Chalmers, 1998; Leroi-Gourhan, 1993; Malafouris, 2013). The relevant distinction is therefore not between pure internal cognition and contaminated technological cognition, nor between a human teacher and an automated one.

In his philosophical anthropology, Arnold Gehlen shows that technical artefacts can substitute for missing capacities, strengthen existing ones, and relieve human beings of effort and worldly and environmental constraints (*Entlastung*) (Gehlen, 1980, pp. 3–4). Delegation is thus not intrinsically pathological. The educational question is which burdens are relieved, at what stage of formation, and with what consequences for the capacities that remain. A calculator used after arithmetic competence has been acquired poses a different formative problem from a device that performs the operation through which that competence was supposed to emerge.

To shed light on this problem, it is necessary to turn to the notions of epistemic virtue and virtue epistemology. Virtue epistemology does not designate a single theory, but a family of approaches that place the abilities and dispositions of the epistemic subject at the centre of the acquisition of knowledge. From a reliabilist perspective, epistemic virtues are truth-conducive cognitive faculties or competences, such as perception, memory, introspection, and inference. Responsibilist approaches, by contrast, understand them as cultivated excellences of intellectual character—such as open-mindedness, intellectual courage, perseverance, and awareness of one's fallibility—that enable agents to pursue truth intentionally and to assume responsibility for the formation and evaluation of their beliefs (Croce, 2017). Shannon Vallor's account of technomoral wisdom provides a complementary ethical perspective on these technological relations (Vallor, 2016, pp. 154–155). Vallor defines *technomoral wisdom* as the technologically situated form of practical wisdom: not a discrete virtue or a body of technical competence, but the integrated moral expertise required to exercise the other virtues intelligently, authentically, and contextually amid technological change. It enables agents not merely to use technological systems effectively, but to judge which ends those systems should serve and what forms of human flourishing their use may enable or obstruct. In the present

argument, this ethical perspective complements epistemic agency by directing attention to the ends and formative consequences of cognitive delegation.

The responsibilist strand of virtue epistemology thus adds a dispositional dimension to epistemic agency. The intellectual virtues concern how a person pursues epistemic goods, rather than the correctness of an isolated output (Zagzebski, 1996). In education, this directs attention to habits of careful checking, willingness to revise, and recognition of one's limits. The epistemology of testimony also prevents responsible inquiry from being identified with verifying everything oneself: Zagzebski (2012) argues that epistemic authority can be compatible with autonomy. The extension to LLM-mediated inquiry proposed here concerns the assessment of dependencies, not an assumption that a model occupies the same testimonial position as an accountable human speaker. Learners need grounds for deciding when to rely on an output, which sources or specialists can support it, what to ask and when judgement should be suspended.

The teacher's role changes accordingly; it is redistributed rather than abolished and aims to foster epistemic virtues. Pressey and Skinner already justified automation by promising to release teachers from routine work; CAI and ITS added courseware design, supervision, and interpretation of learner data. With LLMs, a further function becomes salient: maintaining and validating a shared epistemic situation in which generated discourse can be questioned, verified, contextualised, and connected to the learner's own reasons. Gerd Gigerenzer's account of intelligent engagement with digital systems is relevant here: competence in using an algorithmic tool does not entail the capacity to recognise when its output should be distrusted or when uncertainty should remain unresolved (Gigerenzer, 2022). The educational issue is therefore how epistemic labour is divided among learner, teacher, and machine, and which human capacities that division forms or erodes. This distinction now also appears in policy: the 2026 OECD–European Commission AI Literacy Framework distinguishes AI literacy from mere tool use and places critical evaluation, ethical use, and informed decision-making at its centre (OECD & European Commission, 2026).

Alongside disciplinary expertise, the teacher's role as curator, critic, and guarantor of the epistemic situation entails responsibility for the conditions of inquiry. It does not guarantee that every generated statement is true. Curation requires identifying the capacity an activity is intended to develop, selecting suitable materials, and deciding which operations may be delegated. Criticism requires checking claims against accessible disciplinary sources, reconstructing an inference, testing an example or counterexample, and distinguishing factual error from legitimate interpretative disagreement. Asking the same model to confirm its answer is not independent corroboration. When verification exceeds available expertise, the appropriate response may be to consult a specialist or preserve uncertainty rather than endorse a fluent explanation.

These practices must also become learnable. A teacher can ask students to compare an output with an assigned source, identify unsupported steps, document why they accepted or rejected feedback, and defend a revision in discussion. Assessment can then combine the assisted product with evidence of the learner's contribution, such as an explanation under questioning or a related transfer task. This operational interpretation is consistent with UNESCO's emphasis on AI pedagogy and professional learning as well

as ethics and technical knowledge (Miao & Cukurova, 2024). It is a proposal derived from the preceding argument, rather than a claim that a literacy framework has empirically validated these particular activities. Professional development should provide time for teachers to evaluate outputs together against disciplinary standards and examine student reasoning. The institutional obligation is to resource this work; otherwise the demand for human oversight merely transfers an unmanageable verification burden to individual teachers.

This risk is institutional and anthropological, rather than technologically inevitable. Yet under cost-minimising logics, scalable automation can turn the promise of relieving teachers into a rationale for substituting professional judgement and treating educational labour as an avoidable cost. TICCIT already made lower-cost instruction an explicit objective; LLMs widen the range of intellectual tasks open to such substitution. The issue is whether efficiency remains subordinate to educational aims.

The following table summarises the argument as a redistribution of functions. Each configuration automates a principal operation, makes a particular figure of the learner technically legible, and reconfigures rather than simply abolishes teaching labour. The final row does not represent the arrival of an autonomous artificial educator; it marks the entry of linguistic generation into the educational division of epistemic labour.

System	Principal function automated	Figure of the learner	Reconfigured role of the teacher
Pressey	Scoring, correction, and conditional progression	Examinee represented by selectable responses	Interpreter, motivator, and designer of tests
Skinner	Sequencing, response shaping, and reinforcement	Behaviourally modifiable organism	Programmer of contingencies and instructional frames
Cybernetic systems	Feedback, branching, and adaptation	Self-regulating informational system	Designer of interaction and regulatory criteria
Computer-assisted instruction	Storage and selection based on performance history	Information-processing subject with a data history	Supervisor, author of courseware, and institutional mediator
Intelligent tutoring systems	Diagnosis, learner modelling, and prediction	Computational learner profile	Orchestrator and interpreter of learner data
Large language models	Linguistic generation and conversational assistance	Participant in a hybrid epistemic process	Expert, curator of tasks and sources; critic of outputs; organiser of accountable inquiry and learning environments

CONCLUSION: TEACHING MACHINES AND HUMAN LEARNING

The present genealogy makes visible how learning is itself reorganised through automation. Its contribution is to connect each delegated function to a representation of the learner and to a redistribution of authority over educational judgement. This extends the historical critique of technologies into an epistemological problem: improving a recognised performance may also weaken its value as evidence of formation.

The term "learning machine" has ambiguously referred to two different processes: machines designed to support human learning and systems trained to learn from data. The distinction is important. Parameter optimisation does not by itself entail understanding,

participation in a form of life, or responsibility for belief. Yet machine-learning systems now intervene directly in the practices through which human beings acquire and display knowledge. This is not a neutral development; it is already changing learning and educational practices.

Large language models bring the old dream of the automatic personalised tutor closer to realisation while exposing its central contradiction. The more effectively a system performs operations associated with explanation, argument, and judgement, the less educational success can be inferred from the quality of the output alone, but rather from the entire process. The value of generative AI cannot be assessed solely by correctness, usefulness, or fluency, but also by whether its use helps form subjects capable of understanding, contesting, verifying, and taking responsibility for what is said. The decisive question is not whether education should be technologically mediated, but how mediation should be organised so that assistance does not become substitution and performance does not displace formation or erode epistemic agency. These choices are also political, economic, and normative: they redistribute educational labour and mediate socially situated patterns of representation. This is a criterion for comparing educational arrangements, not an empirical verdict on every use of generative AI. Assistance can form agency when it enables inquiry, reflection, autonomy, curiosity and progressively more responsible reliance. The question remains relevant even where a model exhibits substantial competence, because the capacities of the assisting system and the formation of the learner cannot be inferred from one another.

REFERENCES

Anders, G. (1956). *Die Antiquiertheit des Menschen* [The antiquity of man] (Vol. 1). C. H. Beck.

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science, 228*(4698), 456–462. https://doi.org/10.1126/science.228.4698.456

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185–5198). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.463

Benjamin, L. T., Jr. (1988). A history of teaching machines. *American Psychologist, 43*(9), 703–712. https://doi.org/10.1037/0003-066X.43.9.703

Bini, G., & Quattrociocchi, W. (2026). Epistemia in the classroom: The problem of evidence of understanding in education in the age of generative AI. *AI & Society*. https://doi.org/10.1007/s00146-026-03227-y

Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5454–5476). ACL. https://doi.org/10.18653/v1/2020.acl-main.485

Callahan, R. E. (1962). *Education and the cult of efficiency*. University of Chicago Press.

Chowdhury, S. P., Zouhar, V., & Sachan, M. (2024). AutoTutor meets large language models: A language model tutor with rich pedagogy and guardrails. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale (pp. 5–15)*. Association for Computing Machinery. https://doi.org/10.1145/3657604.3662041

Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Clark, R. E., & Snow, R. E. (1975). Alternative designs for instructional technology research. *AV Communication Review, 23*(4), 373–394. https://doi.org/10.1007/BF02898561

Coeckelbergh, M., & Gunkel, D. J. (2025). *Communicative AI: A critical introduction to large language models*. Polity Press.

Croce, M. (2017). Epistemologia delle virtù [Epistemology of virtues]. *APhEx, 15*.

Crowder, N. A. (1960). Automatic tutoring by means of intrinsic programming. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning: A source book* (pp. 286–298). National Education Association.

Crowder, N. A. (1963). On the differences between linear and intrinsic programming. *Phi Delta Kappan, 44*(6), 250–254.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. Teachers College Press.

Doignon, J.-P., & Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies, 23*(2), 175–196. https://doi.org/10.1016/S0020-7373(85)80031-6

Erlwanger, S. H. (1973). Benny's conception of rules and answers in IPI mathematics. *Journal of Children's Mathematical Behavior, 1*(2), 7–26.

Feenberg, A. (2017). *Technosystem: The social life of reason*. Harvard University Press.

Foucault, M. (1977). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans.). Pantheon.

Foucault, M. (1988). Technologies of the self. In L. H. Martin, H. Gutman, & P. H. Hutton (Eds.), *Technologies of the self: A seminar with Michel Foucault* (pp. 16–49). University of Massachusetts Press.

Frais, J. (2024). Künstliche Intelligenz im Philosophieunterricht: Wie kann ChatGPT kompetenzorientiert eingesetzt werden? [Artificial intelligence in philosophy lessons: How can ChatGPT be used in a competence-oriented way?] (Master's thesis). University of Vienna. https://doi.org/10.25365/thesis.77553

Gehlen, A. (1980). *Man in the age of technology* (P. Lipscomb, Trans.). Columbia University Press.

Gigerenzer, G. (2022). *How to stay smart in a smart world: Why human intelligence still beats algorithms*. Allen Lane.

Goffey, A. (2008). Algorithm. In M. Fuller (Ed.), *Software studies: A lexicon* (pp. 15–20). MIT Press.

Jeoung, S., Ge, Y., & Diesner, J. (2023). StereoMap: Quantifying the awareness of human-like stereotypes in large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 12236–12256). NLP. https://doi.org/10.18653/v1/2023.emnlp-main.752

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports, 15*, 17458. https://doi.org/10.1038/s41598-025-97652-6

Kumar, H., Xiao, R., Lawson, B., Musabirov, I., Shi, J., Wang, X., Luo, H., Williams, J. J., Rafferty, A., Stamper, J., & Liut, M. (2024). Supporting self-reflection at scale with large language models: Insights from randomized field experiments in classrooms. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (pp. 86–97). Association for Computing Machinery. https://doi.org/10.1145/3657604.3662042

Landers, C. (2024). ChatGPT, the CUPID model, and low-stakes writing. American *Association of Philosophy Teachers Studies in Pedagogy, 9*, 188–204. https://doi.org/10.5840/aaptstudies202472389

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444. https://doi.org/10.1038/nature14539

Lemasters, R., & Hurshman, C. (2025). A shift towards oration: Teaching philosophy in the age of large language models. *AI and Ethics, 5*, 1203–1215. https://doi.org/10.1007/s43681-024-00455-0

Leroi-Gourhan, A. (1993). *Gesture and speech* (A. B. Berger, Trans.). MIT Press.

Lewis, B. N., & Pask, G. (1965). The theory and practice of adaptive teaching systems. In R. Glaser (Ed.), *Teaching machines and programmed learning II* (pp. 213–266). National Education Association.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*, 9459–9474.

Li, K., Hopkins, A. K., Bau, D., Viégas, F., Pfister, H., & Wattenberg, M. (2023). Emergent world representations: Exploring a sequence model trained on a synthetic task. https://doi.org/10.48550/arXiv.2210.13382

Loru, E., Nudo, J., Di Marco, N., Santirocchi, A., Atzeni, R., Cinelli, M., Cestari, V., Rossi-Arnaud, C., & Quattrociocchi, W. (2025). The simulation of judgment in LLMs. *Proceedings of the National Academy of Sciences of the United States of America, 122*(42), e2518443122. https://doi.org/10.1073/pnas.2518443122

Lumsdaine, A. A. (1959). Teaching machines and self-instructional materials. *Audio-Visual Communication Review, 7*(3), 163–181. https://doi.org/10.1007/BF02767482

Malafouris, L. (2013). *How things shape the mind: A theory of material engagement*. MIT Press.

McDonald, J. K. (2003). *The rise and fall of programmed instruction: Informing instructional technologists through a study of the past* [Master's thesis, Brigham Young University].

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391104

Mouser, R. (2024). Writing with ChatGPT. *Teaching Philosophy, 47*(2), 173–191. https://doi.org/10.5840/teachphil2024429197

Nietzsche, F. (2016). *Anti-education: On the future of our educational institutions* (D. Searls, Trans.). New York Review Books.

OECD & European Commission. (2026). *Empowering learners for the age of AI: An AI literacy framework for primary and secondary education*. OECD Publishing. https://doi.org/10.1787/65cd27d4-en

Ortiz, A., Roldán, N., Díaz, S., Insaurralde, Y., & Mastrocola, P. (2025). Entre reflexión crítica y procesamiento de datos: Una experiencia de estudiantes con IA en filosofía [Between critical reflection and data processing: A student experience with AI in philosophy]. *Desarrollos en Ciencias Sociales Computacionales, 1*(1), e9. https://revistadesarrollos.uflo.edu.ar/index.php/RDCSC/article/view/4

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P. F., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems, 35*, 27730–27744. https://doi.org/10.52202/068431-2011

Pask, G. (1975). *Conversation, cognition and learning: A cybernetic theory and methodology*. Elsevier.

Pasquinelli, M. (2023). *The eye of the master: A social history of artificial intelligence*. Verso.

Pasquinelli, M., & Joler, V. (2021). The Nooscope manifested: AI as instrument of knowledge extractivism. *AI & Society, 36*(4), 1263–1280. https://doi.org/10.1007/s00146-020-01097-6

Petrina, S. (2004). Sidney Pressey and the automation of education, 1924–1934. *Technology and Culture, 45*(2), 305–330. https://doi.org/10.1353/tech.2004.0066

Pressey, S. L. (1926). A simple apparatus which gives tests and scores—and teaches. *School and Society, 23*, 373–376.

Pressey, S. L. (1932). A third and fourth contribution toward the coming "industrial revolution" in education. *School and Society, 36*, 668–672.

Puech, R., Macina, J., Chatain, J., Sachan, M., & Kapur, M. (2025). Towards the pedagogical steering of large language models for tutoring: A case study with modeling productive failure. In *Findings of the Association for Computational Linguistics: ACL 2025* (pp. 26291–26311). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.findings-acl.1348

Quattrociocchi, W., Capraro, V., & Perc, M. (2025). Epistemological fault lines between human and artificial intelligence. arXiv. https://doi.org/10.48550/arXiv.2512.19466

Redaelli, R. (2025). Intentionality gap and preter-intentionality in generative artificial intelligence. *AI & Society, 40*(4), 2525–2532. https://doi.org/10.1007/s00146-024-02007-w

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Russo, N. (2023). Para physin. Spunti aristotelici, e non, intorno al concetto di macchina. *Bollettino Filosofico, 38*, 98–106. https://doi.org/10.6093/1593-7178/10380

Saettler, L. P. (2004). *The evolution of American educational technology* (2nd ed.). Information Age Publishing.

Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Hambro, E., Zettlemoyer, L., Cancedda, N., & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. *Advances in Neural Information Processing Systems, 36*, 68539–68551. https://doi.org/10.52202/075280-2997

Skinner, B. F. (1958). Teaching machines. *Science, 128*(3330), 969–977. https://doi.org/10.1126/science.128.3330.969

Skinner, B. F. (1974). *About behaviorism*. Alfred A. Knopf.

Stroupe, D. (2014). Examining classroom science practice communities: How teachers and students negotiate epistemic agency and learn science-as-practice. *Science Education, 98*(3), 487–516. https://doi.org/10.1002/sce.21112

Suppes, P. (1966). The uses of computers in education. *Scientific American, 215*(3), 206–220.

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190498511.001.0001

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems 30* (pp. 5998–6008). Curran Associates.

Watters, A. (2015, February 4). The automatic teacher. Hack Education. https://hackeducation.com/2015/02/04/the-automatic-teacher

Watters, A. (2021). *Teaching machines: The history of personalized learning*. MIT Press.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). Emergent abilities of large language models. Transactions on Machine Learning Research. arXiv.https://doi.org/10.48550/arXiv.2206.07682

Wenger, E. (1987). *Artificial intelligence and tutoring systems: Computational and cognitive approaches to the communication of knowledge*. Morgan Kaufmann.

Wiener, N. (1961). *Cybernetics: Or control and communication in the animal and the machine* (2nd ed.). MIT Press.

Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. SAGE Publications.

Winner, L. (1980). Do artifacts have politics? *Daedalus, 109*(1), 121–136.

Zagzebski, L. T. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press. https://doi.org/10.1017/CBO9781139174763

Zagzebski, L. T. (2012). *Epistemic authority: A theory of trust, authority, and autonomy in belief*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199936472.001.0001

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Article 437). ACM. https://doi.org/10.1145/3544548.3581388

Zheldibayeva, R., de Oliveira Nascimento, A. K., Castro, V., Cope, B., & Kalantzis, M. (2026). Generative AI feedback, English writing and teacher rubrics: A multiple-case study of CyberScholar. arXiv. https://doi.org/10.48550/arXiv.2605.17055

СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR

Лоренцо Де Стефано, lorenzo.destefano@unina.it
ORCID 0000-0003-3339-2541

Lorenzo De Stefano, lorenzo.destefano@unina.it
ORCID 0000-0003-3339-2541

Статья поступила 5 июля 2026
одобрена после рецензирования 14 августа 2026
принята к публикации 21 сентября 2026

Received: 5 July 2026
Revised: 14 August 2026
Accepted: 21 September 2026