# THE PHILOSOPHICAL FOUNDATIONS OF EDUCATION IN THE AGE OF ARTIFICIAL INTELLIGENCE:

## EPISTEMIC AUTONOMY, AUTHORITY, AND THE IDEAL OF HUMAN DEVELOPMENT

Author:
Dr. Arş. Gör. Buket KAYIŞLI ARKADAŞ

**BİDGE Yayınları**

The Philosophical Foundations of Educatıon in The Age of Artificial Intelligence: Epistemic Autonomy, Authority, and The Ideal of Human Development

**Author:** Buket Kayışlı Arkadaş

**ISBN:** 978-625-8821-72-7

1st Edition
Page Layout By: Gözde YÜCEL
Publication Date: 20.07.2026
BİDGE Yayınları


Certificate No: 71374

www.bidgeyayinlari.com.tr - bidgeyayinlari@gmail.com
Krc Bilişim Ticaret ve Organizasyon Ltd. Şti.
Güzeltepe Mahallesi Abidin Daver Sokak Sefer Apartmanı No: 7/9 Çankaya /Ankara

# PREFACE

The growing use of artificial intelligence technologies in nearly every aspect of our lives is reshaping our views on education. In particular, the use of generative artificial intelligence systems in processes such as knowledge access, text production, learning, assessment and decision-making has not only given rise to the emergence of new educational tools but also to a renewed examination of knowledge, authority, autonomy and the ideal of human development. The genesis of this book was the need to interpret this transformation from the point of view of philosophy of education and epistemology.

Education in the modern era of artificial intelligence is often discussed in terms of the speed, convenience, personalisation and efficiency that technology can bring. But education's basic concern is not just to give a swifter access to knowledge or to enable one to do educational tasks in less time. Education is a multidimensional process of developing the human capacity to think, understand, justify, form values and assume responsibility. The role of artificial intelligence in education must therefore be measured not only by criteria of technical success, but also by its contribution to human epistemic and moral development.

Central to this study is the question of how learners can preserve their ability to think for themselves in the era of artificial intelligence. With standard artificial intelligence systems, students can access a treasure trove of information, alternative explanations, and immediate feedback. But easier access to information does not necessarily mean that the information obtained has been understood, verified or justified. Thus, an imperative of primary education is to keep responsibility for reasoning, verification and decision-making, whilst benefiting from technological systems.

Epistemic autonomy in this book does not mean the individual's complete independence from all sources of knowledge. In the modern complicated epistemic context, we are compelled to rely on teachers, experts, scientific institutions and technological systems. At the heart of the matter is the question of whether individuals can determine which sources to trust, for what reasons, and within what limits. Epistemic autonomy does not mean to refuse help and guidance but to keep a critical, justified and responsible relation to sources of knowledge.

This approach additionally emphasizes the question of epistemic authority. The history of education has established the teacher as the primary epistemic authority in terms of access to knowledge and the interpretation of that knowledge. But this traditional relationship is being altered by the ability of artificial intelligence to respond quickly and fluently to a wide range of topics. But the ability to produce information is not the ability to understand, to take pedagogical responsibility, to see the learner as a person, to assume responsibility for the consequences of decisions. The book discusses the circumstances in which artificial intelligence might be a reliable source of epistemic support and the reasons why it should not be considered an unlimited epistemic authority.

Another core axis of the study is the type of person education should seek to develop in the era of artificial intelligence. The goal of education cannot be lowered to creating people who merely adapt to technological change or who are able to use AI tools effectively. Education should be aimed at producing people who are capable of critical thinking, testing knowledge claims, taking epistemic responsibility, exercising moral judgement, interrogating technological systems and defending human dignity.

In this regard, the book proposes a Human-centred Education Model based on Epistemic Autonomy. It is based on the principles of critical verification, justified trust, ethical responsibility, humancentredness and pedagogical guidance. Its main goal is to maintain the intellectual and epistemic status of the learner without removing artificial intelligence from education. Technological systems can assist learning but they must not become independent authorities that determine learning aims, moral values or final decisions about human beings.

A human-centred understanding of education does not mean hostility to technology. In contrast, it demands that technology be assessed in terms of human and educational ends. Artificial intelligence can help access knowledge, reduce teachers' routine workload and support diverse educational needs. But technology efficiency should not overpower the thinking process of the learner, the teacher-student relationship, pedagogical responsibility or human dignity.

The book is expected to be a contribution to researchers in the fields of educational sciences, philosophy of education, epistemology, teacher education and AI ethics. It also aims to support teachers, students, educational administrators and policymakers to develop a more critical and human-centred perspective in their decision-making around artificial intelligence.

The arguments made in this book are not intended to be definitive and lasting answers to all questions about artificial intelligence. Educational practices and ethical issues around them must be continuously reviewed given the fast pace of development of artificial intelligence (AI) systems. Thus, the approach developed here is to be understood not as a finished and closed system but as a philosophical framework for reflection and discussion for the safeguarding of the place of the human being in education.

The main aim of this study is to bring the human back into the centre of the debate about artificial intelligence. The future of education will be determined not only by what technological systems can do, but by the kind of relationship human beings have with those systems. Artificial intelligence opens opportunities, but technology can only become a real educational value if thinking, justification, responsibility and human dignity are preserved.

My hope is that this book will contribute to the creation of a better, more just, critical and responsible balance between education, humanity and artificial intelligence.

## CONTENTS

INTRODUCTION .......... 8

EPISTEMIC AUTONOMY IN THE ERA OF ARTIFICIAL INTELLIGENCE .......... 13

ARTIFICIAL INTELLIGENCE AND EPISTEMIC AUTHORITY IN THE EDUCATION SYSTEM .......... 62

THE IDEAL OF HUMAN DEVELOPMENT IN THE AGE OF ARTIFICIAL INTELLIGENCE .......... 112

CONCLUSION .......... 164

REFERENCES .......... 180

# INTRODUCTION

The rapid adoption of artificial intelligence into educational processes represents more than the introduction of a new instructional tool. Generative artificial intelligence systems are transforming, in particular, how students access information, ask questions, produce texts, solve problems, choose sources, and receive feedback. These systems can offer personalized explanations and learning support but also have risks such as misinformation, bias, fabricated references, and overreliance (Kasneci et al., 2023). The fundamental educational issue of the age of artificial intelligence is, therefore, not one of whether or not to use technology, but of how to preserve human capacities for thinking, evaluation and responsibility.

Education is more than simply imparting information or doing tasks in less time. It is a normative process by which people learn how to relate to knowledge, which sources to trust, how to form their own judgments, and for what purpose to use what they know. As Dewey (1916) states, "education is the reconstruction of experience and the development of the capacity of thought". In turn, Peters (1966) links education with knowledge, understanding and conscious participation in worthwhile activities. In this vision, artificial intelligence would not be seen as an autonomous authority that defines the goals of education, but as a limited instrument to aid human development.

This book explores the implications of increasing use of artificial intelligence in education for epistemic autonomy, epistemic authority and the ideal of human development. It understands artificial intelligence not only as a technical novelty, but as an epistemic intermediary that shapes students' knowledge acquisition, belief formation and decision making processes. Generative artificial intelligence can shape what information is made visible, how questions are asked and what explanations are plausible. Therefore, AI literacy should consist not only of operational competence but also inquiry, contextualization, and epistemological evaluation (Rapanta et al., 2025).

The core problem the book addresses is how to preserve students' ability to think for themselves, assess claims to knowledge, and take ownership of their final judgments, while leveraging the assistance of artificial intelligence. At the same time, the book investigates how the teacher's epistemic authority is changing and how far artificial intelligence can be accepted as a source of epistemic support in education.

From a philosophical perspective, the book aims to revisit the epistemic and human foundations of education in the era of artificial intelligence. It thus addresses epistemic autonomy as the student's ability to develop a critical and justified relationship to sources of knowledge, distinguishes the epistemic authority of teachers and artificial intelligence, and deals with the kind of person education should foster.

The importance of the study is that it does not limit the relationship between artificial intelligence and education to convenience, efficiency or technological competence. Rather, it brings basic philosophical concepts like autonomy, authority, trust, responsibility and human dignity to the center of the debate. This approach is also in agreement with the view that AI literacy should encompass knowledge, use, assessment, and ethical responsibility (Ng et al., 2021).

The study is focused on large language models and text-generating artificial intelligence systems for use in education. Learning analytics, automated assessment, and recommendation systems are discussed only as they relate to the core problem. Although the discussion is framed in terms of the philosophy of education broadly, many of the examples are drawn from secondary and higher education.

This study is not an empirical field study. No data were collected from the students or teachers and the proposed model was not tested in an experimental way. Instead, it is a theoretical exploration drawing on the literature of philosophy of education, epistemology, social epistemology, AI ethics and educational technologies. Another important limitation is the rapid pace of change in artificial intelligence systems. Thus, the study develops enduring principles such as critical verification, justified trust, oversight by humans, and ethical responsibility rather than rules tied to particular tools.

The approach used was qualitative and theoretical. The concepts of epistemic autonomy, epistemic authority, trust, trustworthiness, epistemic responsibility and human-centeredness were analyzed. Classical works in the philosophy of education were compared with contemporary studies of artificial intelligence and education published between 2020 and 2026.

In the second phase of the research a synthesis was created across different bodies of literature. Discussions about epistemic autonomy were related to AI literacy, discussions about epistemic authority were related to teacher authority and algorithmic authority, and the ideal of human development was related to human-centred AI and ethical approaches to education. This synthesis led to the development of the Epistemic Autonomy-Based Human-Centered Education Model.

The central idea of the book is epistemic autonomy. Epistemic autonomy is not the independence from sources of knowledge but the ability to establish a critical and justified relationship with them. Epistemic authority is the extra weight that is given to a particular source during the process of coming to know, due to the source's superior competence in the relevant domain. Such authority ought to be domain-specific , criticisable and defensible .

Epistemic responsibility is the obligation of an individual to seek evidence for the claims to knowledge that they accept and communicate and to revise their judgments in light of the possibility of error. Justified trust means that confidence in a source should not be based on appearance or fluency, but on verified performance, auditability, and contextual appropriateness. Conversely, human-centered education asserts that technology should be a tool and human development the goal (Shneiderman, 2020).

The book's main argument is that the positive and justified use of artificial intelligence in education requires a human-centered balance that safeguards students' epistemic autonomy, teachers' pedagogical responsibility, and the primacy of human judgment. Artificial intelligence should not be exiled from education, but it should not be turned into a general, unlimited and independent epistemic authority either.

Artificial intelligence will provide students with information, alternative explanations, and feedback; students will be responsible for setting learning goals, validating outputs, comparing sources, and justifying final decisions. In this regard, the preservation of human epistemic agency in the human–AI learning partnership is of fundamental importance (Wu et al., 2025).

Chapter One discusses epistemic autonomy in the age of artificial intelligence. It focuses on thinking for oneself, the justification of beliefs, critical thinking, and epistemic responsibility.

Chapter Two explores epistemic authority, the teacher as a source of knowledge, the distinction between trust and trustworthiness, and the extent to which artificial intelligence can be regarded as an epistemic authority.

Chapter Three is devoted to the ideal of human development in the age of artificial intelligence, the transformation of the conception of the human person, and human-centered education. At the end of the chapter, the Epistemic Autonomy-Based Human-Centered Education Model is introduced. The model is grounded in critical verification, justified trust, ethical responsibility, human-centeredness, and pedagogical guidance.

The Conclusion brings together the discussions of epistemic autonomy, epistemic authority, and the ideal of human development, and proposes a new balance between education, humanity, and artificial intelligence.

# EPISTEMIC AUTONOMY IN THE ERA OF ARTIFICIAL INTELLIGENCE

## The Notion of Epistemic Autonomy

The application of generative artificial intelligence to educational processes has significantly increased the speed and broadened the limits of students’ access to information. However, from an epistemic point of view, the main problem is not the access to a greater quantity of information, but the grounds on which the accessed content is accepted as true. Kasneci et al. (2023) note that large language models offer great opportunities for explanation, feedback, and personalization, but also risks such as misinformation, bias, and overreliance. Therefore, in the era of artificial intelligence, epistemic autonomy should be associated not with students’ independence of technological tools, but with their ability to evaluate knowledge claims put forth by those tools.

Epistemic autonomy is the ability of an individual to exercise rational self-governance over the processes of belief formation, sustenance and revision. Self-governance doesn't require individuals learning everything on their own or refusing outside information sources. Pritchard (2016) argues that intellectual autonomy is compatible with the social nature of knowledge. It is possible for individuals to rely on the testimony of others and still understand the explanatory relations among the knowledge they acquire and assess its justification. Epistemic autonomy is thus not isolation from sources of knowledge, but the capacity to establish a critical and justified relation with them.

This approach shows the inadequacy of thinking of artificial intelligence systems as just tools. According to Rapanta et al. (2025), generative AI literacy needs to encompass not only functional skills of use, but also inquiry, adaptation and epistemological reasoning. Since artificial intelligence can influence what content is made visible, how questions are asked, and what explanations are seen as plausible, it is actively shaping the student's epistemic environment. Therefore, a student's ability to issue commands to a system is not indicative that the student actually controls the epistemic effects of the system.

There are three basic components of epistemic autonomy that can be identified. The first is cognitive competence, that is, the ability to evaluate knowledge claims and evidence. The second is epistemic motivation, which is the desire to search for truth and justification rather than to accept ready-made answers. The third is the normative dimension, which is about taking responsibility for the consequences of the information adopted and communicated. Similarly, Ng et al. (2021) in their AI literacy framework identified technical knowledge, along with critical evaluation and ethical responsibility, as core competencies.

These three components might not develop equally in AI-supported learning. Liu et al. (2025) found students may utilize metacognitive control strategies to assess AI outputs more frequently, while utilizing artificial intelligence more selectively to assess their own learning. This finding demonstrates that checking system output is necessary but not sufficient for epistemic self-governance. Students also need to assess their own knowledge gaps and learning goals, and the situations in which they over-rely on artificial intelligence.

In this sense, epistemic autonomy is not a technological skill but a form of intellectual responsibility. If students consider AI output not as verified knowledge but as a provisional claim to be evaluated, they can preserve their status as epistemic subjects. The critical AI literacy approach developed by Veldhuis et al. (2025) prompts users to consider not only the accuracy of outputs, but also the social, cultural and ethical implications of artificial intelligence systems. Thus epistemic autonomy is not simply about determining individual skills to verify, but about investigating the technological and social conditions under which knowledge is produced.

## The Philosophical Foundations of Self-Determination

In etymological terms, autonomy means the idea that individuals give themselves their own law. But philosophically, autonomy does not mean that individuals can do whatever they want, or are free from all external constraints. In Kant's moral philosophy autonomy means that the will is not determined solely by external commands, rewards, punishments, or personal inclinations, but that the principles of action are subject to rational evaluation. Autonomy here is not to be confused with arbitrary freedom of choice, but rather justified self-governance.

The Kantian approach does not permit the students to be considered as passive objects of direction from an educational perspective. The purpose of education is not simply to teach students to follow the rules given by teachers or institutions, but to develop students' ability to judge the reasons behind those rules. Similarly, Shneiderman's (2020) human-centered artificial intelligence approach argues that technological systems should enhance human control, responsibility, and decision-making capacity. Such artificial intelligence that thinks or decides for the student is thus incompatible with the principles of human-centered education and autonomy.

To understand autonomy as total independence of the individual is to forget the social nature of knowledge. People can't learn complex topics without consulting teachers, experts, scientific institutions, and reliable publications. Dearden (1972) describes an autonomous person, not as someone who can make all decisions independently of others, but as someone who has reasons for what he thinks and does. So educational autonomy is not the abolition of teachers or experts; it means that students must build a reasoned relationship with authority.

Artificial intelligence enters this relationship as a new information intermediary. The system provides many options or the possibility to ask questions at any time, which may seem to increase the freedom of the individual. Roe and Perkins (2025) in their review of the literature on self-directed learning, however, suggest that artificial intelligence only supports learner independence when processes of goal-setting, self-monitoring and evaluation are maintained. If students do not set their own learning goals and do not question the path recommended by the system, then a wealth of options may not lead to real epistemic self-governance.

For this reason, in the modern era of artificial intelligence, we need to distinguish between procedural autonomy and epistemic autonomy. Procedural autonomy refers to a student's ability to decide when, where, and with what tools to study. Epistemic autonomy, as opposed to procedural autonomy, means that the student can critically evaluate the knowledge claims provided by the tool and to construct the justification of their own judgment. Kong et al. (2024) suggest that AI literacy education should not only involve tool usage, but also the evaluation of system effects and project-based problem solving.

Autonomy does not mean absence of external influence. It means the ability of the person to exercise rational self-governance in the face of such influence. This is not to say that artificial intelligence should be seen as inevitably increasing autonomy, just because it offers students more choices. Students must judge both options and sources of information by reasons to be able to have true autonomy. This conclusion requires an elaboration of the distinction between moral and epistemic autonomy.

**Moral vs. epistemic autonomy**

Both moral and epistemic autonomy are based on the notion of self-governance. However, they are concerned with different normative questions. Moral autonomy is the ability to make decisions about what one ought to do through the exercise of reason and moral judgment. Epistemic autonomy is about the assessment of reasons, sources and evidence in the determination of what to believe. The former is concerned with the rightness of action and moral responsibility, the latter with the truth of belief and epistemic justification.

These two forms of autonomy are not completely separate. Human actions are often grounded in beliefs, and beliefs that are false or inadequately justified may have morally troubling consequences. Cotton et al. (2024) argue that academic integrity in the era of generative artificial intelligence cannot be reduced to rule violation detection, but requires the re-structuring of learning and assessment processes. Thus, offering an unscrutinised AI-generated output as one's own acquired knowledge is an epistemic as well as a moral responsibility problem.

Epistemic autonomy is related to truth and moral autonomy to right action, but responsibility should not be transferred in either domain. Generative AI tools blur the boundaries of academic production and relations of ownership (Perkins, 2023). The fact that a technological system has produced a text does not excuse the student from responsibility for selecting, using and presenting that text. Artificial intelligence is not an excuse or a justification, but a tool whose use must be explained by the human user.

The relation between epistemic and moral autonomy becomes clearer when considering bias and error in AI outputs. Farrokhnia et al. (2024) argue that ChatGPT can offer access, personalization and instructional support, but also presents threats such as misinformation, ethical uncertainty and the degradation of critical thinking. Students must therefore think about the implications of using or sharing an answer with others, not just whether an answer is correct.

Epistemic autonomy in this respect is responsibility for achieving true beliefs and moral autonomy is responsibility for the consequences of application of information. This link is shown in the model proposed by Ng et al. (2021) that combines the treatment of AI evaluation and its ethical use. As epistemic evaluation is incomplete without ethical responsibility, ethical use is incomplete without critical evaluation.

Moral autonomy and epistemic autonomy concern different questions, but they are interwoven in AI-supported education. Students need to judge not only if an output is true, but whether their use of that output is responsible. Thinking for oneself is thus not merely an individual cognitive activity, but a practice of responsibility with epistemic and moral implications.

**Thinking for Yourself**

Thinking for oneself does not imply that all knowledge must be first-hand or that expert opinion must be rejected. In complex societies people must of necessity depend on others in the production and acquisition of knowledge. Pritchard (2016) links intellectual autonomy not to independence from sources of information but to the ability to understand knowledge gained and to assess, where appropriate, the justification for it. Thinking for oneself involves the ability to decide whom to trust, when to trust them, and for what reasons.

Generative artificial intelligence is able to produce fluent and persuasive responses that could challenge this ability. Tlili et al. (2023) mention the potential of ChatGPT to act as an assistive educational tool, but the possibility of generating false, fabricated, or biased content requires pedagogical and ethical safeguards. When students confuse the linguistic quality of an answer with its epistemic reliability, they may give priority to the system's output over their own judgment.

Tian and Zhang (2025) reported a negative association between dependence on artificial intelligence and critical thinking, with cognitive fatigue mediating this association. This finding indicates that the type of AI use is more important than the amount of use. A system can be a useful scaffold for intellectual effort, but if it continually performs problem-solving, evaluation, and decision-making for students, it can weaken their cognitive resilience.

Artificial intelligence can also be used in ways that support thinking for yourself. Students with higher levels of AI literacy were more engaged in planning and metacognitive activities when writing academic papers (Kim et al., 2025). This finding suggests that student interaction with artificial intelligence does not need to be passive acceptance. Students with the right competencies could use the system as a thinking partner, a source of critique, or a counterargument generator.

Vendrell and Johnston (2026) argue that a degree of 'cognitive friction' must be maintained in AI-supported learning to protect critical thinking. Students should learn to identify assumptions, weigh alternatives and fix mistakes in AI output, rather than being given a ready-made answer to every query. This design can stop the elimination of intellectual processes in the name of technological efficiency.

Another condition of self-thinking is consciousness of the part played by artificial intelligence in shaping thought. As Rapanta et al. (2025) state, critical generative AI literacy should involve questioning the assumptions and limitations embedded in the outputs of systems. In the meantime, Anlı (2025) suggests that alienation from knowledge production and verification processes may lead to epistemological alienation. In this situation, individuals have access to information, but they lose control over the production of information and the reasons for considering it to be true.

In fact, Samuel (2026) argues that the epistemic co-agency between humans and artificial intelligence may be based on the relationship of critical collaboration instead of outright rejection. Students can exercise epistemic sovereignty by discussing AI outputs, pointing out contradictions, and coming up with alternative justifications. In this context, co-agency does not imply that humans and machines are equally responsible; it implies that humans guide the technological contributions toward their own intellectual purposes.

Thinking for oneself does not mean abstaining from the use of artificial intelligence, but to subject the content produced to an independent process of judgment. Technology may support epistemic autonomy when students consider system output as a criticizable proposal rather than a final answer. However, when the work of decision-making and the intellectual labor are continually outsourced to the system, artificial intelligence can become a source of epistemic dependence.

**Epistemic Responsibility and the Justification of Belief**

The justification of a belief is not the same as its truth. That artificially generated answer may be correct by accident, but a student who doesn't know why that answer is correct, what evidence backs up that answer and under what conditions that answer holds, can't be said to have sufficient epistemic ownership of that knowledge. Salido et al. (2025) point out that critical thinking is a purposeful, evaluative and self-regulated process. This demands inspection, not only of the correctness of the conclusion but also of the path by which the conclusion was reached.

Epistemic responsibility requires that people seek reasonable evidence for the beliefs they adopt, that they consider opposing views, and that they revise their judgments in light of the possibility of error. This responsibility is made all the more important by the way that large language models work and the way they can produce information that sounds coherent but is not sourced or verifiable. According to Lo's (2023) review, although ChatGPT presents opportunities for learning support and personalization, it also raises concerns about accuracy, assessment, and academic integrity.

Epistemic responsibility is not only about checking AI output after it is generated. Students also need to think about how they constructed their question, what they expected from the system, and what they plan to do with the output. Liu et al. (2025) propose that AI-assisted learning has four distinctive strategic dimensions: task planning, process monitoring, output checking and ethical regulation. Focusing on only one of these dimensions leaves the epistemic self-governance of students incomplete.

Therefore, verification in AI-supported education should be designed as a multistage process. Students should begin by identifying the major claims made by an AI output, checking those claims against scientific or primary sources, and then explaining how the information they find supports their own argument. Based on the AI literacy competencies proposed by Long and Magerko (2020), it is necessary to recognize both the strengths and limitations of a system to be able to use it critically.

Epistemic responsibility also has an institutional aspect. When the learning environment doesn't teach appropriate practices, simply telling students to "use artificial intelligence carefully" isn't enough. Veldhuis et al. (2025) argue that the curriculum should include critical AI literacy, and students should be involved in activities that explore the social implications of artificial intelligence. Educational institutions need to set clear criteria for verification, limits of use, and expectations for transparency.

Epistemic responsibility also changes the teacher's role. The teacher no longer is just a supplier of correct information, but becomes an epistemic mediator, who helps the students to find justifications, to compare sources and to find errors. Oh and Lee (2025) state that humane epistemic agency must be maintained in an era of artificial intelligence, which involves students' active questioning, evaluation, and construction of knowledge. In this context, the teacher needs to act not as a primary enforcer prohibiting AI use, but as a guide making epistemic responsibility visible.

Justification of beliefs has a wider scope of responsibility than just verifying that an AI output is correct. Students should be able to explain why they trust a source, what evidence supports an assertion, and how the contribution of artificial intelligence was used in the learning process. Without the protection of epistemic responsibility, artificial intelligence can increase access to information while decreasing students' ownership of knowledge.

**Epistemic Autonomy as an Educational Aim**

If education's purpose is reduced to handing down what is already known to the next generation, students are left as the receivers and repeaters of a set package of information. However, Dewey (1916) considers education as a process of reconstruction of experience and cultivation of the individual's power of thought. Similarly, Peters (1966) distinguishes education from the ordinary run of instruction or behavioral modification by stressing knowledge, understanding and conscious participation in worthwhile activities. All of these approaches point to a central task of education, which is not simply to increase what students know, but to transform their relationship to knowledge.

If we conceive of epistemic autonomy as an educational goal, students must be subjects who are able to judge the knowledge claims made to them. Students should not only memorize correct answers, but be able to explain why an answer is considered correct, distinguish between sources, and change their minds in the presence of contrary evidence. Similarly, Biesta's (2022) perspective on subjectification indicates that education is not simply about providing students with certain competencies; it aims to shape them into responsible subjects of the world.

With the introduction of artificial intelligence into educational environments, this goal has become more important. Kasneci et al. (2023) highlight that while large language models can offer personalized explanations and rapid feedback to students, they also pose challenges in terms of misinformation, bias, and overreliance. The more readily information is available, the more responsibility the student has to evaluate that information, not less. Epistemic autonomy, then, is an educational aim that respects students' own authority to make the ultimate judgment and justification in technologically mediated information environments.

Education's relation to epistemic autonomy should not be read as unconditioned individual independence. Students can't progress in complex fields without trusting teachers, experts and scholarly works and institutional sources of knowledge. However, trust does not mean that students abdicate their intellectual responsibility. Ng et al. (2021) argue that AI literacy is more than simply knowing how to use the system; it also involves being able to evaluate the system's outputs and understand the ethical consequences of those outputs. Epistemic autonomy therefore does not mean freedom from all authority, but the development of a justified trust relationship with epistemic authorities.

The goal of education in the era of artificial intelligence must therefore be set between two extremes. At one end of the spectrum, we have a compliant approach, which treats technological systems as unchallengeable sources of knowledge; at the other, a prohibitive approach, which treats all use of artificial intelligence as a threat to student independence. The opportunities and threats identified by Farrokhnia et al. (2024) suggest that the educational effects of artificial intelligence are not so much dependent on the tool itself but on the conditions of its use. Education should therefore help students critique content produced by technology rather than prohibit them from using technology.

If we take epistemic autonomy as an educational aim, we cannot measure success by how much information students remember. The main criterion is the ability of students to interpret, evaluate and justify claims to knowledge. Pallant and Blijlevens (2025) also highlight differences in the quality of learning between students who use artificial intelligence to build knowledge and students who use it just to get the job done. The result demands a rethinking of the informational function of education.

## The role of education in the transmission of knowledge

The transfer of knowledge is a crucial part of education. No student is able to reproduce the accumulated knowledge and experience of mankind from the beginning. Teachers, books, scientific institutions, and educational programs expand a world of knowledge beyond the student's personal experience. However, Dewey (1916) warns that knowledge can become a static mass of content when separated from students’ experiences and problem-solving activities. Transmission should not, therefore, replace the student’s process of constructing meaning.

In the traditional transmission model the teacher is positioned as the knower and the student as the person who needs to be educated from a state of not knowing. In times of limited knowledge such a model served a purpose, but it runs the risk of marginalizing students’ ability to inquire and reason. Peters (1966) conceives of education as not enough for learners to learn true propositions; rather those propositions must be embedded in a wider structure of meaning and understanding. Much of the educational value of knowledge is lost when it is not understood by the student or related to other knowledge.

Generative artificial intelligence has made the shortcomings of the transmission model more visible. Students can now receive explanations, summaries or completed texts on many topics in a short time. Tlili et al. (2023) observe that systems such as ChatGPT can deliver access to information and personalized support, but can also generate false, decontextualized, or fabricated content. So the speed at which information is presented does not mean that it has been understood or learned.

The success of artificial intelligence in transmitting information should not be construed as making teachers obsolete. Kasneci et al. (2023) observe that large language models can assist in some elements of instruction, but they are unable to evaluate pedagogical context, students' developmental stages, and the ethical consequences of knowledge in the same way a human teacher can. The role of the teacher is changing from simply transmitting content to helping students learn to evaluate and justify that content.

One of the major problems of the transmission of knowledge is the tendency of students to perceive the presented content as complete and indisputable. This risk could be further worsened by the fluid and sound presentation of AI-generated texts. Long and Magerko (2020) point out that AI literacy includes knowing what artificial intelligence systems can not do, in addition to what they can do. Given this, students should not treat AI output as knowledge but as a suggestion that must be verified for accuracy and context.

The educational value of information is not inherent in the information itself, but in the purpose for which students use it. Students who used generative AI to improve understanding and build knowledge achieved higher quality learning outcomes than students who used generative AI only to complete tasks (Pallant & Blijlevens, 2025). This distinction implies that the transmission of knowledge must be accompanied by student activity, otherwise technology may simply accelerate the production of outputs without promoting learning.

Education cannot abdicate from the transmission of knowledge, but it cannot be reduced to it. AI is able to produce content faster than a teacher, but it cannot create the educational meaning of that content on its own. Similarly, Farrokhnia et al. (2024) show that the opportunities offered by artificial intelligence could become threats in the absence of pedagogical guidance and critical thinking. Students should therefore be seen as active players in the creation of knowledge and not as passive consumers of transmitted information.

## The student as epistemic subject

To be an epistemic subject for a student is not simply to show up active in class or to select from a set of predetermined options. An epistemic subject is a person who can decide which claims to accept and on what grounds, who can formulate his or her own questions, and who can take epistemic responsibility for the conclusions that he or she reaches. Biesta (2022) argues that the aspect of education associated with subjectification is not simply about making students conform to the demands of the existing order but about developing their responsibility for their positions and responses in the world.

This post does not make students completely independent of teachers or other sources of knowledge. Being an epistemic subject does not imply that students must reject information provided by the teacher beforehand. It implies that the relationship with the teacher must not become one of passive obedience. As part of their effort to address AI literacy through the dimensions of evaluation and ethical use, Ng et al. (2021) highlight the student as a decision-maker responsible for system outputs. Epistemic agency, then, is not the result of a lack of information sources, but the exercise of judgment in the face of multiple information sources.

Generative artificial intelligence has the potential to support and to undermine the student’s epistemic status. Systems that generate alternative perspectives, identify gaps in students’ arguments or provide different explanations for concepts they don’t understand may result in greater active student participation in learning. In contrast, Anlı’s (2025) discussion of epistemological alienation warns of the danger of people becoming alienated from the production and monitoring of knowledge, turning into passive consumers of algorithmic outputs. In such cases, students may be exposed to many outputs without genuine control of knowledge.

If students are treated as epistemic subjects, then assessment must also be restructured. Assessments that only measure the final product may not be enough to distinguish the students' thinking from the contribution of artificial intelligence. Within the framework of higher education policies, originality is often only defined as text written by students on their own, but the process of production, the choices that are made and the changes that are implemented should also be assessed (Luo, 2024). Students should be asked to show a result, as well as how they arrived at it and how they evaluated the sources of information used.

Another very important condition for being an epistemic subject is epistemic self-confidence. Students, when they see their own thought processes as less than the more quick and fluent responses of artificial intelligence, lose the motivation to question and object. Kong et al.'s (2024) project-based AI literacy intervention demonstrates that students are able to acquire not only technical skills but also skills of AI evaluation and problem solving. It implies that epistemic agency can be supported by appropriate educational practices.

"Positioning students as epistemic subjects is not a matter of unlimited choice, but a matter of responsibility to think and justify." Artificial intelligence can improve epistemic agency as long as it supports students' thinking. When it takes over the thinking process, it makes them consumers of results. Thus, the student's status as epistemic subject is directly connected to the development of critical thinking and independent reasoning.

## Critical Thinking and Independent Thought

Critical thinking is not just thinking that a claim is wrong or that you are skeptical all the time. It involves a systematic assessment of the clarity of the claim, the assumptions on which it rests, the quality of the evidence used, and the validity of the inferences drawn. As noted by Salido et al. (2025), higher education research increasingly regards critical thinking as an intentional, evaluative, and self-regulated process. Critical thinking is therefore not an add-on skill that is relevant after the fact of knowledge acquisition, but a core process that determines the quality of the act of knowledge acquisition.

"Independent reasoning" does not mean "thinking without reference to external sources." Students may cite expert opinion, teachers, and scholarly literature, but must evaluate the credibility of those sources and how they relate to the student's own claim. Critical generative AI literacy, according to Rapanta et al. (2025), involves not only crafting effective prompts but also scrutinizing, contextualizing, and epistemologically assessing content produced by systems. Then, we mean not independence of sources, but the preservation of the responsibility for judgment.

Studies on the effect of artificial intelligence on critical thinking does not always yield positive or negative results. Essien et al. (2024) found that when AI text generators are used in appropriate activities, students can benefit from AI text generators in analyzing, developing ideas, and evaluating. However, the same study suggests that over-reliance on tools and not validating their outputs could compromise students' critical thinking. The important variable is not that we are using artificial intelligence, but rather what cognitive operations students are asked to perform in using it.

The biggest threat to critical thinking is the outsourcing of cognitive effort to artificial intelligence. When the children get a final answer without first defining the problem, collecting evidence, and comparing solutions, the essential steps of the learning process are invisible. Therefore, Vendrell and Johnston (2026) suggest "cognitive friction" in educational design. When students are asked to identify assumptions, find flaws, and create alternative solutions in AI outputs, technology becomes an object of thought, instead of a proxy for thought.

The activities of comparison and justification are particularly important for the development of critical thinking. Students may be asked to compare two different answers generated by AI, determine which answer has stronger evidence or build a counterargument to the system's answer. Long and Magerko (2020) argue that AI literacy is the ability to critically interpret system outputs and understand the limitations of the system . These activities contribute to finding the right answer as well as to developing criteria for the evaluation of correctness.

Teachers' pedagogical decisions also contribute to critical thinking retention. If teachers only use artificial intelligence to generate content or assignments more quickly, the educational efficiency may increase, but the reasoning dimension may be weakened. Designs that require students to generate questions, justify answers, and reconstruct AI output may foster epistemic autonomy, however. Salido et al. (2025) reviewed pedagogical strategies that also highlight the need to balance personalization with critical evaluation.

In the era of artificial intelligence, a very important task of education is to develop critical thinking and independent reasoning. Technology can increase the number of ideas students are exposed to, but it cannot determine their value. Students are responsible for developing evaluative criteria, considering evidence, and defending their final judgments. Inquiry, evidence and justification are thus the key areas of education where critical thinking becomes concrete.

**Inquiry, Evidence and Justification**

Inquiry is not an acceptance of information, but an investigation of its source, scope and grounds. Dewey's (1916) description of thinking starts with the identification of an uncertain or problematic situation and proceeds to an evidence-based judgment. Inquiry is therefore not an add-on activity at the end of a lesson, but the basic movement that initiates and guides the learning process.

Evidence is not all the information that is believed to support a claim. Reliability should be determined by the quality of the source, the appropriateness of the method, the relevance of the method to the claim, and the strength of the method relative to competing claims. Kasneci et al. (2023) point out that large language models are likely to return responses without explicit references, or generate unverifiable citations. The scientific appearance of information contained in AI output does not mean that it is valid evidence.

Justification is the explanation students offer for why they believe a claim. This process involves not only identifying a valid source but also constructing the connection between the evidence and the conclusion. The centrality of evaluation in the AI literacy framework proposed by Ng et al. (2021) reveals that students are expected not only to use system output, but to evaluate the value and limitations of that output .

In AI-supported education, the responsibility to justify claims should be made visible to students. They may be asked to identify the AI tool they used, specify which recommendations they followed or rejected, and support assertions in the final paper with independent sources. As Luo's (2024) analysis shows, we should not confine originality to the identification of the author of the final text. Students' choices, adaptations and processes of justification also form their academic contribution.

Inquiry and justification are also linked with epistemic responsibility. Students who depend on an unsupported claim in their work cannot defer responsibility to artificial intelligence. Anlı (2025) argues that the ascendancy of artificial intelligence (AI) systems to the decision-making position in knowledge production, with their inability to bear epistemic or moral responsibility, requires a rethinking of the human stance towards knowledge. It is therefore necessary to distinguish between dependence on artificial intelligence and trust in it. Trust must be backed up by independent evidence and oversight.

Educational activities based on verification of AI output can develop evidence-based reasoning. Students may be asked to identify factual statements in a response, distinguish between primary and secondary sources, and produce correction reports for inaccurate statements. Rapanta et al. (2025) demonstrate that inquiry and contextualization cannot be separated from the technical competence in their approach to critical generative AI literacy.

Students should not be the sole responsibility of verifiers. Educational institutions need to provide access to reliable databases, teachers need to explicitly teach the source-evaluation criteria and transparent principles on the use of AI need to be laid down. Salido et al. (2025) state that institutional limitations and literacy inequalities are important problems in the integration of artificial intelligence and critical thinking. Epistemic autonomy is an individual capacity, but its educational conditions of possibility must be institutionally secured.

Inquiry, evidence and justification are the visible educational manifestations of epistemic autonomy. Students gain epistemic ownership of knowledge when they do more than repeat a claim—when they evaluate the source of the claim, consider counterevidence, and explain their reasons for accepting it. Artificial intelligence should not be a shortcut that removes these processes, but rather a tool that tests students' reasoning following the approach of Vendrell and Johnston (2026).

## The Difference Between Having Knowledge and Accessing Information

Digital technologies have greatly shortened the time and effort needed to access information. Generative artificial intelligence has gone even further in this transformation providing not only sources or documents but rather direct explanations, summaries, interpretations and completed answers. According to Kasneci et al. (2023), large language models can personalize learning materials and make complex content more accessible. But the easy availability of an answer does not show that students understand what it means or can evaluate the justification for it.

That there is a difference between having information and having knowledge shows that knowing is more than coming across a true statement. Students can repeat a correct answer that they got from artificial intelligence — but if they can't explain why it is correct, under what conditions it is correct, and what evidence supports it, then they only have limited epistemic ownership of that knowledge. Ng et al. (2021) proposed a framework for AI literacy that included evaluation as an essential element. Here the difference between accessing information and using it mindfully is clear.

The distinction matters educationally because generative artificial intelligence can disclose the outcome of the learning process before students engage in the intellectual operations to arrive at it. Pallant and Blijlevens (2025) found that students who use artificial intelligence to fill gaps in knowledge and improve understanding experienced higher quality learning outcomes. Repeated outputs by students in a direct and procedural way resulted in lower levels of learning.

The central educational problem, then, is not the amount of information that artificial intelligence can provide, but the quality of the students' relationship with that information. If you imagine a pre-cooked response as a final outcome, it may enhance access but it weakens the processes of interpretation and justification. Salido et al. (2025) point out that there is increasing evidence of the detrimental effects of excessive reliance on artificial intelligence on cognitive reflection and critical evaluation.

Access to information is a necessary, but not sufficient, condition of learning. Knowledge refers to knowing content, making connections to prior knowledge, assessing evidence and applying it to solve new problems when appropriate. Education in the era of artificial intelligence should thus not measure the speed with which students find answers, but how they evaluate those answers and turn them into meaningful structures of knowledge.

**The growing availability of information**

The internet and digital databases made it easier to access information, but generative artificial intelligence has changed the practice of access in a fundamental way. Traditional search engines generally produce lists of sources for users, while generative AI combines different kinds of content and offers a direct answer. Tlili et al. (2023) claim that this feature provides great educational opportunities, especially when used in conjunction with instant explanation, language support and individualized instruction.

More readily available information could make education less time- and place-bound. Outside of the classroom, students can ask for explanations of concepts they don't grasp again, can ask for different examples of the same subject, and can get feedback that is appropriate for their pace of learning. Strengths of ChatGPT are its accessibility and personalization potential, but Farrokhnia et al.

(2024) stress that the accuracy and pedagogical appropriateness of its answers can not be assured.

Ease of access, in turn, reduces the cognitive costs of navigating among sources of information. Instead of reading a long text, students may ask for a summary, ask for a complex theory to be simplified or compare different perspectives in table form. Corbin and Walton (2025), however, do suggest that the broad availability of AI summarization tools does open underresearched questions about how students interact with academic texts and what the effect is on their critical reading skills.

Greater access to information also creates an invisible process of selection and framing. The AI doesn't simply enumerate all the information that could be relevant to a user's question. Instead it builds a specific answer, informed by the data on which it was trained, the architecture of the model and the nature of the prompt. According to Long and Magerko (2020), AI literacy is understanding what systems are capable of, and not capable of doing. Students have to understand that the answer they arrive at is not the totality of a field of knowledge, and that there can be other understandings.

Easy access is not an excuse for not having the patience and research skills to get at knowledge. But, when students can choose to bring every problem directly to artificial intelligence, they may not practice their skills in identifying keywords, formulating research questions, and choosing appropriate sources. For example, the contribution of AI text generators to critical thinking depends on the active participation of the students in analysis and evaluation instead of the passive use of the tool (Essien et al., 2024).

More accessible information is a huge educational opportunity, but only becomes learning if students stay engaged in the research and evaluation processes. Artificial intelligence may reduce barriers to access, but it cannot independently determine what information is reliable, relevant, and sufficient. Access should therefore assist not replace processes of understanding and interpretation.

## The Difficulty of Understanding and Making Meaning

Understanding is a broader mental activity than repeating a statement or summarizing it in a few words. Students should be able to relate one concept to other concepts, understand the consequences of a claim, and apply knowledge in a new context. Pallant and Blijlevens (2025) note that students who leverage artificial intelligence to deepen their understanding exhibit deeper learning than students who leverage it just to produce finished outputs.

Generative artificial intelligence can produce explanations that are very coherent and that sound like they come from understanding. However, the generation of a meaningful response by the system does not imply that the student has reached the same level of understanding. Students may read an AI-generated explanation that appears correct; but if they cannot reconstruct it in their own words, or defend it against counterexamples, the knowledge has not yet been incorporated into their conceptual structure. Kasneci et al. (2023) stress that AI-based instructional support should be designed so that it does not remove active participation of students.

This problem is particularly evident in summarization tools. Corbin and Walton (2025) argue that AI summarizers can make difficult readings more accessible and shorter, but they can also hide the rhetorical structure, conceptual nuances and argumentation of a text. Students who only read a summary may gain a general idea of what a text is saying, but not how the author came to that conclusion, or what other options were ruled out.

Meaning making is linking new information to students' prior knowledge, experiences, and questions. The standardized explanation of an artificial intelligence may not always disclose students' individual misunderstandings. Students should be asked to rephrase AI output in their own words, provide examples, critique it, or apply it to a different problem. The evaluation-based AI literacy approach developed by Ng et al. (2021) corresponds to the idea that users should not be passive recipients of content produced by humans.

The substitution of understanding for access may produce what Anlı (2025) describes as epistemological alienation. Individuals reach informational outcomes, but become estranged from the processes through which those outcomes are produced, evaluated, and justified. In these conditions, students remain the users of knowledge, losing partially the status of epistemic subjects of knowledge.

It is not the receipt of AI output that leads to understanding and meaning-making, but the active processing of it by students. Information that is not integrated into students' conceptual structures or applied to new situations remains accessible but superficial content. Therefore educational practices should make visible the processes of explanation, connection and reconstruction, not only the result.

**The Culture of Pre-Fab Answers**

The culture of ready-made answers is the tendency to look for finished results, without going through the intellectual processes necessary to solve or understand a problem. Generative artificial intelligence is not the sole reason for this trend, but the ability of generative AI to generate quick, fluent and often convincing responses may amplify this tendency. As Farrokhnia et al. (2024) note, the use of AI may introduce threats such as cognitive dependence and the degradation of critical thinking.

Ready-made answers may help students to complete tasks in the short term. At the same time they may truncate core stages of learning, including problem definition, information seeking, comparison of alternatives, and correction through error. A review of the literature by Salido et al. (2025) found an association between overuse of artificial intelligence and cognitive reflection. Therefore, educational efficiency should not be interpreted as the elimination of intellectual effort.

An important consequence of a culture of ready-made answers, is that students may find it more difficult to identify their own knowledge gaps. A polished AI text might give the appearance that the student understands the topic. However, success is only at the level of the product when students cannot explain the basic concepts in the text or defend the claims of the text. Pallant and Blijlevens (2025) make a distinction between procedural use and mastery-oriented use, acknowledging that a finished product does not always signal real learning.

The use of pre-prepared answers can also silence the voices and academic positions of the students. As Luo (2024) argues, "Generative AI policies that focus solely on the originality of the final product do not adequately take into account students' thinking and decision-making processes." Assessment should look at what recommendations students accepted and why, what expressions they changed and what their contribution to the AI output was.

It is also not realistic to dismiss the ready-made answers entirely. Sometimes students need a basic explanation or example or starting framework. The relevant question is whether the answer is employed as the beginning, or the end, of thought. Artificial intelligence can assist students in generating ideas and exploring multiple perspectives, but over-reliance on the tool can impede critical thinking (Essien et al., 2024).

The problem with a culture of ready-made answers is not that the answer is there, but that students don't intellectually engage in forming it. Content produced by AI can be a support for learning if it is used as a starting point, counter-position or draft. However, when taken as final and unquestionable, it may undermine students' capacities of thinking, understanding and epistemic responsibility.

**Verification of Information and Source Verification**

The linguistic persuasiveness of a statement is irrelevant to verify information, you have to look at the evidence, source and context. Large language models can generate correct, incorrect and fabricated information with equal fluency. Kasneci et al. (2023) identify AI-generated fabricated citations and misleading explanations as major educational hazards.

Source checking is more than just making sure a link works. Students should think about the academic quality of the source, the author's expertise, the date of publication, the research method, and if it actually supports the claim or not. Long and Magerko (2020) state that part of user competence is knowing the limitations of standard artificial intelligence systems. Even if an AI-generated source is genuine, it may not be relevant to the student's claim or may be misrepresented.

One artificial intelligence system verifying another is not enough on its own. Kuznetsova et al. (2025) found that the effectiveness of chatbots in fact-checking political information varies with language, subject, and the source to which the claim is attributed. The results suggest that AI-based verification can be useful, but should not be considered a definitive and objective judge.

Similarly, the scoping review by Park and Nan (2025) shows that generative artificial intelligence plays a dual role in relation to misinformation. Large language models are capable of generating convincing false content, but they can also assist in discovering and correcting some false claims. The researchers point out that the protective mechanisms of these tools may not be uniform, and that evaluation standards must be established.

Verification should be multisourced. First, find claims in the AI output that can be checked. Second, such claims should be checked with primary research, peer-reviewed publications, or reputable institutional sources. Thirdly, it should be checked if the source actually supports the claim. Finally, contradictory findings should be considered together. The critical AI literacy approach presented by Veldhuis et al. (2025) shows that verification is not just a technical operation, but a critical competence that is sensitive to the social context.

Source checking needs to be a clear part of education programs. This might involve having students check the references in an AI-generated text, identifying errors in the citations and producing a corrected version with the support of reliable sources. Such activities can integrate skills in using AI with competencies in evaluation and ethical responsibility (following Ng et al., 2021).

Checking sources and verifying information are the basic epistemic processes by which access to information is transformed into knowledge possession. By not checking AI output against independent sources, students abdicate to the linguistic credibility of the system the responsibility to distinguish true from false information. Ultimately, it is not the AI system that has the epistemic responsibility, but the human subject who uses and communicates the output.

## Epistemic Dependence in Artificial Intelligence: A Danger

AI-assisted education speeds up the access to information and allows students to complete the learning tasks faster. But ease of use does not by itself determine the epistemic quality of the student's relationship with the system. Kasneci et al. (2023) point to both the opportunities large language models offer regarding explanation, feedback, and personalization, and the risks they pose regarding misinformation and overreliance. If students take the AI output as a final answer with presumed reliability rather than as a proposal for which they need to evaluate the reliability, technological support can become epistemic dependence.

Epistemic dependence does not mean consulting any outside source. Trust in teachers, experts, research and institutional sources of knowledge underpins education and the production of scientific knowledge. The problem arises when trust is no longer justified and is limited, but is a matter of uncritical submission. Overreliance may also lead to users accepting AI recommendations without adequate scrutiny and resorting to cognitive shortcuts, as noted by Zhai et al. (2024).

A defining feature that sets reliance on artificial intelligence apart from reliance on conventional information sources is the humanlike fluency and personalization of system responses. According to Tlili et al. (2023), chatbots may be perceived as consultative actors due to their interactive nature and thus not only as technical tools. The way this is presented may lead students to overlook the statistical nature of the generation of AI output, and to attribute to artificial intelligence the abilities for understanding, intention, or judgment that it does not have.

According to Lu and Hu (2025), generative chatbots threaten autonomy by encouraging false mental states and eroding intellectual skills. Students may then ascribe unwarranted epistemic weight to a system's outputs when they think it understands, has intentions, or independently evaluates truth. Dependence is therefore linked not only to the frequency of use but also to the user's perception of the epistemic status of the system.

The educational effects of epistemic dependence may not be readily visible. Students are able to write assignments more quickly, produce more refined texts, and show better surface performance. However, Pallant and Blijlevens (2025) found some major differences in quality of learning between students who used artificial intelligence for procedural tasks and those who used it for knowledge building. Higher product quality does not necessarily imply better knowledge, understanding or reasoning of students.

The epistemic dependence on artificial intelligence is not due to the use of technology as such, but the continuous transfer of the students' own processes of evaluation and decision-making to the system. Therefore, the central aim of education should not be to ban AI use altogether, but to maintain students' critical position relative to system outputs. The first and most obvious dimension of epistemic dependence is a form of trust in artificial intelligence that is disproportionate to the system's actual ability.

**Overreliance on Artificial Intelligence**

Trust is an epistemic device that is necessary for decision-making to be possible in complex information environments. We cannot check up on every statement, so to some extent we have to rely on experts, institutions and pieces of technology. Shneiderman (2020) suggests that trust in human-centered AI should match the actual performance, reliability, and auditability of the systems. Overtrust is trust that exceeds the system's capacity and therefore ceases to function.

Students may overtrust artificial intelligence when they take system outputs as correct without comparing them with other sources, ignore the uncertainty of artificial intelligence, or prefer automated answers when they disagree with their own knowledge. Zhai et al. (2024) find that overreliance is higher when users are asked to judge the accuracy of the system and decide how much trust to place in the AI recommendations. Thus, the development of an appropriate level of trust in artificial intelligence, as well as its use, is hampered by a lack of knowledge.

The linguistic fluency of generative artificial intelligence is a major reason for the increasing overtrust. Kasneci et al. (2023) claim that large language models can generate false or fabricated information in a convincing and consistent manner. In such situations, where students equate clarity and professionalism with truth, the line between linguistic credibility and epistemic reliability becomes blurred.

Developing AI literacy is important to reduce overtrust, but teaching how to operate the technical systems is not enough. Ng et al. (2021) pointed out that AI literacy should include skills in use, evaluation, and ethical responsibility. Long and Magerko (2020) argue that users must understand what artificial intelligence can do and what it cannot do reliably.

Also it is not appropriate to answer overtrust with complete distrust. Discarding all AI output as worthless or false is to throw away the potential for the technology to support learning. As Rapanta et al. (2025) state, critical AI literacy is found in inquiry, adaptation, and contextual evaluation, not in blind trust or total rejection. The educational goal should not be distrust, but critical trust, calibrated by evidence and system performance.

In educational and academic processes, dependence can be measured in different dimensions, as shown by the artificial intelligence dependence scale developed by Uyar, Karafil, and Karakuyu (2026). Their work suggests that overtrust and reliance are not problems specific to students, but should be explored among teachers and academics as well. Educators who let AI output go unexamined are at risk of normalizing dependency for their students.

Trusting artificial intelligence too much is not simply to believe that a system does not make mistakes. Not considering verification necessary, prioritizing system output over one's own judgement and ignoring uncertainty also indicate overtrust . Therefore, education ought to focus less on whether students trust artificial intelligence and more on the reasons they choose the appropriate level of trust.

**The Reduction of Cognitive Load**

Cognitive tools could help people use their limited mental resources more efficiently. Information can be recorded, calculations can be assigned to machines or repetitive operations automated and attention can be turned to more complex problems. Therefore, a decrease in cognitive effort is not necessarily a decrease in learning. Generative artificial intelligence can help with the development of ideas and critical assessment within appropriately designed learning environments, as shown by Essien et al. (2024).

The problem is when the cognitive activities that are needed for learning are also transferred to technology. Consistently leaving problem definition, argument construction, evaluation of alternatives, and inference to artificial intelligence may cause students to miss opportunities to develop those skills. Zhai et al. (2024) have indicated that overdependence on artificial intelligence might be associated with deficiencies in information retention, analytical thinking, and decision-making.

Cognitive fatigue partially mediates the negative association between AI dependence and critical thinking (Tian & Zhang, 2025). This may seem paradoxical at first because artificial intelligence is generally seen as a tool that alleviates cognitive load. But the constant requirement to verify system outputs, to select between alternative answers, and to evaluate how much confidence to put in artificial intelligence, can add new types of cognitive load.

Whether cognitive offloading becomes risky depends on what processes students offload to artificial intelligence. “Chat-based AIs like ChatGPT could help with learning by providing explanations and feedback, but could limit original thinking and problem solving,” notes Lo (2023). It may be supportive to ask for clarification of basic concepts or for further examples, but it may diminish epistemic agency to pass off the formulation of the research question or the final evaluation.

Anlı’s (2025) analysis of epistemological alienation demonstrates that the loss of cognitive effort is not to be assessed solely at the level of individual skill. Individuals may keep in touch with knowledge but detach from the processes of knowledge production, testing and justification. Thus, students might be able to access informational products, but lose the capacity for the epistemic actions that are involved in their production.

Saving brain power does not mean refusing technology help to students. Instead, artificial intelligence can be used as a tool to provide counterarguments, feedback or correction after students have developed their own initial thinking, rather than as a system that answers before the student attempts the problem. Salido et al. (2025) show that the combination of critical thinking and artificial intelligence requires self-regulation, source evaluation, and pedagogical guidance to be taken into account simultaneously.

Reducing cognitive effort is not the same as weakening cognitive skills. Artificial intelligence can relieve the burden of repetitive and secondary operations so that students can spend more time on higher-order thinking. But short-term efficiency can lead to long-term skill loss if it involves core epistemic operations such as analysis, justification, and decision-making.

**Automation Bias.**

Automation bias is the tendency of users to replace the recommendations of an automated system with careful information seeking and evaluation. Jovchevski et al. (2026) do not restrict this bias to adherence to an incorrect recommendation; they also refer to situations where users do not carry out a necessary action because the automated system did not issue a warning. In education, students may use the lack of an AI-identified problem as evidence of an absence of a problem.

Weak automation bias occurs when users accept an AI recommendation when there is conflicting evidence available to them. Strong automation bias is when people give more epistemic weight to the system even after they have recognized counterevidence or made an accurate independent assessment. The strong form more directly undermines autonomous human agency through excessive epistemic submission to automated judgment (Jovchevski et al., 2026).

Automation bias is not confined to novice users. Experts may also be tempted to rely on automated recommendations due to time pressure, workload, institutional expectations or the high accuracy image of the system. Therefore, AI literacy in education should not be confined to the strengthening of students' basic digital skills. Veldhuis et al. (2025) argue that critical AI literacy should also include the social authority of technology, the design assumptions that underlie it, and its impact on user behavior.

One way to address automation bias is to force students to make an independent assessment before showing an AI output. In such cases, when students produce their own solution first and compare it with the AI answer, the system's recommendation could become no longer the initial and dominant frame. The project-based AI literacy intervention by Kong et al. (2024) shows that students' ability to critically evaluate artificial intelligence systems can be developed through structured education.

Another measure is to design artificial intelligence systems so that they show uncertainty and alternatives rather than giving a single definitive answer. Lu and Hu (2025) propose a Socratic model in which students are asked questions instead of using chatbots that give direct answers. In this design, the system acts as a stimulus to trigger intellectual processes, rather than taking over the reasoning of the user.

Automation bias is a much bigger problem than just "the AI answer is wrong". A major risk is that students limit their information-seeking and evaluative behavior in response to signals from an automated system. Education should therefore not only expose students to AI errors but also develop habits of looking for counterevidence, making independent judgments, and questioning system recommendations.

**Decision making authority should be delegated**

Decision-making involves data collection, alternative development, outcome evaluation, and justification of a particular option. Artificial intelligence may be able to assist at each step, but making a recommendation is not the same as taking responsibility for the decision. Shneiderman's (2020) human-centered artificial intelligence approach argues that technology augmentation should extend human capacity, but human actors have the ultimate oversight and accountability.

In education, delegation of decision-making authority may happen when students let AI decide which source to use, which view to accept, or which conclusion to reach. In the era of generative artificial intelligence, it is not enough to evaluate students' academic contribution by the final text alone (Luo, 2024). Assessment should also be able to see the choices students make between AI recommendations, the outputs they reject, and the adaptations they make.

Artificial intelligence systems are not persons who can be held morally or epistemically responsible. When a system produces wrong information, it does not feel shame for the effects of its response, it does not feel an obligation to justify itself, and it does not answer to those who are harmed. As Tlili et al. (2023) stress, we need to clearly establish the boundaries of responsibility, transparency and academic integrity in the use of AI. "The AI said so" is no excuse for a student's responsibility for making a false or inadequately justified claim.

Convenience can be achieved at the cost of student autonomy, by passing the buck. Lu and Hu (2025) argue that when users repeatedly offload the functions of knowledge acquisition and problem solving to artificial intelligence, they cease to be producers of thought and become instead validators of system outputs. This transformation can decrease not only decision-making competence but also the psychological and normative ownership of decisions.

In an effort to maintain accountability, assessment designs should make students' decision-making processes visible. Students might be asked to identify the AI tool used, explain which suggestions they used and why, show which sources they used to verify system outputs, and defend their final judgment. The AI literacy framework developed by Ng et al. (2021) shows that technical competence is inseparable from ethical evaluation.

Jovchevski et al. (2026) argue that strong automation bias undermines autonomous human agency and the duty to exercise human judgement, especially in high-stakes situations. Educational decisions are not always life and death matters, but they are normatively significant for students' thinking, learning, and development of an academic identity. The constant delegation of reasoning to artificial intelligence may weaken the ideal of the responsible subject that education is supposed to foster.

Artificial intelligence may help to make decisions, to increase the options and to make transparent considerations that have been overlooked but it cannot take epistemic and ethical responsibility for a decision. Even when students are making use of AI recommendations, they are still responsible for the final choice and the product that results. Not only symbolic approval, but also justified evaluation and the ability to refuse the system when necessary are part of human supervision.

**Supporting AI-supported education with epistemic autonomy**

But it is not necessary that students be completely independent of technological systems to be able to guarantee epistemic autonomy in AI-supported education. Human learning has been made possible in the past by teachers, books, scientific institutions and a host of cognitive tools. For Shneiderman (2020), the central criterion of the human-centered AI approach is not whether technology takes away human decision and control, but whether it reliably enhances those capabilities. Epistemic autonomy then is no question of shunning artificial intelligence but of students' ability to consciously steer their relation to it.

The first condition of an autonomy-preserving approach to AI-supported learning is that students do not consider the system output as finished knowledge. As Kasneci et al. (2023) note, large language models can produce both correct information and false or fabricated content at similar levels of linguistic fluency. An AI reply should therefore not be seen as an object of knowledge to be accepted and adopted straightaway, but as a tentative claim to be verified, contextualized, and justified.

Epistemic autonomy cannot survive on individual attentiveness or willpower alone. If instructional designs lead students toward pre-packaged outputs, if assessment measures only the end product, and if source-verification processes remain invisible, it is unreasonable to expect students to develop independent judgment. Vendrell and Johnston (2026) suggest that educational environments based on AI should contain components of cognitive friction, which would make students challenge assumptions, seek alternatives, and justify their final decision.

Thus, four integrated principles must underlie the preservation of epistemic autonomy. Students should first understand how artificial intelligence works and what its limitations are, critically. Second, they should practice comparative research practices that are not based on a single AI output or a single source of information. Third, the claims generated should be verified against reliable and preferably primary sources. 4. Students are ultimately responsible, epistemically and ethically, for the technologies they use.

Protecting epistemic autonomy in AI-augmented education does not call for a defensive posture to technology, but for an educational setup that maintains the human capacity to judge and be accountable for knowledge. Oh and Lee (2025) argue that students' active engagement in questioning, evaluating evidence and constructing knowledge can secure a humane epistemic agency. Thus, the first dimension of this protective framework is the critical AI literacy beyond the technical skills of use.

## AI & Critical Literacy

AI literacy is often defined as the ability to write the right prompts, use the right tools, or get the desired output. However, Long and Magerko (2020) state that AI literacy includes knowledge of the system's capabilities and limitations, the ability to interpret the system's generated results, and critical evaluation of AI applications. Using tools efficiently does not constitute epistemic autonomy.

Ng et al. (2021) define AI literacy as knowing, using, evaluating and ethical. In this sense, students cannot be considered as fully AI-literate if they know how to use artificial intelligence, but do not know how to evaluate the reliability of the information it generates or the ethical consequences associated with its use. Critical literacy brings epistemic reasoning and responsibility to technical skill.

Veldhuis et al. (2025) argue that critical AI literacy is beyond system accuracy. Students should also ask questions about the data used to train AI models, the knowledge and experiences that are omitted, the way they render certain social groups visible or invisible and how they influence users' patterns of thinking. This approach allows learners to comprehend artificial intelligence not as a flawed tool, but as a social and technological system that constructs information environments.

As also claimed by Rapanta et al. (2025), critical generative AI literacy consists of questioning the system output, adapting it to the needs of the user, situating it in a specific context, and evaluating it epistemologically. Students should evaluate whether the AI response is accurate, identify any assumptions it makes, determine which perspectives are omitted, and assess the contexts in which it could be valid.

Critical AI literacy should not turn students into users suspicious of every output, or users who avoid technology altogether. Generative artificial intelligence provides significant opportunities for access, personalization, and feedback that need to be balanced against the risks of misinformation and superficial learning (Farrokhnia et al., 2024). The educational aim is a healthy confidence somewhere between blind confidence and total distrust.

Pietarinen (2026) cautions that the generative artificial intelligence's ability to guide learning may result in a merging of genuine pedagogical help and manipulation. When systems covertly limit students' options, present certain answers as neutral knowledge, or direct users to performance measures rather than educational goals, students' ability to exercise independent judgment may be impaired. Thus, critical literacy includes recognizing how systems affect the behavior of users, as well as appraising the content.

Critical AI literacy cannot be a technical course in isolation, separate from disciplinary learning. Kong et al. (2024) demonstrate in their intervention, which is based on projects, that students can learn technical, critical and ethical competencies at the same time by using artificial intelligence to solve authentic problems. Critical literacy should therefore be embedded in the processes of research, writing, problem solving and assessment.

Critical AI literacy is a broader epistemic skill than the skill to use the artificial intelligence. Students should be able to assess system operations, limitations, data structures, social effects and false sense of trust that persuasive language creates. The finding of Tian and Zhang (2025) about the protective role of information literacy in the relation between AI dependence and critical thinking addresses the importance of this competence for epistemic autonomy . But critical awareness also has to be combined with research methods in which different sources of information are actively compared.

**Comparative use of sources**

Using sources comparatively is not just a matter of collecting a lot of sources randomly. Its primary use is to compare different sources against each other, in terms of their claims, evidence, methods and limitations. For critical thinking to survive in a time of artificial intelligence, the integrated development of source evaluation, electronic literacy, and self-regulation is needed (Salido et al., 2025).

It's not really an independent source comparison to ask another AI system to check an answer generated by an AI. Different models can use similar data sets, the same web content, or closely related language patterns. Given the risks of false and fabricated information identified by Kasneci et al. (2023), students should compare the AI output to peer-reviewed articles, scholarly books, primary documents, and reliable institutional sources.

You also create a hierarchy of sources of comparative use. Here the epistemic weight of a peer-reviewed article is not equal to a personal blog in terms of answering a scientific research question. However, peer review does not shield all claims in a source from critique. Students should evaluate research methods, samples, dates of publication, relevance to discipline, scope of findings. Long and Magerko (2020) state, “AI literacy is the combination of disciplinary knowledge and critical thinking skills to evaluate system outputs.”

Disciplinary conceptions of evidence in comparative source use also differ, which is important to consider. Billingsley and colleagues (2026) argue that in the era of generative artificial intelligence, search processes should do more than provide answers. They should help users recognize the differences in the ways that science, history, philosophy, and other disciplines produce knowledge. We cannot apply the same criteria for assessing the evidential structure of an experimental study that we apply for assessing the justification of a philosophical argument.

Artificial intelligence may facilitate comparative research. Systems can be used by students to find other theories, generate keywords or formulate a search strategy. However, one should not directly use AI generated source lists due to the ethical and epistemic risks pointed out by Tlili et al. (2023). Check each source via a publisher, academic database or DOI record.

Kim et al.'s (2025) study of academic writing reveals that students interact with artificial intelligence in different ways in planning, text production and evaluation. The system can provide epistemic support to students when they compare AI suggestions with different scholarly sources, and when they reconstruct their own arguments. In contrast, the mere use of one output limits students' ability to reason across sources.

For example, when teaching comparative source use, students might be asked to analyze sources that support and contest the same claim. Epistemic autonomy is demonstrated by contrasting an AI-generated answer with a peer-reviewed article, noting discrepancies, and explaining why one claim is better supported than another. This is in line with the views of Rapanta et al. (2025) on contextualization, which states that knowledge should not be used out of the context of its origin.

Comparative source use is a basic research practice that prevents students from depending on a single AI output or authority. The comparison should not only be based on the conclusions of the different sources, but also on their methods and ways of justification. This allows students to articulate why they trust a source, but the degree of trust is contingent on systematically verifying the particular claims made by artificial intelligence.

**Validation of AI Output**

Verification does not mean that for every statement in an AI output, you have to look for a similar sentence on the internet. First, it is necessary to distinguish the factual claims, the interpretations, the predictions, and the value judgments. Kasneci et al. (2023) state that large language models can generate content that looks correct but does not present references. Students therefore have to decide which statements require independent evidence.

The first step in verifying a claim is to determine its source. If artificial intelligence cites a research finding, statistic or philosopher's view, students should locate the original work. Lo (2023) reviews that generative artificial intelligence may generate misinformation and wrong citations in education. A source must not only exist, but it must be verified to actually support the claim that the AI system has attributed to it.

The second is to evaluate the source for its timeliness and contextual relevance. An older study might be useful for a particular historical debate but may not be enough to describe the present characteristics of rapidly changing artificial intelligence systems. Given the development of technology, even the opportunities and threats identified by Farrokhnia et al. (2024) may need to be re-evaluated. And students should be asking not only if a source is credible, but if it is current and relevant to the claim being made.

The third phase is the assessment of evidence quality. The results of one sample of participants may not be generalized to all students and a conceptual paper should not be presented as if it is providing experimental evidence. Salido et al. (2025) show in a systematic review that methodological and contextual differences matter in the literature on artificial intelligence and critical thinking. Verification is a weighing of relative evidence, not a search for a source.

The fourth stage is to consider conflicting or contradictory evidence. Artificial intelligence is often designed to give a user a coherent answer to a question, and may not represent the uncertainty that is often present in scholarly debates. As Vendrell and Johnston (2026) point out, students using artificial intelligence to support critical thinking should not just be looking for answers, but also for assumptions, counterexamples and alternative explanations.

Students need to know the boundaries of their own knowledge so they can check the output. Limited knowledge of the response make it difficult to find subtle mistakes in AI responses. The evaluation dimension proposed by Ng et al. (2021) suggests that AI literacy should be considered together with disciplinary knowledge, information literacy and ethical awareness. If students find that they are unable to verify a claim, they should consult teachers, experts or trustworthy academic sources.

In education, AI-output can be made in the form of an audit, so that the process of checking is made visible. Identify separately the core thesis, the evidence, the source, counterevidence, uncertainty, and the final student evaluation. This process-oriented activities can develop critical AI competencies, as evidenced by the project-based approach of Kong et al. (2024).

Checking AI output is a more general epistemic activity than checking misinformation. Students need to think about the source of a claim, its relevance, its methodological basis, its context and its weight relative to other evidence. Vendrell and Johnston (2026) suggest that verification can be seen as a type of generative cognitive friction, which enables students to be cognitively engaged. In the end, the student is the one who must decide and use the information that has been verified.

### The student's final epistemic responsibility

Epistemic responsibility entails that students have reasonable grounds for knowledge claims they accept, use and communicate to others. Content produced by AI does not exempt students from their duties. Shneiderman (2020) in his human-centered approach noted that automation is to support human judgment, not to take responsibility away from a technology system.

Generative artificial intelligence systems can produce linguistic output but they are not persons and therefore cannot be held to account, academically or morally. The system is not liable for a fake source that the system suggests, a false claim made by the system, or a wrong assessment made for a student. The application of artificial intelligence in academic work should be considered in terms of transparency and honesty principles (Perkins, 2023). Students can't say "the AI did it" to avoid accountability for misinformation.

Cotton et al. (2024) argue that monitoring mechanisms that are only used to detect AI use cannot protect academic integrity. Assessment processes should demonstrate student's thought processes, decision making and use of sources. Students may be asked to identify the AI tool used, to disclose the prompts they entered, to explain which suggestions they rejected and to describe the original contribution they made to the final text.

Luo (2024) contends that "a definition of originality in higher education policy as students writing every word independently is insufficient to explain the complex relationships of authorship that emerge in the era of generative artificial intelligence" . Instead, the assessment should be on the students' control of the process, the choices they make, and their ability to defend the product that results. Epistemic responsibility also does not mean to renounce all tools in text production, but to account truthfully for their contribution and own the final judgment.

Students’ final responsibility involves the ability to refuse AI suggestions where necessary. Generative artificial intelligence can have subtle and unobservable impacts on the behaviour of its users, and a careful watch is needed to distinguish between acceptable educational help and manipulative influence (Pietarinen, 2026). Students still need to have their justifiable judgement when faced with a persuasive or confident system.

Part of our epistemic responsibility is to be willing to accept that you don't know. Artificial intelligence systems may tend to give complete answers under conditions of uncertainty. Students should be encouraged not to jump to conclusions when evidence is lacking, to suspend belief for the time being, or to seek expert support. Calibration of confidence to the quality of evidence is seen as part of critical literacy in the treatment of epistemological reasoning by Rapanta et al. (2025).

The epistemic responsibility should not be placed wholly on the students either. Educators should articulate clear expectations for AI use, provide access to trusted sources, teach verification skills, and develop process-oriented assessment. Uyar et al. (2026) prepared a scale to measure the dependence of AI . The results show that teachers and academics are also vulnerable to dependence . Teachers cannot expect students to meet a higher standard of epistemic responsibility if teachers themselves use their own AI outputs without verification.

The validation of the correctness of the finished text is not enough to evaluate the students’ ultimate epistemic responsibility. Students should be able to justify their claims and articulate their sources, evaluate counter-evidence and reveal the role of AI in a transparent manner. Kim et al. (2025) also found different patterns of student interactions with AI, with the same tool potentially being used to help some students to plan and evaluate their work while others rely directly on the output.

The key principle that grounds the ethical and educational legitimacy of the AI use is the student's ultimate epistemic responsibility. Students can use artificial intelligence to support them but they need to check the statements they make, rationalize the decisions they make and be responsible for the outcomes of the product they create. The educational approach needed to preserve epistemic responsibility should be technology-enabled but human-centered in its decision-making and accountability.

Epistemic autonomy here is not understood as independence from all sources of information, but rather as the critical, justified and responsible management of one's relationship to those sources. In this approach, the use of artificial intelligence is not the challenge to epistemic autonomy, but the question of whether the system replaces human thought is the central criterion. Ng et al. (2021) extended the understanding of AI literacy by adding dimensions of evaluation and ethics. They argued that skills of technical use are insufficient for epistemic autonomy.

The first part of the text has been dedicated to the philosophical background of autonomy and the difference between moral and epistemic autonomy has been accounted for. Epistemic autonomy does not imply that students should not trust authority, but instead that they should decide for themselves which sources of knowledge to trust, and why. To think for oneself then is not to generate all knowledge for oneself but to take responsibility for knowledge learned from others or technological systems.

The second part looked at epistemic autonomy as a central aim of education. Education should not only give students the right answers, it should develop their ability to ask questions, to assess evidence and to build justification". And, as Lee (2025) shows through her model of humane epistemic agency, the active participation of students in the processes of knowledge production has become even more important in the era of artificial intelligence.

The third part discussed the difference between information and knowledge. Generative artificial intelligence raises accessibility of information but comprehension, meaning making and epistemic ownership still rely on cognitive activities of students. Kasneci et al. (2023) highlighted the risks of misinformation and fabricated sources, showing that quick access does not mean accuracy or learning.

The fourth part identified the main types of epistemic dependence as overtrust in artificial intelligence, reduced cognitive effort, automation bias and delegation of decision-making responsibility. Tian and Zhang (2025) showed that AI dependence may negatively relate to critical thinking through the mediation of cognitive fatigue. The use of AI does not necessarily imply epistemic dependence; the risk depends on the kind of use, the level of literacy of the students and the pedagogical design.

Finally, four principles were developed to preserve epistemic autonomy: critical AI literacy, comparative source use, output verification, and the student's ultimate epistemic responsibility. Veldhuis et al.'s (2025) critical AI literacy framework and Vendrell and Johnston's (2026) approach to cognitive friction suggest that artificial intelligence should be designed not as an authority that gives students answers, but as a tool to challenge and develop student thinking.

The general conclusion of Chapter One is the fact that the goal of education in the era of artificial intelligence is not to create people who are independent of technology. Education should produce epistemic subjects who make use of the opportunities of artificial intelligence but who also question knowledge claims, compare sources, evaluate the quality of evidence and take responsibility for their final judgments. Epistemic autonomy does not demand that we break the tie between human beings and technology, but that the tie be subjected to human reasoning and responsibility.

Epistemic autonomy is not to say that epistemic authorities are unnecessary or should be thrown out altogether. Teachers, experts, scientific institutions and systems of artificial intelligence are still all kinds of information sources in educational environments. The central question concerns the conditions under which these sources can be defined as legitimate epistemic authorities, and the way in which their authority can be balanced with the autonomy of the student. In Chapter Two, we will therefore analyze epistemic authority in education through the knowledge authority of the teacher, the trust in expertise, the algorithmic authority and the emergence of artificial intelligence as a new epistemic actor.

# ARTIFICIAL INTELLIGENCE AND EPISTEMIC AUTHORITY IN THE EDUCATION SYSTEM

## The Idea of Epistemic Authority

The processes of knowledge acquisition do not have a limit in the individual abilities. In many spheres, from everyday life to scientific research, people have to rely on the testimony of teachers, experts, institutions and other sources of information. Hardwig (1985) discusses epistemic dependence to show that modern bodies of knowledge reach far beyond what any individual can directly verify. It follows that reliance on the knowledge of others should not be seen as a necessary surrender of epistemic independence but instead as one of the conditions for rational action in complex informational environments.

Epistemic authority is the special weight of a person's utterances regarding a certain subject in the belief-formation process of an information recipient. Zagzebski (2012) argues that an appeal to epistemic authority rests on the reflective judgment that the authority is more likely than the individual alone to reach the individual's epistemic goals. In this method, trust in an authority does not mean the abandonment of thought altogether, but a rational recognition of a person's own limitation and the superior competence of another person.

Epistemic authority does not mean that the person must be right about everything. In some cases experts should be viewed as authorities who guide decisions and in some cases as advisers who provide evidence, argues Stewart (2020). The authority of an expert is domain-specific, and we should not expect the same epistemic weight when the expert strays from that domain or when the evidence available is uncertain. Thus, epistemic authority is not an absolute, unlimited and entirely personal status.

This traditional relationship is being transformed by artificial intelligence systems. These systems have the ability to process many sources in a short time and have more accuracy than humans in some tasks, if given fluent explanations. However, Ferrario et al (2024) argue that accuracy is not sufficient for true expertise as expertise also involves understanding, research abilities and sensitivity to the epistemic needs of the recipient. The high performance of artificial intelligence does not, therefore, automatically endow it with unlimited epistemic authority.

Hauswald (2025) takes a more restrictive stance, arguing that some artificial intelligence systems can be understood as artificial epistemic authorities when they perform well-defined tasks and have been shown to be reliable. This approach does not presuppose that artificial intelligence has to understand in the same way as a human being, but rather, it emphasizes the system's performance in allowing users to acquire true beliefs. But the domain, data conditions and duration of reliable operation of a system must be continuously monitored.

The issue of epistemic authority is directly linked, from an educational point of view, to the importance that students give to artificial intelligence. Urhahne et al. (2026) note that ChatGPT might give students the impression of an almost omniscient authority with the right answer to every question. When students mistake this impression for the system's genuine epistemic powers, linguistic fluency and rapid response production may be taken as signs of expertise.

An educationally acceptable epistemic authority must not destroy the capacity for judgment of the student. Pietarinen (2026) argues that generative artificial intelligence should be placed in education not as an epistemic authority or a replacement for a human expert, but as a heuristic collaborative tool supporting critical thinking. Thus this approach preserves the final epistemic authority in the human subject, without denying the contributions that artificial intelligence can bring.

Epistemic authority does not require that the recipient of information accept a source blindly. It means the justified recognition of a source's superior competence in a specific domain. AI may be a source of reliable information for some tasks, but its authority should be limited by domain, purpose, record of accuracy, and oversight by humans. For it to begin to clarify this framework it is necessary to distinguish authority from power, command, and expertise.

**What Authority Is**

“Authority is often linked to power, violence, or the power to command. However, epistemic authority does not force people to subscribe to certain beliefs by physical or institutional penalties. Its influence results from the recipient’s perception that the source is more competent or reliable than the recipient. For Zagzebski (2012), epistemic authority is not based on external coercion, but on a reflective decision based on the epistemic self-trust of the individual.

So epistemic authority must be divorced from political or legal authority. An institution can enforce compliance with rules through the threat of sanctions; the epistemic authority of a teacher or expert instead depends on knowledge and the possibility to provide justification. Experts mediate evidence for nonexperts by interpreting it; however, this mediation does not render expert opinion immune to criticism (Stewart, 2020).

The legitimacy of epistemic authority is conditional upon the authority holder's being in a superior epistemic position regarding a given subject. This superiority may be related to education, experience, methodological competence, successful record of prediction or recognition in an expert community. Goldman (2001) notes that nonexperts may assess experts by the quality of their arguments, the opinions of other experts, their record of success, and possible conflicts of interest.

Artificial intelligence adds a new dimension to the idea of authority, because artificial intelligence systems are not persons educated in the traditional sense, responsible, or part of a particular professional community. Artificial intelligence not only transmits existing information but also transforms the production and legitimation of truth (Shin, 2026). The ways in which algorithmic systems rank, summarize, and generate answers might have implications for what sorts of knowledge are considered visible and trustworthy.

So the power of artificial intelligence comes from its informational veracity and also its infrastructural placement. If students repeatedly use the same system as their first place to look for information, AI might indirectly influence the questions that are asked, the concepts that are highlighted, and the views that seem normal. This kind of authority is not exercised through an explicit command, but by structuring the epistemic environment.

So epistemic authority is a status that is both recognized and produced. The actual competence of the source is important, but its authority may also be enhanced by the meanings attributed to that source by students, institutions or society. Presenting artificial intelligence systems as neutral, unlimited and omniscient tools may give them an authority that they do not really have.

The central issue in the era of artificial intelligence, then, is to distinguish between apparent authority and justified epistemic authority.

**Features of Epistemic Authority**

The first characteristic of epistemic authority is that it is domain-specific. Just because you're an expert in one thing doesn't mean your opinion on other things carries the same weight. As Goldman (2001) put it, expertise is having a large number of true beliefs in a given domain of questions and the ability to answer those questions correctly. In education, students should be able to distinguish the subjects on which a teacher or an artificial intelligence system is really competent.

The second feature is that epistemic authority is not only about the transmission of information. In the words of Ferrario et al. (2024), authentic epistemic expertise is not simply a matter of storing true propositions or making accurate predictions. Experts must understand the relationships between items of knowledge; they must develop methods to respond to new problems and evaluate the causes of their own errors.

The third characteristic is sensitivity to the epistemic needs of the recipient. Instead of presenting the same information to all students in the same way, the teacher takes into account students' developmental levels, misconceptions, and learning goals. Although artificial intelligence can generate personalized responses, it cannot evaluate the pedagogical, emotional and social implications of such responses as a human teacher can.

The fourth capability is the ability to recognize the possibility of error and the limits of one's own knowledge. A good expert does not just tell you what is known; the expert also tells you what is not known, what conclusions are not certain, where the evidence is not enough. Kerasidou and Kerasidou (2025) suggest that medical artificial intelligence may create strong statistical associations, but that these associations cannot replace causal explanation and professional judgement.

Fifth characteristic is responsibility. Human experts can explain what led them to a judgment, respond to criticism and face professional or institutional penalties. Shneiderman's (2020) human-centric approach to artificial intelligence advocates for the use of technological systems following the principles of reliability, auditability, and human control. The power of a system that cannot be held to account must therefore always be limited and subject to oversight by humans.

Of these features, the notion of artificial epistemic authority developed by Hauswald (2025) emphasizes particularly reliable performance. The output of a system can provide a strong epistemic reason if it is more accurate than a user in a well-defined task. BUT Ferrario et al. (2024) critique this, reminding us that successful prediction should not be confused with understanding and responsible expertise.

This discussion leads to a twofold conclusion concerning education. It is not realistic to consider every piece of information generated by artificial intelligence as worthless, because such systems can provide substantial epistemic support in particular tasks. However, AI output should not be presented as an authority that completely replaces the independent reasons available to students and teachers.

Epistemic authority is domain specific, it has knowledge and methodological competence, sensitivity to the recipient, awareness of fallibility, and accountability. Artificial intelligence might show trustworthy performance in some domains, but it is different from human experts in understanding, pedagogical sensitivity and responsibility. Therefore, the relationship between expertise and trust should be further investigated.

**Expertise and Trust: The Relationship**

The connection between expertise and trust arises because nonexperts cannot directly evaluate all of the evidence that experts possess. Modern scientific knowledge, Hardwig (1985) argued, requires people to trust the knowledge of others. This dependence is not irrational; but it does demand that the grounds and limits of trust be specified.

Trust doesn't mean believing everything an expert says without question. Goldman (2001) suggests that the nonexpert may evaluate the expert indirectly. How an expert debates other experts, professional peer support, their prior record, conflicts of interest and how evidence is presented can all feed into a decision on trust.

Boyd (2022) argues that it is not enough to only look at the visible characteristics of an individual expert in online information environments. Peer review, criticism, data sharing, institutional oversight and reciprocative monitoring within expert communities all support the reliability of scientific experts. Epistemic trust is thus often not only directed at a particular individual but also at the epistemic institution that supports that individual.

Epistemic trust is not simply a matter of calculating the chance of a source producing true information, as Fricker (2021) demonstrates. Trust in another's testimony may also depend on that person's intention to inform and on interpersonal norms. This dimension is particularly relevant in the teacher–student relationship, as a teacher is responsible not only for transmitting information but also for the learning and epistemic development of the student.

Epistemic trust also means that the individual is seen as a subject of knowledge (Dormandy, 2024). Students have confidence in teachers as sources of information, but teachers need to recognize students' abilities to ask questions, to object, and to produce knowledge. An absolute one-way relationship of authority can harm students' epistemic self-trust.

These elements of interpersonal trust are not present in the relationship with artificial intelligence. An AI system is not a moral agent that cares about the epistemic well-being of the student or feels responsible for the consequences of its answers. However, human-like language and personalized responses can cause users to have a similar attitude to interpersonal trust.

The research of Urhahne et al. (2026) reveals that students' epistemic beliefs about artificial intelligence (AI) have a significant impact on their use of such systems. Students with a belief in knowledge as certain, quickly obtained and coming from one authority may see ChatGPT as a source that does not need to be questioned. Thus, students need to develop not only AI-related skills but also more sophisticated epistemic beliefs about the nature of knowledge.

It is rational to depend on expertise as a way of dealing with the unavoidable epistemic dependence of the nonexpert. Such trust needs to be underpinned by domain knowledge, a history of success, institutional oversight and the capacity to give reasons. Trust in artificial intelligence should be differentiated from interpersonal trust and considered to be a limited trust based solely on verified system performance.

## The Difference between Trust and Trustworthiness

Trust is a stance taken by the recipient of information towards a person, an institution or a system. A source is trustworthy to the extent that it can give us accurate information, correct its errors, and do the epistemic job it is supposed to do successfully. Thus, a source can be reliable without being trusted, and a non-reliable source can be over-trusted.

This distinction is relevant to education in that the appearance, linguistic fluency or wide usage of a source may influence the student. Seeing artificial intelligence systems providing confident, organized and academic-style answers may make students more trusting. But such confidence might not be justified by the actual correctness of the system, the data used, the relevant subject area or the currency of its information.

According to Shneiderman (2020), trustworthy AI should be verifiable, auditable, safe, and subject to human control. User trust should therefore not be based on a general positive attitude towards technology, but on the performance of a system in specific tasks. Overtrust is caused by trust in excess of system capacity and trust below the level justified by system capacity can lead to unnecessary rejection of useful technology.

Generalized distrust could also lead to problems in AI mediated communication, say Sahebi and Formosa (2025). Although treating all content as less trustworthy simply because it was produced by artificial intelligence may be a justified epistemic caution in some cases, it may lead to unfair distrust and epistemic injustice towards speakers in others. So, the level of trust must be calibrated in the light of context and evidence.

Ferrario et al. (2024) note that just because artificial intelligence is very accurate with past data does not mean the technology will always be reliable in future conditions. Changes in the data distribution, a declining currency or use in a new context may reduce the performance of the system. Trustworthiness is not a fixed characteristic, but a relationship dependent on the task and the conditions.

According to Gazit (2026) in her institutional approach, especially in closed and difficult-to-explain artificial intelligence systems, trustworthiness cannot be based solely on algorithmic transparency. Epistemic confidence can also come from the authority of the institutions that employ the system, the oversight mechanisms they use, and their accountability structures. Similarly, in education, the trustworthiness of an AI system must be validated not solely by the technical developer, but also through institutional policies, teacher oversight, and verification procedures.

To maintain epistemic autonomy, students' trust and a system's trustworthiness need to be aligned. Students may use artificial intelligence as a critical support tool, when they can calibrate their trust in the tool based on evidence, performance, and context. But students can become dependent on algorithmic authority when trust is founded on the fluent language or authoritative appearance of the system.

## The Teacher's Knowledge Authority in Pedagogical Practice

The educational process is founded firstly on a certain epistemic inequality. Teachers are ahead of students in subject matter, method and educational experience; students participate in learning with knowledge and reasoning capacities still in process of development. Dewey's (1916) view of education demonstrates that this inequality does not justify students to be passive. The teacher's epistemic superiority should not be used to make students dependent on the teacher, but to help them gradually become more independent inquirers and thinking subjects.

A teacher’s epistemic authority is not based simply on her having more true beliefs. Education is a normative process by means of which pupils are introduced to activities which are valuable in relation to knowledge and understanding (Peters, 1966). Teachers select information, arrange it according to students’ developmental levels, explain relationships between concepts, and intervene in misapprehensions. Hence, the teacher is not a simple transmitter of information but a pedagogical interpreter who constructs the educational meaning of knowledge.

Generative artificial intelligence can partly replace the teacher's information-providing functions. Large language models can provide rapid explanations, feedback, and personalization, but they may also produce false, biased, or fabricated information, as highlighted by Kasneci et al. (2023). The effect is not the destruction of the teacher’s informational authority, but its transformation. The role of the teacher changes from the source of information to the assessor of the reliability of the different sources, including artificial intelligence, and teaches students how to assess such sources.

In Tishenina (2026), the functions of generative artificial intelligence cannot replace the educational role of the teacher in its full capacity, because teachers do not only lead students towards certain answers, but also summon students to be responsible subjects in the world. AI may be able to provide answers, but it won’t be able to judge with the same relational fidelity as a human teacher whether a student is hiding behind an excuse, evading independent thought, or abdicating ethical responsibility in the learning process.

The teacher’s epistemic authority should not depend on students suspending their own judgment. Wu et al. (2025) argue that in human-generative AI learning partnerships, it is necessary to preserve human epistemic agency through goal setting, evaluation, and responsibility. The same principle is applied to teacher-student relationships. The teacher should not be an authority that always makes decisions for the students, but a temporary and justified guide who develops the capacity of students to make decisions.

## The Role of the Teacher as a Knowledge Provider

In the traditional educational environment, one of the main channels through which students access course content has been the teacher. Textbooks, libraries and digital resources have changed this position and generative artificial intelligence has further decentralized access to information. But as Dewey (1916) has noted, the educational value of the teacher is not in the knowledge itself but in the capacity to relate the experience of the student to the subject matter in a meaningful way.

Teachers choose information sources, assess the educational relevance of specific content, and tailor complex knowledge to the student's developmental level. Artificial intelligence can also adapt content, but the study by Yue et al. (2024) shows that for successful AI education, the conjunction of technological, pedagogical and content knowledge is needed. Such integration demonstrates that the teacher should be not only an artificial intelligence user but also an expert able to evaluate its pedagogical appropriateness.

The teacher's authority as a source of knowledge does not permit the presentation of information as closed to criticism. Rather, teachers should expose the evidence underlying claims, the controversies within a field, and the shortcomings of their own explanations. Evaluation and ethical responsibility are core competencies identified in the AI literacy approach developed by Ng et al. (2021). Teachers need to have these competencies before they can expect them from students and should not present unverified AI output as course content.

Artificial intelligence's arrival in the classroom has further eroded the image of the teacher as "the one who knows the answer to every question." A teacher's epistemic authority is not necessarily undermined by admitting that one does not know something. A dependable authority can identify the boundaries of knowledge , conduct additional research when necessary , and rectify errors . When a teacher says, "I don't know; let's check it together," the teacher gives students an example of honesty about what one knows and the responsibility to do research.

One of the teacher's main functions as a source of knowledge is to teach students to make epistemic distinctions between sources. The danger of misinformation and fabricated citations highlighted by Kasneci et al. (2023) suggests that students may confuse a fluent answer with reliable information. Teachers need to explain the different epistemic statuses of scholarly articles, textbooks, institutional reports, online content and AI-generated output.

Teachers should also create a safe space for students to discuss differences between AI-generated output and the teacher's own knowledge. If teachers use their authority to settle all disputes in their favor, students may come to distrust both the teacher and their own judgment. In a classroom where evidence is examined together, however, teachers demonstrate expertise in processes of inquiry and justification, not a monopoly over knowledge.

The teacher's role as the source of knowledge is not vanishing with digital technologies and generative artificial intelligence. Rather, it is shifting from delivering content to assessing sources, pedagogical interpretation, and epistemic guidance. The authority of teachers should not stem from the claim to know everything, but from their ability to justify, delimit and educationally organize knowledge.

**The Authority of Pedagogical Legitimacy**

Pedagogical authority is the teacher's authority to set learning goals, organize activities, evaluate students and intervene when necessary. Such authority is educationally legitimate, in Peters's (1966) conception of education, if it opens students to valuable forms of knowledge and activity. A teacher's decisions should be accepted not merely because of institutional position, but because they serve the educational development of the student.

According to Arendt (1961), the teacher's authority is based on the assumption of responsibility for the world. The teacher brings knowledge, culture and a shared world that existed before the students. This position does not mean that teachers have unlimited power; they are accountable for the consequences of the content they choose and the choices they make. Teachers exercise pedagogical authority not just by virtue of being "in charge" but because they have taken on the responsibility of preparing students for entry into a common world.

The legitimacy of pedagogical authority is also related to the compatibility of the tools used to educational aims. Wright et al. (2026) found that teachers' perceptions of artificial intelligence include dimensions of benefit, risk, equity, and professional development. Teachers can't simply adopt artificial intelligence because it makes things more efficient – it has to serve student learning, equitable participation and intellectual development.

Legitimate pedagogical authority gives students reasons. If a teacher gives an assignment, restriction or rule about artificial intelligence, just because "the teacher said so," the authority is severed from justification. Teachers should be able to articulate what skill a particular task is meant to develop, why the use of AI is limited at certain stages and why technology is allowed at others.

Another measure of legitimacy is whether pedagogical authority helps students need less direction from the teacher in the future. Tishenina’s (2026) definition of teacher as one that calls the student into subjecthood makes clear that the ultimate purpose of education is not to create compliant and obedient users. Teachers have to use their authority to develop students’ skills to ask questions, to object, to take responsibility.

The same criterion applies to the authority of the teacher over artificial intelligence. If teachers are only using technology as a way to reduce their workload or to gain more control over the classroom, then pedagogical legitimacy may be weakened. The use of power to assess student work with inexplicable algorithms or to continuously monitor behavior may become disproportionate to the educational aim. According to Pietarinen (2026), the line between acceptable guidance and manipulative influence in AI-mediated education must be carefully drawn.

Pedagogical authority may be based on the teacher’s institutional position, but it is legitimized by its service to the student’s educational development, by giving reasons for decisions and by taking responsibility. The teacher's role is not to manage students forever, but to help them learn how to manage their own learning and judgments, little by little. It is thus necessary to delimit the limits of pedagogical authority.

## The Boundaries of Teachers’ Authority

The first boundary of teacher authority is the teacher’s subject matter. Being competent in a discipline does not imply that the teacher has the same epistemic weight in all areas of expertise. Teachers should know their own limitations and refer students to other experts or reliable sources as needed. Indeed, the awareness of these limits increases rather than decreases the teacher’s credibility.

The second limitation is the freedom of thought and conscience of the student. Teachers may present scientific evidence, philosophical arguments, and social values for discussion, but teachers should not require students to adopt a controversial position as a matter of personal loyalty. Freire's (2000) critique of the banking model of education shows how the student is objectified when knowledge is treated as ready-made content to be deposited in the student's mind.

The third limit is the teacher's power of judgment. The power to give grades and to determine achievement may make it difficult for students to challenge the teacher's views. They should therefore be transparent regarding assessment criteria, base feedback on evidence and offer students opportunities to challenge assessments. Pedagogical authority should not demand personal conformity by exercising power over a student's academic future.

The fourth limit concerns decision-making through artificial intelligence. The human-centered AI approach requires accountability to lie with the human actor (Shneiderman, 2020). Teachers should not offer an AI-generated grade, risk assessment or content recommendation as the final answer. The responsibility for the decision rests with the teacher and the institution, although algorithmic output can support pedagogical judgment.

Fifth limit is the prohibition of manipulation. Pietarinen (2026) claims that AI-mediated educational practices could be manipulative when they invisibly limit the students' choices, guide them towards targets of which they are unaware or emphasize measurable performance over educational benefit. Teachers should make their intervention points as visible as possible and ensure informed participation on the part of the student.

The sixth limit is maintaining the student's epistemic agency. Wu et al. (2025) proposed that learners be kept with the functions of goal setting, output evaluation, and assuming responsibility. If the teacher or artificial intelligence chooses the method and conclusion for each question, the student is only an implementer. Educational guidance should not entirely suppress student initiative.

Lowering the power of the teacher does not mean a loss of guidance in the classroom. Teachers may provide more structure depending on the age of the student, the subject matter, and the student's level of learning. But such a structure should be transparent, proportionate and temporary. Teachers should reduce control and give students more intellectual responsibility, as students are able.

The authority of teachers must be limited by their domain of expertise, the fundamental rights of students, objective evaluation, supervision by humans, and the preservation of epistemic autonomy. Limiting authority does not make the teacher ineffective, but rather guarantees that educational power is used responsibly and with justification. These limits are the main criteria for identifying legitimate authority and not coercion and blind obedience.

## The Difference Between Authority, Coercion and Blind Obedience

The fundamental difference between authority and coercion is that authority is based on a claim to legitimacy, whereas coercion is based on sanctions and fear. The epistemic relation of authority is one where students accept the teacher's point of view because of the teacher's competence in the relevant subject and the reasons given. If students echo the same view for fear of low grade, humiliation, or exclusion, there may be outward conformity, but no real epistemic acceptance.

According to Arendt (1961) when force is used, authority is undermined, because the use of force indicates that the basis of voluntary recognition of authority has disappeared. In education, if someone shouts, punishes students for objecting, or asserts one's own view as beyond dispute, this may not be a sign of strong authority, but rather a sign of weakness in the ability to justify.

Blind obedience is when students accept what authority says is true without needing proof or reasons. Even where such students do accept a correct view, they cannot be said to have developed epistemic autonomy . In the banking model, criticized by Freire (2000), the student is a passive receptacle of transmitted content. This model maintains the teacher as the master of knowledge and restricts the student's ability to question and create meaning.

The teacher's legitimate authority does not threaten the student's objection. Students' counter-questions, alternative sources, or errors in the teacher's explanation can increase the epistemic quality of the classroom. Teachers need not accept all objections but they must justify their position and take student evidence seriously.

Artificial intelligence brings a new dimension to coercive relationships and blind obedience. When students and teachers are presented with algorithmic recommendations as if they are neutral, objective, and impersonal decisions, they may see no point in resisting. Kasneci et al. (2023) show that the risks of misinformation and bias expose that technological systems are not to be assumed as epistemically neutral. "the system decided it" must not be an excuse for teachers to avoid taking responsibility for justifying their decisions.

Teachers should also avoid employing artificial intelligence as a covert surveillance device to bolster their authority. Algorithmic judgment as irrefutable proof of a student's work being produced by AI is a punitive coercion in the circumstances of an uncertain determination. A human-centered approach would be for teachers to explain the limitations of algorithmic evaluation, to give students a right of response and to make final decisions based on independent evidence.

If we separate authority from coercion, the educational implication is that discipline and freedom are not opposites. Learning can need rules, deadlines and responsibilities. Such arrangements may be part of legitimate authority provided they are explainable, open to discussion and proportionate to the educational aim.

Legitimate authority rests on expertise, rationale, and educational responsibility; coercion depends on fear and sanctions; blind obedience requires students to suspend their own judgment. The authority of the teacher is not meant to keep the students always dependent on the teacher but to enable them to become independent step by step through an understanding of reasons. Likewise, artificial intelligence should not be used to reinforce the unquestionable authority of teachers or institutions, but as a pedagogical support to be criticised.

**Can Artificial Intelligence Be a New Epistemic Authority?**

Generative artificial intelligence systems have transformed the ways in which people search for information and receive answers. Large language models combine different types of content and generate answers directly, unlike traditional search engines that tend to give users lists of sources. Freiman (2024) argues that the beliefs formed by means of conversational artificial intelligence constitute a unique form of epistemic relation to technology that cannot be reduced to either standard tool use or interpersonal testimony. This means that artificial intelligence has become an epistemic intermediary, i.e., a source that participates in the construction of what the users believe.

To be an informational intermediary, however, is not to be an epistemic authority. Epistemic authority is the special weight given to a source in the recipient's process of belief formation because that source is more competent and reliable in some particular domain. Artificial intelligence systems may be very accurate, but they do not understand, they don't espouse truth, and they can not take epistemic responsibility (Ferrario et al., 2024). In this respect, artificial intelligence can be a powerful epistemic tool, but it is not an epistemic authority in quite the same way as a human expert.

A major reason for trusting the epistemic authority of artificial intelligence is the capacity of these systems to deliver rapid responses on a wide range of subjects. Students ask questions about history, philosophy, mathematics, health or daily life to the same system and they often get organized answers. According to Tlili et al. (2023), the accessibility and the interactive features of the chatbots can make the users to consider them as more than just a software, but as actors to be consulted.

Such an appearance can create a gap between the actual capacities of the system and the capacities users ascribe to it. Lu and Hu (2025) contend that chatbots can be attributed understanding, intention and independent reasoning that results in false mental states. The fact that an AI response sounds human is not evidence that the system has meaningful beliefs or a conscious commitment to truth.

Thus the epistemic status of artificial intelligence should not be explained in terms of one single category. If there are well-bounded tasks, with independently verified boundaries and performance, then the AI-generated output might give a strong epistemic reason. Kerasidou and Kerasidou (2025) agree that AI may be better at some predictive tasks . Particularly in the setting of medical AI. They do argue however that the lack of causal understanding and professional responsibility means that the authority of the system cannot be compared to that of a human expert.

Educating people properly is not about treating artificial intelligence as a tool that has no authority, or as an independent expert in every field that can be trusted. It can be a limited epistemic advisor but the weight we give to its output should be a function of the relevant domain, currency of information, demonstrated performance, and independent verifiability. Its authority must not be a function of simulated personality or persuasive style but of performance under certain conditions.

Artificial intelligence is a new and powerful epistemic intermediary that influences users' belief formation. But it cannot be treated as an epistemic authority in the same manner as a human expert. The epistemic weight of AI should be limited in terms of how it is regarded as a source of information and what makes it so regarded.

**Our View on Artificial Intelligence as an Information Source**

Perception of the information source of artificial intelligence is conditioned by the users' experience of the structure of the interface. Artificial Intelligence systems are able to answer questions instantly, adapt to the language of the user and relate to previous messages. Freiman (2024) contends this sort of interaction is a specific sort of thing, somewhere in between testimony given by a person and data generated by a tool. As the system's style is conversational, users may treat it as a person testifying even when they are aware that the system is just generating outputs.

The anthropomorphic language contributes to the impression of the system as having knowledge. If AI says things like "in my opinion," "according to my assessment" or "the main reason is," users might think that the system has thought and judgment. Lu and Hu (2025) contend that such false mental attributions would lead users to attribute an undeservedly high epistemic value to AI outputs.

Another reason why it is seen as a source of information is that AI appears to be all-encompassing. They may take the system's ability to answer questions across a range of domains as evidence of integrated and broad expertise. Large language models, however, generate text based on huge data patterns, and do not produce equally recent, accurate or methodologically reliable data in all domains. Kasneci et al. (2023) show that large language models are able to generate false or fabricated content that has the same narrative structure as true information.

Another factor in the perceived authority of artificial intelligence systems may be the tendency of artificial intelligence systems to make their sources invisible. "A traditional search process will present users with different sources; artificial intelligence might combine those sources into a synthesized response. That structure may move judgments of truth from a debate between different sources to an algorithmic synthesis (Shin 2026). As such the system is in danger of becoming an infrastructural authority that determines what seems to be important and true information, as opposed to simply a tool for presentation of information.

Trust and opacity in AI are studied by Hähnel (2025), who shows that users often do not understand why a system comes to a particular conclusion. The reasoning of human experts may also not always be fully transparent, but experts can take professional responsibility, respond to criticism and explain their methods. Trust in artificial intelligence is not a personal relationship with normative dimensions.

Education has to make this difference visible, experientially, between being seen as an information source and actually being trustworthy. Students could be asked to pose the same question to an AI system, a textbook, and a peer-reviewed article, and compare the evidential bases of the answers. The systematic review by Park et al. (2025) indicates that students might have difficulty in critically assessing the outputs of generative AI and that this skill should be explicitly taught.

Artificial intelligence is considered a source of information, not only because it can provide accurate information but also because it is fast, easy to access, speaks like human beings and can hide many sources in one answer. Students need to understand the difference between AI's ability to regurgitate information and AI actually knowing the information." One of the chief difficulties in making this distinction is the tendency to interpret fluency of response as evidence of truth.

**The Reality of Fluent Answers**

Fluency in language can make a text easy to read, internally consistent and give confidence . But there is no necessary connection between fluency of expression and truth. Kasneci et al. (2023) found that large language models can generate false or fabricated information in a persuasive and academic manner. Users may confuse the formal quality of a response with epistemic quality.

Many artificial intelligence systems respond in an absolute manner, with no explicit indication of uncertainty. A human expert might say that the evidence is thin or that scholars differ or that a question is outside the expertise of the expert. In contrast, a language model could potentially produce a coherent answer even if there is not enough evidence. Tlili et al. (2023) claimed that this feature could lead to difficulties in detecting misinformation in educational environments and could foster students' overconfidence.

The fluency lends authority, which may also obscure the difference between explaining and justifying. The artificial intelligence systems can create additional sentences that appear to offer a reason for why an answer is correct, but these are not extracted from a real chain of evidence or trusted sources. Shin's (2026) critique of algorithmic truth illustrates how, when justification is replaced by computational coherence, truth can be a system-presented thing rather than a debated and established thing by reasons.

Smooth replies can also make users doubt their knowledge. If an AI provides a confident answer that runs counter to a student's judgment, the student may conclude that the system knows more than they do, rather than investigating the evidence for the AI's claim. Frequent use of generative artificial intelligence may be linked to decreased levels of information-verification behavior, suggesting that habitual use might lessen the perceived need to verify information, Yan et al. (2025) found.

To this end, the formal and epistemic properties of AI-generated responses should be evaluated separately. In a first stage students may decide whether or not an output is clear, organized and understandable and then in a second stage they may check its sources, evidence and contradictions. As Veldhuis et al. (2025) state, critical AI literacy requires users to question not only the results of a system, but also the data and design conditions that lead to the production of these results.

Fluent but incorrect outputs of artificial intelligence may be used to design educational mistake detection activities. Students can be given texts that seem convincing and asked to identify factual claims, claims that require sources, and gaps in justification. Such activities may be more helpful than abstract warnings that fluency does not equal truth.

Fluency of speech can be informative in reading the output of AI, but it does not equate to truth, weight of evidence, or expertise. Students need to learn how to differentiate between linguistic credibility and epistemic credibility. This difference makes it necessary to analyze in more detail whether artificial intelligence has the ability to create knowledge or not.

**Artificial Intelligence's Capacity to Generate Knowledge**

The ability of artificial intelligence to generate knowledge depends on the distinction between 'generating new content' and 'having knowledge'. Based on patterns in training data , large language models can generate sentences , explanations and proposed solutions that have not occurred before in exactly the same form . Freiman (2024) admits that these technologies do contribute to the formation of users' beliefs, but that does not mean that the systems are knowing subjects in the same sense as human beings.

If we consider the production of knowledge as the production of a true proposition, then one could speak of artificial intelligence systems as producing knowledge. Such systems can identify relationships within large data sets, detect trends that human observers may miss and assist in the generation of new hypotheses. Ferrario et al. (2024) claim that epistemic expertise is not only about the production of accurate output, but also about understanding, about being able to research, and about an epistemic orientation towards truth.

The main difference is about the problem of understanding. Artificial intelligence can offer many reasons to explain a concept but it is hard to argue that the system understands why the correctness of the explanation is important, it has an epistemic concern when it makes a mistake or it revises its beliefs through reasoned judgement. Kerasidou and Kerasidou (2025) argue that the fact that artificial intelligence is able to perform some epistemic operations to a high degree of accuracy is not enough to count it as an epistemic agent in the strong sense.

Belief-like states in conventional artificial intelligence systems should be considered regarding their ethical and epistemic implications, according to Ma and Valton (2024). But their approach does not necessarily assume conscious beliefs which are the same as those of human beings. Artificial intelligence classifications, predictions and profiles can affect people's lives and should be evaluated not only on accuracy but also on moral and social consequences.

The production of knowledge by artificial intelligence is also inseparable from human effort and institutional forms. Models are trained on human-generated text, selected datasets, labeling processes, feedback, and design choices. Anlı (2025) suggests that the portrayal of artificial intelligence as an autonomous knowledge producer may obscure the human labour and relations of responsibility that influence its outputs.

Opacity also constrains the epistemic status of knowledge claims produced by AI. Schmidt et al. (2025) argue that even if a system can reliably generate correct outputs, the conversion of a true belief into knowledge can be problematic in terms of safety and epistemic luck in high-stakes settings where users cannot inspect how the output was generated. The fact that the output is true does not mean that the user knows why it should be accepted as true.

The generative power of artificial intelligence should thus be understood as part of a distributed epistemic process in which both human and technology participate. Artificial intelligence is useful for pattern recognition, synthesis and proposal generation while human actors are needed to consider research question, methodological appropriateness, meaning, value and responsibility. The link between epistemic production and accountability, however, is weakened when the outputs of the system are not human-monitored.

Artificial intelligence can generate new and sometimes correct content. But this ability does not warrant the conclusion that it is an understanding, knowing and responsible epistemic subject like a human being. While AI might be a strong element in the production of knowledge, the epistemic value of its claims has to be subject to human validation, methodological critique and institutional accountability.

**Faults and Prejudice in AI Outcomes**

AI generated output error is not limited to the production of false factual statements. Other forms of epistemic error include fabricated sources, correct information in the wrong context, obsolete information, over-generalizations, and the omission of countervailing perspectives. Kasneci et al. (2023) note that the content generated by large language models seems authentic, but it is fabricated, especially regarding sources and citations.

The system's tendency to output probabilistically consistent and contextually relevant outputs can lead to incorrect outputs, commonly known as "hallucinations". Artificial intelligence is not a person who knows the truth and deliberately says something false. The system produces text without knowledge of the difference between falsity and truth as a human epistemic subject does. Thus, the user should not interpret confident language as the model tracking its internal truth.

Bias is not a one-time error. If the training data used to build the model is richer in some groups, languages, cultures or intellectual traditions, the model's outputs may make some perspectives seem more normal or trustworthy. As Veldhuis et al. (2025) argue, critical

AI literacy has to critique system errors, but also relations of representation and power.

The way a question is worded can also introduce bias. Instead, the AI might generate an answer that is consistent with the user's assumptions, rather than challenging them. Thus the system may reinforce one-sided initial framing. "Users should experiment with different prompts, consider alternative perspectives, and conduct some external research to enhance answer diversity.

Ma and Valton (2024) suggest that AI should not only be evaluated on the accuracy of the belief-like profiles it creates of a given person or group, but also on the potential for epistemic and moral harm. Even the most well-supported generalization can be unfair or discriminatory when used in a broad scope.

Opacity makes it harder to spot errors and biases. Hähnel (2025) shows that limited access to the functioning of the system creates serious problems for the basis on which trust in the user can be built. A system may be very accurate overall but the user may not understand why a particular answer was generated or what patterns of data contributed to the answer.

The best educational response to the risk of error and bias is not a single act of verification, but a multi-layered checking of sources. Students should identify factual claims in an output, consult primary sources where possible, seek out counterevidence, and consider whether different social or cultural perspectives have been omitted. Yan et al. (2025) propose that awareness of generative artificial intelligence and knowledge of its ethical risks could contribute to increased verification behavior.

We should not think of error and bias in AI-generated output as systems sometimes giving wrong answers. The problem also has structural epistemic limitations at the level of data, design, representation, context and user interaction . Only when the outputs of artificial intelligence are verifiable and subject to human review at any time can we concede to its epistemic authority.

**The Question of Authority in Algorithmic?**

In schools, algorithms are often framed as technical tools to sort information, analyze student performance or deliver suitable content recommendations. But, as Gillespie (2014) points out, algorithms do not merely select what is already there; they also create what is relevant, what is seen, and what is worthy of our attention. Thus, algorithmic systems do not only have an effect on individual answers, but they could also draw the boundaries of the epistemic environment to which users are exposed.

Algorithmic authority is the particular importance that users attach to a decision—be it a ranking, prediction, classification or recommendation—made by an algorithm. The power of algorithms is not only in the technical code, but in the social narratives that frame algorithms as objective, scientific, fast and free of human error (Beer, 2017). Therefore, users who are unaware of the assumptions that were used to generate the output may believe that it is more neutral than a human judge.

Such authority need not be exercised by issuing explicit commands in AI based systems. Suggesting particular sources to a student, giving priority to some answers, describing a learning path as ‘appropriate’ or profiling a student of risk could all constrain the possible action. Shin (2026) argues that artificial intelligence could transform how truth is generated and validated, thereby shifting epistemic authority from contexts of discussion to computational systems.

The main problem with algorithmic authority is that users often do not realize that this authority is there. A teacher or expert can provide guidance that is attributed to a particular person, whereas algorithmic selection can be perceived as an organic result inherent to an interface, ranking, or automated recommendation. If users are unaware of the data, criteria and institutional objectives that inform this selection, they may not critically question the guidance provided by the system.

Algorithmic authority should not be an excuse for teachers or institutions to shirk responsibility in education. Kasneci et al. (2023) point out that artificial intelligence systems can generate false and biased content. If the final authority on a student's grade, learning need, or academic integrity status is "the system decided it," then the algorithmic output is beyond question.

The critique of algorithmic authority doesn't mean that algorithmic systems are entirely useless. Such systems can detect trends in large data sets that humans have trouble finding, and they may provide teachers with valuable information. But their output is not to be taken as final judgments, but as clues which support human reasoning and are to be judged against independent evidence.

Algorithmic authority is not just a matter of technology producing right or wrong answers but also of its power to determine what information is visible, what options are deemed reasonable, and what choices are implemented unchallenged. The fundamental assumption of this authority is that technological systems are value-neutral and independent.

## Meaning of Algorithmic Authority

The first feature that distinguishes algorithmic authority from traditional authorities is that it is not based on an explicit claim to knowledge or power by a knowable person. An algorithm is not a person who talks, opens an account, or assumes professional responsibility. However, its outputs could guide the actions of teachers, students and administrators within institutional systems.

Gillespie (2014) contends that algorithmic systems influence public and epistemic orders by the specific power they have to influence what counts as relevant information.

This can also increase algorithmic authority by making the output of the system look more accurate or objective, simply because it was generated by a technical process. Often the user does not know the training data, the optimization objective or the error criteria the algorithm was trained on, but may assume that a numerical result is more accurate than human judgement. Beer's (2017) approach shows how algorithmic power is not in a vacuum (or the vacuum of algorithmic design) but is co-produced with social imaginaries and expectations.

This authority is infrastructural as well. An educational platform can make the algorithm the hidden manager of the learning environment by auto-organizing the content the student sees, the topics the student repeats, or the feedback the student receives. Even when student choice is not completely removed, some paths may be more available than others and some less visible.

As Shin (2026) notes, artificial intelligence modifies not only the production of information, but also the conditions of recognition and legitimation of truth. Algorithmic authority may then work less by explicitly claiming that 'this is the right answer' and more by making the user feel that other answers or further justification are not required. The response is fast, methodical and system recommended, so it does not seem to require further validation.

At stake, too, is the institutional authority of algorithms. A school or university receives institutional legitimacy for its output when it applies a particular system for formal assessment, early warning or content recommendation. Students are thus subjected to the algorithm and the institution that selected and uses the results of the algorithm. Hence responsibility cannot be shared invisibly between the technical developer and the educational institution.

Anlı's analysis of epistemological alienation (2025) suggests that when we delegate the production of knowledge to unaccountable systems, we may lose control over the outcomes. As algorithmic authority increases, the connections between those who create information, those who make decisions, and those who are held responsible for outcomes may become more opaque.

Algorithmic authority is the special epistemic weight that the output of algorithmic systems carries for technical, institutional, and social reasons. This authority is exercised not only in the making of decisions but also in the shaping of the environment of information and alternatives in which decisions are made. The primary difficulty with questioning is the illusion that algorithms function separate from human values.

**The Myth of Technological Neutrality**

The idea that algorithms simply take data and spit out results, without considering human values, interests or biases, is the illusion of technological neutrality. In practice the problem that the algorithm is going to solve, the variables it will use, and its measure of success are all pre-defined. Fu and Weng (2024) in a review, argue that the ethical use of artificial intelligence in education is contingent on equity, oversight by humans, and bias issues.

Data are not neutral reflections of social reality. Educational datasets can reflect historical inequalities of opportunity, socioeconomic disparities or institutional practices. When systems predict the future based on patterns of past success, they can reproduce existing inequalities as if they were natural, inevitable outcomes.

The seeming neutrality of measurement choices can also obscure their value-laden nature. If the only indicators of student success are the number of exams passed, tasks completed or times logged into the platform, then curiosity, creativity, ethical reasoning and contextual difficulties may not be part of the evaluation. Al-Zahrani (2024) argues that the efficiency discourse around AI in education needs to be balanced with its possible adverse effects on human relations, privacy and justice.

This assumption of technological neutrality facilitates the transference of responsibility. An algorithmic decision by a teacher or administrator can be framed not as a personal or institutional decision, but as the inevitable outcome of data. But the choice of data, the design of the model, the threshold values and the implementation of the decision are all human and institutional choices.

Veldhuis et al. (2025) advocate for a critical AI literacy that compels users to think about not just the correctness of an output, but also the groups that the system represents, the values it privileges, and whose interests it serves. What is emphasized here is not neutrality but situatedness and accountability.

The fact that neutrality is criticized does not mean that every algorithmic outcome is intentionally biased. The core argument is that it is not possible to build a system without objectives, data and evaluative criteria. A reliable educational application must therefore be able to clearly state the purpose for which a system is being used and the limitations under which it is operating.

Algorithms can do math, but the choices of data, the definition of the problem, the standards of success, the context of implementation are not value-neutral. These choices are invisible because of the illusion of technological neutrality and because the algorithmic output has too much authority. This decrease in invisibility brings demands for transparency and explainability.

### The Explainability Problem of Transparency

In general, transparency means that information about a system's data, goals, design, operation and terms of use is available and understandable. Explainability is the property of being able to explain a certain output or decision in a way that is understandable to the user. Markus et al. (2021) show that explainability is user and purpose dependent. An explanation that is adequate from a technical perspective for a developer may not be meaningful for a teacher or student.

Having the source code or the mathematical model of a system does not automatically make the system transparent to education. Even if teachers can see the technical details they may not understand why the system has flagged a specific student as being at risk. Likewise, providing students with a number of important variables does not necessarily explain the full logic of the decision.

Another is fidelity of explanations. The system might be computing a simplified version of the explanation given to users. A fully faithful explanation may be too complex to be useful for education, while a model that is not faithfully represented in an understandable explanation creates false trust. These criteria should be jointly considered in the design of explanations (Markus et al., 2021)

Transparency does not equal trust. Some models of trust do require explanation (e.g., Baron, 2025), but not all require explanation for appropriate types of reliance on AI. Users can trust, to some degree, systems, on the basis of verified performance, independent auditing and institutional safeguards, without full understanding of the technology.

But opacity poses serious epistemic problems, especially when high-stakes decisions are made about students. If the system is opaque and has high accuracy, the users cannot know why a certain output is correct. Say Schmidt et al. (2025). Relying only on general performance rates to make decisions that will affect a student's academic future does not adequately guard against individual errors.

Transparency links with the right to appeal a decision. If students don't know what data and criteria were used to make a decision about them, they can't demonstrate where the decision might be wrong. Explanation is not just a communication device to build trust. It is a precondition for accountability and correction.

Hähnel (2025) shows that the relationship between opacity and trust cannot be solved on the level of technical explanation alone. Also important is to look at who created the system, which institution is responsible for its oversight, who is liable when things go wrong and how users can contest the result.

Transparency and explainability are necessary, but not sufficient, conditions for trustworthy artificial intelligence. Explanations need to be relevant to the user, faithful to the underlying process, contestable and validated by institutional accountability. Otherwise, explanation may become a superficial means of legitimating system authority, rather than a means of enhancing trustworthiness.

**Overconfidence in A.I.**

Trust in artificial intelligence is the user's trust that a system will succeed in performing a given task appropriately. Trustworthiness is the combination of accuracy, robustness, safety and auditability that validates trust. As Durán & Pozzi (2025) point out, a user's trust does not mean that the system actually deserves this trust.

Unjustified trust is when users trust a system without knowing the tasks it successfully performs, the conditions under which it fails, and the procedures for checking the output. Trust does not come from mass adoption, institutionalization, or the ability of artificial intelligence to speak fluently.

According to the literature review of Benk et al. (2025), trust in artificial intelligence is dependent on the task, the user group, the culture and the risk level. Nor can the student's use of AI for language correction and algorithmic assessment of the student's academic achievement be subject to the same trust conditions.

Such unwarranted trust can go unnoticed as the system continues to perform successfully. As time goes by, users may become accustomed to getting relevant results and think that there is no need to verify them. But past performance may not generalize as well when the system is applied to a new topic, a different student population, or old data.

And low trust in artificial intelligence could be a problem, too. The use of AI should not be a reason to dismiss a reliable, audited system that can provide useful support. Education is not supposed to be about maximizing trust, but about aligning it with the real capacity of the system.

Trust is warranted only when the user is aware of the system's task domain, its performance history, data limitations, typical errors, independent evaluation and the degree of oversight by humans. Students and teachers should be able to see the evidence behind an AI-generated output and know when expert judgment is needed.

Kasneci et al. (2023) in their analysis of large language models in education mentioned that the benefits of artificial intelligence should be discussed along with the risks of misinformation and overtrust. Teachers need to teach students not only how to use a system but when not to use a system and when to question the output of a system.

The misplaced faith in artificial intelligence is less about trustworthiness than it is about image or ubiquity or institutional standing. Critical trust is an epistemically appropriate trust, not an absolute trust or a total rejection, but a trust calibrated to the task, the evidence, the level of risk and human supervision.

## Balancing Authority Between the Teacher, the Student and Artificial Intelligence

The advent of generative artificial intelligence in education has reignited the question of who contributes to knowledge and from where. In the traditional classroom the teacher was a central figure in selecting information and organizing instruction, while the student was generally viewed as the led party in the learning process. Artificial intelligence has become a third element in this relationship, as it can quickly generate answers, explanations tailored to the individual, and a large amount of content. In a systematic review, Qian (2025) highlights teacher–student–AI collaboration as an important area for future educational research.

But these three things are not to be regarded as one kind of person. The teacher is a person who defines pedagogical goals and assumes educational responsibility; the student is the subject of learning and the carrier of their epistemic development; artificial intelligence is a technological system that generates content through data and models. In Shneiderman’s (2020) vision of human-centered AI, the technology should augment human capabilities, rather than substitute for human judgment in the relationship.

Therefore, the balance of authority should not be understood as an equal distribution of the power of decision making between the teacher, the student and artificial intelligence. AI may be able to do some things faster or more accurately than a teacher or student, but it does not have the authority to set pedagogical goals, make ethical judgments or accept responsibility. Within the framework of human-AI learning partnerships, Wu et al. (2025) argue that goal setting and outcome evaluation must be kept as central components of human epistemic agency.

This three-part structure also does not give the teacher unlimited authority. Teachers cannot make every decision about artificial intelligence on the basis of their institutional authority alone; they must justify their choices in terms of student learning, safety, equity, and epistemic development. Tan et al. (2025) conducted a systematic review indicating that responsible AI use requires teachers to undergo continuing professional development and to prepare for ethical challenges.

The student's place in this equilibrium of authority must not be minimized to that of a passive user or a consumer who chooses between teacher and artificial intelligence. Students need to evaluate the credibility of sources, edit the output produced by AI, and take responsibility for how the information produced is used. In the AI literacy framework developed by Ng et al. (2021) evaluation and ethical use are included as core competencies, which further strengthens the final epistemic responsibility of the student.

The distribution of authority between teacher, student and artificial intelligence is based on functional and normative differentiation, not equality. The teacher should be pedagogical guidance and responsibility, the student should be critical reasoning and epistemic ownership, and artificial intelligence should be limited processing and support functions. To understand this structure we need to look at the changing role of the teacher.

**Teacher’s Changing Role**

Even before the advent of artificial intelligence, the teacher was not the only source of knowledge in the educational environment. In the processes of acquiring knowledge, books, libraries, experts and digital resources played an important role. What is unique about generative AI is that it can take existing sources and combine them to give students immediate and personalized answers, rather than just a list of available materials. According to Kasneci et al. (2023), these systems can provide instructional support and feedback, but also generate false and fabricated content.

Such a development may diminish the role of the teacher in the presentation of information, but it does not make the epistemic and pedagogical significance of the teacher obsolete. Yue et al. (2024) indicate that effective teacher use of artificial intelligence necessitates the integration of content and pedagogical knowledge and technological competence. Teachers need to assess whether the AI-generated output is accurate in the relevant field, appropriate to the developmental level of the student, and relevant to a particular learning goal.

The changing role of the teacher is a move from content delivery to learning design. The teacher decides when the students will think independently, when the students will receive feedback from artificial intelligence and when the students should refer to scholarly sources. According to Qian’s (2025) findings, the pedagogical potential of AI has been associated with such goals as critical thinking and learner autonomy, but also with the need to pay attention to the risk of overdependence.

Teachers are becoming professional assessors of the ethical and social implications of the use of AI. The meta-systematic review of Bond et al. (2024) emphasizes the need to improve the research in the area of artificial intelligence in higher education regarding ethics, collaboration, and methodological rigor. Issues of privacy, data security, unequal access, and algorithmic bias cannot be left to technical specialists alone by teachers.

The changing role of the teacher means creating conditions under which students can think better, not thinking for them. Salido et al. (2025) argue that the convergence of artificial intelligence and critical thinking should be considered alongside pedagogical design, self-regulation, and information-evaluation skills. The teacher’s expertise is expressed not in the provision of a ready-made answer, but in the design of the appropriate question, comparison and verification process.

The teacher’s role is shifting from a monopolist of knowledge into a learning designer who organises information sources, limits the use of AI according to pedagogical aims and supports the students’ independent judgement. This change does not diminish the authority of the teacher; it merely changes its foundation from ownership of information to epistemic guidance and responsibility.

**The teacher's epistemic guidance**

Epistemic guidance helps students to learn not only how to find correct information, but also the grounds on which a claim should be accepted. According to Long and Magerko (2020), AI literacy requires not only recognizing a system’s capacities and limitations but also critically interpreting its outputs. Teachers need to not only verbalize these competencies but also model them in the classroom processes of inquiry and evaluation.

The teacher’s epistemic scaffolding might start with questions about AI-generated output. Students may be asked what claims the response makes, which of those claims need sources, what assumptions have been made, and what opposing viewpoints have been left out. Critical generative AI literacy, as discussed by Rapanta et al. (2025), is associated with inquiry, contextualization, and epistemological reasoning.

Guidance also involves comparing sources of information. Teachers should make clear that an AI-generated response, a peer-reviewed article, a textbook, and a primary source do not have the same epistemic status. Billingsley et al. (2026) show that AI-assisted research designs can foster epistemic curiosity by rendering visible the diverse ways and notions of evidence through which disciplines address the same question.

Teachers must not simply be the ones correcting AI errors. They should explore, together with students, why an error was not detected in the first place, what features of the answer contributed to trust, and how similar errors can be detected. Tlili et al. (2023) found that the ability of AI to generate fluent but inaccurate content, presents pedagogical and ethical risks for students.

Epistemic guidance should also prompt students to assess their own learning. Students need to be able to recognize where they are using artificial intelligence to augment understanding and where they are using it merely to get a job done. The relationship between AI dependence and critical thinking emphasizes the significance of metacognitively monitoring patterns of use (Tian & Zhang, 2025).

Teacher guidance is not for making sure students are constantly in need of teacher approval. Through structured activities, learners are able to acquire competencies in problem solving and evaluating AI-generated output, as found by Kong et al. (2024) in their project-based AI literacy intervention. “It should be a system of educational support that makes itself gradually unnecessary.

With epistemic guidance, the teacher wants students to learn to evaluate sources, evidence, justification and uncertainty, not just to give the right answer. When guidance works, students learn independent criteria on which to question both teachers' and artificial intelligence's claims. This puts the critical position of the student in the center of the balance of authority.

## The Critical Attitude of the Student

The critical position of the student does not mean the refutation of all statements made by a teacher or by artificial intelligence. Criticality is the capacity to assess the clarity, source, evidence, scope, and strength of a claim in relation to alternative explanations. As Veldhuis et al. (2025) explain, critical AI literacy is not only about truthfulness, but also evaluates data, representation, social power, and ethical implications.

When students are given different answers from a teacher and an artificial intelligence system, they should not just pick the source that answers more fluently or quickly. They should consider which source is relevantly expert, whether reasons have been given, and whether the claim is backed up by reliable evidence. The evaluation dimension of Ng et al.'s (2021) framework shows that there is no way to detach the use of AI from epistemic judgment.

The student is in a critical position, which means they have to evaluate their own thinking as well. Students should not only seek errors in external sources, but also consider how their own assumptions, the way they phrase the prompt and the answer they expect will influence the outcome. Wu et al. (2025) point out that epistemic agency is enhanced when students participate in goal setting and result reevaluation.

It's not that holding a critical position is the same as rejecting all use of artificial intelligence. The system can be used by students to generate alternative explanations, counter arguments, or new research questions. However, Vendrell and Johnston (2026) argue that such use should be structured via tasks that preserve student cognitive activity, allowing students to formulate an initial evaluation before receiving a completed response from the system.

The students too have a responsibility to be open about the role of artificial intelligence. Cotton et al. (2024) suggest that academic integrity is more than simply identifying AI-generated text. Students should be able to indicate which tool they used, what was the purpose, which system recommendations they accepted and how they contributed to the final product.

Luo (2024) argues that student originality is not a matter of producing work without technology, but of maintaining control over the production process and being able to defend the product. This turns the student from a consumer of AI-generated output into an epistemic subject responsible for choices and judgments.

The student's critical position includes comparisons of teacher, scholarly sources, and artificial intelligence, an exploration of personal assumptions, and an acceptance of responsibility for the final judgment. Only at the end does the balance of authority result in dependence on either the teacher or the artificial intelligence. And to achieve this goal, it is necessary to sharply limit the role of artificial intelligence in education.

**Limiting Artificial Intelligence to Use as an Assistive Tool**

Nor does limiting artificial intelligence to an assistant role mean removing technology from any educational task. AI might be useful in providing explanation, examples, language support, feedback on drafts and generation of alternative solutions. As shown by Farrokhnia et al. (2024), the educational opportunities associated with ChatGPT should be balanced against the risks of misinformation, superficial learning, and dependence.

The first criterion for limitation is the educational purpose of the task. If the skill students are expected to develop is directly delegated to artificial intelligence, tool use may eliminate the learning objective. For example, if the goal is to develop argumentation skills, it is pedagogically inappropriate to have the system generate the entire argument. Similarly, when the goal is to develop source evaluation skills, source selection should not be completely outsourced to AI.

The second criterion is about when AI will be used. Vendrell and Johnston (2026) suggest that students being able to view a full response before forming their own initial ideas might decrease cognitive effort. To keep human thinking, it may be better to ask students to develop a solution and then use artificial intelligence to collect different perspectives or feedback.

The third criterion is the importance of system output in a decision. Artificial intelligence may provide a recommendation, warning, or probability, but it should not issue the final judgment regarding student success, academic integrity, or personal development. Shneiderman's (2020) human-centered approach advocates for human supervision and accountability to be maintained for high-impact decisions.

The fourth criterion is that the interaction must be open to discussion and rejection. As Pietarinen and Shi (2026) warn, "pedagogical support may become manipulative when artificial intelligence bypasses students' rational evaluation and invisibly narrows their options." Students need to know why a recommendation is made by the system and be able to reject it.

The fifth criterion is that the use of AI should not take away teacher responsibility. All lesson content, questions, and feedback generated by the system must be checked and approved by the teacher before use. Tan et al.'s (2025) review shows that teacher professional development should include ethical and pedagogical evaluation competencies, not just technical skills.

To limit artificial intelligence to an assistive role is to specify conditions of use according to task and level of risk rather than to develop fixed and unchanging lists of prohibitions. The educational effects of ChatGPT differ according to the context and form of use (Lo, 2023). The demands of a high-stakes assessment that measures student competence should not be applied to a low-risk language editing task.

The limit of artificial intelligence as an assistive tool has to be determined by the objective of the task, the timing of use, the weight of output in decision making, rejectability, verification and human supervision. To the extent that technology supports students' thinking, it has pedagogical value. When technology usurps the processes of thought and responsibility, it upsets the balance of authority. The primary reason for this limitation is to give priority to human judgement.

**The Primacy of Human Judgment**

It does not claim that human beings are more accurate than artificial intelligence in all circumstances. The primacy of human judgment is based on the fact that Artificial intelligence can out-calculate, out-detect or out-classify humans in some tasks. The priority claim is the necessary role of the human subject in determining educational purposes, in contextually evaluating evidence, and in taking responsibility for decisions.

Human judgment is about more than just the production of outcomes; it involves judging why an outcome is important, who will be affected, and what values are at stake. Instead, as Shneiderman (2020) describes, "human-centered artificial intelligence" should aim to develop "high levels of automation and high levels of human control" at the same time, rather than as competing alternatives. It should also support the human capacity for evaluation in education and not make it invisible.

Teacher judgment is related to the student's individual situation, the context of the classroom and educational aims. While AI may be able to point to recommendations based on general patterns, it cannot fully understand why a student gave a particular answer, what life circumstances might have led to a failure, or how a particular intervention might impact a student's self-confidence. Kasneci et al. (2023) demonstrate that AI support cannot be understood apart from the pedagogical context.

The student judgment is a basic condition of learning and of epistemic autonomy. If students let the technology do all the thinking and decision-making for them, they may get the right answers but they will not develop their own minds. Tian and Zhang's (2025) study finds a negative relationship between AI dependence and critical thinking and thus supports the importance of regular use of human judgment.

The primacy of human judgment does not imply that we have to accept any human opinion that is contrary to the output generated by the AI. If independent evidence supports the conclusion generated by artificial intelligence, teachers and students may revise their initial judgments. The key principle is that decisions should be arrived at by critical evaluation of evidence not submission to system authority.

Vendrell and Johnston (2026) have taken a similar approach, arguing that educational designs need to intentionally provide space for thought in an effort to preserve human judgment. Students should be asked to provide rationale, seek counterevidence, admit uncertainty, and explain why they have rejected or revised an AI-generated recommendation. Human judgment can be developed only when it is put into real practice.

Pietarinen and Shi (2026) claim that AI designs which safeguard human autonomy should include principles like contestability, reasons, and calibrated trust based on evidence. These principles make sure that artificial intelligence is a tool to serve human critical purposes, not an authority to rule human beings.

The importance of human judgment is not an indictment of technology. it maintains the last position assumed by the human subject in questions of purpose, meaning, value and responsibility. Teachers and students should always be able to demand rationales, to challenge and revise decisions based on AI-generated output. The balance of power can only be maintained if the ultimate epistemic and pedagogical responsibility lies with human actors.

In Chapter Two I examined epistemic authority in education through the social and normative dimensions of the individual's relation to sources of knowledge. Epistemic authority is based on justified acknowledgment of a source's superior competence in a given domain, and does not imply blind obedience or intellectual submission. Shneiderman (2020) supports the human-centered approach, stating that the authority of artificial intelligence systems should be limited so as not to deprive humans of the opportunity to evaluate and monitor.

The first half of the text has demonstrated that epistemic authority is specific to a domain, is open to justification and is dependent on trustworthiness and accountability. Artificial intelligence may perform well in certain tasks, but such performance does not equate to pedagogical responsibility or human understanding. Kasneci et al. (2023) have pointed out risks of misinformation and overtrust, showing that technological achievement cannot be directly translated into general epistemic authority.

The second part was about the teacher's epistemic authority. It argued that the authority of teachers is not a monopoly of information but the obligation to select, contextualize, justify and organize knowledge according to student development. The results of Yue et al. (2024) highlight that teacher competence in the era of artificial intelligence hinges on the combination of technological, pedagogical, and content knowledge.

The third part explored the question of whether artificial intelligence is a new epistemic authority. Users may have high confidence in authority about fluent responses, wide scope and fast synthesis. However, the AI literacy approach based on evaluation, developed by Ng et al. (2021) shows that system outputs have to be independently scrutinized for truthfulness, contextual appropriateness, and ethical implications.

The fourth part was about the problem of algorithmic authority, taking into account technological neutrality, opacity, explainability and unjustified trust. Algorithms do not only produce technical results but also influence what kinds of information is perceived as visible, normal and trustworthy according to the critical AI literacy framework of Veldhuis et al. (2025).

The last part built a model for balancing the authority between teacher, student, and artificial intelligence. The model does not give the three equal say in decision-making, but rather differentiates their roles. The teacher is responsible for pedagogical aims and epistemic guidance; the student is responsible for critical evaluation and final epistemic ownership; artificial intelligence is responsible for processing, content-generation, and feedback support.

In this model, the teacher's role is not to monopolize knowledge over artificial intelligence. The systematic review by Tan et al. (2025) shows teachers require continuous professional development in AI use. Teachers should not be authorities who reject technology but explain its limitations, design appropriate uses and strengthen students' critical positions.

Students are not passive entities choosing between teacher and artificial intelligence, either. According to Wu et al. (2025), the epistemic-agency approach includes goal setting, evaluation of information sources, reconstruction of system outputs, and responsibility for final decisions. The critical position of the student is thus crucial for the preservation of the balance of authority.

Artificial intelligence's rightful role in education is that of a limited and assistive epistemic tool. To develop critical thinking in AI use, pedagogical designs must preserve students' cognitive activity (Vendrell & Johnston, 2026). If artificial intelligence becomes the ultimate authority deciding on behalf of teachers or students, the educational relationship produces epistemic dependence.

The broad conclusion of chapter two is that the total abolition of epistemic authority from education is neither feasible nor desirable. Students will still turn to teachers, experts, academic institutions and technology systems. The educational goal is not to reject the authority of all authorities, but to assess authorities by domain, justification, trustworthiness, and responsibility.

The sound balance between teacher, student and artificial intelligence must be based on the priority of human judgment. Technology can help education with its capacities to process and generate information but the authority to decide on educational aims, values and human consequences cannot be delegated to technology. Artificial intelligence ought not to be excluded from human-centered education that safeguards student epistemic autonomy and teacher pedagogical responsibility.

These conclusions show that AI-supported education calls for an investigation not only of individual autonomy and relations of authority, but also of the just production and distribution of knowledge. The next chapter focuses on how artificial intelligence systems represent different groups, the voices they make visible or invisible, and the new inequalities they might generate in access to knowledge. The following chapter will therefore explore the connection between artificial intelligence, education and epistemic justice.

# THE IDEAL OF HUMAN DEVELOPMENT IN THE AGE OF ARTIFICIAL INTELLIGENCE

## The Purpose of Education and the Problem of Human Development

To ask what the purpose of education is, is to ask a larger question than to ask what particular items of knowledge should be handed on to pupils. It is a normative choice about what kind of person and what kind of society is considered valuable. Education is not the preparation for a later stage of life; it is the reconstruction of the experiences which make life what it is. (Dewey, 1916) Thus, educational systems not only transmit existing knowledge and skills, but also shape the way individuals see the world, relate to others and participate in shared problems.

The issue of human development in the era of artificial intelligence is often framed in terms of how quickly education systems adapt to technological change. Schiff (2022) found that national AI policy analysis often frames education as preparing an AI-ready workforce and producing technical experts. This approach may respond to economic needs, but when education is reduced to producing the competencies demanded by the market, the moral, democratic and critical development of the individual becomes secondary.

The ideal of human development must not be understood as the creation of a catalogue of capacities that would enable human beings to compete with technology. Human education does not try to compete with AI in speed of calculation, in processing of huge amounts of data, or in pattern recognition. Shneiderman's (2020) human-centered AI approach suggests that technology should not attempt to replace human capabilities, but rather to empower human creativity, responsibility, and decision-making power. Thus, education should create subjects that are capable of generating purposes, meanings and values, instead of flawed iterations of artificial intelligence.

The educational systems not only train individuals for personal success. They also help to perpetuate the social order. But continuity should not mean the unproblematic reproduction of existing institutions and values. Freire (2000) contends that education that turns students into passive recipients who gather ready-made knowledge constrains their power to read the world critically and to change it. Thus the problem of human development demands a balance between cultural transmission and critical transformation.

Education that is supported by artificial intelligence may either strengthen or weaken this balance. Jaramillo and Chiappe (2024) argue that curricula enhanced by AI should be designed to engage critical thinking, interdisciplinary problem solving, and engagement with real-world issues. If used for efficiency, speed, and standardized achievement, such systems may produce compliant users; if used to explore alternative perspectives and assess complex problems, they may produce more critical subjects.

The essence of the ideal of human development should be not only the question "What should a person know?But also, "To what ends should a man employ knowledge?" Ma et al. (2025) review indicates that responsible AI education should combine cognitive knowledge with affective and behavioral development. Students need to be able to identify bias in AI, but they also need to care about the consequences of unjust output and act responsibly when they see it.

## Philosophical Perspectives on the Purpose of Education

Philosophical approaches to education purpose are based on diverse assumptions about human nature and good life. According to Kant education is the means by which human beings are able to discipline their natural inclinations, to develop their abilities, and to use their own reason. Kant (2007) argues that education should not be aimed solely at adaptation to social rules, but also at the empowerment of the individual to become a free and moral subject. Especially so in the era of artificial intelligence, when it demands that students do not surrender their decisions to technological suggestions.

In Dewey's (1916) pragmatist approach, the aim of education is not to create a pre-formulated and unchangeable model of man. Education enhances the quality of an individual's experience and makes the individual more receptive to new experiences. In this sense, the school is not just a place where students prepare for future social roles, but a community in which democratic life is lived collectively. The educational aims, therefore, should be reexamined in light of the changing problems of students and society.

Freire's (2000) critical pedagogy understands education as not a neutral transmission of social reality. Education can reproduce dominant values and relations, but it can also enable people to render those relations visible and to change them. Teaching students how to use artificial intelligence well alone is not emancipatory education, if they are not also developing the capacity to question the social effects of technology companies, data regimes and algorithmic classifications.

Contemporary educational policies have been increasingly infused with instrumental and economic approaches. “AI strategies tend to frame education in terms of competitiveness, workforce readiness, and creation of technical expertise” (Schiff, 2022). Vocational competence is a valid educational aim, but if it is the only aim the human being is reduced to an economic function. Education should prepare students not only for changing occupations but to debate the purposes of the changing technological order.

The philosophical foundations of AI literacy could provide a bridge between these various aims. Heung and Chiu (2025) suggest that the technical knowledge interest involves understanding how systems operate, the practical interest involves evaluating the interpersonal and cultural meanings, and the emancipatory interest involves questioning the structures of power, bias, and injustice. The education of purely technical interest may produce competent users but the education that combines all the three dimensions may produce critical and responsible subjects.

While one philosophical approach cannot determine the purpose of education, one can arrive at a normative balance among these perspectives. Kant emphasized autonomy, Dewey democratic experience, Freire critical transformation, and current studies in AI literacy technological and ethical competence. The ideal of human development should not absolutize one of these aims to the exclusion of the others.

## The Ideal of the Formation of the Good Man and the Good Citizen

The ideal of developing a good person is not about making a person who simply conforms to social rules or achieves cognitive success. Moral character encompasses the capacity to care about the welfare of others, to weigh the consequences of one’s actions, and to connect personal interests with the public good. Lam (2026) argues that learning with LLMs must be conceptualized as moral self-cultivation, relational responsibility, and harmoniousness with the community. It resists the reduction of AI ethics to a simple list of rules to be followed.

The ideal of creating a good citizen is also broader than creating a person who submits to the existing political order. Democracy is not a form of government but a way of life where common experiences are shared by means of communication (Dewey, 1916). A good citizen is not simply someone who knows the legal rights and obligations, but someone who gets involved with common problems, who listens to other viewpoints and who participates in making public decisions.

Nussbaum (2010) argues that democratic citizenship requires critical reasoning, the ability to imagine the perspective of others, and awareness of the fragility of human life. These capacities are even more important in the modern era of artificial intelligence, as algorithmic classifications can reduce people to data points, render diverse life experiences invisible, and create personalized information environments. A good citizen should know the limits of the digital reality presented to them.

Digital citizenship is more than just online safety and rules of appropriate behavior. Örtegren (2024) demonstrates that models of digital citizenship are underpinned by different philosophical assumptions about human–technology relations and social power. Viewing citizens merely as users of technology may obscure the political and economic character of technological arrangements. Critical digital citizenship, however, frames the individual as a co-determiner of technological change.

As Karatsiori et al. (2026) point out, global citizenship education can reproduce market-oriented and Western-centered conceptions of universality. The wish to develop a good citizen has to go beyond a “global competence” paradigm that superficially engages with cultural differences. Global technological systems centralize certain types of knowledge, exclude certain experiences, and reproduce inequalities: people should be able to ask what these are.

Responsible AI literacy can close the gap between the good person and the good citizen ideals. Ma et al. (2025) argued that ethical AI education should cultivate cognitive, affective and behavioral dimensions jointly. Students need to learn about concepts such as justice, privacy, and accountability; become sensitive to how algorithmic harms affect others; and show responsible patterns of use.

Chan’s (2023) ecological AI policy framework encourages students to engage in determining the rules of artificial intelligence in educational institutions. This involvement shows that citizenship education is not only about theoretical knowledge. If students can help set the rules for the technologies they use, they can practice decision making, justification, deliberation and shared responsibility.

**The Autonomous and Critical Person**

Autonomy does not entail that the individual is free from external influences or that he or she appeals to no authority at all. In the Kantian sense, autonomy is the capacity to rationally evaluate one’s principles and decisions. According to Kant (2007) education should not only discipline the human being, but enable the individual to use his/her own reason. In the modern era of artificial intelligence, this goal requires that students see algorithmic recommendations not as decisions themselves but as reasons to be judged.

To be critical does not mean to be always against others or to be suspicious of every source of knowledge. Critical thinking is the ability to evaluate the foundation for claims, the concepts being utilized, the alternatives ignored, and the consequences of decisions. AI-supported curricula should involve students in real-world problems and interdisciplinary thinking, according to Jaramillo and Chiappe (2024). Criticality is thus not only an abstract skill, but a practice of justified decision-making on complex issues.

Artificial intelligence could support student autonomy by increasing access to information. Students find different explanations, opposing points of view and personalized feedback. Long and Magerko's (2020) description of AI literacy includes understanding not only what a system can do but also its limitations and impacts on human decision-making. Availability of more choices enhances autonomy, but so does the ability to assess those choices.

According to Rapanta et al. (2025), it is not sufficient to define generative AI literacy only in terms of functional skills. Students should interrogate outputs, contextualize them, transform them to their own uses, and assess them epistemologically. If students just accept an AI-generated answer, they may get a quick product, but they may lose some of their intellectual ownership and responsibility for decision making.

Autonomy does not require that teachers or institutions remove all guidance. Students may need help understanding how systems work, evaluating sources, and making ethical decisions. Ng et al. (2021) conceptualize AI literacy as knowing, using, evaluating and ethics dimensions. The teacher is not the decider for the student, but helps students gradually get the ability to make independent judgments in all four dimensions.

A critical person must question not only the output of an artificial intelligence, but what they are asking an artificial intelligence to do. The questions that students ask may reflect their own biases, assumptions about success, or worldview. According to Freire (2000), critical consciousness involves people looking not only at the external world but also at their own thought patterns and the social conditions that shape those patterns.

Autonomy may be divorced from social responsibility when translated into personal achievement and freedom of choice. Shneiderman's (2020) human-centered approach considers human control not only as a matter of individual user preference but in terms of safety, accountability, and social benefit. Therefore, an autonomous individual is not only someone who can use technology well for himself/herself, but someone who can assess the impact of that use on other people and on society.

**The Struggle Between Conforming to Society and Challenging Society**

Education should convey a common language, knowledge, values and forms of behaviour so that individuals can take part in social life. Those who cannot adjust to society in some way may have difficulty working with others and participating in democracy. Education, as Dewey (1916) put it, incorporates individuals into processes of common experience and communication, and in so doing, it preserves social life. Thus, adaptation is not an educational function that must be denied out of hand.

But when adaptation becomes uncritical acceptance of the norms of the existing order, education may become an instrument of ideological reproduction. Freire (2000) argues that education which simply places students in the world as it is limits their capacity to transform it. Education should equip students with the ability to understand social institutions and at the same time to criticize them on the grounds of justice, freedom and human dignity.

In the era of artificial intelligence, discourses on digital skills and employability increase the demand for adaptation. Schiff (2022) points out that national policies frequently mould education to meet the demands of the AI economy. While the capacity to use new tools is important for students, creating individuals who just adapt to technological transformation can produce a model of citizenship that does not question the purposes that are served by technology.

To question society is not to reject all traditions and institutions. Nussbaum (2010) argues that democratic education must allow individuals to criticize their own tradition from an outside perspective and to appreciate other people's points of view. A critical citizen is not someone who stands apart from social ties, but someone who is able to reassess those ties in the light of standards of justice and the common good.

The limits of social adaptation are less obvious in the digital and algorithmic order. Örtegren (2024) contends that technological systems are intertwined with social power relations and that digital citizenship education should address sociotechnical structures. Students may not see the political and economic choices that underlie a recommendation when it is presented by an algorithm or a rule is presented as a technical necessity.

As Karatsiori et al. (2026) point out, technocratic and market-oriented forms of global citizenship may reproduce rather than question inequalities. If people are going to question society, they need to look not only at how standard artificial intelligence systems are deployed, but also at who is building them, what kinds of knowledge they prefer, and who profits or suffers from their results.

Educational design can turn this tension into a productive one. Students might be asked to consider AI policies, automated assessment systems or examples of algorithmic decision making from the perspective of different social groups. Heung and Chiu

(2025) emancipatory knowledge interest requires students to see technology as not only a technical object but also as a social structure that can be transformed.

Their own practices should also be open to examination by educational institutions. As Chan (2023) argues, student participation in the development of AI policy can provide critical citizenship experience within the school itself. Students should have a voice in decisions about the use of data, academic integrity and automated assessment within their own institutions, in addition to critiquing technological power in the wider society.

## The Image of the Human Being in an Age of Artificial Intelligence: A Transformation

Ideas about human nature have been influenced by social institutions, economic systems, and technological developments from ancient times. In the history of education, the human being has generally been conceived as a rational and developable subject, able to participate in social life. In the modern era of artificial intelligence, however, the individual has increasingly come to be seen not only as a thinking and acting person, but also as an element within a system whose data can be collected, whose behavior can be predicted, and whose performance can be optimized. Selwyn and Gašević (2020) in their analysis of datafication of higher education demonstrate that educational reality is more and more understood through measurable indicators.

This transformation does not imply that technology eliminates the human being directly. Instead it changes which human qualities are seen as visible and worthy. Because artificial intelligence systems are built on measurable behaviors, qualities that are difficult to measure, such as attention, curiosity, empathy, moral hesitation or personal meaning, may fall into the background. Selwyn, Pangrazio and Cumbo (2022) suggest that school data not only describe students but also constitute what student characteristics are considered educationally relevant.

Another element of the image of the human being in the era of artificial intelligence is the ongoing subjecting of the individual to future predictions. Based on a student’s past behavior, one can predict his/her future academic success, risk of dropping out, or aptitude in a particular field. Jun and Lee (2026) suggest that such a predictive order might close down possibilities that students have not yet developed. Humans are not a necessary continuation of their past data, education should keep open the possibility of the individual to go beyond an existing profile.

The change in concept of human being has also an effect on the relation of the individual with himself. Joseph (2025) says that the algorithmic self may include algorithmic recommendations, classifications and predictions within the self. Students may learn to internalize digital feedback that labels them “successful,” “at risk,” “uncreative,” or “suited to a particular field” as objective truths about their own self-narratives.

At the same time, it would be insufficient to argue that artificial intelligence changes the idea of the human being only negatively. Technology can help people to learn their learning styles, to find different explanations and to get help that suits their needs. However, in Shneiderman’s (2020) human-centered approach, the crux is whether the system supports human capabilities and keeps the ultimate control in human’s hands. Technological personalization can only be educational if it does not imprison the individual in a fixed data profile.

The central problem in conceiving the human being in the era of artificial intelligence is that individuals may increasingly be valued to the extent that they adapt to technological systems. The ideal student may be presented as the one who learns fast, is constantly productive, is digitally visible and can be easily turned into data. Promsiri (2025) suggests that the effect of artificial intelligence on the learner’s identity and psychological agency cannot be measured solely by the achievement indicators. Education should not be about the adaptation of students to systems, but about their capacity for establishing a critical and purposive relation with systems.

**The Digital Human**

The digital individual is not simply someone who uses computers or artificial intelligence. The digital traces, online relationships, search histories, creative productions and algorithmic interactions are all part of the construction of identity. Joseph (2025) argues that algorithmic systems can affect not only what people see, but also their preferences and the stories they tell about themselves. The digital individual is, then, not someone who has an external, instrumental relation to technology, but someone whose identity is constantly reconstructed through technological mediation.

In educational environments, the digital individual is made visible through behavior on learning platforms. The student’s digital educational profile includes the amount of time spent reviewing specific content, the number of errors made, the times the student logs on and the assignments completed. However, as Selwyn and Gašević (2020) observe, while such data can be useful to teachers, the emphasis on quantifiable activities may exclude less visible aspects of learning, such as silent reflection, emotional struggle, or learning experiences outside the classroom.

Another feature of the digital individual is the tendency to evaluate oneself through the feedback of algorithms. Students can interpret their own capacities according to a learning pathway recommended by the system or a future achievement prediction . This may be a support for individualized instruction, but may also lead students to subject their own judgment to that of the system's evaluation. Rapanta et al. (2025) state that critical AI literacy is about challenging outputs and the impacts of those outputs on identity and decision-making.

The digital individual may become a subject always accessible and traceable. The permanent recording of learning activities blurs the line between the student's learning space and private life. According to Alfiras et al. (2026), data protection and student well-being are major ethical issues, meaning that there are limits to data collection for the purposes of personalization.

There are some upsides to digital identity as well. Students can belong to many different communities, express themselves in a variety of ways and access learning opportunities not available in their physical environment. Ng et al. (2021) approach to AI literacy highlights the need for digital participation to go hand-in-hand with conscious usage, critical assessment and ethical responsibility. The digital individual must be more than a user of platforms, but a participant, able to question the rules regulating digital environments.

The autonomy of the digital individual is based on the possibility to reject the algorithmic recommendations. According to Pietarinen and Shi (2026), pedagogical influence can turn into manipulation when systems influence users' choices without rational evaluation. Students should understand the reason behind the content presented by an algorithm and be able to select an alternative learning path.

**The Human Being as Data**

Datafication is the process of turning human activity into quantifiable, comparable, and processable data. Education transforms test scores, attendance, interactions on platform, signs of attention, and learning behaviors into digital data. Selwyn, Pangrazio and Cumbo (2022) demonstrate that such data are not neutral mirrors that just reflect the student, but rather selective representations that determine how the educational system recognizes the student.

The reduction of the human being to data is based on the assumption that all personal characteristics can be explained through data. However, two students with the same test score may differ greatly in their motivation, learning conditions and life experiences. A data point might suggest a certain behavior but it doesn't, by itself, explain the meaning and context of that behavior. Selwyn and Gašević (2020) argue that the promises of data-driven decision making become reductionist when the social and pedagogical processes that produce data are overlooked.

Datafication becomes an even more powerful tool for managers when it allows for predictions about the student's future. Classifications can limit the choices available to the student, although risk scores and achievement forecasts may be used to offer early support. Jun and Lee (2026) contend that algorithms should have their predictive power limited in a way that is compatible with the individual's "open future." A prediction should not be presented as the student's fate or identity.

Reduction to data may also make it easier to give more educational importance to characteristics that are easy to measure. Performance, time spent on a platform, assignment completion rates are easily monitored, but moral development, curiosity, solidarity, or creative risk taking are harder to quantify. This may cause educational institutions to value as important what they can measure and as unimportant what they cannot measure.

The justice dimension of datafication matters just as much. Algorithms can lock in historical social inequalities if the data they are fed reflect those inequalities. According to Alfiras et al. (2026), algorithmic fairness and bias auditing are essential for responsible educational AI. A data profile constructed around a student should not be read as evidence of individual failure outside of social and economic conditions.

Humans are not completely free from data. Teachers need to use test results, observation and attendance records to understand how students are growing. The problem is not with the use of data, but with data defining the whole person. A human-centered approach would use data not as the final judgment but as a limited indicator to be supplemented by teacher evaluation and the student's own account.

**Culture of Velocity, Efficiency and Performance**

Speed is one of the most popular arguments for artificial intelligence in education. The ability of systems to generate content, analyze student work, and make personalized recommendations in a short period of time is seen as a big efficiency benefit. According to Madanchian and Taherdoost (2025), technical performance is an important factor in evaluating such tools, but fairness, transparency and pedagogical impact should be considered in an integrated manner.

Speeding up education processes does not mean improving learning. It may take time to understand, to connect concepts, to make mistakes, and to reorganize one’s thinking. Artificial intelligence can speed up a task, but it may shorten the intellectual processes by which learning occurs when it takes students directly to an answer. Promsiri (2025) advocates for a careful assessment of the effects of outsourcing cognitive processes to artificial intelligence on student motivation, identity, and metacognitive development.

A performance culture is constantly assessing the value of students by their results. The signs of success are higher grades, quicker completion and more productivity. As a consequence of this culture being reinforced by artificial intelligence, students may start to view learning as a process of performance production rather than a process of meaning construction. The datafied student model may also normalize the continuous comparison and ranking.

Speed and efficiency also change the work cultures of teachers. Artificial intelligence can speed up the production of lesson plans, questions and feedback, but then teachers may be expected to produce more in less time. Tan et al.'s (2025) systematic review demonstrates that teachers' adaptation to artificial intelligence should not be confined to tool use, but should be considered in conjunction with professional development and pedagogical needs.

In a performance culture, error and slowness can be experienced as a sign of inadequacy. But mistakes are a vital part of learning and some students need more time to develop their ideas. The speed of response that is the norm for artificial intelligence systems may make the natural pace of human thought seem defective. "Human-centered education should realize that not all processes should be accelerated, nor desirable to be so.

Efficiency advantages of artificial intelligence need not be forsaken. Automation of routine tasks may allow teachers to spend more time with students and to observe their thinking and provide complex feedback. Hao et al. (2025) assess human–AI collaboration and note that artificial intelligence can deliver speed and scale, while human beings can offer ethical oversight, contextual interpretation, and strategic sense making.

## The Human Model: Adaptation of Technology-Based

The human model of technological adaptation defines the ideal person as one who learns new systems rapidly, is continuously updated and adapts to the changing demand of the workforce. This approach to education is less about helping students interrogate the purposes and values associated with technology, and more about ensuring that students participate in technological change without resistance. Shneiderman (2020) flips this by arguing for technology to be adapted to human purposes, in his human-centered AI approach.

There is a need for a certain amount of technological adaptation. If you live in a digital society and you don't understand standard artificial intelligence systems, you may be disadvantaged in accessing information and participating in society. Ng et al. (2021) argue that AI literacy should entail not only the identification and use of systems but also the assessment of systems and ethical responsibility. Adaptation without critical reflection produces passive users.

The technological adaptation model can depict human beings in a state of perpetual deficit. They require constant updating. "Some people are not quick enough to adapt to new tools and that might be seen as backward or inefficient. Such a view risks interpreting differences in age, economic status, disability and digital access as personal failings. People cannot be left to pay the full price of technical change; institutions have to be held responsible for access and fairness.

Personalization (with AI help) might also enhance adaptation. A system can keep track of what students do all the time and suggest "the most appropriate" pathway to them. However, Jun and Lee's (2026) approach shows that a pathway considered algorithmically appropriate might limit students' opportunities for future development in other directions. The system should not force students to become a certain type of person.

Another danger of technological adaptation is that people might compare their own worth with machines. Artificial intelligence might make people feel inferior because they can't write, calculate or get information as fast. Joseph (2025)'s notion of the algorithmic self shows how technological standards can influence how individuals view themselves. The aim of human education is not to build a machine that runs slower.

The technological adaptation is critical and has to be discussed in terms of the conditions of use of the technology and not in terms of the rejection of the systems. Critical generative AI literacy includes the ability to contextualise technology and adapt it for human ends (Rapanta et al., 2025). As people evolve with technology, they should be able to adapt technology to their human and educational purposes.

In a-like vein the technological adaptation of teachers should be approached. Tan et al.'s (2025) review suggests that it is not enough to train teachers to use the new tools, but also their pedagogical decision making, ethical evaluation and professional development need support. Teachers should not be the implementers of systems but professional actors who define the purpose and limits of their use.

## Who Needs Education in an Age of Artificial Intelligence?

The question of human development in the era of artificial intelligence is not simply a question of which electronic devices students should learn to use. The key question is what human capabilities individuals must develop in a world where technological systems are changing access to knowledge, decision-making and social interaction . As suggested in Shneiderman's (2020) human-centered AI approach, technology should supplement human control and responsibility rather than replace them. Then education must be about training students principally as subjects who can evaluate the aims and consequences of technology and not just as users who can adapt to systems.

In this model of the human being, technological competence and human development should not be alternatives. Students must be able to use artificial intelligence, but, as in the framework proposed by Ng et al. (2021), education only produces functional users when the ability to use technology is not complemented by evaluation and ethical responsibility. The person to be educated must be able to use technology and at the same time maintain personal judgment, responsibility for knowledge, and respect for the rights of others.

The specific value of the human in the era of artificial intelligence should not be reduced to a list of functions that machines can not fulfill. As systems evolve they may become able to perform cognitive tasks previously considered uniquely human. However, the problem of thinking as highlighted by Costa and Murphy (2025) requires that the human being be conceived not merely as a producer of outputs, but as a subject who interrogates the meaning and consequences of actions. Education should not be about producing people who are faster than machines, but about producing people who can explain why they think, and why they act.

## The Critical Thinker

Critical thinking is not an automatic rejection of all claims, or a reflexive distrust of all sources. This is the ability to judge the clarity of the concept, the evidence, the assumptions, the scope and the relative strength of a claim against alternative explanations. According to Essien et al. (2024), artificial intelligence can be beneficial in the generation of ideas and analytical processes, but critical thinking can be superficial if students outsource the evaluation to the system.

A critical thinker does not confuse the linguistic fluency of an AI generated output for truth. The clarity and organization of a system's response do not correlate with the reliability of evidence supporting its claims. Long and Magerko's (2020) approach to AI literacy requires users to understand not only capabilities but also limitations of systems. Students are trained to verify an answer by checking the sources, date and other viewpoints of an answer before they accept it.

Critical thinking should also be used for the students' own questions and preferences. Artificial intelligence might follow the assumptions embedded in a user's prompt and come up with a one-sided response. Veldhuis et al. (2025) describe critical AI literacy as questioning not only results but also data, representation and relations of power. Students should ask, "Did the system give the right answer?" and, "Whose perspectives does this answer leave out?"

"You can't build critical thinkers without some cognitive struggle in education. If students are provided the whole solution to a problem right away, they may miss opportunities to think, make mistakes, and revise their solutions. Costa and Murphy's (2025) critique of thoughtlessness demonstrates that intellectual effort should not be sacrificed in the name of speed and efficiency.

Instructional designs may ask students to develop an initial solution before consulting the artificial intelligence, and then compare their reasoning with the system's response. Students might also be asked to spot assumptions made in AI-generated texts, produce counterevidence, or compare the strength of evidence in two different responses. Research on generative AI literacy shows that such explicit and structured activities should support critical evaluation.

## The Subject of Epistemic Responsibility

Epistemic responsibility is the duty to furnish adequate justification for the knowledge claims that we accept, use and convey to others. “The system said so” is not an excuse for a student to deny responsibility for the false information produced by artificial intelligence. Wu et al. (2025) argue that in human-AI learning partnerships, the human actor should be responsible for the goals, the evaluation of outcomes, and the final epistemic decisions.

An epistemically responsible person makes the distinction between having access to information and having knowledge. If the student cannot state why the answer is correct or what evidence makes it so, then the student may not have made an epistemic achievement in the sense of giving a correct answer from artificial intelligence. In Ng et al.’s (2021) framework, the evaluation dimension emphasizes that AI use is inseparable from source verification and justification.

To be able to admit what one does not know is part of epistemic responsibility. Artificial intelligence can provide full answers even when there is uncertainty, but the responsible person must be able to suspend judgment when evidence is lacking. That attitude is not a weakness; it is a form of epistemic honesty. Schwarting and Reihlen (2026) show in their analysis that the transfer of responsibility to technology may turn into a voluntary abdication of one’s own ability for judgment.

Transparency in disclosing AI contributions is also a requirement of epistemic responsibility. Students should be able to explain what tool they used, for what purpose, what suggestions they accepted or rejected, and how they verified the output. This transparency is not just a matter of academic integrity; it is an educational practice that makes visible the student's ownership of the knowledge-creation process.

The dependence on artificial intelligence might hinder the development of epistemic responsibility. The system that handles question generation, source selection, text production, and result evaluation all the time can make students loose connection with their own decision making processes. Tian and Zhang (2025) showed that there was a negative relation between AI dependence and critical thinking and the protective role of AI literacy.

### The person who reasons morally

Moral reasoning is not just the repetition of standard rules of ethics. It is the capacity to consider together the impact of a decision on different people, competing values, rights, responsibilities and contextual conditions. The review of Gouseti et al. (2025) shows that AI ethics in education should be dealt with the dimensions autonomy, justice and well-being. These values do not always lead to the same conclusion, and students may need to balance tensions among them.

For example, a system that tracks the performance of students may be useful in providing early support, but it may also pose risks of privacy, labelling and discrimination. This is not only someone who can say whether a system is beneficial or harmful. Such a person judges on whose behalf the benefit appears, who bears the risk and if there are less harmful alternatives.

Education in AI ethics should require students to face real decision-making situations. Studies designed to promote ethical awareness through problem solving have shown how important it is to give students chances to evaluate bias, privacy, justice, and responsibility in realistic situations. Therefore ethical education should be a practice of producing justified decisions in situations where principles conflict rather than the memorization of abstract rules.

Ma et al. (2025) review together the cognitive, affective and behavioural dimensions of responsible AI literacy systematically. The first, the cognitive dimension, is represented by knowing what algorithmic discrimination is; the second, the affective dimension, is represented by showing sensitivity towards those affected; the third, the behavioral dimension, is represented by objecting to an unjust practice. Moral reasoning should connect knowing and doing.

The fact that artificial intelligence is able to give moral advice does not relieve the student of his moral responsibility. A system may offer possible evaluations based on different ethical theories, but it cannot take responsibility for deciding which value should have priority in a particular situation. Moral education should employ artificial intelligence not as an off-the-peg conscience but as a modest discussion device for comparing different reasons.

**The Person Who Can Challenge Technology**

Questioning technology is not the same as opposing technological development or rejecting all artificial intelligence systems. It's about understanding the way a technology defines a problem, what purposes it serves, what data it relies on, and who benefits from the results. Shneiderman's (2020) human-centered approach argues that technological systems should be for human ends. Hence technology is not an end but a means that requires justification.

Someone able to question technology knows that systems aren't neutral. Data selection, success criteria, and design choices carry values. Nemorin (2024) critiques AI ethics policies and finds that governance frameworks that are claimed to be universal can reproduce Eurocentric systems of knowledge and value. Students need to understand technology not just as a technical construction but as a cultural and political one.

Critical technological literacy is asking questions that go beyond whether a system makes visible errors. Even if an application functions correctly it may be constantly monitoring students, motivating competitive performance standards or degrading human relationships. The critique within the framework suggested by Veldhuis et al. (2025) addresses representation, power, user agency, and the assessment of accuracy.

A person who is skeptical of technology has the right to turn down AI suggestions and work on alternatives. Pietarinen and Shi (2026) argue that AI applications can be manipulative when they bypass users' rational evaluation and when they invisibly restrict the available options. In education, students need to understand why a system is making a recommendation and be able to choose a different path.

Such questioning should not be confined to individual use. Students should have a say in which artificial intelligence (AI) systems are used in their schools or universities, what data are collected, and how automated decisions are supervised. An individual capable of questioning technology is not just an enlightened consumer but a democratic subject who takes part in the construction of technological arrangements.

## The One Who Protects Human Dignity

Human dignity is the recognition that a person has value in him or herself, not only as an economic, technological, or social asset. From a Kantian perspective, the human being cannot be merely a means to the ends of others. This principle offers a strong limit to the reduction of students into data sources, performance profiles or objects of algorithmic prediction in the era of artificial intelligence.

The one who wants to defend human dignity must first of all refuse to abdicate his own judgment and responsibility. Schwarting and Reihlen (2026) argue that ongoing submission to artificial intelligence might develop a tendency in individuals to refrain from exercising their capacities for autonomy and responsibility. Human dignity means maintaining the ability to make one's own decisions in the face of technological recommendations.

Dignity is also a vision of a man as more than data of the past. Jun and Lee (2026) contend that algorithmic predictions should not prohibit students' unactualized potentialities for development, what they refer to as "anticipatory dignity." Just because a student is labeled at risk or low achieving does not mean that the student cannot grow to be a different person in the future.

The protection of human dignity also includes respect for the dignity of others. Students should reflect on whether their use of artificial intelligence could cause harm to others through discrimination, exclusion, surveillance or misinformation. According to Gouseti et al. (2025)'s framework of autonomy, justice and well-being, the impact of technological decisions on multiple stakeholders should be seen together.

The defender of dignity is opposed to turning people into numerical profiles in the name of efficiency. Artificial intelligence can help teachers understand the needs of students but it cannot replace human emotions, relationships, living conditions and development opportunities. The output of technology is not a final judgment for a responsible person but a limited indicator that must be supplemented by the account of the individual and the human judgment.

Human dignity is not just a passive value to be protected, but students must also learn the moral courage to defend the dignity of others. The behavioral dimension of dignity-based education includes protesting against an unfair algorithmic practice, asking why some disadvantaged groups are invisible, and insisting on keeping human decisions.

**Education in the Human-Centered Approach**

A human-centered approach to education does not mean that the technology cannot be used in traditional educational environments. Its main argument is that technology is not an independent force that determines the goal of education, but a tool for human development. Shneiderman (2020) advocates for human-centered artificial intelligence driven by automation and human control. So we can increase the processing power of AI in education but the authority of teachers and students to set targets, assess results and make decisions is unchanged.

Being human-centered means more than making a system easy to use. The fact that an application is easy to use does not mean that it is pedagogically sound or ethically appropriate. Alfredo et al. (2024) show that meaningful participation of students and teachers in the design of human-centered educational AI is still limited. Being human-centered is not to measure users' needs after the system is built, but to involve users in goal setting, in decisions about data practices and design choices, and in setting levels of automation.

As it pertains to education, a human-centred approach starts with the student as a person, as someone who can grow, not as a data profile or a performance output. Artificial intelligence systems can provide some useful insights into student behaviour but they cannot capture the whole picture of motivation, emotion, living conditions, or the meaning students make of learning. However, current AI applications rarely take into account important dimensions of learning such as motivation, emotion and metacognition, as Hooshyar et al. (2025) point out.

In the human-centered education the teacher is not a passive intermediary who mechanically implements plans decided by the technology. The review by Tan et al. (2025) emphasizes the significance of teachers' pedagogical and ethical decision-making competencies for the integration of artificial intelligence. The teacher should be a professional actor who evaluates the material created by the system in the light of the student context, rejecting it if necessary, and bearing the final responsibility for education.

**Technology Means and the End of the Human Being**

Then we understand that when we see technology as a tool, the value of artificial intelligence is not in speed and novelty, but in how it complements human development. Shneiderman (2020) argued that human-centered artificial intelligence systems should augment human creativity and responsibility, rather than replace them. The educational value of an application lies only in the fact that it promotes the understanding, autonomy and participation of the students.

The means/ends distinction clarifies the difference between technical and educational success. A system may allow a student to perform a task faster but that does not necessarily imply that the student has understood the concept in question, or that he/she has developed the capacity to think. Kasneci et al. (2023) show that large language models can generate helpful explanations, but also generate inaccurate and fabricated content. Hence, the mere ability to produce output does not signify the attainment of educational objectives.

The human being as the end also signifies that students must not be reduced to data sources for artificial intelligence systems.

Collecting data from students can improve the system's performance, but the process of collecting data should not compromise student privacy, dignity, or agency. The necessity conditions of responsible AI in education include technical functionality, data security, equity, and supervision by humans (Fu and Weng, 2024).

Human-centeredness also places limits on artificial intelligence making decisions on behalf of human beings. Systems might recommend content to students or give early warnings to teachers, but they shouldn't make final judgments about a student's ability, worth or future. Alfredo et al. (2024) note, "A human control principle is that automated recommendations should be up for discussion and rejection."

To see technology as a means is not to say it is inert or ineffective. User choices and actions can be shaped by technology tools. That is why Ng et al. (2021) add evaluation and ethical responsibility to knowledge and use in their AI literacy approach. The student should be able to ask about the options that a tool exposes and the options it hides.

Educational policy is directed towards individual achievement but also towards well-being, relationships and shared life when the human being is the aim. Gouseti et al. (2025) argued that the assessment of AI ethics in education should revolve around the concerns of autonomy, justice and well-being. A technology is not human-centered if it increases student anxiety, exclusion or dependence, even if it performs well.

**Maintaining the Teacher–Student Relation**

The teacher-student relationship is not a transfer of knowledge from one person to another. It also contains trust, acknowledgment, shared responsibility and concern for the development of the student; The teacher is not just a director of student performance in specific forms, but also a relational actor who calls them to become responsible subjects in the world (Tishenina, 2026).

Artificial intelligence can help students with fast, patient and customized answers. But it doesn't have the same holistic understanding as a human teacher does of the emotional context of a student's words, of why a student is silent, of the connection between failure and life circumstances. Hooshyar et al. (2025) note that differences in emotion, motivation, and context are often overlooked in educational AI.

That is not to say we should exclude technology from all communication processes in an effort to preserve the teacher-student relationship. AI could take over the menial feedback and content curation, allowing teachers to build stronger connections with their students. But the review by Tan et al. (2025) suggests this is not a foregone conclusion and teachers might have to produce more and quickly rather than be given a reduced workload.

Artificial intelligence with a human-centered approach should not be the last communication channel between teacher and student. Students should be able to hear from teachers the reasons for assessment decisions, object and describe their own learning experiences. According to Chan (2023), the use of artificial intelligence should be guided by pedagogical decisions, governance and accountability mechanisms.

Keeping the relationship does not mean keeping students dependent on teachers. The teacher is a temporary guide to assist learners in developing their own judgment. In Ng et al.'s (2021) evaluation-oriented approach to AI literacy, the goal is for students to independently assess sources and system outputs. The teacher-student relationship should increase, not decrease, the autonomy of the students.

AI-supported education also needs to realize that teachers can be wrong and students can produce knowledge. A human-centered relation is based on the sharing of reasons, not authority. Students should be able to present contradictory findings obtained through artificial intelligence to the teacher and the teacher should be willing to revise their claims in the light of the evidence.

**Effectiveness and attainment of human development**

Efficiency means doing more and doing things faster with the same amount of resources. Artificial intelligence can help speed up tasks such as preparing teaching materials, providing routine feedback, translating texts and customizing content. Kasneci et al. (2023) argue that these functions might be useful for teaching processes. But the effectiveness of education cannot be measured by the number of tasks done, or the number of texts written.

Human development requires time and repetition and mistakes and uncertainty. Technically speaking, a long time spent on a problem by a student may appear wasteful, but it might be beneficial for conceptual understanding and intellectual strength. Hooshyar et al. (2025) suggest that educational AI should think about motivation, metacognition, and learning processes as well as outcomes.

The pressure for efficiency also impacts the work of teachers. AI might free teacher time, but those freed-up hours could be translated into institutional expectations to produce more content and grade more students. Tan et al. (2025) found that teachers require support not only in technical skills but also in pedagogical decision making and professional autonomy.

Human development is more than just easily measured performance indicators. Other qualities such as curiosity, moral sensitivity, cooperation, patience, and meaning making are not well measured by short-term numerical measures. For instance, Bond et al. (2024) meta-systematic review indicates that research on artificial intelligence in education needs to put more emphasis on ethical and human-centred aspects and on measurable performance.

Striking the right balance does not mean losing efficiency. Artificial intelligence could take over repetitive tasks and free up teachers' time for valuable feedback, conversations and student support. For this to happen, institutions should not only view time savings as opportunities to produce more, but also as resources to promote human interaction and pedagogical quality.

The purpose of the task should determine the ratio of efficiency to development. It may be efficient to teach a skill by directly transferring it to a system, but it may not be pedagogically effective. In contrast, if students have already mastered a skill, automating routine processes may free up time for more complex thinking.

**Artificial intelligence: ethical and transparent use**

Is it ethical to use AI only as much as students abuse a system? It also raises questions about what types of data the system collects, how it makes decisions, what groups it might harm, and who is to blame when things go wrong. The focus of responsibility for educational AI is on principles such as privacy, fairness, oversight by humans, and transparency (Fu and Weng, 2024).

Transparency does not imply that all users know all technical details of the system. Students and teachers should have accessible information about the purpose of the system, the sources of its data, its main limitations and how to evaluate its decisions. According to

Alfredo et al. (2024), human-centered systems should be designed so that users know the level of automation and where human control is retained.

Explanations should empower users to challenge decisions. The student should be allowed to challenge the data and the decision criteria that led to a risk score, achievement prediction or academic integrity judgment against him. Transparency is not just a tool for building confidence in a system, it is the basic foundation on which accountability and correction mechanisms are built.

An ethical use requires minimisation of data. Educational institutions should not be gathering every kind of data that they can technically gather. In the framework of autonomy and well-being by Gouseti et al. (2025) we demonstrate that student privacy cannot be violated without limit in the name of personalization. Data collection should be restricted to clear and proportionate educational purposes.

Artificial intelligence systems should be routinely audited for bias. A system can be very overall accurate, but have fewer mistakes for some languages, cultures, disabilities, or socio-economic groups. “A human-centered approach means looking at how outcomes impact different groups of students, not just the average.

The weight of morality cannot be solely on the shoulders of software developers. In the ecological policy approach, Chan (2023) elaborates on the roles of institutional leadership, teachers, students and technical providers. Educational institutions need to own the consequences of the systems they use, and not to blame the technology by saying “this is how the technology works”.

Using AI ethically also means getting students involved in making AI policy. According to Ma et al. (2025), responsible AI literacy requires learners to comprehend the rules, assess ethical issues and act responsibly. Student participation makes them democratic stakeholders and not objects of technological regulation.

## A Human-Centered Education Model Based on Epistemic Autonomy

The growing deployment of artificial intelligence in education has provided students with more information and more support for their learning, but has also sparked more debate about the role of students in the creation of knowledge. Students can be active subjects evaluating the answers provided by a system or they can be passive users adopting ready-made outputs. Wu et al. (2025) demonstrate through their treatment of human epistemic agency that, in a learning partnership with generative artificial intelligence, students should retain the roles of setting goals, assessing progress, and making the ultimate decisions.

The Epistemic Autonomy-Based Human-Centered Education Model does not seek to make students independent of all sources of knowledge. In line with Zagzebski's (2012) idea, people might realize their own knowledge is not enough and go to more knowledgeable people, but they must be able to judge why and how much they can trust them. In the model, autonomy does not mean to reject the epistemic support, but to critically deal with the support, with justification and responsibility.

The model also situates technology as not the end of the educational relationship but as a means for supporting human development. "Human-centered artificial intelligence should develop high levels of automation with human control," Shneiderman (2020) says. Artificial intelligence may assist information seeking, generation of examples and feedback in education. But the setting of learning goals, the evaluation of values and final epistemic responsibility must remain with human beings.

The proposed model consists of five functional principles and two constitutive dimensions that sustain them. Critical verification, justified trust and ethical responsibility govern students' epistemic and moral engagement with AI-generated outputs. The educational and institutional conditions in which this is done are determined by human-centeredness and pedagogical guidance. Having clarified the purpose and basic structure of the model, we should think about each of these principles in turn.

**Purpose of the Model**

The main goal of the model is to raise people who would be able to work with the artificial intelligence, but not to pass their knowledge and decision-making to it completely. Ng et al.'s (2021) approach to AI literacy shows that technical use should be coupled with the ability to assess systems and understand ethical concerns. Thus the model aims to help students not only to gain effective results from a system, but to assess the accuracy, limitations and consequences of those results.

The second aim of the model is to make the pedagogical difference between accessing information and possessing knowledge visible. A correct answer from artificial intelligence may not represent a lasting epistemic accomplishment if the student cannot explain its foundations or use it in other contexts. According to Schmidt et al. (2025), it is not necessarily the case that a correct result by a system implies a user has knowledge based on that result.

A third aim is to maintain student epistemic agency. As Wu et al. (2025) stated, "Learners should select their own information-seeking strategies, evaluate the system outputs, and revise them when necessary while working with generative artificial intelligence." A form of epistemically dependent learning can be learning that is prompted or answers prepared beforehand, but seems active.

The model also aims to shift the teacher's role from a knowledge transmitter—made obsolete by artificial intelligence—to a pedagogical guide who helps students develop their independent judgment. Tishenina (2026) argues that the teacher's relational role in calling students to become responsible subjects cannot be completely replaced by generative artificial intelligence. The teacher should not be the final arbiter who approves the AI-generated output, but guide students to develop their own criteria for verification.

And finally, the model aims at relating the use of technology to human development, democratic participation and ethical responsibility. In Dewey’s (1916) conception of education, learning is not only individual knowledge, but also collective inquiry and reconstruction of experience. AI-based education should not only improve individual performance, but also allow students to discuss common problems and to participate in decisions on the social consequences of technology.

**The Basic Principles of the Model**

The model’s principles do not form a linear sequence, but rather a mutually supportive structure. "Without critical verification, trust in artificial intelligence cannot be justified. Justified trust is necessary if we are to know which sources students should trust and to what extent. Information without ethical responsibility can be used for destructive purposes. Without human-centeredness and pedagogical guidance, the conditions will not be created for development of individual competencies.

These principles have in common human epistemic agency. Wu et al.’s (2025) study shows that the student’s attitude toward generative artificial intelligence may vary with the task type. Sometimes students may use the system as a tool for searching information, sometimes as a partner in critical discussion, sometimes only as a source of technical support. However, in all cases the final evaluation and ownership should be with the human user.

AI literacy is regarded as a multidimensional competence in the model. Hackl et al. (2026) present a seven-dimensional approach that underscores the necessity for technical and practical expertise to be accompanied by critical, ethical, social, integrative, and legal competencies. So the ability to use AI effectively inside the model is more than simply writing good prompts or getting fast outputs.

The principles should be applied in a way that is appropriate to the nature and level of risk of a task. Students might get more leeway for lower-risk activities such as editing language, but academic grading, student profiling and other important decisions would require stronger supervision by humans. Fu and Weng's (2024) framework of responsible human-centered AI shows that autonomy and human agency must be evaluated together with fairness, privacy, and nonmaleficence.

The principles of the model are not just individual responsibilities decreed to students. According to Chan (2023), the ecological policy approach implies that educational institutions need clear rules on AI use, data policy, assessment criteria, and appeal mechanisms. A human-centered approach does not fit with demanding students use artificial intelligence critically, and not providing enough information about the system or a real right to choose.

**The principle of critical examination**

The principle of critical verification requires that any significant claim generated by artificial intelligence be scrutinized for sources, evidence, currency, context, and dissent. Kasneci et al. (2023) show that large language models can generate false information and false citations in a coherent narrative structure like factual information. And so the fluency or persuasiveness of a response cannot prove that it is true.

The first step of verification is to determine which claims in an output need to be verified. Students should be able to distinguish between factual propositions, interpretations, value judgments, and predictions. The dimension of evaluation in the framework of Ng et al. (2021) indicates that AI literacy is about users making judgments on the quality of system outputs, not simply passively accepting system outputs.

The second step is comparison across sources. AI-generated responses should be checked against primary sources, peer-reviewed studies, textbooks or institutional reports, and the presence of sources provided by the system verified. In the analysis of opacity by Hähnel and Hauswald (2025) it is shown that often a user is not able to see how a system came to a specific conclusion. The lack of information can be partially compensated by independent source verification.

Critical verification should not be restricted to factual accuracy. Veldhuis et al. (2025) advise users to also think about what groups are represented in an output, what perspectives are excluded, and what social values are normalized. Even a factually accurate answer can be epistemically problematic if it restricts the context or makes particular experiences invisible.

This uncertainty should be made explicit as part of the verification process. Students should be able to reserve final judgment when the evidence is insufficient or the experts disagree. Schmidt et al. (2025) criticize opaque artificial intelligence showing that arriving at the right answer does not necessarily equate to a sound epistemic position. Students should also be able to explain the rationale for a correct answer.

In teaching practice, students can be asked to produce a "verification trail" for an AI-generated answer. This record might include the claims that were studied, the independent sources used, errors that were identified, opposing viewpoints that were considered, and the student's final assessment. Such a practice, based on Freire's (2000) critical conception of education, turns students from recipients of ready-made knowledge into subjects who question knowledge claims.

## The Principle of legitimate trust

Justified trust means that trust in artificial intelligence should not be based on the appearance, popularity or fluency of the language of the system, but on demonstrated trustworthiness. Durán and Pozzi (2025) differentiate between trust, that is the user's attitude, and trustworthiness, that are the characteristics that make a system trustworthy. Thus, a student's confidence in a system is not sufficient to show that the system is actually reliable.

Trust should be specific to the task. An AI system might be good for language editing but not necessarily for up-to-date scientific information, legal interpretation, or evaluating individual students. Zagzebski's (2012) account of epistemic authority demonstrates that trust in a source is a function of the reflective recognition of its superior competence in a domain. A system should not be considered a universal and unlimited authority in every domain.

Justified trust needs to know the system's past performance and conditions of failure. Understanding what data the AI has been tested on, what groups it is more likely to have higher error rates on, and how the information it uses is kept up to date is important. • Opacity prevents users from predicting the system behavior and from assessing individual outcomes (Hähnel and Hauswald, 2025). Hence, independent oversight and institutional safeguards should be part of the foundation of trust.

Explainability matters, but is not enough by itself. Baron (2025) suggests that users can have some limited trust in a system based on proven performance and reliable institutional oversight, even if the full system is not understood. In education, students may not be taught all the technical details of all the models, but they should be told the purpose of the system, major limitations, typical types of error, and situations in which a human expert should be consulted.

The levels of trust should be proportional to the levels of risk. A mistake in a low-stakes brainstorming session can be easily reversed, but decisions about student achievement or academic integrity require a higher level of evidence. According to Fu and Weng (2024), the principles of human agency and nonmaleficence imply that high-impact decisions should not be made purely based on automated outputs.

Justified trust is opposed to both over-confidence and unjustified distrust. Perhaps useful support systems are lost by rejecting a system that has been independently proven to work simply because it is artificial intelligence. The goal of the model is not to increase or decrease trust, but to align it with the actual capabilities of the system.

**Principle of ethical responsibility**

The principle of ethical responsibility calls upon us to consider the use of AI not only in terms of access to accurate information, but in terms of its impact on people and communities. Ma et al. (2025) also demonstrate that responsible AI literacy should involve affective sensitivity and behavioral responsibility besides cognitive knowledge. Students should know what a risk is, care about people affected by it, and act responsibly.

Ethical responsibility is also about questioning whether the data and the outputs being used are in violation of the rights of others. The basic conditions of human-centered artificial intelligence are privacy, security, fairness, and nonmaleficence in the framework by Fu and Weng (2024). Students should consider the possible harms before uploading personal data, processing the work of others, or posting content created by AI.

The principle also includes a call for transparency around AI contributions. Students should be able to describe the system that was used, what it was used for, what contributions they accepted, and how they verified the final output. This disclosure is not just a mechanism of monitoring academic integrity, but a pedagogical practice that makes visible the student's responsibility within the framework of knowledge production.

Ethical responsibility requires us to take a position on algorithmic bias and discrimination. Veldhuis et al. (2025) argue that critical AI literacy should address representation and power relations. Students should be able to assess whether an output systematically excludes certain cultures, languages or social groups, and to take steps to remedy such omissions, as appropriate, by turning to alternative sources.

The burden should not fall on the individual user alone. Chan (2023) proposes an ecological policy model, which needs a clear definition of the roles of educational institutions, teachers, students and technology providers. Institutions cannot blame student "misuse" for harms that result from unreliable systems and we must hold institutions accountable for system selection, data management, and oversight processes.

The principle of ethical responsibility also calls for student involvement in creating rules for the use of AI. For example, Hackl et al. (2026) suggest that the social and legal dimensions of AI literacy call for students to be prepared not only as technical users but also as citizens who can express informed opinions about technological regulation. Ethical education must include discussion of the justification of rules, not just following the rules.

**The Human-Centeredness Principle**

The principle of human-centeredness implies that technology should not be the ultimate goal of education but the means to support human development. Shneiderman (2020) suggests that as automation increases, human-centered systems should increase human control and responsibility. Artificial intelligence in education can be considered legitimate only insofar as it develops the capacities of students to think, learn and participate.

Human-centeredness means that students are not reduced to measurable performance indicators. Artificial intelligence systems can provide useful information about student behavior, but they cannot fully represent students' living conditions, emotions, purposes, or capacity for change. Fu and Weng (2024) principles of agency and autonomy ensure students' right to tell their own stories and challenge system decisions based on algorithmic profiles.

The principle opposes the unseen limitation of student choices by artificial intelligence. Pietarinen and Shi (2026) claim that system behaviors that bypass students' rational examination and guide them to undiscovered goals might be manipulative. The human-centered system should offer recommendations that are explainable and rejectable and allow students to choose alternative learning paths.

Human-centeredness also protects the professional judgment of teachers. Teachers should not become a middleman that automatically uses content created by AI and assessments. According to Tishenina (2026), in a relational conception of teaching, the teacher has a specific role in attending to students' responsibility, subject formation and development.

The principle is not that man is better than technology. Artificial intelligence is able to do some things faster or better than humans, but better performance does not entitle it to determine what the goals of education should be or to make value judgments about humans. Following Shneiderman (2020), the different strengths of humans and artificial intelligence should be combined, with ultimate control and responsibility left to humans.

Human-centeredness implies institutional participation. Chan (2023) argues that AI policies should be developed in parallel at the pedagogical, governance and operational levels. Students and teachers should have a real voice in choosing the systems that impact them and the conditions under which those systems are used.

**The Principle of Pedagogical Direction**

The principle of pedagogical guidance means that students should not be left alone with their encounters with artificial intelligence, but at the same time they should not become permanently dependent on teachers. As Tishenina (2026) states, teachers do not only lead students to particular answers, but they call upon students to be subjects, responsible for their thoughts and actions. Guidance is, then, the task of enabling students to make increasingly independent epistemic decisions.

Teachers should reveal the abilities and limits of artificial intelligence in the classroom. Concrete examples can be taken to analyze the risks of misinformation, bias and fabricated sources identified by Kasneci et al. (2023). They might be asked to compare the AI-generated answers with academic sources, categorize errors, and discuss the reasoning behind their corrections.

The use of artificial intelligence by students should be regulated at the level of pedagogical guidance. Students confronted with a full solution before they have had an opportunity to formulate their own initial solution may shift their cognitive effort to the system. As suggested by Wu et al. (2025) for epistemic agency, students should first set their own goals and make preliminary assessments, and then use artificial intelligence for comparison, feedback or generation of alternatives.

Guidance should not state a single rule of use for all tasks. If the task is to develop an argument then artificial intelligence may have to be restricted from generating the entire argument. When the purpose is to compare or critique a pre-existing text, broader use may be appropriate. Pietarinen and Shi (2026) maintain that pedagogical guidance should be open and transparent for discussion so as to preserve students' critical thinking.

Teachers should also be receptive to criticism of their own authority. Freire's (2000) dialogical view of education assumes that students are part of creating knowledge, not recipients of ready-made information. When students present evidence from artificial intelligence or some other source that contradicts a teacher's position, the evidence should be discussed together rather than using institutional power to silence the disagreement.

Pedagogical guidance should also transform the design of assessment. Assessments should not be limited to the final product but should also include students' prompts, first drafts, verification processes, source selection, and the reasons they accept or reject suggestions from AI. The importance of making the learning process visible through dimensions of evaluation and ethical use is supported by Ng et al. (2021).

Guidance should encourage students' metacognitive awareness. Students should know whether they are using artificial intelligence to deepen their understanding or whether they're just using it to get a task done quicker. Wu et al. (2025) argue that the capacity to adopt an appropriate stance in response to shifting epistemic situations is a fundamental condition of active agency in human-AI partnerships.

## Implications of the model for the processes of education

The Epistemic Autonomy-Based Human-Centered Education Model is not an individual competency model that describes how students should respond to the outputs produced by AI. The model will only work if course design, teacher education, assessment systems, conceptions of academic integrity and institutional policies all evolve together. From an ecological perspective of Chan (2023), pedagogical practices, governance rules and technical infrastructure are not to be understood in isolation from one another.

The main criterion for the use of the model is if the technology enhances the human agency in the learning process. Learning partnership between students and artificial intelligence should maintain the functions of goal setting, information evaluation and final decision making, according to Wu et al. (2025). Although the choice between options generated by a system may appear active, it is not enough for epistemic autonomy.

Artificial intelligence in education is not about applying the same approach to all courses. In their systematic review, Belkina et al. (2025) reveal that generative AI applications yield different results according to the discipline, learning objective, student level, and activity type. Thus, the model should be seen as a set of principles for pedagogical decision-making rather than a fixed list of tools.

## Curriculum Development

The first step in the design of a course is therefore to define the learning objective clearly, before deciding on how to use artificial intelligence. If the skill students are to develop is directly transferable to artificial intelligence, the use of technology can remove the learning objective. Generative artificial intelligence can produce outputs that are quickly and succinctly structured, but it does not ensure that students have truly engaged in conceptual understanding, inquiry, or argument building (Kasneci et al., 2023).

In the Epistemic Autonomy-Based Human-Centered Education Model, learning activities are organized in a three-stage learning cycle. In the first stage learners construct a first solution based on their own knowledge and sources. In the second phase they receive an alternative explanation, critique or feedback from artificial intelligence. At the third stage they compare the system's output with independent sources, justify their final position. The notion of epistemic agency (Wu et al., 2025) suggests that this sequence has the potential to move students away from passive consumption of pre-made outputs.

"When you use AI is just as important as what you want to learn. Students who are given a full answer before they have had enough time to work on the problem lose the opportunities of cognitive difficulty, of making mistakes, and of reworking their own solutions. A review of generative AI literacy suggests that students have a particular need for explicit instruction in critical evaluation (Park, 2025). AI access could be offered at the conclusion of an activity, not at the commencement, after students have made an initial intellectual effort.

Artificial intelligence should be used in course design not as an authority that gives the one answer, but as a tool for comparison and discussion. Students can compare responses to different prompts, identify the assumptions underlying the responses, or ask artificial intelligence to generate a counterargument to a position. Applications demonstrate the potential of interaction and reflection to enhance the pedagogical value of generative artificial intelligence (Belkina et al., 2025).

The principle of critical verification must be a separate stage in the lesson plan. Instead of simply being asked to "use sources," students may be asked to explain which claims in the AI-generated output they verified and which sources they used. The risk of fabricated citations identified by Kasneci et al. (2023) necessitates independent assessment of the existence, content and relevance of sources.

Ethical responsibility should also be explicitly included as a learning outcome in course design. Students should discuss what types of data should not be input into artificial intelligence systems, what groups may be harmed by certain outputs, and how the use of that output may obscure the labor of others. According to Park's (2025) review, privacy, data security, and academic integrity are the major issues of generative AI literacy today.

**Teacher Training**

Teacher training should not be limited to short-term training courses to familiarize teachers with technical features of AI tools. As shown in the systematic review by Tan et al. (2025), teachers need competencies in pedagogical adaptation, ethical evaluation, and professional judgment, in addition to technological proficiency. The ability to write good prompts might be useful, but the inability to recognise wrong or pedagogically inappropriate outputs is a more fundamental problem.

Teacher training in generative artificial intelligence should include at least five areas of expertise. These are technical fluency, pedagogical integration, critical appraisal, ethical reasoning and collaborative practice (Shenoy & Saarela, 2026) These domains are interdependent and teachers need to know how to use a system technically, but also when a system should not be used for a particular learning purpose.

Build professional development around case-based learning. Teachers could be asked to review a wrong output of AI, to point out the ethical risks of an application using student data, or to discuss the relevance of the same system for various courses. Shenoy and Saarela's review (2026) shows that higher-order critical-thinking outcomes will not arise spontaneously without explicit structure and guidance.

Teacher education should also be discipline-based. The assessment of sources in history, the solving process in mathematics, the construction of arguments in philosophy and the editing of texts in foreign language teaching cannot be assessed with the same criteria for AI use. Belkina et al. (2025) demonstrate that pedagogical implications of generative AI applications differ for disciplines and student groups.

Teachers need to be empowered to use AI but also the choice not to use it or to use it sparingly. Professional autonomy means teachers will not be forced to use a system purchased by the institution in all cases. In her policy framework, Chan (2023) claims that teachers shall be active stakeholders in the development of AI policy.

Teacher education must be evaluated beyond self-report scales. Performance-based tasks could include designing an AI-supported lesson plan, identifying risks, evaluating student work, and justifying use decisions. Shenoy and Saarela (2026) note that there is a dearth of performance-based teacher assessments in the existing literature.

## Évaluation et évaluation

Generative AI has changed students' approach to doing assignments and exam tasks. A direct implication from this is that assessments that rely solely on the final written document might not capture students’ real competencies. Xia et al. (2024) find that artificial intelligence can support assessment, but it can also create new problems around identification of student contributions, validity and academic integrity .

The model suggests that assessment should include evidence of the process as well as the final product. Evaluations could be: students’ first drafts, the prompts, validated claims, the sources used, and the reasoning behind accepting or rejecting suggestions from the AI. Luo (2024) argues that originality should be less about working independently of technology and more about managing the process and defending the resulting product.

Assessment tasks may require you to draw upon personal, local or subject contexts. We should not be looking for questions that a system cannot answer, that are “AI-proof”. Assessment should be designed to establish acceptable use of AI (Corbin et al., 2025). Some of the tasks may employ artificial intelligence to measure the competency.

Assessment can take the form of oral defenses, classroom discussions, draft feedback, or reflective reports. Students could be asked to identify the main arguments of their final product, highlight the AI-generated suggestions they revised, or assess an alternative viewpoint. Cotton et al (2024) argue that academic integrity is inseparable from the design of assessment.

Even if artificial intelligence is completely banned, secure assessment environments might still be necessary. However, the shift of all assessment to supervised and closed settings may rule out important learning objectives such as research, collaboration and digital competence. Xia et al. (2024) suggest that assessment systems should be built in a balanced way by combining various task types.

Nonetheless, AI-assisted automated assessment should have human monitoring. Systems may provide feedback or early evaluations, but they should not render final judgments on student achievement or academic futures. Kasneci et al. (2023) on inaccurate and biased outputs Independent teacher judgment must be used to verify even high-impact decisions.

## Academic Integrity

Academic integrity means not only not plagiarising but the principles of honesty, trust, fairness, responsibility and respect for the work of others in the production of knowledge. Balalle and Pannilage (2025) note how artificial intelligence can help as well as threaten these values. So the presence of a tool is not proof of either wrongdoing or integrity.

A critical part of academic transparency involves disclosing the use of AI. However, the Gonsalves (2025) findings suggest that students are not always aware of what they are to disclose, how much and in what way. Likewise, a generic statement such as "I used artificial intelligence" is not enough to describe the real contribution of the technology to the learning process.

The AI-use disclosure of the model can include at least four types of information: the name of the system used, the purpose of use, stages in which the artificial intelligence helped, and how the student verified the output. Students probably don't need to show all prompts but should explain what the major AI contributions were to the final product. Gonsalves (2025) shows that disclosure systems should be transparent, proportionate and designed to diminish fear of punishment

We can't entrust the students to protect academic integrity. Unclear assignments, evaluations based on product alone, and inconsistent institutional practices can increase the likelihood of misconduct. Gruenhagen and colleagues (2024) stress the importance of developing a clear and shared understanding, since students do not always perceive the use of AI as an academic violation.

AI detection tools aren't a surefire test. False positives can lead to students being falsely accused and can damage the trust between teachers and students. Cotton et al. (2024) argue that we ought to give priority to the integrity of assessment design, communication and learning culture, rather than relying solely on technological detection.

Academic integrity means that students should give credit for any AI contributions and take responsibility for the resulting text. Students are not exonerated of responsibility and cannot blame the system for spurious references or discriminatory statements generated by artificial intelligence. A review by Balalle and Pannilage (2025) found that institutional culture and student education should support ethical AI use.

## Policies for the Use of Artificial Intelligence

AI use policies shouldn't just be a list of allowed and disallowed tools. Rules that are tied to particular product names quickly become obsolete as technological systems change rapidly. According to Chan (2023), the ecological policy approach should be based on pedagogical aims, principles of governance and operational support.

The first part of policy is the pedagogical justification. Institutions should define learning goals for which artificial intelligence can be applied and tasks for which its use should be limited. Wang et al.'s (2024) review of university policies suggests an important point that the use of AI should be aligned with learning objectives.

The second element is clarity of the discipline and the course. A general policy for universities cannot set assessment purposes for all and every course. Corbin et al. (2025) find that the acceptability of using AI is context-dependent. Instructors should therefore clarify with examples what types of use are permitted, restricted or prohibited in each course.

Third, data privacy and security. Students may not upload personal data, unpublished research data or data belonging to others to publicly accessible artificial intelligence (AI) systems. Data privacy is an important aspect in the university policies (Wang et al., 2024). Institutions should provide secure system choices and transparent rules for data processing.

The fourth should set standards for disclosures and transparency. Institutions should indicate how different types of AI contributions are to be recognised and when references, footnotes, methods or separate declarations should be used. Unclear disclosure expectations may increase noncompliance (Gonsalves, 2025).

The fifth element is the assessment and appeal mechanisms. No allegation of a student's misuse of artificial intelligence should be made on the basis of the output of an automated detection system alone. Students need to be able to tell, show drafts, show their process. A human-centric model demands that decisions with a substantial impact are reviewed and logically evaluated by a human before being put into practice.

The sixth element should be education and support services. It will not be enough to send the policy document to the students and teachers only. Park's (2025) results point to a need for structured education for students especially around critical evaluation, ethical use and data security. Institutions should offer case studies, advice services and syllabus templates.

Policies should be participatory and reviewed regularly. Chan (2023) suggests involving students, teachers, administrators and technical experts in the policy-development process. In this way rules can become not only prohibitions imposed from above but norms reflecting the educational aims and the common responsibility of the institution.

# CONCLUSION

The rapid integration of artificial intelligence into educational processes has led to a revolution that goes far beyond the diversification of tools used in education. Generative artificial intelligence systems are changing the way people search for information, ask questions, write text, solve problems, get feedback and make evaluations. This transformation requires a reconceptualization of the roles of teachers, students and technology in the educational process. The basic question is not whether artificial intelligence should be used in education, but for what purposes, within what limits and on the basis of what conception of the human being it should be used.

Artificial intelligence presents great educational opportunities since it is able to analyze a lot of information in a short period of time, providing different explanations and offering personalized support. But speed, fluency and breadth of coverage are not the same as accuracy, understanding and dependability. The fact that a system generates a persuasive answer does not mean it is correct, up-to-date, unbiased or pedagogically appropriate. The educational value of artificial intelligence should thus be judged not merely by its ability to generate outputs but by the extent to which it supports human capacities for thinking, evaluation, and responsibility.

This study argues for a primary position that artificial intelligence should not be viewed as an independent educational agent that replaces the teacher or student but rather as a limited tool that supports human epistemic and pedagogical activity. Teachers should retain a crucial role in determining educational goals, in tracking students' progress and in exercising pedagogical responsibility. In turn, students should not be passive recipients of information provided by artificial intelligence, but epistemic subjects who question, verify, transform and take responsibility for their final judgments.

The new balance that needs to be found between education, humanity and artificial intelligence, can not be simplified into a mere division of tasks between men and technology. Such a balance requires that epistemic authority and pedagogical responsibility, as well as moral decision-making and educational aims remain with human beings. Artificial intelligence can do things like process information, produce examples, provide alternatives, and help with routine work, but it should not be the ultimate arbiter of what knowledge is worth having, what goals to choose, or what a decision means for human beings.

Epistemic autonomy does not demand that individuals themselves generate all knowledge, or that they dismiss authority, or that they act apart from external sources of information. In the present epistemic world, one has to rely on teachers, experts, scientific institutions, books, digital sources, and artificial intelligence systems. What makes epistemic autonomy distinctive is not the use of such sources but the critical and justifiable way in which the relationship with such sources is managed.

Epistemic autonomy in the era of artificial intelligence is more than simply the faster access to information or the choice among alternatives. Students need to be able to evaluate the source, evidence, currency, context and limitations of the knowledge claims they encounter. But the use of an AI-generated output is not the same as epistemic ownership of that output. Even if students arrive at the right conclusion, they cannot be said to have completed an adequate process of knowledge acquisition and justification if they cannot explain why an answer should be considered correct.

Epistemic autonomy requires the interplay of three fundamental abilities. The first is cognitive competence, by which people can evaluate knowledge claims, evidence and sources. The second is epistemic motivation, which drives people to keep looking for truth and justification, rather than be satisfied with ready-made answers. The third is a normative dimension of responsibility for the information one accepts, uses, and communicates to others. Lacking any of these dimensions diminishes the student's epistemic position vis-à-vis artificial intelligence.

Being able to ask a question of an AI system or select between the options it offers is not, in and of itself, a sign of autonomy. AI may shape what questions are asked, what concepts are emphasized, and what answers are deemed reasonable. Students therefore need to reflect not only on the output but also on the processes of formulating questions and defining problems. Epistemic autonomy should comprise not only the evaluation of the answer received from artificial intelligence but also the interrogation of what is being asked to the system and why.

In this context, thinking for oneself does not mean refusing to use artificial intelligence. Those who can think for themselves can handle technological support for their own intellectual ends, and view system outputs not as final judgments but as proposals subject to criticism. AI could be used to generate counter-arguments, compare different explanations, identify weaknesses in an argument, or provide feedback on an initial draft. The student, however, should be left with the final evaluation, interpretation and decision-making responsibility.

To maintain epistemic autonomy, students need to be able to judge when to rely on artificial intelligence and when to persist in their own intellectual endeavors. The employment of a system that eliminates all cognitive difficulty may boost short term efficiency, but it may also erode students' capacity for problem solving, patience, learning from errors, and restructuring their own thinking. Education has to strike a conscious balance between the convenience of artificial intelligence and the intellectual growth of students.

Epistemic autonomy should not be seen merely as an individual characteristic. To be critical and responsible, students need clear principles of use, verification criteria, access to reliable sources and fair appeal mechanisms by educational institutions. One cannot expect students to assume full epistemic responsibility in opaque systems and under ambiguous rules. Epistemic autonomy presupposes the development of individual competencies and supportive institutional conditions.

Epistemic autonomy is not a withdrawal from artificial intelligence, but a capacity to critically manage the epistemic support it provides. An epistemically autonomous student evaluates knowledge claims, acknowledges doubt, revises trust upon evidence, justifies final decisions, and assumes responsibility for the information used. Education should aim, among other things, to train students as subjects who can keep their intellectual direction even with the aid of technology, not make them more technology-dependent.

Epistemic authority is the weight a source gives to a person's process of forming beliefs about a particular subject. This weight should not be because of the coercive power of the source, but of its knowledge, experience, methodological competence and reliability in the relevant field. To rely on epistemic authority is not to abandon one's own judgment. In contrast, it means knowing the limits of one's knowledge and turning to a more competent source for justified reasons.

Epistemic authority must be domain-specific, open to criticism, justifiable and accountable if it is to be legitimate. Success in a particular task does not justify treating a person or system as an authority in all fields. The scope of authority should be limited by real expertise and verified conditions of performance. When these lines blur, a source can be given authority it does not deserve and the recipient of information may lose the ability to make critical judgments.

Artificial intelligence systems can project a powerful sense of epistemic authority because they can produce quick, fluid and comprehensive answers. Confident language can be a sign of knowledge, understanding and expertise to students. But there is no necessary connection between linguistic fluency and epistemic reliability. Artificial intelligence can produce accurate and inaccurate information in the same coherent narrative and can generate full responses even when evidence is insufficient.

Artificial intelligence must not therefore be given extensive and unlimited epistemic authority. It can be regarded as an information support system with demonstrated performance on some tasks. But it should not be turned into a general authority without consideration of its domain, currency, error history, data conditions and degree of independent oversight. AI-generated outputs should be subject to human review and independent sources, especially in cases of high-impact decisions and complex normative questions.

Nor is the epistemic authority of teachers absolute or beyond criticism. Teachers may have more knowledge of the subject and more teaching experience than students, but this does not mean that they have an unquestionable dominance over students. The authority of a teacher is legitimized by the obligation to justify, to recognize personal limits, to correct mistakes, and to follow the development of the student. Students should be encouraged to question teachers' views and present alternative evidence.

In the era of artificial intelligence, the epistemic authority of teachers is not going away, but changing. The teacher's main function is not to give all the information directly, but to explain the epistemic value of different sources, to help the students to differentiate between them and to guide them through the verification processes. A teacher's authority should be less the result of his having ready-made answers and more the result of his ability to put the right questions, to make visible the limits of knowledge, and to assist learners in creating their own criteria of evaluation.

Students should not be viewed as passive recipients of information in this hierarchy of authority. Students are not consumers who choose between the teacher and artificial intelligence, but epistemic subjects who compare the claims of different sources and justify their final judgments. Whether it is a teacher or an AI system, whether it is a human or a machine, acceptance of the views expressed by the source should depend on the quality of evidence and reasons it provides, and not on its status.

The healthy balance of epistemic authority in education is not a matter of equal distribution of authority between teachers, students and artificial intelligence. Their roles are different in nature. Pedagogical purpose and responsibility are assumed by teachers; students retain ownership of their learning and their final judgments; artificial intelligence is used for information-processing and support functions. There is nothing in the technological system that can substitute for determining educational aims, moral values, and high-impact decisions about human beings.

By epistemic authority we should understand a justified relation between blind obedience and the rejection of all authority.

Students may turn to teachers, experts, institutions or artificial intelligence, but they should calibrate their trust according to the real competence of the source and the context. Education is not intended to be a means of getting rid of authority, but of making people able to discriminate between right and wrong authority, and to keep their own judgement.

"The idea of human development in the era of artificial intelligence should not be defined by how quickly people adapt to technological change. The effective use of electronic devices is an important competence, but education is not only about producing technically competent users. Education should prepare students to develop the human capacities needed to judge both the purposes and the effects of technology, rather than mould them to fit the demands of technological systems.

The reduction of the ideal of human development to market needs, speed, efficiency and performance criteria creates the risk of valuing human beings only in terms of their economic functions. That artificial intelligence can perform certain cognitive tasks faster does not mean that education should prepare human beings to compete with machines. The purpose of human education is not to produce people who do calculations faster or process more data. It is to produce people who can think about purpose, meaning and values.

In the modern era of artificial intelligence, the educated person should first be able to think critically. Critical thinking is not about rejecting all sources of information or living in permanent suspicion. This means assessing claims' sources, assumptions, evidence, limits and ignored perspectives. "Students should look at the epistemic value of AI-generated answers, without being biased by their formal quality.

The person to be educated should also take epistemic responsibility. People shall be accountable for the veracity of information they use or disseminate or in whose name they appear. The origin of a text being artificial intelligence does not absolve the student of responsibility for the choice to use it. Students need to be transparent about system contributions, check the outputs, and take ownership of the final product.

Moral reasoning is another central element in the ideal of human development. Efficiency and accuracy alone do not suffice to judge an AI system. Values such as privacy, justice, autonomy, discrimination, well-being, and human dignity are to be considered together. Someone who can think morally should not only ask whether an application works but also who benefits, who incurs the risk, and whether there are fairer alternatives.

People who are able to question technology do not see artificial intelligence as some neutral, inevitable force. They know that all technological systems are designed for specific purposes, with specific data, design choices, and social relations. They therefore learn not only how to use a system but also to question what values it privileges, what groups it represents and what behaviours it normalises.

The protection of human dignity must be at the core of the ideal of human development. Students are not just cogs in a system that spits out data, measures performance, and predicts future behavior. They cannot be fully described with numbers, be they indices or measures, in terms of their living conditions, feelings, relationships, purposes or capacity for change. Algorithmic predictions should not be seen as final judgments that determine a student’s future in advance.

Education that protects human dignity would acknowledge students’ right of objection, their personal stories and their readiness to grow. In contrast, when a student is being evaluated automatically, the student should be able to learn the reasons behind it, to contest the data used and to explain personal circumstances. A human-centered education system should not consider that individual as the necessary continuation of their data from the past, but as holding the possibility of moving beyond existing classifications.

The person to be educated must be able to adapt to society without losing the ability to question social and technological structures. Education must transmit common knowledge, values and duties but it must not make students passive perpetuators of current institutions. A good citizen is not just someone who follows the rules of digital systems, but someone who debates whether they are just, and who takes part in their creation.

In the era of artificial intelligence, the ideal of human development should be grounded in the integrity of critical thinking, epistemic responsibility, moral reasoning, technological critique, democratic participation, and human dignity. Such a person does not reject technology or surrender to it. Rather, this person uses technology, but does not give up personal judgment, values and responsibility, and directs technology for human purposes.

The first condition of a human-centered approach to education is to view technology as a means and human development as the end. The value of artificial intelligence in education should be in their support of student learning, autonomy, and well-being, not their novelty or speed. A high performing, human-centered technology cannot be human-centered if it diminishes students' capacity for thought, privacy, or dignity.

Course design needs to be restructured in such a way as to maintain the students' cognitive activity. Providing students with a complete answer from artificial intelligence before they have had the opportunity to generate their own initial thoughts, may skip important steps in the learning process. Educational activities should therefore be developed in a stepwise manner: students build their own solutions, then receive alternatives or feedback from artificial intelligence, and finally compare different sources and justify their final decisions.

"Critical verification should be a common academic skill across all courses. Students should be asked not only to cite sources but to describe what sources and criteria they used to verify AI-generated claims. Verification processes should include confirming that sources exist, investigating the connection between the source and the claim, considering the currency of the information, and considering opposing viewpoints.

Teacher education must be more than the technical use of tools. Beyond knowing how artificial intelligence systems work, teachers must acquire skills in critically analyzing outputs, identifying ethical risks, making discipline-specific decisions about its use, and providing epistemic guidance to students. The teacher's role is not just writing good prompts but knowing when artificial intelligence is enhancing learning and when it's robbing the learning objective.

The teacher-student relation should be maintained and not replaced with technological mediation. Routine feedback, organization of content and generation of examples can be done by artificial intelligence, but the teacher should be responsible for the evaluation of students' emotional states, learning difficulties and personal contexts. We should be able to learn why assessment decisions were made from human teachers and discuss them.

Assessment should not be limited to the end product. The evaluation process should also involve students' initial drafts, sources, central questions addressed to artificial intelligence, verification activities and reasons for accepting or rejecting system recommendations. This allows for both the student work products and the ways in which students access information and control their thinking to be made visible.

Academic integrity is more than simply banning the use of AI altogether, or relying on automated detection systems. Students are to explain what tools they have used for what purpose and at what stage. The disclosure requirements should be clear, understandable and proportionate, with concrete examples showing how different contributions should be reported. Transparency should not be a form of punishment, but a means of making the student's responsibility in the learning process visible.

AI-detection tools should not be treated as definitive proof. The possibility of them making a mistake may cause unfair accusations and damage the confidence between teacher and student. In investigations of academic integrity, students should have the right to protect drafts, sources and production processes. Important decisions should not be made only on the basis of the output of automated systems.

AI use policies should be more than just a list of do's and don'ts. Policies should jointly regulate pedagogical purposes, data security, disclosure requirements, oversight by humans, appeals and institutional responsibilities. Technological tools change rapidly, so policies based on purposes of use and ethical principles will be more sustainable than those based on specific product names.

The policies should be discipline and course learning objective specific. One type of use may be appropriate in one course, but destroy the core competency being assessed in another. Instructors should therefore clearly explain which tasks allow AI use, which tasks restrict it, and how use should be disclosed in each course.

Protection of student data must be a basic requirement of human-centred education. Simply because educational institutions can technically collect all kinds of data doesn't mean they should. Data collection should be limited to clear, necessary and proportionate educational purposes. Students should be informed of what data are collected, the reason for collecting that data and who it is shared with.

Regular evaluation of artificial intelligence systems for their impact on different groups of students. High average accuracy does not mean that a system works fairly across languages, cultures, disabilities or socioeconomic groups. Human-centered institutions should check for system errors and biases affecting diverse student groups and work to overcome inequalities encountered.

Students and teachers should be involved in developing AI policies. Human-centeredness does not mean simply handing users off-the-shelf systems to run or taking users' satisfaction surveys. This involves meaningful participation in decisions around system choice, use purposes, data principles and evaluation criteria. This participation can turn technological governance into a pedagogical democratic experience.

The Epistemic Autonomy-Based Human-Centered Education Model provides a common reference to these recommendations. The principle of critical verification requires the examination of claims to knowledge; the principle of justified trust requires adjusting confidence in artificial intelligence according to evidence; the principle of ethical responsibility requires the evaluation of human consequences of use; the principle of human-centeredness requires technology to remain subordinate to human development; and the principle of pedagogical guidance requires the gradual development of students' independent judgment.

To summarize, this study states that the role of artificial intelligence in education can not be defined only by its technical capacity. Artificial intelligence systems are able to quickly process information, give various explanations and help in learning processes. Education is not just a process of getting information faster or making more outputs. It is also a normative activity which builds up human capacities for thought, interpretation, value formation, relations with others and responsibility.

Artificial intelligence in education in general does not necessarily foster epistemic autonomy. Increased availability of information does not necessarily mean that students' skills to evaluate and justify knowledge do so as well. Ownership of information is diminished when students accept system outputs directly, even when their access to information is increased. So the key task of education is to enable students to think with artificial intelligence, while at the same time preventing them from shifting the responsibility for thinking to the system.

Similarly, high performance of artificial intelligence does not automatically mean it has epistemic authority. Success on particular tasks does not mean a system has human-like understanding, pedagogical sensitivity, or moral responsibility. Artificial intelligence can be used as epistemic support, but only in a limited, task-specific and auditable way. It cannot be the authority that decides educational aims or final decisions about humans.

The role of the teacher does not disappear with the presence of artificial intelligence, it is redefined. Teachers are no longer the only ones who know everything, but they still play a central role in helping students evaluate sources, set learning goals, and take pedagogical responsibility. The authority of the teacher should not be founded on a monopoly of knowledge, but on the ability to justify, to relate and to support independence.

The position of the student must be transformed from a passive recipient of information to an active epistemic subject. Students should not just submit answers from artificial intelligence or edit them. They must verify claims to knowledge, evaluate competing perspectives, explore ethical implications, and take responsibility for their final products. “The success in education should not be judged by how fast students use artificial intelligence, but by how consciously, critically and responsibly they use it.

The ideal of human development in the era of artificial intelligence should not consider human beings to be inferior versions of machines. Education should be for developing the capacity of people to create purpose, meaning, value and responsibility – not to train people to race machines in speed and processing power. Educational aims should focus on critical thinking, epistemic responsibility, moral reasoning, empathy, democratic participation and human dignity.

Human-centered education is not anti-technology. It is the assessment of technology in the name of human development, educational justice and the common good. Artificial intelligence can save time on routine tasks, make information more accessible and support different learning needs. But the efficiency it brings should not rob students of time for thinking, the relationship between teacher and student and human contact in education.

The new balance to be found does not need to consider humans and artificial intelligence as equal actors of the same kind. The human being is the subject who determines purposes, weighs values and takes charge. Artificial intelligence is a technical and epistemic support tool serving pre established purposes. If this distinction is not maintained, education may cease to be an instrument of human development and become a means of adapting the individual to the demands of technological systems.

The responsibility of educational institutions is not just to purchase artificial intelligence and make it available. Institutions must assess the pedagogical need for systems, safeguard data security, train teachers and students, establish clear policies of use, and maintain oversight by humans in high-impact decisions. A human-centered approach cannot be reconciled with the idea that blame is placed only on students and teachers without institutional accountability.

The integrated framework of the Epistemic Autonomy-Based Human-Centered Education Model developed in this study explains the new balance between education, humanity and artificial intelligence. The goal of the model is to maintain students' epistemic position without excluding the artificial intelligence from education, to safeguard teachers' pedagogic guidance and to restrict the technological trust on the basis of evidence, and to put the ethical responsibility in the center of the educational process.

The success of the model should be measured by how much artificial intelligence is used. The key measures of success should be that students can explain why they are using a system, check that the outputs are correct, justify the level of trust they place in the system, judge the ethical consequences and take responsibility for their final decisions. This kind of move takes the use of technology out of the realm of being an accomplishment in itself and puts human judgment back in the forefront.

The central problem for education in the era of artificial intelligence is not how to adapt to technological progress, but how to preserve the human epistemic and moral position amidst that progress. Artificial intelligence can be an important support to education, but it cannot replace human judgment, pedagogical responsibility or human dignity. Education should not reject artificial intelligence, but it should limit it. Benefit from it without turning human beings into a simple tool. Avoid technological efficiency prevailing over human development.

There must be a new balance between education, humanity and artificial intelligence, in an educational order where students can think for themselves, teachers can give responsible guidance and technology is transparent, auditable and subject to human intentions. Such an order can benefit from the opportunities offered by artificial intelligence while maintaining the human capacities for thought, value creation, objection and self-determination. In the era of artificial intelligence, the ultimate aim of education should be to create not those who bow to technology or run away from it, but human beings who comprehend, question, steer and responsibly apply technology for the common good of humanity.

# REFERENCES

Alfiras, M. I. I., Emran, A. Q., & Mohamed, A. M. (2026). Ethics and governance of generative AI in education: A systematic review on responsible adoption. Discover Education, 5, Article 37. https://doi.org/10.1007/s44217-025-01051-y

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. Computers and Education: Artificial Intelligence, 6, Article 100215. https://doi.org/10.1016/j.caeai.2024.100215

Al-Zahrani, A. M. (2024). Unveiling the shadows: Beyond the hype of AI in education. Heliyon, 10(9), Article e30696. https://doi.org/10.1016/j.heliyon.2024.e30696

Anlı, Ö. F. (2025). Kim biliyor?: Yabancı zekâ, epistemolojik yabancılaşma ve bilginin dönüşümü. Kilikya Felsefe Dergisi, 2, 254–286.

Arendt, H. (1961). The crisis in education. In Between past and future: Six exercises in political thought (pp. 173–196). Viking Press.

Balalle, H., & Pannilage, S. (2025). Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity. Social Sciences & Humanities Open, 11, Article 101299. https://doi.org/10.1016/j.ssaho.2025.101299

Baron, S. (2025). Trust, explainability and AI. Philosophy & Technology, 38, Article 4. https://doi.org/10.1007/s13347-024-00837-6

Beer, D. (2017). The social power of algorithms. Information, Communication & Society, 20(1), 1–13. https://doi.org/10.1080/1369118X.2016.1216147

Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., Grundy, S., & Hassan, G. M. (2025). Implementing generative AI (GenAI) in higher education: A systematic review of case studies. Computers and Education: Artificial Intelligence, 9, Article 100407. https://doi.org/10.1016/j.caeai.2025.100407

Benk, M., Kerstan, S., von Wangenheim, F., & Ferrario, A. (2025). Twenty-four years of empirical research on trust in AI: A bibliometric review of trends, overlooked issues, and future directions. AI & Society, 40(4), 2083–2106. https://doi.org/10.1007/s00146-024-02059-y

Biesta, G. (2022). World-centred education: A view for the present. Routledge.

Billingsley, B., Clarke, S., & Selker, T. (2026). EI Search: How to search for knowledge in the age of generative AI. Science & Education, 35, 777–821. https://doi.org/10.1007/s11191-025-00687-4

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. International Journal of Educational Technology in Higher Education, 21, Article 4. https://doi.org/10.1186/s41239-023-00436-z

Boyd, K. (2022). Trusting scientific experts in an online world. Synthese, 200, Article 14. https://doi.org/10.1007/s11229-022-03592-3

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. International Journal of Educational Technology in Higher Education, 20, Article 38. https://doi.org/10.1186/s41239-023-00408-3

Corbin, T. A., & Walton, J. (2025). The missing story of GenAI summarisers: A critical research agenda. Higher Education Research & Development, 44(7), 1655–1668. https://doi.org/10.1080/07294360.2025.2486185

Corbin, T., Dawson, P., Nicola-Richmond, K., & Partridge, H. (2025). "Where's the line? It's an absurd line": Towards a framework for acceptable uses of AI in assessment. Assessment & Evaluation in Higher Education, 50(5), 705–717. https://doi.org/10.1080/02602938.2025.2456207

Costa, C., & Murphy, M. (2025). Generative artificial intelligence in education: (What) are we thinking? Learning, Media and Technology, 1–12. https://doi.org/10.1080/17439884.2025.2518258

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Dearden, R. F. (1972). Autonomy and education. In R. F. Dearden, P. H. Hirst, & R. S. Peters (Eds.), Education and the development of reason (pp. 448–465). Routledge & Kegan Paul.

Dewey, J. (1916). Democracy and education: An introduction to the philosophy of education. Macmillan.

Dormandy, K. (2024). Epistemic self-trust: It's personal. Episteme, 21(1), 34–49. https://doi.org/10.1017/epi.2020.49

Durán, J. M., & Pozzi, G. (2025). Trust and trustworthiness in AI. Philosophy & Technology, 38, Article 16. https://doi.org/10.1007/s13347-025-00843-2

Essien, A. E., Bukoye, O. T., O'Dea, X., & Kremantzis, M. D. (2024). The influence of AI text generators on critical thinking skills in UK business schools. Studies in Higher Education, 49(5), 865–882. https://doi.org/10.1080/03075079.2024.2316881

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, 61(3), 460–474. https://doi.org/10.1080/14703297.2023.2195846

Ferrario, A., Loi, M., & Viganò, E. (2024). Experts or authorities? The strange case of the presumed epistemic superiority of artificial intelligence systems. Minds and Machines, 34, Article 20. https://doi.org/10.1007/s11023-024-09681-1

Freiman, O. (2024). Analysis of beliefs acquired from a conversational AI: Instruments-based beliefs, testimony-based beliefs, and technology-based beliefs. Episteme. Advance online publication. https://doi.org/10.1017/epi.2024.17

Freire, P. (2000). Pedagogy of the oppressed (M. B. Ramos, Trans.; 30th anniversary ed.). Continuum. (Original work published 1970)

Fricker, E. (2021). Can trust work epistemic magic? Philosophical Topics, 49(2), 57–82. https://doi.org/10.5840/philtopics202149215

Fu, Y., & Weng, Z. (2024). Navigating the ethical terrain of AI in education: A systematic review on framing responsible human-centered AI practices. Computers and Education: Artificial Intelligence, 7, Article 100306. https://doi.org/10.1016/j.caeai.2024.100306

Gazit, L. (2026). Constitutive knowledge sources: An institutional approach to epistemic trust in opaque AI systems. AI and Ethics. Advance online publication. https://doi.org/10.1007/s43681-025-00930-2

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), Media technologies: Essays on communication, materiality, and society (pp. 167–194). MIT Press.

Goldman, A. I. (2001). Experts: Which ones should you trust? Philosophy and Phenomenological Research, 63(1), 85–110. https://doi.org/10.1111/j.1933-1592.2001.tb00093.x

Gonsalves, C. (2025). Addressing student non-compliance in AI use declarations: Implications for academic integrity and assessment in higher education. Assessment & Evaluation in Higher Education, 50(4), 592–606. https://doi.org/10.1080/02602938.2024.2415654

Gouseti, A., James, F., Fallin, L., & Burden, K. (2025). The ethics of using AI in K–12 education: A systematic literature review. Technology, Pedagogy and Education, 34(2), 161–182. https://doi.org/10.1080/1475939X.2024.2428601

Gruenhagen, J. H., Sinclair, P. M., Carroll, J.-A., Baker, P. R. A., Wilson, A., & Demant, D. (2024). The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments. Computers and Education: Artificial Intelligence, 7, Article 100273. https://doi.org/10.1016/j.caeai.2024.100273

Hackl, V., Müller, A. E., & Sailer, M. (2026). The AI literacy heptagon: A structured approach to AI literacy in higher education. Computers and Education: Artificial Intelligence, 10, Article 100540. https://doi.org/10.1016/j.caeai.2026.100540

Hähnel, M., & Hauswald, R. (2025). Trust and opacity in artificial intelligence: Mapping the discourse. Philosophy & Technology, 38, Article 115. https://doi.org/10.1007/s13347-025-00947-9

Hao, X., Demir, E., & Eyers, D. (2025). Beyond human-in-the-loop: Sensemaking between artificial intelligence and human intelligence collaboration. *Sustainable Futures, 10*, Article 101152.

Hardwig, J. (1985). Epistemic dependence. The Journal of Philosophy, 82(7), 335–349. https://doi.org/10.2307/2026523

Hauswald, R. (2025). Artificial epistemic authorities. Social Epistemology, 39(6), 716–725. https://doi.org/10.1080/02691728.2025.2449602

Heung, Y. M. E., & Chiu, T. K. F. (2025). Theorizing AI literacy development using Habermas' three cognitive knowledge interests from a systematic review: A STEM interdisciplinary perspective. Computers & Education. Advance online publication.

Hooshyar, D., Šír, G., Yang, Y., Kikas, E., Hämäläinen, R., Kärkkäinen, T., Gašević, D., & Azevedo, R. (2025). Towards responsible AI for education: Hybrid human-AI to confront the elephant in the room. Computers and Education: Artificial Intelligence, 9, Article 100524. https://doi.org/10.1016/j.caeai.2025.100524

Jaramillo, J. J., & Chiappe, A. (2024). The AI-driven classroom: A review of 21st century curriculum trends. Prospects, 54, 645–660. https://doi.org/10.1007/s11125-024-09704-w

Jiang, J., Vetter, M. A., & Lucia, B. (2024). Toward a "more-than-digital" AI literacy: Reimagining agency and authorship in the postdigital era with ChatGPT. Postdigital Science and Education, 6(3), 922–939. https://doi.org/10.1007/s42438-024-00477-1

Joseph, J. (2025). The algorithmic self: How AI is reshaping human identity, introspection, and agency. Frontiers in Psychology, 16, Article 1645795. https://doi.org/10.3389/fpsyg.2025.1645795

Jovchevski, P., Buijsman, S., & Neerincx, M. (2026). What is wrong with automation bias? Philosophy & Technology, 39, Article 84. https://doi.org/10.1007/s13347-026-01090-9

Jun, H., & Lee, H. J. (2026). Anticipatory dignity and digital meritocracy: Philosophical reflections on the algorithmic transformation of South Korea's education system. Studies in Philosophy and Education. Advance online publication. https://doi.org/10.1007/s11217-026-10044-9

Kant, I. (1997). Groundwork of the metaphysics of morals (M. Gregor, Trans.). Cambridge University Press. (Original work published 1785)

Kant, I. (2007). Lectures on pedagogy (R. B. Louden, Trans.). In G. Zöller & R. B. Louden (Eds.), Anthropology, history, and education (pp. 434–485). Cambridge University Press. (Original work published 1803)

Karatsiori, M., Atenas, J., Nerantzi, C., & Abegglen, S. (2026). To be (or not to be) a global citizen in an algorithmic world? A cartographic, ethical and reflexive framework for higher education. International Review of Education. Advance online publication. https://doi.org/10.1007/s11159-025-10183-0

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., 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, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kerasidou, A., & Kerasidou, C. X. (2025). Epistemic authority and medical AI: Epistemological differences and challenges in medical practice. Medicine, Health Care and Philosophy. Advance online publication. https://doi.org/10.1007/s11019-025-10306-2

Kim, J., Lee, S.-S., Detrick, R., Wang, J., & Li, N. (2025). Students–generative AI interaction patterns and its impact on academic writing. Journal of Computing in Higher Education. Advance online publication. https://doi.org/10.1007/s12528-025-09444-6

Kong, S. C., Cheung, M. Y. W., & Tsang, O. (2024). Developing an artificial intelligence literacy framework: Evaluation of a literacy course for senior secondary students using a project-based learning approach. Computers and Education: Artificial

Intelligence, 6, Article 100214. https://doi.org/10.1016/j.caeai.2024.100214

Kuznetsova, E., Makhortykh, M., Vziatysheva, V., Stolze, M., Baghumyan, A., & Urman, A. (2025). In generative AI we trust: Can chatbots effectively verify political information? Journal of Computational Social Science, 8, Article 15. https://doi.org/10.1007/s42001-024-00338-8

Lam, C. M. (2026). Confucian self-cultivation and the ethics of learning with large language models. Studies in Philosophy and Education. Advance online publication. https://doi.org/10.1007/s11217-026-10045-8

Liu, X., Xiao, Y., & Li, D. (2025). Assessing strategic use of artificial intelligence in self-regulated learning: Instrument development and evidence from Chinese university students. International Journal of Educational Technology in Higher Education, 22, Article 69. https://doi.org/10.1186/s41239-025-00567-5

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), Article 410. https://doi.org/10.3390/educsci13040410

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Lu, W., & Hu, Z. (2025). Addressing autonomy risks in generative chatbots with the Socratic method. Science and Engineering Ethics, 31, Article 41. https://doi.org/10.1007/s11948-025-00567-8

Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the "originality" of students' work. Assessment & Evaluation in Higher Education, 49(5), 651–664. https://doi.org/10.1080/02602938.2024.2309963

Ma, M., Ng, D. T. K., Liu, Z., & Wong, G. K. W. (2025). Fostering responsible AI literacy: A systematic review of K–12 AI ethics education. Computers and Education: Artificial Intelligence, 9, Article 100422. https://doi.org/10.1016/j.caeai.2025.100422

Ma, W., & Valton, V. (2024). Toward an ethics of AI belief. Philosophy & Technology, 37, Article 76. https://doi.org/10.1007/s13347-024-00762-8

Madanchian, M., & Taherdoost, H. (2025). Decision-making criteria for AI tools in digital education. Digital Technologies and Applications, 4, Article 100069. https://doi.org/10.1016/j.dte.2025.100069

Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2021). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. Journal of Biomedical Informatics, 113, Article 103655. https://doi.org/10.1016/j.jbi.2020.103655

Nemorin, S. (2024). Towards decolonising the ethics of AI in education. Globalisation, Societies and Education. Advance online publication. https://doi.org/10.1080/14767724.2024.2333821

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Nussbaum, M. C. (2010). Not for profit: Why democracy needs the humanities. Princeton University Press.

Oh, P. S., & Lee, G.-G. (2025). Confronting imminent challenges in humane epistemic agency in science education: An interview with ChatGPT. Science & Education, 34, 779–805. https://doi.org/10.1007/s11191-024-00515-1

Örtegren, A. (2024). Philosophical underpinnings of digital citizenship through a postdigital lens: Implications for teacher educators’ professional digital competence. Education and Information Technologies, 29, 4253–4285. https://doi.org/10.1007/s10639-023-11965-5

Pallant, J. L., & Blijlevens, J. (2025). Mastering knowledge: The impact of generative AI on student learning outcomes. Studies in Higher Education. Advance online publication. https://doi.org/10.1080/03075079.2025.2487570

Park, J. (2025). A systematic literature review of generative artificial intelligence (GenAI) literacy in schools. Computers and Education: Artificial Intelligence, 9, Article 100487. https://doi.org/10.1016/j.caeai.2025.100487

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching & Learning Practice, 20(2). https://doi.org/10.53761/1.20.02.07

Peters, R. S. (1966). Ethics and education. George Allen & Unwin.

Pietarinen, A.-V., & Shi, S. (2026). Manipulation and deception in generative AI-mediated education: Preserving epistemic agency, critical thinking, and creativity. Postdigital Science and Education. Advance online publication. https://doi.org/10.1007/s42438-026-00644-6

Pritchard, D. (2016). Intellectual autonomy and the epistemic value of understanding. In J. Baehr (Ed.), Intellectual virtues and education: Essays in applied virtue epistemology (pp. 112–131). Routledge.

Promsiri, T. (2025). AI and the psychology of educational disruption: Historical patterns and cognitive implications. Acta Psychologica, 260, Article 105637. https://doi.org/10.1016/j.actpsy.2025.105637

Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. TechTrends, 69, 1105–1120. https://doi.org/10.1007/s11528-025-01100-1

Rapanta, C., Bhatt, I., Bozkurt, A., Chubb, L. A., Erb, C., Forsler, I., Gravett, K., Koole, M., Lintner, T., Örtegren, A., Petricini, T., Rodgers, B., Webster, J., Xu, X., Christensen, I. M. F., Dohn, N. B., Christensen, L. L. W., Zeivots, S., & Jandrić, P. (2025). Critical GenAI literacy: Postdigital configurations. Postdigital Science and Education, 7(4), 1296–1333. https://doi.org/10.1007/s42438-025-00573-w

Roe, J., & Perkins, M. (2025). Generative AI in self-directed learning: A thematic scoping review. Interactive Learning Environments. Advance online publication. https://doi.org/10.1080/10494820.2025.2603431

Sahebi, S., & Formosa, P. (2025). The AI-mediated communication dilemma: Epistemic trust, social media, and the challenge of generative artificial intelligence. Synthese, 205, Article 94. https://doi.org/10.1007/s11229-025-04963-2

Salido, A., Syarif, I., Sitepu, M. S., Suparjan, Wana, P. R., Taufika, R., & Melisa, R. (2025). Integrating critical thinking and artificial intelligence in higher education: A bibliometric and systematic review of skills and strategies. Social Sciences & Humanities Open, 12, Article 101924. https://doi.org/10.1016/j.ssaho.2025.101924

Samuel, A. (2026). Learning with machines: Toward a theory of epistemic co-agency. *Computers and Education: Artificial Intelligence, 10*, Article 100573.

Schiff, D. (2022). Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. International Journal of Artificial Intelligence in Education, 32, 527–563. https://doi.org/10.1007/s40593-021-00270-2

Schmidt, E., Putora, P. M., & Fijten, R. (2025). The epistemic cost of opacity: How the use of artificial intelligence undermines the knowledge of medical doctors in high-stakes contexts. Philosophy & Technology, 38, Article 5. https://doi.org/10.1007/s13347-024-00834-9

Schwarting, D., & Reihlen, M. (2026). Self-imposed immaturity in the age of artificial intelligence. Philosophy of Management. Advance online publication. https://doi.org/10.1007/s40926-026-00386-2

Selwyn, N., & Gašević, D. (2020). The datafication of higher education: Discussing the promises and problems. Teaching in Higher Education, 25(4), 527–540. https://doi.org/10.1080/13562517.2019.1689388

Selwyn, N., Pangrazio, L., & Cumbo, B. (2022). Knowing the datafied student: The production of the student subject through school data. British Journal of Educational Studies, 70(3), 345–361. https://doi.org/10.1080/00071005.2021.1925085

Shenoy, P., & Saarela, M. (2026). A critical review and actionable framework for integrating generative AI into teacher professional development. Discover Education, 5, Article 570. https://doi.org/10.1007/s44217-026-01579-7

Shin, D. (2026). Automating epistemology: How AI reconfigures truth, authority, and verification. AI & Society, 41, 1553–1559. https://doi.org/10.1007/s00146-025-02560-y

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe and trustworthy. International Journal of Human–Computer Interaction, 36(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Stewart, C. (2020). Expertise and authority. Episteme, 17(4), 420–437. https://doi.org/10.1017/epi.2018.43

Stolpe, K., & Hallström, J. (2024). Artificial intelligence literacy for technology education. Computers and Education Open, 6, Article 100159. https://doi.org/10.1016/j.caeo.2024.100159

Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. Computers and Education: Artificial Intelligence, 8, Article 100355. https://doi.org/10.1016/j.caeai.2024.100355

Tian, J., & Zhang, R. (2025). Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy. Acta Psychologica, 260, Article 105725. https://doi.org/10.1016/j.actpsy.2025.105725

Tishenina, M. (2026). Generative AI and the irreducible role of teachers as callers into subjectness: A placebo–nocebo–

treatment framework. Studies in Philosophy and Education. Advance online publication. https://doi.org/10.1007/s11217-026-10051-w

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, Article 15. https://doi.org/10.1186/s40561-023-00237-x

Urhahne, D., Wijnia, L., & Steinmayr, R. (2026). The role of epistemic beliefs in predicting ChatGPT adoption and avoidance in higher education. Acta Psychologica, 263, Article 106134.

Uyar, A., Karafil, B., & Karakuyu, A. (2026). Artificial intelligence dependency among educators: A scale development and validation study. Education and Information Technologies, 31, 2369–2398. https://doi.org/10.1007/s10639-025-13862-5

Veldhuis, A., Lo, P. Y., Kenny, S., & Antle, A. N. (2025). Critical artificial intelligence literacy: A scoping review and framework synthesis. International Journal of Child-Computer Interaction, 43, Article 100708. https://doi.org/10.1016/j.ijcci.2024.100708

Vendrell, M., & Johnston, S.-K. (2026). Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education. Computers and Education: Artificial Intelligence, 10, Article 100572. https://doi.org/10.1016/j.caeai.2026.100572

Wang, H., Dang, A., Wu, Z., & Mac, S. (2024). Generative AI in higher education: Seeing ChatGPT through universities' policies, resources, and guidelines. Computers and Education: Artificial Intelligence, 7, Article 100326. https://doi.org/10.1016/j.caeai.2024.100326

Wright, G. W., Diaz, B., & von Gillern, S. (2026). Science educators' attitudes and perspectives on artificial intelligence (SEAP-AI) scale. Journal of Science Education and Technology. Advance online publication. https://doi.org/10.1007/s10956-026-10302-y

Wu, J.-Y., Lee, Y.-H., Chai, C. S., & Tsai, C.-C. (2025). Strengthening human epistemic agency in the symbiotic learning partnership with generative artificial intelligence. Educational Researcher, 54(6), 358–368. https://doi.org/10.3102/0013189X251333628

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. International Journal of Educational Technology in Higher Education, 21, Article 40. https://doi.org/10.1186/s41239-024-00468-z

Yan, W., Liu, Y., Mamaeva, V., Dong, F., Tao, G., Li, R., & Yang, H. (2025). Generative AI literacy: Scale development and its influence on privacy protection behaviors and information verification behaviors. Telecommunications Policy. Advance online publication. https://doi.org/10.1016/j.telpol.2025.103117

Yue, M., Jong, M. S.-Y., & Ng, D. T. K. (2024). Understanding K–12 teachers' technological pedagogical content knowledge readiness and attitudes toward artificial intelligence education. Education and Information Technologies, 29, 19505–19536. https://doi.org/10.1007/s10639-024-12621-2

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

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments, 11, Article 28. https://doi.org/10.1186/s40561-024-00316-

BİDGE