Transformative Learning in the Age of Artificial Intelligence: A Critical Theory Debate

Ted Fleming (ORCID: 0000-0001-5996-6174)
Teachers College, Columbia University, New York, USA

Abstract

This paper examines transformative learning in light of recent developments in artificial intelligence (AI) in education. Drawing on critical theoretical perspectives, particularly the work of Jürgen Habermas and Axel Honneth, it explores how AI might enrich key aspects of transformative learning, including disorienting dilemmas, critical reflection, discourse, and recognition, which remain central to any meaningful integration of AI in education. The paper argues that current approaches to AI in education tend to be instrumental, emphasizing efficiency and performance over meaning-making and emancipatory learning, and raises concerns about distorted discourse, reduced agency, and forms of misrecognition in AI-enriched learning environments. At the same time, it suggests that AI can expand access to information, introduce new perspectives, and support critically reflective inquiry. A model of AI-supported transformative learning is proposed that foregrounds learner agency, critical digital literacy, and ethical dialogue, while also outlining how business and industry, through corporate responsibility plans and charters, can harness AI to foster more democratic and egalitarian workplaces that align technological innovation with workers' interests in democratic citizenship and emancipatory education.

Keywords: Transformative learning; artificial intelligence; critical theory; AI-mediated education; recognition theory

Introduction

Artificial intelligence (AI) has expanded across education and workplace learning, reshaping how knowledge is created, taught, and applied. Generative AI tools are increasingly embedded in classrooms and career and technical education (CTE) programs, enabling new ways of working with learners that may be less personal but more efficient (Holmes et al., 2022; Zawacki-Richter et al., 2019). Governments and institutions are investing in AI to address skills shortages and lifelong learning needs in rapidly changing economies and societies (OECD, 2023). These developments signal growing demands to reconsider how we think, how we relate to others, and how we engage with technology in learning environments.

AI in education, particularly through large language models (LLMs), can be understood as operating through pattern recognition and the probabilistic linking of words and concepts. This contrasts with more critical conceptions of education that emphasize paradigm shifts, imagination, and critical reflection. Much current discourse on AI in education remains focused on technical and instrumental forms of learning. Research and practice frequently prioritize optimization, performance metrics, automation, and efficiency, framing AI as a neutral tool that enhances outcomes (Selwyn, 2019). This instrumental orientation reflects what Habermas (1987) critiques as the dominance of system rationality, where technological and economic imperatives shape educational processes. Such dominance risks marginalizing communicative and emancipatory dimensions of learning, as well as narrative approaches to meaning-making.

Transformative learning theory, grounded in critical reflection and democratic discourse, has received limited attention in AI-focused scholarship (Mezirow, 2000; Cranton & Taylor, 2013). Consequently, AI integration in education may reinforce instrumental rationality while neglecting deeper learning processes related to meaning-making, social justice, and shifts in perspective.

Transformative learning theory (Mezirow, 2009) describes how individuals critically examine and revise the assumptions underlying their meaning-making through reflection, leading to more inclusive understandings of the world (Mezirow, 1991, 2000). Critiques of transformative learning theory have highlighted its overreliance on rationality and its limited attention to social and relational dimensions of learning (Brookfield, 2005; Cranton & Taylor, 2013). Critical theorists such as Habermas and Honneth offer conceptual resources to extend this framework, including communicative action, intersubjectivity, recognition, and democratic participation (Habermas, 1987; Honneth, 1995). While these ideas have informed developments in transformative learning, their implications for AI-mediated learning remain underexplored.

It is therefore notable that both the transformative potential and risks of AI in education have received limited critical attention. AI systems not only deliver content but also shape interactions, participation, and learners' experiences of recognition and agency (Williamson & Eynon, 2020). A critical perspective invites examination of how AI may contribute to the "colonization of the lifeworld" by privileging functional imperatives over dialogue and mutual understanding (Habermas, 1987, p. 325). This raises important questions about whether and how AI can support critical reflection, recognition, and democratic engagement when thoughtfully designed and implemented.

This article reappraises the implications of AI for transformative learning through the lenses of Habermas and Honneth. By bringing critical theory into dialogue with contemporary developments in AI, it explores how learning is being reconfigured in digital and algorithmic contexts, particularly in relation to discourse, critical reflection, recognition, and emancipation. The article makes three contributions: First, it extends transformative learning theory by examining the opportunities and challenges posed by AI. Second, it engages critical theory and AI scholarship to propose an initial conceptual framework for understanding the social implications of AI in education. Third, it situates AI within broader debates about democracy, emancipation, and citizenship, arguing that a living theory of transformative learning must grapple with the complexities of contemporary AI landscapes.

Rethinking Transformative Learning for an AI World

Transformative learning theory (Mezirow, 2000) describes the process by which individuals critically reflect on problematic and previously unquestioned assumptions, often in dialogue with others. It typically begins with a disorienting dilemma—an experience that challenges existing frames of reference and prompts learners to reconsider taken-for-granted assumptions (Taylor, 2017). Through critical reflection, learners examine the validity of assumptions shaped by life experiences and sociocultural influences (Mezirow, 2009). Discourse, in this context, refers to the communicative process through which individuals test and negotiate meaning under

conditions of openness and rational exchange. Together, these elements constitute key stages in transformative learning.

Mezirow's framework, grounded in critical theory, remains essential for understanding the broader implications of transformation. Habermas's theory of communicative action provides a normative model for discourse, emphasizing mutual understanding achieved through the force of the better argument rather than coercion or authority (Habermas, 1984). Transformative learning encompasses both individual cognitive development and social learning processes. Habermas's distinction between the lifeworld and the system is particularly relevant: the lifeworld represents shared meanings and interpretive frameworks, while the system refers to institutional and economic structures governed by instrumental rationality. Tensions emerge when system imperatives colonize the lifeworld, thereby constraining opportunities for critical dialogue and reflection (Habermas, 1987).

Axel Honneth (1995) extends this perspective by emphasizing that individuals are formed through intersubjective relations structured by recognition. He identifies three forms of recognition: self-confidence, rooted in relationships of care; self-respect, grounded in legal and moral rights; and self-esteem, derived from social valuation of individual contributions. These dimensions highlight that learning and transformation are not solely cognitive processes but are also shaped by social relationships and experiences of recognition. Transformative learning can therefore be understood as both personal and social, conditioned by recognition within families, legal systems, and workplaces.

Critics argue that transformative learning theory privileges rational critical reflection while underestimating the roles of emotion, embodiment, and power (Dirkx, 1997; Hoggan, 2016). Although Mezirow acknowledged social change, critics contend that his framework insufficiently addresses structural inequalities, cultural differences, and collective forms of transformation (Brookfield, 2018). In response, scholars have called for more relational and socially grounded approaches that draw on insights from critical theory.

Engagement between transformative learning theory and AI-mediated learning remains limited, and its assumptions about discourse, reflection, and interaction may require reconsideration (Jandrić et al., 2018). For example, the conditions for discourse identified by Habermas (1984)—equality, openness, and freedom from coercion—are more difficult to sustain in environments shaped by algorithms and platform logics. Similarly, recognition may be altered when interactions are mediated by digital systems rather than face-to-face engagement.

In light of these challenges, revisiting transformative learning is both necessary and timely. Habermas's analysis of communicative action and system colonization, alongside Honneth's theory of recognition, offers generative conceptual tools for addressing these limitations and opening new avenues for understanding learning in contemporary contexts. Without abandoning its foundational insights, this reappraisal seeks to extend transformative learning theory so that it remains responsive to the possibilities and challenges presented by AI.

AI is Reshaping Learning

Education is undergoing rapid transformation under the influence of artificial intelligence (AI), reshaping how learning is designed, delivered, and experienced. Advances such as adaptive learning systems and learning analytics have introduced new forms of interaction that extend beyond traditional educational practices. AI provides immediate feedback and offers forms of personalized support (Kasneci et al., 2023). These systems respond dynamically to learner performance and draw on vast datasets to predict outcomes and optimize interventions (Baker & Siemens, 2022). As a result, learning environments are increasingly data-driven and AI-mediated, shifting away from interpersonal models of teaching and learning. Traditionally grounded in dialogue, feedback, and shared experience, learning is now often structured through algorithmic interactions. AI is also embedded in curriculum design, tutoring, assessment, and aspects of academic advising (Luckin et al., 2016; Xia et al., 2024). These developments raise important questions about agency, authorship, and knowledge production in digital learning environments.

This shift introduces new tensions in educational debates. AI enables individualized learning experiences, delivering content and feedback tailored to each learner. However, systems built on standardized models, large datasets, and performance metrics may constrain diversity and creativity in thinking. Personalization itself may be limited by algorithmic structures, raising concerns about power and the extent to which learners are supported in developing new perspectives.

Tensions also emerge between efficiency and meaning-making. AI systems optimize speed and performance, and may restrict access to resources through algorithmic filtering. While they can enhance access and productivity, they may also reduce learning to measurable outcomes, potentially inhibiting critical reflection and transformative learning. Learning risks becoming focused on information processing rather than on deeper experiences involving questioning, discourse, and reflection. These concerns echo longstanding critiques of instrumental approaches to education, where “how” takes precedence over “why.”

Such developments reflect what Habermas (1984) describes as instrumental rationality, emphasizing control and efficiency. Transformative learning, by contrast, often requires time and deliberation, processes that may be at odds with the speed of AI systems. In AI-led environments, algorithms prioritize optimization and performance, raising the further colonization of the lifeworld, where communicative and interpretive dimensions of learning are subordinated to system imperatives (Habermas, 1987). Although AI-mediated environments offer new opportunities, they may also constrain the conditions necessary for critical discourse and reflection (Alagic & Byun, 2026).

AI does not determine educational outcomes; rather, learning is shaped by how technologies are designed, implemented, and assessed. AI holds potential to support critical and reflective learning and to address ethical concerns and learner agency (Birhane, 2021), yet its role in promoting social change remains uncertain. Proponents argue that AI can support both efficiency and critical inquiry, as well as inclusive participation aligned with transformative learning.

Learning with AI therefore presents both opportunities and challenges. While access to learning may expand, questions persist about the nature of learning as an interpersonal and meaning-making activity. Critical theory provides a valuable lens for interrogating these developments and for reimagining learning in ways that balance efficiency with the pursuit of emancipation.

Can AI Join Human Dialogue?

When examining AI in education through Habermas's conceptual lens, a central question emerges: how can AI participate in discourse? For Habermas, discourse involves the redemption of four validity claims—truthfulness, rightfulness, sincerity, and clarity (Habermas, 1984). These claims are typically taken for granted in everyday communication and become visible only when challenged. Among them, sincerity and the moral grounding of rightfulness present particular challenges for AI. Discourse is not merely an exchange of information but a normative process requiring intentionality, reflexivity, and ethical engagement. Whether AI can contribute to such forms of authentic communication remains contested.

First, AI relies on dominant linguistic, social, and cultural patterns that may reproduce conventional assumptions without critical interrogation. Its outputs can shape both the direction and content of dialogue (Bender et al., 2021). For educators, this raises concerns about the integrity of knowledge generated through AI. Habermasian discourse aspires to egalitarian, open, and critical dialogue guided by the force of the better argument. AI-generated contributions risk distorting this process and undermining its emancipatory potential, which is grounded in the intention to challenge falsity, insincerity, and irrationality.

Second, AI lacks consciousness, intentionality, and lived experience. Its probabilistic outputs are based on pattern recognition rather than self-reflection or meaning-making. While AI can simulate dialogue, it cannot sustain the ethical and sincere engagement required for genuine discourse. Interactions may appear dialogic but lack the intersubjective depth central to Habermas's theory of communicative action. For learners, AI may function as a conversational partner, yet it cannot achieve the reciprocity of human understanding. This raises questions about the authenticity of AI-mediated communication (Floridi & Chiriatti, 2020). Such simulations risk being mistaken for genuine discourse, potentially hollowing out the role of human interaction in learning.

Third, the possibility of critical dialogue within AI-mediated environments is uncertain. AI can generate rapid and seemingly authoritative responses, and while Mollick (2024) characterizes AI as a form of co-intelligence, it falls short in supporting critical engagement. Its efficiency may discourage deeper analysis and reflection, limiting exploration of sociocultural assumptions and alternative perspectives. Moreover, AI cannot sustain the extended dialogue necessary to resolve real disagreements in learners' lived contexts. These limitations risk reinforcing passivity and encouraging a consumerist approach to information rather than an active struggle to make meaning. Habermas's recent extension of his work on the public sphere acknowledges similar concerns in the context of evolving media environments (Habermas, 2023).

From a critical theory perspective, these developments reflect the growing influence of system imperatives on the lifeworld. Habermas (1987) distinguishes the system, governed by

instrumental rationality, from the lifeworld, grounded in shared meanings and cultural values. AI, embedded in institutional and economic structures, operates according to system logic. When integrated into educational interactions, it risks privileging efficiency over understanding, thereby colonizing the lifeworld. If discourse becomes governed by algorithmic processes, it risks losing its critical and emancipatory function.

At the same time, some scholars argue that AI can support reflective dialogue when thoughtfully integrated into learning environments. It may prompt questions, introduce alternative perspectives, and facilitate collaborative inquiry (Holmes & Tuomi, 2022). In this sense, AI need not replace discourse but can serve as a tool to scaffold it. The challenge lies in ensuring that AI enhances rather than constrains the conditions for critical engagement.

The key issue, therefore, is not whether AI can participate in discourse, but whether it supports or undermines the discursive practices central to transformative learning. Such learning depends on critical reflection, open dialogue, and the interrogation of assumptions. AI-mediated interactions may reduce opportunities for genuine critique, obscure bias, and prioritize efficiency over meaning, thereby constraining transformative processes. Nonetheless, proponents suggest that AI holds potential to support dialogue and critical inquiry, though this potential remains largely unrealized. Ongoing critical examination is required to determine how AI can sustain the conditions necessary for discourse. With appropriate design and ethical consideration, AI may, in time, play a more meaningful role in supporting learning and transformation.

Recognition and Learning in the Age of AI

Since Rousseau, recognition has been understood as foundational to social life, and Axel Honneth's reinterpretation of Hegel has advanced a contemporary critical theory account of its significance. For Honneth, recognition operates through intersubjective relations of respect that enable individuals to develop self-confidence, self-respect, and self-esteem (Honneth, 1995). In adult education, these ideas have been applied to teacher–student relationships and to issues of access, where recognition is enacted through affirmation, guidance, evaluation, and feedback that shape learners' identities and potential for success. Peer interactions likewise offer opportunities for mutual recognition and the negotiation of meaning (Noddings, 2013). Institutional forms of recognition, such as recognition of prior learning (RPL), further establish the social conditions under which individuals feel valued and capable of engaging in critical reflection and discourse (Noddings, 2013).

In AI-mediated environments, however, recognition is reconfigured. Learners increasingly receive feedback through automated grading, personalized recommendations, and conversational agents that simulate recognition (Zawacki-Richter et al., 2019). These systems offer immediate, consistent, and scalable responses that may enhance support, particularly in resource-constrained contexts. Yet such feedback lacks the emotional depth, contextual sensitivity, and ethical intentionality that characterize interpersonal recognition. While AI can evaluate performance and generate encouragement, it cannot "recognize" in the intersubjective sense described by Honneth; it lacks the capacity for reciprocity and for engaging learners as moral and social subjects.

Learners now construct identities within hybrid environments that combine human and technological interactions. AI systems shape feedback, map learning pathways, and represent learner abilities through data analytics (Williamson & Eynon, 2020). Predictive models categorize learners based on performance metrics, influencing how individuals understand their own capacities and how they are perceived by others. Although AI participates in processes resembling recognition, it differs fundamentally from legal and institutional forms of recognition, which are grounded in shared social values and normative consensus. AI, as a technological system, cannot fulfill these functions.

Misrecognition remains a persistent risk. Standardized responses may fail to capture the complexity of individual experience, producing feedback that is generic, decontextualized, or insensitive to cultural and emotional dimensions of learning (Selwyn, 2021). Learners may experience themselves as unseen, misunderstood, or reduced to data, potentially undermining motivation and self-confidence. Bias embedded in data or algorithms can further produce misrecognition, particularly for disadvantaged or misrepresented groups (Birhane, 2021), with significant implications for individual development.

At the same time, AI-mediated learning may expand certain dimensions of recognition. Consistent and timely feedback can support learners who might otherwise receive limited attention, while multilingual capabilities and adaptive pathways can enhance access for diverse populations (Holmes et al., 2022). These developments also highlight the need to strengthen educators' knowledge and professional capacity in working with AI (Celik, 2023). Properly designed, AI systems may support human interaction by handling routine tasks and allowing educators to focus on relational and pedagogical engagement.

These possibilities depend on maintaining a clear distinction between technical feedback and the normative foundations of recognition. Attempts to reproduce interpersonal care through technology risk reductionism. While AI can be interactive, it is not yet interpersonal in the full sense required for recognition. Rather than replacing human recognition, AI reshapes the conditions under which recognition is experienced. The task, therefore, is not to redefine recognition to fit AI, but to design AI systems that support rather than undermine it. This foregrounds the importance of ethical frameworks emphasizing transparency, responsibility, and critical digital literacy (Kharbach, 2026).

Honneth's recent work further underscores the stakes of this discussion. He reconceptualizes freedom not as individual autonomy alone but as an achievement grounded in recognition across interpersonal relations, legal institutions, and work practices (Honneth, 2023). Freedom is realized when individuals mutually acknowledge each other's capacities, needs, and aspirations. Education, labor, and social care are central to this process. For AI to contribute meaningfully to such conditions, it must engage with these normative demands. Honneth (2024) also emphasizes the role of work in sustaining democratic life, where fair and transparent divisions of labor enable participation, cooperation, and mutual recognition. These developments pose significant challenges for AI, which must respond to these expectations if it is to support democratic and emancipatory social orders.

The emergence of AI-mediated learning environments therefore presents both opportunities and risks. While AI may extend access and support, it may also introduce new forms of misrecognition. Transformative learning and critical theory must engage these challenges while preserving the relational, ethical, and emancipatory dimensions of learning.

Transformative Learning in AI Contexts

As transformative learning theory begins to engage with the possibilities of AI, AI itself must be examined for its potential role in reshaping frames of reference. Transformative learning involves disorienting dilemmas, followed by critical reflection and dialogic discourse that lead to more inclusive perspectives (Mezirow, 2000). In AI-mediated environments, these phases may require reconfiguration as AI reshapes the conditions under which learning occurs.

Disorienting dilemmas may take new forms in the age of AI, emerging from encounters with technologies that challenge assumptions about knowledge, authorship, and expertise. Learners may experience surprise or disorientation when AI-generated outputs exceed expected human performance, prompting reflection on the value of their own skills and the nature of learning itself (Kasneci et al., 2023). Such experiences may destabilize existing frames of reference, raise questions about originality and authority, and foreground the role of human judgment in knowledge production. In this sense, AI and transformative learning may mutually stimulate further development.

Critical reflection, the second phase of transformation, is similarly affected. AI tools can support reflection by offering feedback, generating alternative perspectives, and facilitating access to information. They may prompt learners to question assumptions, consider counterarguments, and articulate ideas more clearly (Mollick, 2024). At the same time, AI may constrain reflection by presenting authoritative or simplified responses that limit deeper inquiry. The speed and fluency of AI-generated content can create an illusion of understanding, masking the cognitive effort required for critical analysis (Selwyn, 2019). As with any tool, the depth of reflection depends less on access to AI than on how it is used within learning processes.

Discourse, central to transformative learning, also shifts in AI-mediated contexts. Traditionally understood as open, reciprocal dialogue among participants, discourse increasingly takes the form of hybrid human–AI interaction. Such interactions may expand opportunities for engagement by providing access to additional conversational partners and resources (Luckin et al., 2016). However, they also raise concerns about truthfulness, sincerity, and reciprocity. While AI can enhance participation and access to information, it may simultaneously weaken the relational dimensions of transformative dialogue.

A model of AI-augmented transformative learning may therefore emerge that supports, rather than replaces, human learning processes. In such a model, AI functions as a tool for exploration, reflection, and dialogue, while learners retain responsibility for interpretation, critique, and application (Feldstein, 2026). Transformative learning in AI contexts would remain grounded in human cognition and interpersonal engagement, supported by transparent and ethically designed algorithmic systems.

For such transformation to occur, several conditions are essential. First, human agency must remain central. Learners need the capacity to critique AI outputs, make independent judgments, and resist passive reliance on automated systems. Without this, AI risks reinforcing instrumental rather than hermeneutic learning. Second, critical digital literacy becomes increasingly important (Kharbach, 2026). Learners must understand how AI operates, including its limitations, biases, and underlying assumptions. AI literacy thus extends beyond technical skills to include ethical and epistemological awareness (Ng et al., 2021). Third, ethical engagement requires that AI systems are designed and used in ways that promote transparency, equity, data privacy, and fairness.

AI can support learners in questioning assumptions and expanding perspectives, yet its uncritical use may undermine reflection and dialogue. The challenge for educators and institutions is to design learning environments that leverage AI's strengths without foreclosing the possibility of transformative learning. In this sense, AI has the potential to move beyond efficiency and become a resource for fostering deeper critical reflection.

Training Workers or Educating Citizens?

The growing influence of artificial intelligence (AI) in education has significant implications for career and technical education (CTE) and workforce learning. As industries undergo rapid transformation driven by automation and digital technologies, they face increasing demands to retrain and upskill workers. In this context, AI serves both as a driver of change and as a tool for addressing emerging skill gaps (World Economic Forum, 2023). CTE programs, traditionally focused on preparing learners for specific occupations, are now expected to equip them for labor markets shaped by automation, data analytics, and digital systems (Carnevale et al., 2020). AI-enhanced learning environments offer flexible, personalized pathways that respond to evolving workplace demands, requiring a rethinking of training approaches to ensure that learning remains relevant, meaningful, equitable, and potentially transformative.

An ethical approach to education must design learning experiences that foster critical engagement with AI and avoid passive reliance on automated systems. While AI tools support skill development and provide real-time feedback, there is a risk that learners may become dependent on them without developing critical awareness. Educators must therefore emphasize not only technical competence but also the capacity to question, interpret, critique, and evaluate AI outputs (Long & Magerko, 2020). This involves integrating critical literacy for "reading" AI within CTE programs, enabling learners to understand how these systems function, recognize their limitations, and engage with them as informed participants. Such approaches align with broader goals of preparing individuals both for employment and for active citizenship in technologically mediated societies.

Maintaining human-centered recognition within AI-enhanced environments remains essential. Recognition plays a vital role in shaping learner identity, motivation, and engagement, particularly in CTE contexts where learners may encounter structural barriers (Wheelahan, 2015). While AI can provide scalable support, it cannot replicate the relational dimensions of teaching grounded in empathy, understanding, and mutual acknowledgment. Educators must

ensure that AI complements rather than replaces human interaction, enabling teachers to focus on mentoring, coaching, and building meaningful relationships with learners (Holmes et al., 2022).

Beyond classroom practice, AI also has significant institutional implications. Educational institutions increasingly adopt AI to improve efficiency, streamline administrative processes, and enhance decision-making through data analytics (Williamson & Eynon, 2020). While these developments expand access and scalability, they risk narrowing educational priorities to measurable outcomes and productivity. From a critical perspective, this reflects a longstanding tension between efficiency and emancipation. If efficiency dominates, broader purposes of education—such as personal development, critical thinking, and social participation—may be marginalized. In a recent interview, the philosopher C. Thi Nguyen affirmed that standardized tests can only ever function as rough measures of even simple forms of performance and should not become ends in themselves (Meyer, 2026). On this point, Nguyen makes the following observations regarding large-scale metrics (quoted in Mayer, 2026):

> One is that when the demand is that they are highly usable at scale and highly public, they systemically eliminate high-expertise and high-skill judgment.
> The other is that, given that they have to be applied in a stable way, they are systematically bad at highly variable phenomena.

Institutions must therefore adopt a more holistic approach to AI integration that balances economic, social, and humanistic goals. This includes developing policies that promote ethical AI use, transparency, and accountability, while supporting pedagogical innovation that prioritizes learner agency and inclusion (Miao & Holmes, 2023). Practical measures include guidelines for responsible AI use in teaching and assessment, protections for data privacy and equity, and professional development to support effective integration. Such efforts help ensure that technology aligns with educational values rather than dictating them.

As AI reshapes work and career pathways, learners require both new technical skills and expanded opportunities for lifelong learning. CTE programs can support this transition by fostering transferable skills such as problem-solving, collaboration, and critical reflection. AI can enhance these opportunities, but only when integrated in ways that support, rather than constrain, learner autonomy.

Ultimately, the implications of AI for pedagogy and CTE extend beyond technical implementation to fundamental questions about the purpose of education. The challenge is to harness AI's potential for workforce development while preserving education's transformative and emancipatory aims. By designing pedagogies that promote critical engagement, sustain human-centered recognition, and balance efficiency with deeper educational values, educators and institutions can contribute to more equitable and meaningful learning futures.

Toward a Critical AI Pedagogy

As AI continues to expand in education, it raises questions that extend beyond technical adoption and call for deeper critical engagement. A critical pedagogical approach, informed by both critical theory and transformative learning, highlights the need to foreground critical AI literacy.

Such an approach addresses the social, ethical, and organizational implications of AI in education and society (Antal & Sobczak, 2004). AI is not neutral; it is embedded within structures of power that shape and reproduce knowledge.

Critical pedagogy brings attention to how technology can reinforce domination by overemphasizing efficiency and control, while transformative learning contributes a focus on critical reflection that challenges assumptions and strengthens human agency (Mezirow, 2000; Feenberg, 2017). Together, these perspectives support the development of critical awareness about knowledge production and access. Critical AI literacy involves understanding how AI operates, including its strengths, limitations, and often-hidden ethical implications. It extends beyond basic AI literacy by enabling learners to interrogate technical foundations, social consequences, and embedded power structures, including their broader environmental and economic effects (Roe et al., 2024). In doing so, it supports a shift from passive consumption of AI outputs to active, critical engagement with AI as a socio-technical system (Ng et al., 2021).

Three guiding principles follow from this perspective. First, AI design should prioritize recognition, ensuring that learners are valued as participants in meaningful learning processes. Second, democratic discourse—characterized by dialogue, collaborative inquiry, and equitable participation—should remain central in AI-supported learning environments. Third, the tendency for AI to reinforce instrumental approaches must be redirected toward supporting ethical reflection and meaningful engagement. In practice, this involves teaching learners to evaluate AI-generated content, use AI to support collaborative inquiry, and engage in ethical exploration. AI must meet these requirements if it is to function as a critical partner in transformative learning.

Beyond educational settings, social responsibility (SR) frameworks within organizations provide an important context for applying these principles. Through environmental, social, and governance (ESG) strategies, organizations articulate responsibilities to workers and communities, including sustainability, equity, and inclusion. These frameworks, often supported by structural adjustments and performance indicators, can guide how AI is implemented in ways consistent with ethical and democratic values.

This paper proposes that AI-enhanced learning and work environments should explicitly incorporate the core elements of recognition. Drawing on Honneth's framework, three priorities emerge: supporting worker well-being through safe and supportive environments; ensuring respect for workers' rights, including equality and legal protections; and fostering solidarity by recognizing workers as members of communities and as citizens. Within this perspective, transformative learning in AI contexts must engage critical questions about sustainability, environmental responsibility, and the dominance of profit, supported by the critical use of AI.

Honneth (2024) further argues that democratic life depends not only on political rights and participation in public discourse but also on how work is organized. Workplaces function as key sites of recognition, where individuals engage in cooperative activity and develop capacities for democratic participation. When structured around fairness, transparency, and respect, work environments can foster the self-confidence and agency necessary for democratic citizenship. Conversely, alienating and disrespectful conditions undermine both recognition and democracy.

In this context, labor is not merely economic activity but a site where freedom and democracy can either wither or thrive.

From a critical theory perspective, AI has the potential to support democratic workplaces when it reduces repetitive labor, enhances cooperation, fosters creativity, and recognizes workers' contributions. However, it also carries risks. Practices such as excessive surveillance, managerial control, and the reduction of workers to data points can undermine participation, solidarity, and recognition. In this sense, AI can either strengthen or weaken the democratic and emancipatory potential of work.

Conclusion

This article has argued for integrating artificial intelligence (AI) into transformative learning theory by rethinking its assumptions in light of contemporary technological developments. Drawing on Habermas and Honneth, it shows that AI reshapes the conditions for reflection, discourse, and recognition, while also introducing risks associated with instrumental rationality. In producing impersonal responses, AI may contribute to forms of misrecognition. At the same time, when thoughtfully designed, it holds potential to expand access, support inquiry, and enhance learning.

The central challenge lies both in the nature of AI as a technological system and in the need to develop critical pedagogical approaches that guide its use. Transformative learning in the age of AI requires sustained attention to human agency, critical digital literacy, and ethically grounded systems that support interpersonal dialogue and mutual recognition. Only through such deliberately critical and recognition-oriented design can AI contribute to genuinely emancipatory and democratic learning futures.

Disclosure: *The author reports there are no competing or conflicting interests to declare*.

Acknowledgement: The author gratefully acknowledges Dr. Viktor Wang for his valuable insights and contributions, which significantly strengthened the development of this manuscript.

Biographical Note

Ted Fleming graduated with a theoretical science degree from Maynooth University in Ireland where he is now Emeritus Professor of Adult Education. He completed his MA and EdD in Adult Education with Jack Mezirow at Teachers College where he is now Adjunct Professor. He received the Jack Mezirow Theory of Transformative Learning Award for 'original contribution to the theory of learning' writing on Axel Honneth and delivered the first Mezirow Memorial Lecture at TC. He serves as the External Advisor to Greek Municipality of Larissa UNESCO funded Learning City project. He is a member of the International Expert Panel advising the Irish

Government's Further Education and Training Authority (SOLAS) on including student voice in adult education services. Research interests include access for adult students to higher education and critical social theory. He is a member of the Editorial Boards of international journals including American *Journal of Transformative Education,* the Polish *Interdisciplinary Studies in Education & Society; Teraźniejszość - Człowiek – Edukacj; International Journal of Adult Education and Technology (IJAET);* and the Greek *Adult Education - Critical Issues*. Recent publications are available at www.tedfleming.net

References

Alagic, M., & Byun, J. (2026). Transformative learning in the age of generative AI: Educators' perspectives on a paradigm shift. In D. Guralnick, M. E. Auer, & A. Poce (Eds.), *Creativity and new technologies in learning for the workplace and higher education (TLIC 2025)* (pp. 29–41). Springer. https://doi.org/10.1007/978-3-032-09905-1_3

Antal, A. B., & Sobczak, A. (2004). Beyond CSR: Organisational learning for global responsibility. *Journal of General Management, 30*(2), 77–98. https://doi.org/10.1177/0306307004030002

Baker, R. S., & Siemens, G. (2022). Learning analytics and educational data mining. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 259–279). Cambridge University Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns, 2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

Brookfield, S. D. (2005). *The power of critical theory: Liberating adult learning and teaching*. Jossey-Bass.

Brookfield, S. D. (2018). *Becoming a critically reflective teacher* (2nd ed.). Jossey-Bass.

Carnevale, A. P., Smith, N., & Strohl, J. (2020). *Recovery: Job growth and education requirements through 2030*. Georgetown University Center on Education and the Workforce. https://cew.georgetown.edu/cew-reports/recovery2020/

Celik, I. (2023). Towards intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior, 138*, 107468. https://doi.org/10.1016/j.chb.2022.107468

Cranton, P., & Taylor, E. W. (2013). A theory in progress? Issues in transformative learning theory. *European Journal for Research on the Education and Learning of Adults, 4*(1), 33–47. https://doi.org/10.3384/rela.2000-7426.rela5000

Dirkx, J. M. (1997). Nurturing soul in adult learning. *New Directions for Adult and Continuing Education, 1997*(74), 79–88. https://doi.org/10.1002/ace.7409

Feenberg, A. (2017). *Technosystem: The social life of reason*. Harvard University Press. http://www.jstor.org/stable/j.ctv24w64m4

Feldstein, A. (2026). The AI paradox: How efficiency is dismantling the relational infrastructure of transformative learning. In L. Declercq, J. Loeckx, M. De Laat, N. B. Dohn, & T. Ryberg (Eds.), *Proceedings of the international conference on networked learning* (Vol. 15). https://doi.org/10.54337/nlc.v15.11092

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines, 30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Habermas, J. (1984). *The theory of communicative action: Vol. 1. Reason and the rationalization of society*. Beacon Press.

Habermas, J. (1987). *The theory of communicative action: Vol. 2. Lifeworld and system*. Beacon Press.

Habermas, J. (1989). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. Polity.

Habermas, J. (2023). *A new structural transformation of the public sphere and deliberative politics*. Polity.

Hoggan, C. D. (2016). Transformative learning as a metatheory: Definition, criteria, and typology. *Adult Education Quarterly, 66*(1), 57–75. https://doi.org/10.1177/0741713615611216

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education, 57*(4), 542–570. https://doi.org/10.1111/ejed.12533

Holmes, W., Bialik, M., & Fadel, C. (2022). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign. https://curriculumredesign.org/wp-content/uploads/AIED-Book-Excerpt-CCR.pdf

Honneth, A. (1995). *The struggle for recognition: The moral grammar of social conflicts*. MIT Press.

Honneth, A. (2023). *The poverty of our freedom: Essays 2012–2019*. Polity.

Honneth, A. (2024). *The working sovereign: Labor and democratic citizenship*. Polity.

Jandrić, P., Knox, J., Besley, T., Ryberg, T., Suoranta, J., & Hayes, S. (2018). Postdigital science and education. *Educational Philosophy and Theory, 50*(10), 893–899. https://doi.org/10.1080/00131857.2018.1454000

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kharbach, M. (2026). Critical AI literacy vs AI literacy: A short guide for teachers, definitions, frameworks, and questions for your class. *Educators Technology*. https://www.educatorstechnology.com/2026/04/critical-ai-literacy-vs-ai-literacy.html

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). ACM. https://doi.org/10.1145/3313831.3376727

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

Mayer, S. J. (2026). Why measuring everything ruins everything. *John Dewey Society Democracy Initiative*. https://jdsdemocracyinitiative.substack.com/p/why-measuring-everything-ruins-everything

Mezirow, J. (1991). *Transformative dimensions of adult learning*. Jossey-Bass.

Mezirow, J. (2000). *Learning as transformation: Critical perspectives on a theory in progress*. Jossey-Bass.

Mezirow, J. (2009). Transformative learning theory. In J. Mezirow & E. Taylor (Eds.), *Transformative learning in practice* (pp. 18–32). Jossey-Bass.

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing. https://doi.org/10.54675/EWZM9535

Mollick, E. (2024). *Co-intelligence: Living and working with AI*. Penguin.

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

Noddings, N. (2013). *Caring: A relational approach to ethics and moral education* (2nd ed.). University of California Press.

OECD. (2023). *AI and the future of skills: Methods for evaluating AI capabilities* (Vol. 2). OECD Publishing. https://doi.org/10.1787/a9fe53cb-en

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity.

Selwyn, N. (2021). *Education and technology: Key issues and debates* (3rd ed.). Bloomsbury.

Taylor, E. W. (2017). Transformative learning theory. In A. Laros, T. Fuhr, & E. W. Taylor (Eds.), *Transformative learning meets Bildung: An international exchange* (pp. 17–29). Sense Publishers.

Wheelahan, L. (2015). Not just skills: What a focus on knowledge means for vocational education. *Journal of Curriculum Studies, 47*(6), 750–762. https://doi.org/10.1080/00220272.2015.1089942

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology, 45*(3), 223–235. https://doi.org/10.1080/17439884.2020.1798997

World Economic Forum. (2023). *The future of jobs report 2023*. https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf

Xia, Q., Weng, X., Ouyang, F., et al. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education, 21*, 40. https://doi.org/10.1186/s41239-024-00468-z

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0