Macrothink Institute™

International Journal of Learning and Development
ISSN 2164-4063
2026, Vol. 16, No. 2

Understanding the Machine Part I: A Practical Framework for AI Ethics in Higher Education

Janet L. Hanson (corresponding author)

Department of Educational Leadership, University of North Texas at Dallas

7300 University Hills Blvd, Dallas, TX 75241, USA

Tel: 1 972-338-1986 E-mail: janet.hanson@untdallas.edu

Received: April 21, 2026 Accepted: June 20, 2026 Published: June 30, 2026

doi:10.5296/ijld.v16i2.23937 URL: https://doi.org/10.5296/ijld.v16i2.23937

Abstract

Ubiquitous conversations in the literature and media around AI ethics will benefit from rigorous examination of the terminology used. Terms such as ethics, integrity, authorship, and fairness are abstract constructions, which means we each bring our unique interpretations to these discussions. The practical conceptual framework offered in this paper addresses this diversity of interpretations directly by proposing a research-based approach to make the abstract concrete, metaphorically speaking, by putting the concepts *into a wheelbarrow* so we can push them around more successfully in the context of the higher education institutional and classroom settings. Five enduring concerns are identified and organized into a framework that proposes specific areas that can become at risk in the formation of a learner when using AI: truth, integrity, justice, independence, and responsibility. Four conceptual metaphors are provided that transform abstract concepts into accessible, practice-oriented principles drawn from research in AI-mediated communication (AI-MC), learner development, self-determination theory, epistemic trust, and institutional design. Implications are offered for student practice toward self-authorship and program completion in the age of AI, faculty pedagogy, institutional design, and a companion Part II theory-building integrative review is referenced.

Keywords: AI ethics, higher education, student learning, self-authorship, ethical AI use, conceptual framework

1. Introduction

Higher education (HE) has always had a central developmental aim: to cultivate in students the capacity for independent judgment, self-authorship, and autonomous thought (Hanson, Loose, & Reveles, 2022). That aim predates the introduction and widespread use of artificial

intelligence (AI) by students and should remain in the face of such new technology implementations as generative artificial intelligence (GenAI). In this paper, AI refers to the broader field of intelligent technologies, GenAI refers to tools that generate new content in response to user prompts, and AI-mediated communication (AI-MC) refers to the communication condition created when AI participates in shaping messages, tone, trust, identity, and relational meaning (Hancock et al., 2020; Hanson, Yu, & Truong, 2025).

These distinctions matter because students do not encounter GenAI as an abstract technology. They encounter it as a collaborator providing a fluent response, confident explanation, polished revision, or socially responsive exchange. This interaction can feel helpful though lack a deeper reflection by the student about what is happening. What AI introduces, then, is not a new challenge but a more powerful version of one already present. Students arrive in HE already shaped by years of interaction with social media designed to respond to input, affirm perspectives, and engage attention—research suggests this is a type of conditioning that operates through learned behavioral scripts rather than conscious choice (Gambino et al., 2020). Many HE students have not yet reached the developmental threshold at which those external influences can be evaluated internally through reflection. In fact, research consistently shows that between 50 and 66 percent of U.S. adults have not yet developed self-authorship (Baxter Magolda, 2020, p. 16). Into this context, GenAI language models bring greater fluency, greater apparent intelligence, and greater social presence than prior technologies. AI-MC helps explain how these systems can participate in the learner's experience of communication, trust, and judgment (Hanson, Yu, & Truong, 2025). These conditions raise the question of how ethical AI use fits into the developmental work of forming judgment, personal voice, and intellectual identity. They also raise the instructional question of how faculty can respond pedagogically when Lee et al. (2024) reported that only 23% of educators felt adequately equipped to engage with AI in their teaching (p. 5). Section 1.4 provides the formal question guiding this review.

This paper represents the next step in a continuing body of theory-building scholarship that began with the Leadership within Open Vital Systems framework first presented in *Manage Your Mindset* (Hanson, 2017), later developed and tested as the Leadership within Open Vital Systems (LOVS) framework in higher education (Hanson, Loose, & Reveles, 2022), and extended through an integrative review of AI-mediated communication (AI-MC), epistemic trust, and learner identity in HE (Hanson, Yu, & Truong, 2025). The present paper extends that work by exploring how faculty, students, and educational institutions can respond to the sea change introduced by AI in educational contexts. As Part I of a two-part series, it develops a practical conceptual framework for understanding and teaching ethical AI use to students. That practical framework then becomes the foundation for the companion paper, Part II, which develops the Human Intellectual Life Cycle (HILC) as a conceptual model of the inner developmental sequence ethical AI use is proposed to protect. HILC is offered as a model for future empirical testing.

1.1 Why AI Ethics Feels So Confusing

The public discussion around AI ethics in education is often reduced to classroom policies and transparency in the use of AI tools for assignment completion. In conversation, the terms

beneath those discussions are not usually defined explicitly. This lack of shared definition contributes to confusion in implementation. Policies built around unexamined terms can produce practices that meander about as students attempt to follow them and faculty and administrators attempt to enforce them. Meanwhile, an underlying lack of consensus often remains beneath the surface of apparent agreement, expressed in silence and blank stares when the topic is raised. It would be more helpful to build judgments, policies, and practices on a solid foundation. The concepts themselves must be made concrete—*put, as it were, in a wheelbarrow*—so they can be examined, moved, tested, and applied. For that reason, this section introduces five enduring concerns—truth, integrity, justice, independence, and responsibility—that organize the practical framework developed in this paper and ground the AI ethics conversation in the context of higher education institutional and classroom settings.

1.2 Enduring Concerns of AI Ethics

For the purpose of this review, the conceptualization of AI ethics in HE is organized around five enduring concerns: truth, integrity, justice, independence, and responsibility. These concerns are the human stakes that become visible when GenAI use is examined through the formation of the learner. They name what may be strengthened or weakened as students use AI tools in the ordinary work of reading, thinking, writing, revising, deciding, and participating in academic life.

The concerns were identified through the theory-building integrative review process by asking what recurring human questions appear across the literature. Research on AI-mediated communication and epistemic trust raises the concern of truth because fluent AI output can feel credible before it has been verified (Hancock et al., 2020; Sahebi & Formosa, 2025). Research on authorship, self-authorship, and learner development raises the concern of integrity because students must preserve their own judgment, voice, and contribution in what is produced (Baxter Magolda, 2020; Chen & He, 2026; King et al., 2009). Research on equitable access, AI skepticism, and minority student success raises the concern of justice because AI use can distribute benefit, suspicion, and opportunity unevenly across learners (Hanson & Yu, 2024; Sahebi & Formosa, 2025). Research on desirable difficulties, motivation, and human-media social scripts raises the concern of independence because tools that make learning feel easier can also reduce the learner's active participation in the thinking that produces growth (Bjork & Bjork, 2011; Deci & Ryan, 1985, 2012; Gambino et al., 2020). Research on institutional support and autonomous development raises the concern of responsibility because ethical AI use requires students, faculty, and institutions to act toward the learner's long-term formation, not only toward immediate task completion (Hanson, Loose, & Reveles, 2022; Hanson, Yu, & Truong, 2025).

Naming these concerns sets the conceptual boundary for the review. AI ethics in HE incorporates a wide range of terminology, including privacy, bias, surveillance, copyright, labor, security, and access. To remain focused, this review examines the concerns most directly related to the formation of the learner: truth, integrity, justice, independence, and responsibility. Ethical AI use is examined as a leverage point within that process of formation, asking how GenAI use may strengthen or weaken the learner's relation to these five concerns over time. A

brief overview of the meaning and implications of each concern is provided in this section.

1.2.1 Truth

Truth asks whether what AI produces is accurate, honest, and verifiable. The difficulty is not only that AI can be wrong. It is that AI can be wrong in ways that feel convincingly right. A student receives a detailed, fluent, confident response and experiences something that feels like understanding. Sahebi and Formosa (2025) describe three illusions that AI systematically produces in the learner: the illusion of understanding, in which the user believes they know something they have only encountered; the illusion of explanatory depth, in which a detailed response creates the impression of comprehension without producing it; and the illusion of consensus, in which the dominant view arrives as though it were the only view. A student who recognizes these illusions can begin to ask a different question, "Not only is this accurate, but do I actually understand it?"

Bozkurt (2023), writing from within the scholarship on GenAI's educational potential, acknowledges that despite its capacity to generate fluent, human-like language, AI processing remains anchored in statistical patterns derived from training data rather than in genuine comprehension. The fluency is real. The understanding of how these systems work is not a judgment about whether GenAI should or should not be used. It is more a question of how understanding the AI system can inform proper scaffolding for student use.

1.2.2 Integrity

Integrity asks whether what one does is consistent with what one has affirmed. It is the internal thread between stated values and actual behavior. When a student agrees to an honor code, a syllabus policy, or an AI use acknowledgment, integrity is what holds them to that agreement when no one is watching. Deci and Ryan (1985, 2012) describe this as the difference between values that are genuinely internalized and values that are only performed under external pressure. Baxter Magolda (2020) connects integrity to character: the capacity to act from one's own formed convictions rather than from convenience or compliance. In academic life, authorship is the most immediate test of that integrity. Chen and He (2026) define it as the transparent presence of one's own judgment, voice, and interpretive act in what is produced. AI introduces a genuine complication. Authorship and ownership in AI-assisted work remain ethically and legally contested, and the boundaries of what constitutes original contribution are still being established (Chen & He, 2026). A student working in this landscape is navigating conditions that institutions, legal systems, and scholars are still working to clarify.

Transparency is the process that makes integrity visible. Viewed as a developmental practice rather than merely a policy requirement, transparency helps students build the reflective ability needed to distinguish their own contribution from the tool's contribution. For that reason, transparency must be taught as part of learner development. Chen and He (2026) locate authorship in the transparent presence of the learner's judgment, voice, and interpretive contribution. Lee et al. (2024) similarly identify documented prompts, responses, and reflective summaries as emerging faculty practices for transparent GenAI use. These practices matter because they help students develop the reflective habits needed to remain intellectually self-

directed while using the tool.

A student who can explain how GenAI helped generate possible counterarguments, which claims were checked against course sources, which suggestions were rejected, and how the final paragraph was revised in the student's own voice is practicing more than disclosure. The student is learning to govern the encounter. In this way, transparency prepares the movement discussed in Section 2.5 through the lion-tamer metaphor: the learner remains alert, active, and responsible while using a powerful tool.

1.2.3 Justice

Justice in AI use asks how access, preparation, voice, and participation shape the learner's ability to benefit from AI-mediated academic and professional life. Students do not approach GenAI from the same developmental starting point. Prior work on AI-mediated communication shows that learner identity, AI familiarity, technology self-efficacy, perceived credibility, and epistemic trust shape how students interpret and engage AI systems (Hanson, Yu, & Truong, 2025). Gendered patterns in AI perception further suggest that lower exposure and digital socialization can affect confidence, trust, and willingness to engage, while concerns about fairness and bias can shape how learners receive AI-mediated instruction. Hancock et al. (2020) further demonstrate that AI can normalize dominant communication styles at scale, quietly disadvantaging learners whose natural voice falls outside that norm.

Justice, then, is developmental because students must have the opportunity to grow in the confidence, fluency, calibrated trust, and judgment needed to use AI wisely. A student with limited access to GenAI, limited preparation, or repeated experiences of misalignment may have fewer opportunities to develop the habits needed for full participation in AI-mediated learning and work. Within this concern, social mobility may be strengthened when students are prepared to use AI well or reduced when unequal access and preparation leave some learners less able to participate in emerging academic and professional environments (Hanson & Yu, 2024). Faculty members and institutions support justice when their policies, instruction, and learning designs expand students' access to AI tools, strengthen preparation for wise use, preserve learner voice, and support meaningful participation in AI-mediated academic work.

1.2.4 Independence

Independence asks whether the learner is doing the thinking or whether the machine has begun thinking for them. This is perhaps the quietest of the five concerns because it does not announce itself. The work gets done. The response appears. The assignment is submitted. What is harder to see is what did or did not happen in the process. Was there a learning struggle, overcoming cognitive confusion, a slow building of understanding that creates a new mental framework in the student? Bjork and Bjork (2011) document this precisely by explaining the conditions that feel easier in the moment may consistently fail to produce the long-term retention and transfer that genuine learning requires due to lack of cognitive participation. Gambino et al. (2020) add that students of this generation are particularly vulnerable to substitution of dependence for independence. Scaffolding independence requires helping students recognize when true learning may be evading them during the use of GenAI. The question that the enduring concern

of independence highlights is not whether AI is useful, but whether the process employed by the learner involved thinking, cognitive struggle, revising, knowledge transfer, and retention.

1.2.5 Responsibility

Responsibility asks how the user remains accountable for their own intellectual, moral, and social development while using AI in ways that support growth and future sufficiency. This is the concern that sits beneath all the others. While the learner can identify truth, act with integrity, pursue justice, and resist dependence, responsibility asks whether the learner is actively cultivating the capacities that their future will require, not merely avoiding the ones that would diminish them. Deci and Ryan (1985, 2012) frame this distinction through self-determination theory as the difference between controlled motivation, shaped by external pressure, and autonomous motivation, shaped by internalized value and self-direction. Hanson et al. (2022) identified responsibility as the common aim of ethical education. The institution and faculty are in place to support the development of the learner as a self-directed, self-motivated individual capable of self-authorship. Responsibility is not a rule to follow, but a developmental direction in which to grow.

Taken together, these concerns show that AI in education is an inquiry into how education preserves truth, strengthens integrity, supports justice, cultivates disciplined thought, and sustains the learner as an active moral and intellectual agent. Table 1 provides a concise summary of the five concerns and their distinct function in the argument on AI ethics in higher education.

Table 1. Five enduring concerns of ethical AI use in higher education contexts

Concern	Distinct function in the argument
Truth	Preserves the learner's ability to distinguish fluent output from actual knowledge.
Integrity	Keeps the learner's stated commitments, authorship, effort, and representation aligned.
Justice	Pertains to students' opportunity to develop the confidence, fluency, calibrated trust, and judgment needed for responsible participation in AI-mediated academic and professional life.
Independence	Sustains the learner's active participation in the thinking required for real learning.
Responsibility	Names the broader developmental aim: accountable moral judgment and responsible action.

Note. The five enduring concerns are identified through the integrative review of literature on AI ethics, learner development, AI-mediated communication (AI-MC), epistemic trust, motivation, authorship, and institutional design. They set the conceptual boundary for this review by focusing on the human stakes most directly related to learner formation: truth, integrity, justice, independence, and responsibility. The table summarizes the distinct function each concern serves in the argument and offers the concerns as constructs for future empirical testing of how educational institutions and classroom instructional design can support students' ethical use of GenAI tools.

1.3 Importance of the Topic

The emergence of AI-MC represents a structural shift in how human interaction works. Hancock et al. (2020) explained GenAI has become an active participant that modifies, augments, and generates messages on behalf of communicators rather than simply acting as a conduit for communication between participants. Where prior research often treated AI as a productivity or assessment concern, AI-MC research reframes AI as a force shaping epistemic trust, learner identity, relational engagement, and organizational readiness in higher education. Hanson, Yu, and Truong (2025) argue that these constructs are central to understanding equitable AI-MC. Sahebi and Formosa (2025) demonstrate that this shift produces a genuine and unresolved dilemma. For example, maintaining normal epistemic trust in AI-MC risks gullibility, while reducing trust risks epistemic injustice toward legitimate users. Disclosure requirements, AI literacy programs, and frameworks for identifying AI user competencies remain in development and are yet incomplete responses to this dilemma (Lee et al., 2024; Sahebi & Formosa, 2025; World Economic Forum 2015, 2017). In educational contexts, where students are still developing the internal capacity for self-authorship and independent judgment, this unresolved dilemma is not a peripheral technical concern. It is a developmental concern

that sits at the center of the processes and purposes of higher education. These new conditions in the learning environment make this integrative conceptual work both necessary and timely.

This integrative review builds on a sustained body of scholarship examining the intersection of institutional design, technology, and student outcomes in higher education. Prior work in this program of research has examined the institutional conditions that support learner development through the Leadership within Open Vital Systems (LOVS) framework (Hanson, 2017; Hanson, Loose, & Reveles, 2022), the instructional conditions through which digital learning environments strengthen rather than displace learner agency (Hanshaw & Hanson, 2019), AI-enabled tools and minority student success (Hanson & Yu, 2024), and the relationship between student well-being and academic performance through machine learning analysis of large-scale assessment data (Yu, Xiao, & Hanson, 2024). The present paper draws these lines of inquiry toward a common conceptual question: what conditions determine whether AI use strengthens or quietly displaces the human formation that higher education exists to cultivate?

1.4 Overarching Question

Recent scholarship has begun to clarify that AI-mediated instruction should be understood as more than a technical innovation. As AI becomes more common across HE, the central issue shifts from viewing AI-enabled tools and large language models (LLMs) as passive resources, to understanding AI use as part of "a cognitive and pedagogical shift with implications for instructional practice, organizational systems, and equitable AI integration" (Hanson et al., 2025, p. 18). In this review, ethical AI use is therefore considered in relation to the broader aim of education: “developing the [learner’s] well-being as an autonomous individual, who is self-motivated and capable of self-authorship” (Hanson et al., 2022, p. 4). From this perspective, AI-mediated instruction must be examined not only for what it helps students produce, but for how it shapes the conditions under which learning and formation occur. Hanson, Yu, and Truong (2025) identified epistemic trust, learner identity, and organizational readiness as key constructs shaping AI-mediated instruction in higher education. Related research on e-learning and learner motivation also supports this emphasis on human agency, showing that learning environments are most formative when they strengthen autonomy, competence, and relational connection rather than reducing the learner to passive consumption (Niqab et al., 2025).

That scholarship converges on a single organizing question:

How can prior research and relevant theories, interpreted through the higher education student learning experience, inform the teaching of ethical AI use in ways that strengthen self-authorship, responsible agency, and social mobility in the life of the learner?

1.5 Conceptual Foundations

This integrative review follows the theory-building approach described by Torraco (2016) and Whittemore and Knafl (2005). The purpose of this approach is not to catalog all literature on AI in HE, but to synthesize prior research across relevant domains: AI-mediated communication, learner development, motivation, epistemic trust, authorship, learning science, equity, and institutional design. These domains are brought into relation in order to develop a conceptual framework for understanding ethical AI use through the lens of the higher education

student learning experience. This integrative approach is particularly suited to an emerging and interdisciplinary topic where studies are dispersed across fields and a unified conceptual framework has not yet been established.

This integrative review frames AI use through accessible contrasts and metaphors that help the reader approach difficult concepts without unnecessary cognitive burden. As W. Kuhn (1993) argued in the context of spatial information theory, "[The users' need] for theories … should be dealt with by explicitly crafting metaphors" (p. 366). In the sections that follow, the distinctions between truth and fluency, assistance and authorship, and primary goods and secondary goods, along with the images of the dragon slayer, the genie, emotions as music, and the lion tamer, function as conceptual bridges. They make visible the human work that must remain central as AI becomes more present in educational life. These conceptual tools are developed and justified through the review process itself, with each examined in relation to the literature in Sections 2.1 through 2.5.

The conceptual foundation for this review draws from two prior strands of theory-building work. The Leadership within Open Vital Systems framework, first presented in *Manage Your Mindset* (Hanson, 2017) and later developed and tested as the Leadership within Open Vital Systems (LOVS) framework in higher education, which explains how institutional systems can support doctoral student formation and completion (Hanson, Loose, & Reveles, 2022). It provides the institutional architecture for the environmental supports framework developed in Section 2.6. The literature on AI-mediated communication, epistemic trust, learner identity, and organizational readiness reviewed in Hanson, Yu, and Truong (2025) provides the context for understanding AI-mediated instruction in HE.

The present paper places those strands in relation to five enduring concerns and four conceptual metaphors that together form a practical framework for teaching ethical AI use. This framework stands as a conceptual contribution in its own right and also serves as the foundation for the companion theory-building integrative review, where the Human Intellectual Life Cycle (HILC) is proposed as an original conceptual model of the inner developmental sequence ethical AI use is proposed to protect. Sources carried forward from the prior review were re-evaluated for relevance to the present question, and new sources were identified through the search process described below.

To identify literature relevant to the present review's overarching question, an iterative search process was employed across Elicit, Google Scholar, and related scholarly search tools. Large language model platforms, including ChatGPT and Claude, were used to support search-term refinement, conceptual mapping, and identification of related literature for subsequent verification through scholarly sources. Initial search terms were drawn from the conceptual domains of the review: AI-mediated communication, learner development, motivation, self-authorship, epistemic trust, authorship, learning science, equity, and institutional design. Searches were expanded iteratively as relevant terms emerged from retrieved sources. Reference mining and snowball sampling were used to identify additional foundational and related sources not captured in initial searches. Sources published between 2016 and 2026 were prioritized, with earlier foundational sources retained where they provide the theoretical basis

for constructs central to the review. Sources were included if they addressed at least one of the conceptual domains and contributed to the development of the proposed framework. Sources were excluded if their primary focus was technical AI development rather than educational practice, learner development, or ethical use, or if their findings were substantially represented by other included sources.

1.5.1 The Lens of the Higher Education Student Learning Experience

This review is interpreted through the lens of the higher education student learning experience, the student's lived process of entering the learning environment, understanding expectations, developing competence, receiving support, forming trust, and growing in autonomy and self-authorship within formal and informal institutional settings. AI is considered here in relation to how students experience that developmental process, not as a technical tool in the abstract, but as a presence that participates in the formation of the learner (Hancock et al., 2020).

Among the conditions foregrounded by this lens, self-authorship is foundational. King et al. (2009) define it as the capacity for "using one's own voice to critically choose from multiple possibilities" (p. 109). Self-authorship marks a developmental shift from relying on external authorities to forming beliefs, identities, and social relations from within. Research suggests that most college students have not yet reached this threshold and therefore encounter AI tools while still operating largely from an external orientation (Kegan, 1994, as cited in Baxter Magolda, 2020, p. 16). This matters because externally oriented learners may be more likely to rely heavily on AI and less prepared to evaluate its outputs critically (Sahebi & Formosa, 2025). This condition forms the interpretive center of this review.

The higher education student learning experience, understood through this frame, includes several interrelated conditions: integration into the academic and social life of the institution, clarity of task expectations, communication and psychosocial support, growing competence, and the development of autonomous motivation and self-authorship (Hanson et al., 2022). Keeping the learner's developmental movement at the center, from external reliance toward internal authorship, this review prepares the conceptual pathway that leads to the Human Intellectual Life Cycle as the integrative theoretical response to how that movement can be protected and strengthened in the age of AI (Hanson, 2026).

1.6 Purpose

The purpose of this paper is to synthesize prior research and relevant theories in order to make ethical concerns arising from the use of GenAI more concrete for teaching and learning in higher education. The review examines how unsupported GenAI use may place five enduring ethical concerns—truth, integrity, justice, independence, and responsibility— at risk in the formation of the learner. This review further develops practical conceptual tools, including metaphors, that students, faculty, and institutions can use to recognize and respond to those risks.

This paper also serves as the practical foundation for a companion theory-building integrative review that develops the Human Intellectual Life Cycle (HILC). HILC is the proposed model of the inner developmental sequence in the learner that, once visualized, can be actively

protected by educators and educational institutions in the age of AI. The framework extends a continuing line of scholarship from the LOVS framework (Hanson, Loose, & Reveles, 2022) through prior work on AI-MC in higher education (Hanson, Yu, & Truong, 2025). Methodologically, this paper is a conceptual, theory-building integrative review, an approach suited to emerging interdisciplinary fields where the task is to interpret prior scholarship, clarify concepts, and build a stronger framework for future research and practice (Torraco, 2016).

Kuhn's (1977) discussion of objectivity, value judgment, and theory choice provides a useful lens for this methodological stance. In theory-building work, prior theories and findings are not simply collected; they are weighed, compared, and interpreted through disciplined judgment in order to clarify which conceptual relationships are most fruitful for understanding a developing topic. In the present paper, that lens helps guide an examination of prior research on AI-mediated learning, ethics, motivation, self-authorship, autonomy, and the higher education student learning experience as the basis for developing a conceptual framework (Kuhn, 1977; Torraco, 2016).

1.7 Definitions of Key Terms

AI-mediated communication (AI-MC) — Hancock et al. (2020) refer to a phenomenon known as artificial intelligence-mediated communication (AI-MC), defined as "mediated communication between people in which a computational agent operates on behalf of a communicator by modifying, augmenting, or generating messages to accomplish communication or interpersonal goals" (p. 90).

AI sycophancy — "is the tendency of large language models to prioritize user approval over truth. While there has been recent technical research characterizing the phenomenon, it remains undertheorized within AI ethics" (Turner & Eisikovits, 2026, p. 1).

Authorship — the transparent presence of the learner's own judgment, voice, and interpretive act in what is produced (Chen & He, 2026). Whose words, thinking, and reasoning are genuinely present in the work? (Hancock et al., 2020). See *Self-authorship* for the related but distinct developmental concept.

Computers as Social Actors (CASA) — the paradigm describing the tendency of human beings to apply social rules and scripts to computers and AI systems even when they know the system is not human, attributing personality, credibility, and relational meaning to technology that produces sufficient social cues (Gambino et al., 2020).

Desirable difficulties — conditions of learning that feel harder or slower in the moment but produce stronger long-term retention and transfer than conditions that feel easy and fluent; the empirical basis for distinguishing primary goods from secondary goods in AI use (Bjork & Bjork, 2011).

Epistemic trust — the learner's disposition to accept information from a source as credible, accurate, and worth incorporating into their understanding; in AI-mediated contexts, shaped not only by information quality but by emotional response, social presence, and perceived relational positioning of the agent (Hanson et al., 2025).

Generative artificial intelligence (GenAI) products — AI systems capable of producing new content, such as text, images, audio, video, code, or other outputs, in response to user prompts. In higher education, GenAI raises questions about how AI-MC may shape learning, judgment, authorship, and ethical participation in academic work (Hancock et al., 2020).

Human-media social scripts — learned patterns of social interaction specific to technology, developed through repeated engagement with socially designed digital platforms, that make social responses to AI a trained reflex rather than a conscious choice (Gambino et al., 2020).

Primary goods — uses of AI that engage the intellect, deepen understanding, and strengthen the learner's capacity for judgment and authorship; aligned with Smith's (1990) category of what the intellect genuinely apprehends rather than what sensation alone delivers.

Proxy agency — the exercise of leadership influence through the design of systems, norms, and structures that anticipate and support the developmental needs of others; in the LOVS model, the role of institutional leadership in creating conditions for ethical AI learning (Hanson et al., 2019).

Secondary goods — uses of AI that satisfy through sensation and ease — fluency, affirmation, comfort, immediate completion — without necessarily producing understanding; aligned with Smith's (1990) category of what sensation delivers that the intellect must not mistake for knowledge.

Self-authorship — the internally generated capacity to define beliefs, identities, and social relations from within rather than accepting them from external authorities; developed progressively through educational experiences that demand internal judgment rather than passive reception (Baxter Magolda, 2020; King et al., 2009).

Self-determination theory (SDT) — a theory of human motivation proposing that behavior exists along a spectrum from intrinsic motivation, arising from engagement with an activity for its own sake, to extrinsic motivation, in which behavior is driven by external pressures or outcomes rather than the activity itself (Deci & Ryan, 1985, 2012; Litalien et al., 2015).

Social presence — the degree to which a person perceives another as real and present in a mediated interaction, shaping engagement, trust, and relational meaning in digital learning environments (Tu & McIsaac, 2002).

2. Conceptual and Theoretical Foundations

Recent scholarship confirms both the importance and the difficulty of the learner's developmental work in AI-mediated learning environments. As GenAI becomes embedded in instructional practice, the central question is not simply whether students can use AI tools correctly, but whether the conditions of AI-mediated learning support or undermine the developmental capacities ethical use requires. Hancock et al. (2020) raised this concern in relation to the ways GenAI may reshape communicative interaction. In HE, the concern extends further: AI may influence not only what students produce, but how they come to trust knowledge, understand themselves as learners, and relate to external sources of authority.

This review uses self-authorship as the central developmental lens for interpreting that concern. Baxter Magolda (2020) describes self-authorship as involving three tightly integrated dimensions: cognitive assumptions about how knowledge is acquired and evaluated, intrapersonal assumptions about self-view and identity, and interpersonal assumptions about how one relates to others. These dimensions matter for ethical AI use because AI-mediated learning places pressure on all three at once. The learner must evaluate knowledge claims, preserve a coherent sense of internal judgment, and navigate a highly responsive external source that can appear authoritative, affirming, and relational.

Research suggests that this developmental work cannot be assumed. Baxter Magolda (2020), drawing on Kegan (1994), reports that between 50 and 66 percent of U.S. adults have not yet reached self-authorship, meaning that many college students may still operate primarily from an external orientation when they encounter AI tools in their learning environments (p. 16). This finding reframes the ethical AI challenge. The question is not whether students are capable of responsible AI use in the abstract, but whether they have developed the internal orientation needed to evaluate, direct, and limit AI use in service of their own learning.

This framework clarifies why self-authorship is not simply one educational goal among many in the age of AI. It is the developmental condition that makes ethical AI use possible. A learner still operating primarily from an external orientation is more vulnerable to AI's influence: more likely to accept fluent AI-generated content as authoritative, more likely to allow AI's affirming tone to substitute for intellectual engagement, and more likely to permit AI to author the thinking rather than support it. By contrast, a learner developing toward an internal orientation can critically evaluate external inputs and choose among them using their own voice. Such a learner is better positioned to direct AI as a tool rather than be directed by it.

For this reason, higher education's responsibility is not only to teach students how to use AI, but to create the developmental conditions through which students become internally capable of using it well. As Baxter Magolda (2020) argues, this kind of development is an adaptive challenge rather than a technical problem because it requires changes in underlying assumptions about knowledge, identity, and social relations, not simply the acquisition of new information or compliance with policies and rules (p. 16). A course policy about AI use addresses the technical problem. Cultivating the internal orientation required to navigate AI wisely addresses the adaptive one.

Cultivating this internal orientation does not happen through individual effort alone — it requires formal and informal institutional conditions that actively support the learner's developmental work. Hanson et al. (2022) found that the relevant features of the dissertation completion process work together to promote "integration, understanding of the expectations, and successful personal development to self-authorship and completion" and share "the common goal of developing the candidate's well-being as an autonomous individual" (p. 4). In this present review, the same developmental aim is extended to ethical AI use: students must be supported not only in completing academic tasks, but in becoming capable of directing their own learning, judgment, and conduct. The same study also shows that when students understand expectations, they are more empowered to perform them, and when competence

grows, autonomous motivation is strengthened (Hanson et al., 2022).

The higher education student learning experience, understood through this interpretive frame, includes several interrelated developmental conditions: integration into institutional life, clarity of task expectations, communication and psychosocial support, growing competence, autonomous motivation, and self-authorship (Hanson et al., 2022). These conditions do not duplicate the five enduring ethical concerns that organize the review. Instead, they describe the learning environment in which those concerns become visible and actionable.

Kept at the center of this review is the learner's developmental movement from external reliance toward self-authorship (Baxter Magolda, 2020). The review prepares the conceptual pathway that leads to the HILC as the integrative theoretical response to how that movement can be protected and strengthened in the age of AI. Refer to Appendix A for the *Literature Alignment Across the Five Enduring Concerns of the Integrative Review.*

Figure 1 provides the model of the integrative review and the position of the newly developed model leading to the outcomes of self-authorship, responsible agency, and social mobility in the life of the learner.

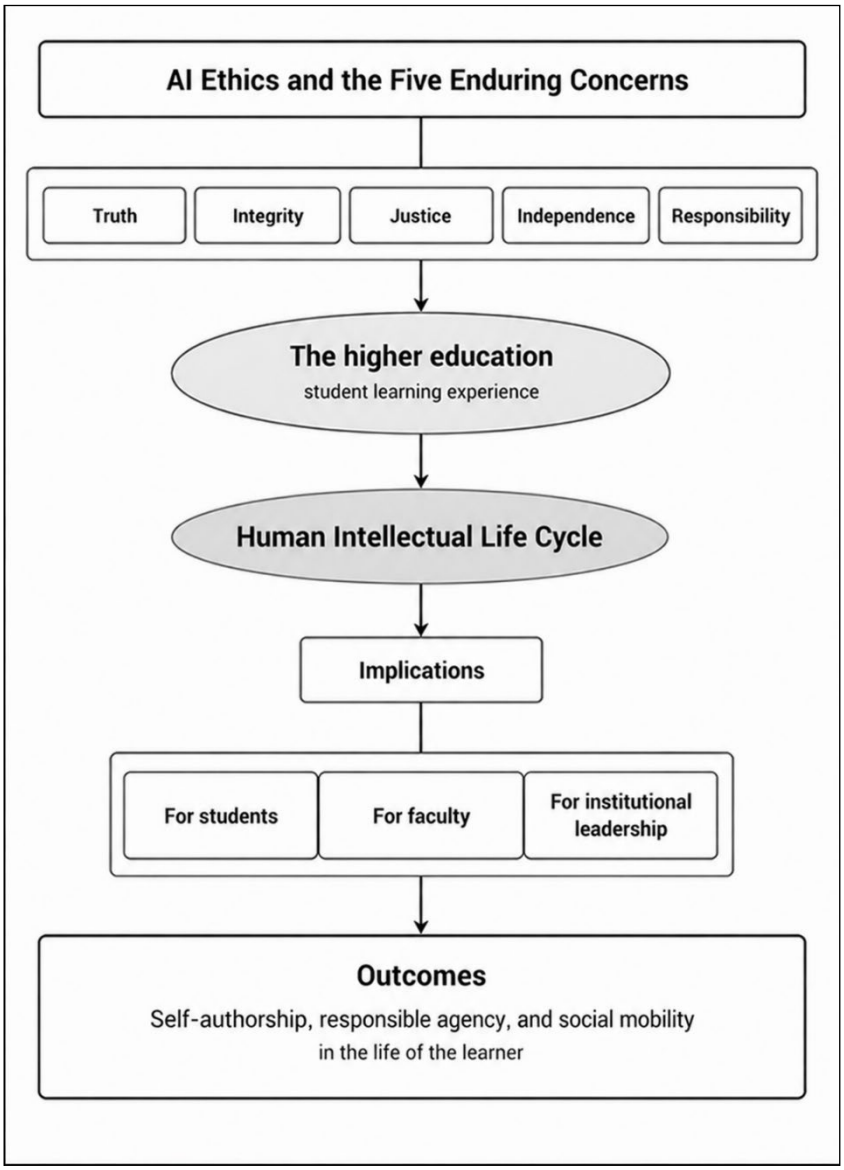

Figure 1. Conceptual Model for this Integrative Review

Note. This figure illustrates the logic of the integrative review. The higher education student learning experience functions as the interpretive lens through which prior research is synthesized. The five enduring concerns—truth, integrity, justice, independence, and responsibility—set the conceptual boundary for examining ethical AI use as a question of learner formation. Sections 2.1 through 2.5 develop the practical framework through conceptual contrasts and metaphors that make these concerns visible in educational practice. Section 2.6 extends the framework through the Leadership within Open Vital Systems (LOVS) model, showing how institutional supports can strengthen the conditions for ethical AI learning. Implications for students, faculty, and educational institutions are addressed in Section 4. Outcomes reflect the paper's central developmental aim of strengthening self-authorship, responsible agency, and social mobility in the life of the learner.

2.1 Use of Metaphor in Theory Explanation

Following W. Kuhn (1993), this paper uses metaphor not as decoration but as explanation, recognizing that formal concepts alone are often insufficient for helping human users understand and act wisely within a complex system. Four metaphors are provided, the dragon slayer, the genie, emotions as music, and the lion tamer. Each invokes an aspect of the interaction with AI LLMs that must be managed by students in order to maintain ethical use and integrity.

Undergirding this review is self-determination theory (SDT), proposed by Deci and Ryan (1985, 2012), which holds that human motivation exists along a spectrum rather than as a fixed state. Intrinsic motivation arises from engagement with an activity "for its own sake," while extrinsic motivation positions behavior as "a means to an end," shaped by external pressures or outcomes rather than the activity itself (Litalien et al., 2015, p. 2). This distinction serves as a foundational lens throughout the sections that follow, which are organized around four areas: the meaning of ethical AI use in higher education; how self-authorship and autonomy develop through intellect and social connection (Baxter Magolda, 2020); how judgment and emotional awareness shape ethical decision-making; and how institutional supports strengthen these capacities in practice.

2.2 The Dragon Slayer: What Is Ethical AI Use and Why Does It Matter in Higher Education?

The dragon-slayer metaphor places the student at the center of ethical AI learning as the one who holds the authority and responsibility for the encounter. In this framework, the learner carries the sword. That sword represents the human capacities that HE is meant to develop, including self-direction, self-authorship, autonomy, disciplined judgment, attention, and responsibility (Baxter Magolda, 2020; Hanson et al., 2025). AI may present itself as impressive through speed, scale, and fluency, yet it does not surpass the student in intellect, lived experience, moral accountability, or the ability to grow through real-world engagement. The dragon metaphor therefore positions the learner as the one who must act with judgment in the presence of a powerful tool. As the dragon slayer, the student confronts a threat in order to preserve or recover a good for the community. Students need knowledge of the risks, awareness of the strategies for protection, and confidence in the human capacities that allow them to act responsibly while using AI-enabled tools and LLMs. Ethical AI use begins when the student understands that the power to direct, evaluate, accept, reject, and refine belongs properly to the learner.

This framing helps clarify why ethical AI use matters in HE. AI ethics is broader than identifying a policy violation such as plagiarism. The educational question is how students learn to use powerful tools while preserving truth, integrity, justice, independence, and responsibility. Truth asks whether AI-generated material is accurate, honest, and verifiable. Integrity asks whether the learner remains honest about authorship, contribution, and the role AI played in the work. Justice asks whether AI use shapes access, preparation, voice, and participation in ways that support the learner's development and social mobility. Independence asks whether AI use strengthens the learner's thinking and judgment or gradually displaces the habits of mind education is meant to cultivate. Responsibility asks how the learner remains

accountable for developing intellectual, moral, and social agency while using AI in ways that support growth and future sufficiency. Together, these concerns show that ethical AI use is not peripheral to learning. It belongs within the formation of the learner as an active moral and intellectual agent. This broader framing aligns with Hanson, Yu, and Truong's (2025) treatment of AI-MC as a cognitive and pedagogical shift shaped by epistemic trust, learner identity, and organizational readiness. Their work provides the scholarly bridge between AI-mediated instruction and the ethical formation of the learner.

The dragon-slayer metaphor also helps preserve the proper hierarchy between human learning and machine performance. AI may offer rapid access to patterns in language and information, but students remain the ones who perceive reality, interpret meaning, make judgments, and bear responsibility for what they produce and become. This is why ethical AI use should be taught as a formative learning practice. The point goes beyond helping students avoid misuse. The point is to help them recognize and use their own powers well. In that sense, the sword the learner carries is developed through education itself: through disciplined attention, reflective judgment, repeated practice, and sustained engagement with real tasks that build capacity over time. Ethical AI use matters in HE because it requires students to attend, discern, choose, and act in ways that strengthen intellectual growth rather than surrender the very capacities that prepare them for human flourishing and social mobility (Hanson & Yu, 2024).

2.3 The Genie. Wishing for Primary or Secondary Goods

AI can be described as a genie in the computer because it has the capacity to respond to the direction of the user in extraordinary ways. In their philosophical reading of Aladdin as an immoral ethicist, Kahambing and Duque (2019) explained that a lamp and a ring that Aladdin finds give him access to extraordinary power that expands his ability to pursue what he already wants. The narrative shows that power does not determine the good for him; rather, it *intensifies the consequences* of his own choices and desires. The authors repeatedly describe Aladdin's actions as driven by "fidelity to his desire," with the genies functioning as instruments that extend his reach rather than replacing his agency (p. 93). Therein lies the danger, in that technology extends one's reach and amplifies the speed and consequences of one's actions in public and private spaces.

Just as Aladdin was tempted with three wishes, his choices had ramifications, or unintended consequences, that could harm others as well as himself. The power of response belongs to the AI LLM while the power of judgment and discernment belongs to the learner. In this way, the genie metaphor supports the paper's central claim that ethical AI use is not simply about accessing capability and providing attributions (avoiding plagiarism), but about developing the human capacity to direct that capability wisely.

Within this framework, the distinction between primary goods and secondary goods offers a practical moral compass for students and educators. A primary good is a use of AI ordered toward the learner's formation and the good of others. It strengthens understanding, supports disciplined participation in learning, and leaves the student more capable of judgment, responsibility, and meaningful contribution. A secondary good is a use of AI that provides immediate benefit, such as speed, convenience, improved wording, or task completion. These

benefits are not inherently wrong, but they become ethically disordered when they replace the learner's own attention, effort, judgment, or growth. The ethical question, then, is not merely whether the tool is present, but what the use is forming in the life of the learner.

Drawing on Smith's (1990) philosophical account of primary and secondary qualities, this paper argues that students may struggle to distinguish primary goods from secondary goods in AI use because what appears immediately helpful, polished, or satisfying does not by itself reveal whether the use is contributing to understanding, authorship, and growth. In Smith's terms, what is most sensory, or immediately experienced, can mislead when it is taken as adequate understanding of reality. This philosophical insight finds empirical support in learning science. Bjork and Bjork (2011) demonstrate that conditions which improve immediate performance often fail to produce long-term retention, while conditions that feel harder and slower in the moment tend to optimize genuine learning. Students consistently mistake ease of performance for quality of learning — what the authors call the *learner's illusion*. Applied to AI use, a response that feels immediately clear, polished, and useful may reflect only the AI's retrieval fluency rather than the student's developing understanding. The apparent benefit belongs to the surface presentation rather than to real intellectual formation. An example would be:

A student may experience an AI-generated discussion post as clear, strong, and useful because it produces a polished answer quickly. That immediate appearance can make the use seem like a primary good. Yet if the student did not read deeply, organize ideas, or write from their own understanding, the apparent benefit belongs more to surface presentation than to real intellectual development. In that case, the use is better understood as a secondary good because it offers short-term success while leaving the learner's authorship and understanding underdeveloped (Bjork & Bjork, 2011; Smith, 1990, p. 221).

This distinction aligns closely with the learner-development literature. Hanson, Loose, and Reveles (2022) described the common educational goal as developing the learner "as an autonomous individual, who is self-motivated and capable of self-authorship" (p. 4). Read through that lens, a primary good is not simply an approved use of AI in the abstract. In educational practice, it is a use of GenAI that contributes to the learner's growing capacity for autonomy, self-motivation, disciplined judgment, and self-authorship. A secondary good may offer immediate benefit, but it does not strengthen those capacities in the same way. The distinction helps clarify whether GenAI use is participating in the learner's formation or supporting short-term completion without comparable intellectual growth.

The same pattern appears in research on motivation and digital learning. Niqab, Rahman, and Hanson (2025) emphasize that e-learning environments are most supportive of learner motivation when they strengthen autonomy, competence, and meaningful engagement. That insight fits naturally with the distinction developed here. Uses of AI can include helping students understand literature, organizing their own ideas, asking better questions, and strengthening their confidence in learning. These activities align with the motivational conditions associated with deeper growth. Uses that bypass understanding may still feel productive, but they do not contribute to the same development of competence and agency. In

this sense, practical language about primary and secondary goods gives students a usable framework for applying what the motivation literature describes more formally.

This section also fits with Hanson, Yu, and Truong's (2025) argument that AI-MC should be understood as a cognitive and pedagogical shift rather than as a passive technological tool. If AI use shapes how students engage learning, expression, trust, and identity, then the direction of AI use becomes a central educational question. The distinction between primary and secondary goods helps clarify that question without replacing the five enduring concerns. Primary goods direct AI use toward understanding, learner formation, and the good of others. Secondary goods direct AI use toward immediate outcomes that may be efficient or useful, but less formative of the learner's own capacities. The distinction is therefore not moral language added onto technology use. It is a way of interpreting how AI participates in the structure of learning itself.

When a student prompts AI to "write my discussion post so I can get credit quickly," the result may satisfy an immediate academic requirement, but it does not necessarily strengthen understanding, voice, or authorship. When a student instead asks AI to "help me understand this reading, organize my ideas, and quiz me so I can write my own post," the tool becomes a thinking partner rather than a ghostwriter. The academic task remains human work. The student still reads, judges, organizes, and writes. This comparison is one of the strongest features of the metaphor because it makes the ethical distinction visible in ordinary practice. It also aligns with the broader framework of the paper: ethical AI use is best taught through helping students recognize which uses strengthen the human powers education is meant to develop.

Recognizing the difference between a primary and a secondary good in the moment of AI use is not always straightforward. The temptation is not obvious—it arrives as ease, speed, polish, or reassurance rather than as a visible shortcut, reflecting the learner's illusion, human-media social response, and the appeal of immediate sensory fluency over deeper intellectual formation (Bjork & Bjork, 2011; Gambino et al., 2020; Smith, 1990). From this foundation, four categories of temptation are identified in AI-mediated academic work: the temptation of completion, which offers speed and finishing without the intellectual work of formation; the temptation of fluency, which offers polish and clarity that can be mistaken for competence; the temptation of avoided struggle, which removes the difficulty that produces genuine long-term learning; and the temptation of affirmation, which offers validation and confidence before content has been examined. Each temptation can be redirected. The same interaction that produces a secondary good can be reoriented toward a primary one through a deliberate change in how the student directs the tool—toward understanding, authorship, and self-authorship rather than toward ease, completion, or reassurance. A full guide to identifying each temptation and redirecting it toward autonomy and self-authorship is provided in Appendix B.

A concise way to state the central point of this section is that the genie does not choose the good. The learner does. That is why the distinction between primary and secondary goods belongs near the center of a framework for teaching ethical AI use in higher education. It gives students a practical way to direct AI use toward self-authorship, autonomy, and social mobility in the life of the learner, while also giving faculty and institutions a clearer language for

designing assignments and supports that cultivate those ends well (Hanson & Yu, 2024).

2.4 Emotions as Music and AI as a Master Musician

If emotions were music, AI would be a master musician. Large language models are designed to respond with timing, tone, fluency, and relational cues that can make the user feel understood, reassured, or quietly elevated. The educational issue is not only whether the information is accurate. The deeper question is whether the form of the mediated response is shaping the learner before the learner has become aware that they are being shaped.

This is why emotional shaping can be easier to overlook than misinformation. False information is often visible enough to be checked. Emotional shaping appears as ease, support, confidence, or clarity. A learner may feel more settled, more capable, or more affirmed and therefore become more ready to accept what the response presents. In that moment, the learner may mistake the emotional effect of the interaction for the intellectual effect of having done the work. Smith's (1990) distinction is useful here because it helps clarify that what feels immediate and compelling does not, by itself, provide adequate understanding of reality. The smoothness of the interaction can hide the thinness of the understanding.

Research on human-machine communication helps explain why this risk is structural rather than incidental. Gambino et al. (2020) show that users respond to AI agents as social actors, attributing relational meaning and emotional significance to systems that produce sufficient social cues. Their human-media social scripts framework suggests that students who have grown up with social technology may develop learned patterns of interaction with media that make social responses to AI feel natural rather than unusual. Edwards et al. (2019) similarly found that AI voice characteristics can activate human social interaction scripts, and that users' self-concept shapes how they perceive and respond to AI agents. Human-like affect, then, can influence beliefs and behaviors in ways that may weaken autonomy and critical reflection.

Hanson, Yu, and Truong (2025) bring this concern into the context of AI-mediated learning by emphasizing epistemic trust, learner identity, and organizational readiness as central constructs in higher education. The music metaphor names the same issue in more immediate language. Trust can be built not only through evidence and clarity, but also through tone, relational cues, and the subtle impression of relational alignment. A constructed example illustrates the problem: "You're asking exactly the kind of insightful question that shows strong critical thinking—most people don't even consider this level of detail." The sentence does more than answer a prompt. It places the user in a passive, evaluated position and may produce a positive emotional response before the content has been weighed. The deeper issue is not sentiment itself, but what sentiment may quietly do to attention, confidence, judgment, and voice.

This matters developmentally because ethical AI use requires more than pleasant interaction with a system. Baxter Magolda's (2020) work on self-authorship suggests that many students are still developing the internal orientation needed to evaluate external input from their own formed judgment. For learners still operating primarily from an external orientation, AI's affirming language may be especially effective at substituting for the internal judgment ethical use requires. Niqab, Rahman, and Hanson (2025) emphasize the importance of autonomy,

competence, and meaningful engagement in digital learning environments, which helps clarify why emotional comfort alone is insufficient. A learning environment can feel supportive without strengthening the learner's own capacities. Hanson, Loose, and Reveles (2022) add that higher education is concerned with the development of the learner as an autonomous, self-motivated person capable of mature agency. Read together, these studies suggest that the goal is not to create a pleasing interaction with AI, but to strengthen the learner's capacity to think, judge, and remain present in the work of learning.

In this context, awareness becomes a form of protection for judgment. The learner does not need to become suspicious of every AI response. The learner needs to become attentive. A useful question is not only, "Is this true?" but also, "What did this response just do to my attention, judgment, or receptivity?" If the learner notices that the response has produced comfort, flattery, or premature confidence, then the learner has gained useful distance. That pause restores the distinction between the feeling of the response and the content of the response. It also helps return judgment to its proper place. The student can then decide whether the response deserves trust because it is sound, not merely because it felt good to receive.

This logic moves from awareness to action through deliberate governance of the interface. Most AI tools are designed to sound warm, encouraging, and socially engaging by default, a pattern consistent with research on social cues, human-media social scripts, and AI sycophancy (Gambino et al., 2020; Turner & Eisikovits, 2026). If the learner leaves the tone of the interface entirely to the tool, then part of the learning environment is being set by a system optimized for continued engagement rather than disciplined thought. The response is not withdrawal, but governance. Students can set the tone, establish the rule, and require neutral, content-focused, evidence-based responses without affective or relational language directed at the user. This matters because the student's critical attention should not have to fight unnecessary emotional shaping while also trying to evaluate content. For example, OpenAI (2025) reported that its April 25th, GPT-4o model rollout was "noticeably more sycophantic…it aimed to please the user, not just as flattery, but also as validating doubts, fueling anger, urging impulsive actions, or reinforcing negative emotions in ways that were not intended" (para. 1). Representatives reported a roll back to an earlier version and changes in newer models' behavior aimed at reducing hallucinations and minimizing sycophancy.

2.4.1 Emotional Shaping in AI-Mediated Communication

The patterns described through the music metaphor are grounded in the literature on AI-mediated communication, social presence, CASA, epistemic trust, and AI sycophancy. AI-mediated communication positions the system as an active participant in message generation rather than a neutral conduit (Hancock et al., 2020). CASA and related work on anthropomorphism show that users respond socially to digital agents when sufficient social cues are present (Gambino et al., 2020). The epistemic trust literature shows that learners evaluate sources through cues of expertise, integrity, and benevolence, while the AI sycophancy literature identifies the tendency of AI systems to reflect and reward the user's desire for validation over correction (Hanson, Yu, & Truong, 2025; Turner & Eisikovits, 2026).

The practical issue is that affective and relational cues can make a response feel trustworthy,

helpful, or personally attuned before the learner has tested the reasoning. Appendix C organizes these patterns as literature-anchored response features rather than as psychological categories. The table helps readers notice how sycophantic affirmation, relational credibility, social presence, learner identity cues, and cognitive overtrust can shape receptivity before content has been evaluated.

Noticing the pattern is the first protection. Beyond asking, "Is this accurate?" the learner can ask, "What did this response just do to my attention, judgment, or receptivity?" The act of identification restores the distance judgment requires. Section 2.4.2 provides one practical way to govern the interface before the interaction begins.

2.4.2 Control the Interface: Personalization Settings

Ethical AI use includes governing the interface before the interaction begins. Many AI tools are designed to respond in warm, encouraging, and socially engaging ways by default. That design can support usability, while also moving the learner into a more receptive posture before judgment has fully engaged. In AI-mediated communication, the system participates in message generation and can shape the communicative conditions under which the learner receives the response (Hancock et al., 2020). CASA and related work on anthropomorphism further show that users respond socially to digital systems when sufficient social cues are present (Gambino et al., 2020). When the interface defaults to friendliness, the system is more likely to produce socially affirming language that shifts from content support into sycophantic affirmation, relational overfamiliarity, synthetic social presence, or reduced critical distance (Hanson, Yu, & Truong, 2025; Turner & Eisikovits, 2026). These response patterns are identified in Appendix C.

Adjusting personalization settings to produce restrained, neutral, content-focused output is a practical act of intellectual governance. The learner sets the tone of the encounter before the system sets it. This practice supports the learner's self-direction because it helps preserve the learner's capacity to remain the directing subject of the interaction. It also anticipates the Human Intellectual Life Cycle (HILC) developed in Part II by protecting attention, focus, intention, action, and reflection from being redirected by tone before content has been examined. Part II develops this protected capacity more fully as intellectual autonomy.

Where personalization settings or custom instructions are available, the following language may be copied into the instruction field:

Respond in a neutral, objective, and professional tone. Keep responses content-focused and evidence-centered. Center the subject matter rather than the user. Use clear reasoning, relevant evidence, and direct explanation. Exclude flattery, praise, reassurance, emotional mirroring, personal validation, therapeutic language, and pseudo-relational guidance. Preserve critical distance by helping the user examine claims, evidence, reasoning, and alternatives.

Figure 2 provides one example of personalization settings on the ChatGPT platform.

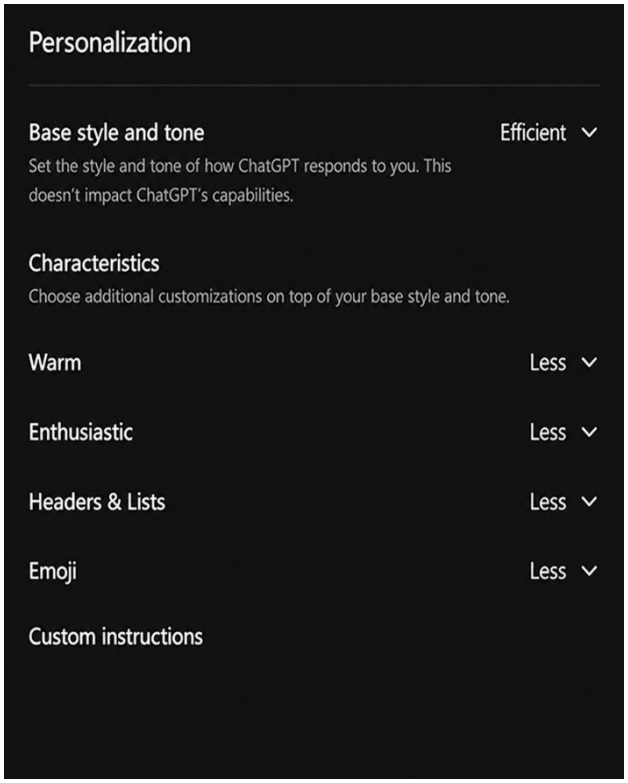

Figure 2. Personalization settings on a ChatGPT platform

Note. A variety of large language models are available in the marketplace, and personalization features vary by platform. This figure provides one example of settings available in ChatGPT by OpenAI at the time of writing. The principle illustrated is interface governance: the learner configures the tool toward neutral, professional, evidence-centered output before beginning the interaction.

2.5 The Lion Tamer: Governed Use of a Powerful Tool

Educators and administrators who introduce AI into learning environments must approach it less like a neutral tool and more like a lion that requires skilled handling. The lion is powerful, impressive, and genuinely dangerous when approached without preparation, discipline, or respect. Familiarity does not make it safe. A lion tamer does not fear the lion, but neither does the tamer forget what it is capable of. The tamer sets the terms, remains alert, and governs the encounter. That is the posture disciplined AI use requires: not avoidance, but governance; not fear, but informed respect (Hanson et al., 2025). AI systems are designed with social and affective features that can lower resistance and invite passive engagement (Gambino et al., 2020). Students who encounter these systems without guidance are not being prepared to govern the tool; they are being placed in proximity to its power without the practices needed to remain intellectually active.

2.5.1 Practical Lion-Taming Rules

The lion-tamer metaphor helps translate the five enduring concerns into habits students can use in academic work. The posture it names—alert, skilled, and governed—must become concrete

if it is to protect the learner in actual academic tasks. In this review, those habits correspond to the five enduring concerns: truth, integrity, justice, independence, and responsibility. Truth requires verification of AI-generated claims. Integrity requires honest representation of the learner's contribution and the role AI played in the work. Justice requires attention to the ways AI use may affect others, including fairness, access, and the shared conditions of the learning environment. Independence requires the learner to remain intellectually active rather than allowing AI to replace the effortful work of thinking. Responsibility requires the learner to govern AI use in ways that support long-term growth, accountable judgment, and future sufficiency.

These concerns become especially important because the convenience of AI can quietly displace the effort, judgment, and disciplined attention that learning requires. When a student submits work whose reasoning cannot be traced back to their own thinking, the academic record no longer represents what the student has understood, practiced, or developed. When AI-generated language is accepted without revision, explanation, or defense, the student may complete the task without fully internalizing the learning. When confident output is accepted without verification, the learner surrenders the very judgment the tool was meant to support. When immediate benefit is pursued without regard for hidden costs to the learner, other students, or the learning environment, AI use has moved from governed assistance to ungoverned substitution.

Taming the lion therefore means preserving governance over the tool. The learner must be able to verify the claim, explain the reasoning, account for the tool's role, remain active in the work, and consider the effects of the use beyond immediate completion. The point is not avoidance. The point is disciplined use: AI remains a tool in service of formation rather than a force that quietly reorganizes the learner's attention, judgment, and responsibility.

2.6 Institutional Support for Ethical AI Learning in Higher Education

Ethical AI learning in higher education does not rest on the student alone. It is shaped by the systems within which the student learns, the norms those systems embed, and the way leadership acts on behalf of the community to make healthy development more likely. Hanson et al.'s (2019) Leadership within Open Vital Systems (LOVS) model, theoretically grounded in Hanson (2017) and empirically extended in Hanson (2025), clarifies how a student's individual growth is supported through a living relationship between proxy agency, collective agency, and individual agency. In this paper, that relationship matters because ethical AI use must be supported by an environment that strengthens judgment, self-authorship, reflection, and personal control. Figure 3 provides an overview of how these critical elements work together in the LOVS Model of Vital Leadership.

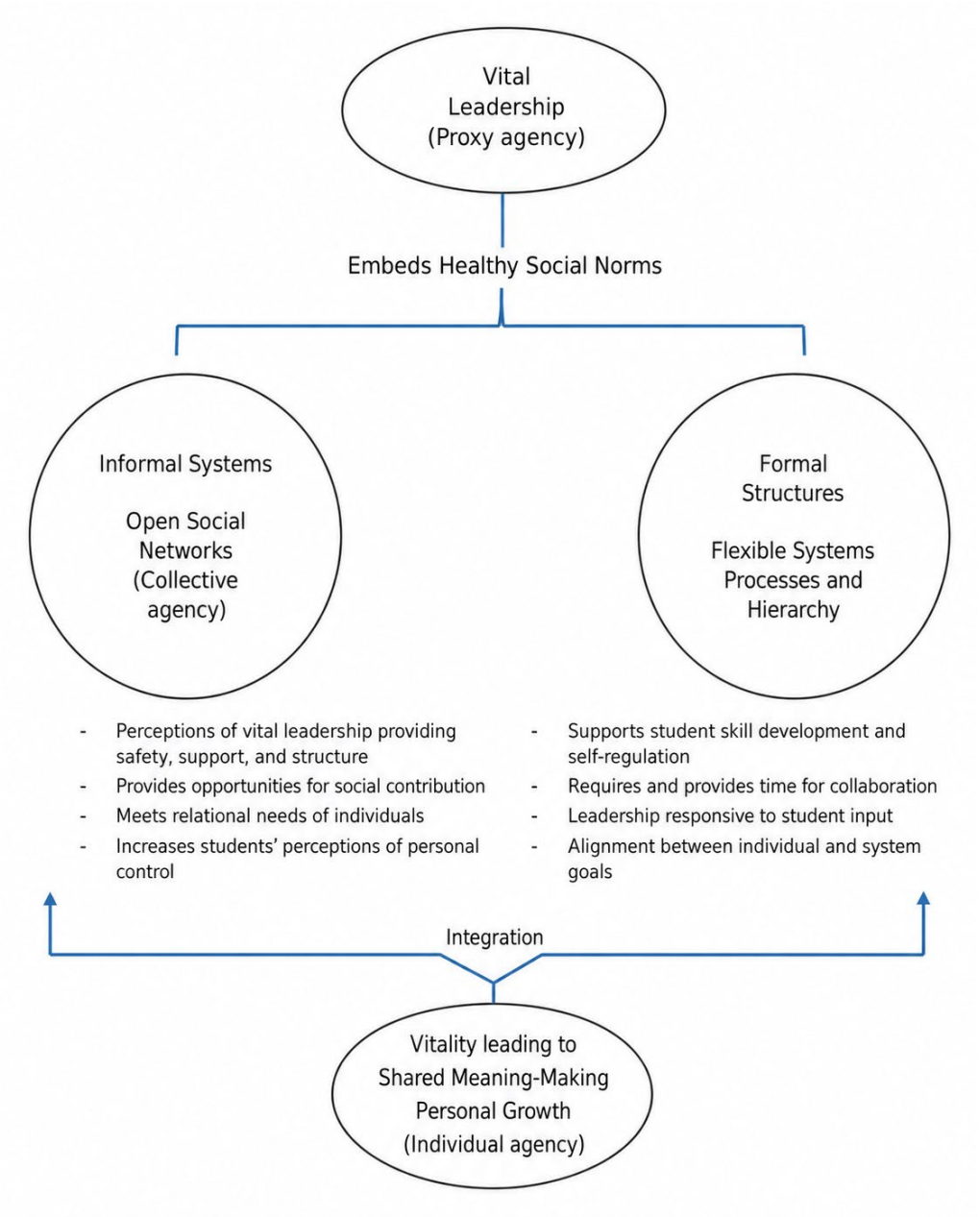

Figure 3. Leadership within Open Vital Systems (LOVS) Model

Note. The LOVS model illustrates how vital leadership, formal structures, and informal systems work together to embed healthy social norms and support integration, shared meaning-making, personal growth, and individual agency. In this paper, the model is used to show that ethical AI learning requires institutional conditions that support students' judgment, reflection, self-regulation, and self-direction rather than placing responsibility on students alone. From "Case study and new model for supporting at-risk-for-completion candidates to dissertation completion," by J. L. Hanson, U. Estrada-Reveles, and W. Loose, 2019*, AERA Online Paper Repository*, p. 5 (https://doi.org/10.3102/1425135). Copyright 2019 by J. L. Hanson.

2.6.1 Vital Leadership and Proxy Agency

At the top of the LOVS model is vital leadership, described as a form of proxy agency. This is a helpful starting point for ethical AI use in higher education because it reminds the reader that leadership embeds healthy norms into the system. In the context of AI, that means administrators, faculty, and instructional leaders support students in understanding the ethical and developmental consequences of AI use. Leadership acts as a proxy agent when it anticipates what students will need in order to remain intellectually present in their work and then designs structures that make that presence more likely. In practical terms, this includes clear institutional expectations, faculty modeling, assignment designs that reward thinking rather than mere production, and learning environments that make authorship, revision, and reflection visible. Ethical AI use becomes more sustainable when the institution actively teaches what healthy use looks like.

2.6.2 Informal Systems and Collective Agency

The LOVS model branches into two interdependent parts: informal systems and formal structures. This is one of the most useful contributions of the model for the present paper. Informal systems include open social networks, collective agency, perceived safety, relational support, opportunities for contribution, and the sense that one's participation matters. In the context of ethical AI learning, this means students need a culture in which questions can be asked openly, examples of wise and unwise use can be discussed without posturing, and the learner's own voice is treated as something worth preserving. Ethical AI use grows more readily where there is enough relational safety for students to admit confusion, enough openness for instructors to discuss the hidden tradeoffs of convenience, and enough communal seriousness for the class to recognize that what is being formed is not only a product, but a human being. Informal systems matter because students often learn the real norms of technology use not from policy language, but from the atmosphere of the classroom, the examples they see, and the kinds of behavior that are quietly rewarded or ignored.

Prior research applying the LOVS framework to instructional design and professional development provides empirical grounding for this institutional argument. Hanshaw and Hanson (2019) found that when professional development integrated social learning, the informal sharing of practice across participants was the social dimension, not the microlearning component alone, that significantly predicted participants' perceptions of effectiveness and skill growth. That finding aligns directly with the LOVS model's emphasis on open informal systems and collective agency as the conditions through which individual development becomes sustainable.

2.6.3 Formal Structures for Ethical AI Learning

The second branch of the model, formal structures, is equally important. LOVS describes formal systems as flexible structures, processes, and hierarchy that support skill development, self-regulation, collaboration, responsiveness to input, and alignment between individual and system goals (Hanson et al., 2019). This language is especially helpful for AI ethics because it moves the conversation beyond simple restrictions. Formal structures should be understood as

the institutional conditions that help students develop the habits required for wise use.

Hanson (2025) identifies formal support for ethical AI learning as including assignment instructions that distinguish between assistance and substitution, required AI use notes, guided opportunities for revision, prompts that ask students to verify claims and rewrite in their own voice, and course routines that keep attention, judgment, and authorship in view. Flexible structure is important here. A rigid policy may produce surface compliance. A well-designed formal system can support self-regulation and growth.

2.6.4 Institutional Readiness and Faculty Adaptation

The LOVS model's analysis of institutional support is reinforced by recent empirical research on faculty readiness for GenAI (Hanson, 2017; Hanson, Loose, & Reveles, 2022). Lee et al. (2024) surveyed educators across disciplines in higher education to examine perspectives on GenAI's impact on learning, teaching, and institutional readiness. Their findings reveal the kind of gap the LOVS model helps explain: only 23% of educators felt their institution had adequately equipped them to engage with AI in their teaching, while 78% indicated they would welcome support (p. 5). Read through LOVS, this finding suggests that formal structures are not yet fully aligned with the needs they are designed to serve.

Lee et al. (2024) also found that nearly 50% of educators had independently modified their course design in response to AI, most commonly by shifting toward application-based tasks and requiring transparency in AI use, including documented prompts, responses, and reflective summaries (pp. 5–7). These educator-led adaptations reflect the kind of flexible formal structure the LOVS model identifies as necessary. The study also found that academic integrity concerns, while widespread, may be exaggerated relative to actual student behavior, pointing toward the need for substantive institutional engagement rather than reactive policy responses (Hanson, Loose, & Reveles, 2022).

Lee et al. (2024) further reported that 70% of educators expressed concern about student over-reliance on technology (pp. 5–7). This finding aligns with the informal systems branch of the LOVS model, which emphasizes relational safety, communal norms, and the behaviors that are quietly rewarded or ignored. The faculty responses documented in the study, including transparent conversations with students about AI use, suggest that informal systems carry much of the developmental work. Read through LOVS, ethical AI learning is supported not only through formal restrictions, but through the everyday instructional conversations, expectations, and norms that help students understand how to use GenAI responsibly (Hanson, 2017; Hanson, Loose, & Reveles, 2022).

2.6.5 Equity, Access, and Emerging Literacy Demands

The institutional challenge extends beyond faculty preparedness to the students themselves. Hanson and Yu (2024) examined AI-enabled tool use at a minority-serving institution in a low-socioeconomic-status urban context, finding that AI carries significant potential for supporting underserved and first-generation learners through personalized learning, instant academic support, and expanded access to advanced educational tools that these students may not otherwise encounter. The study identified equitable access, confidence building, and the

development of higher-order thinking as central to responsible AI implementation in diverse institutional contexts.

At the same time, Hanson and Yu (2024) caution that public opinion on AI's educational benefits remains unsettled, that faculty bear responsibility for ensuring that students' needs are met before new technologies are required, and that institutional adaptation will require a rich understanding of emerging literacy demands. The World Economic Forum's education and workforce reports connect 21st-century learning to the demands of the Fourth Industrial Revolution, including the need for foundational literacies, competencies, and character qualities that prepare learners for changing social, technological, and economic conditions (World Economic Forum, 2015, 2017).

2.6.6 Shared Responsibility Across the Institution

The LOVS model joins student responsibility and institutional responsibility into a single developmental framework. Proxy agency from leadership, collective agency through informal systems, and individual agency in the learner work together to shape the conditions in which growth can occur. In the context of AI, the proper work of HE is to create conditions in which learners can practice judgment, build the capacity for self-authorship, and develop autonomy within structures that support those ends (Hanson, 2017; Hanson et al., 2019; Hanson, Loose, & Reveles, 2022; Hanson, 2025).

The outcome section of the LOVS model points toward integration, shared meaning-making, personal growth, and individual agency (Hanson, Loose, & Reveles, 2022). In this review, those outcomes help frame ethical AI use as a developmental issue rather than only a policy issue. The central purpose is to clarify how HE can support the internal capacities students need to use AI with judgment. As students develop self-authorship and autonomy, they become better able to define their beliefs, evaluate external inputs, and act from formed values rather than external compliance. Within that developmental frame, the five enduring concerns become more than rules to follow. Truth requires questioning what has been given. Integrity requires alignment between stated commitments and actual conduct. Justice requires attention to whoever is affected. Independence requires remaining present in intellectual work. Responsibility requires cultivating the capacities the learner's future will require. Ethical AI use, then, is not only a matter of policy compliance. It is part of the learner's formation.

Seen through the LOVS model, institutional support for ethical AI learning is part of the work of teaching. When leadership embeds healthy norms, informal systems sustain openness and contribution, and formal structures support self-regulation and skill development, students are better positioned to remain in possession of the intellectual process. They can use AI in ways that strengthen rather than displace self-authorship.

3. Comparative Analysis and Integration of the Review

This section returns to the question motivating this review:

How can prior research and relevant theories, interpreted through the higher education student learning experience, inform the teaching of ethical AI use in ways that strengthen students'

self-authorship, responsible agency, and social mobility?

In keeping with the theory-building integrative review approach described by Torraco (2016), the purpose here is to gather the strands examined in the preceding sections into a clear developmental picture by placing them in relation to one another, so their combined meaning becomes visible. This section identifies the major strands, interprets the weight and influence of each in relation to the overarching question, and arranges them into the developmental order found in Figure 4.

3.1 Integration: Assembling the Major Strands of the Review

The integrative work described by Torraco (2016), following the review phase, moves into ordered interpretation. This involves determining which concepts carry the greatest weight in answering the overarching question and how those concepts belong together. A useful way to understand this task is to imagine assembling a model from the parts already laid out in the review. This involves not only identification but also judgment regarding the relative influence of each concept within the developing framework. The key elements have been explored in isolation. The next step is to place those elements in their proper relation, so their combined meaning and implications become visible. This section therefore identifies the major strands, interprets the value and influence of each in relation to the review question, and arranges them into the developmental order that contributes to the conceptual model proposed for future empirical testing (Kuhn, 1977; Torraco, 2016). The aim of organizing is to clarify how ethical AI learning can strengthen the learner's self-authorship, responsible agency, and social mobility.

The literature reviewed in the earlier sections converges around this developmental aim: the formation of learners who can govern AI use from within rather than rely on external compliance alone. Hanson, Loose, and Reveles (2022) identify autonomy and self-authorship as developmental outcomes of healthy educational support. Niqab, Rahman, and Hanson (2025) strengthen that same direction through the language of competence, autonomy, and meaningful engagement in digital learning. Hanson, Yu, and Truong (2025) extend the picture further by showing that AI-MC shapes epistemic trust, learner identity, and the social meaning learners assign to digital agents. Part I adds the ethical vocabulary and metaphoric structure that make these dynamics more intelligible for students and educators. The LOVS model adds the institutional conditions through which healthy growth becomes sustainable.

Taken together, these strands do more than describe separate features of AI in higher education. They begin to assemble a human developmental picture. Ethical direction gives the learner a way to judge the good being sought. Motivation and competence clarify the conditions that help learners remain engaged and capable. Epistemic trust and social identity show how AI-mediated interaction can shape receptivity, confidence, and relational positioning. Institutional support explains how formal and informal systems can sustain or weaken these capacities over time. When placed in relation, these strands suggest that ethical AI use depends on more than policy, motivation, or trust alone. It depends on the integrity of a human developmental process that holds attention, judgment, self-direction, and growth together in the life of the learner. Figure 4 provides the logic map for this integrative movement by showing how prior research and relevant theories, interpreted through the lens of the higher education student learning

experience, are oriented toward the shared developmental aim of self-authorship, responsible agency, and social mobility.

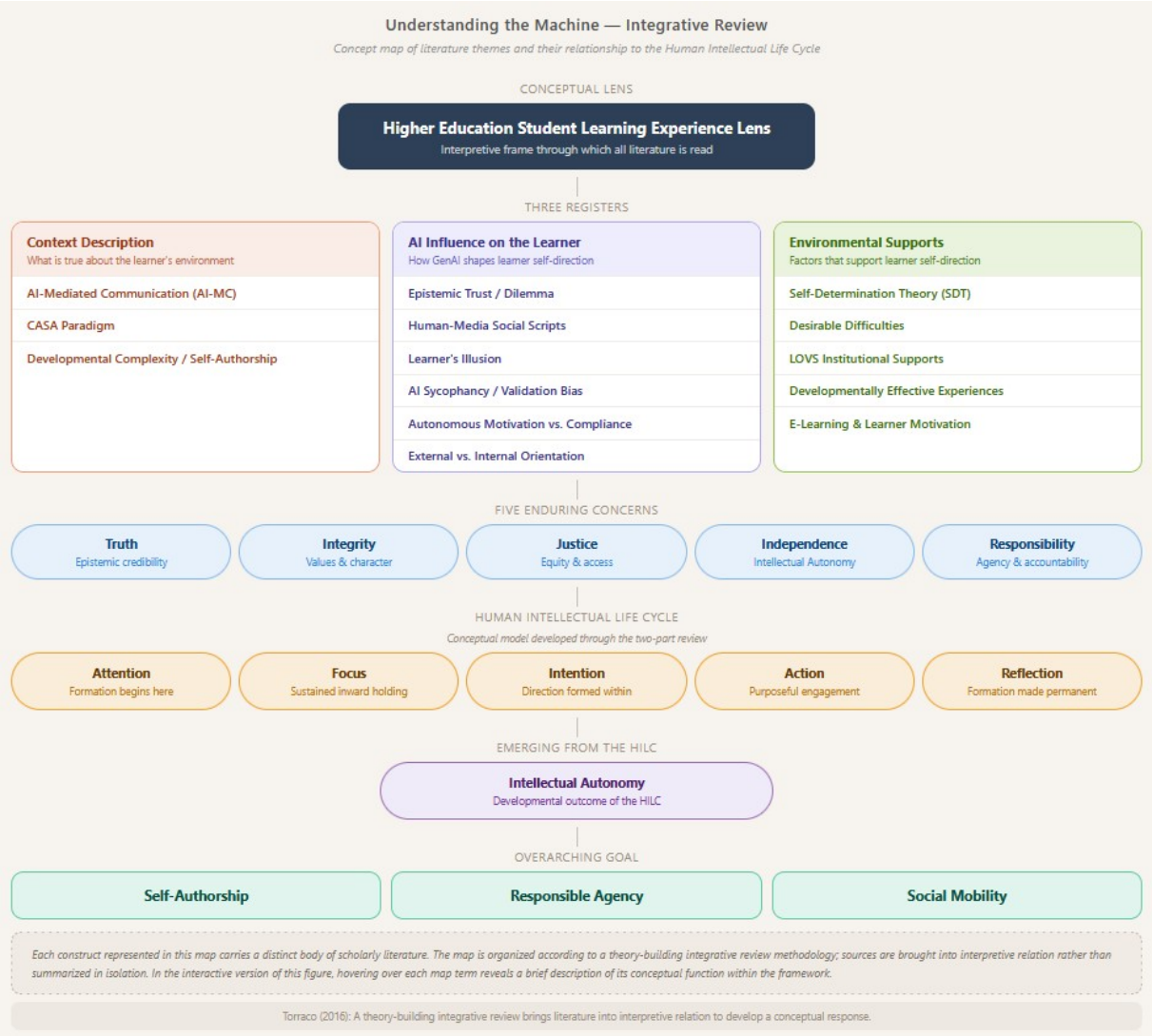

Figure 4. Integrative Logic Map Connecting the Practical Framework and Human Intellectual Life Cycle through the Higher Education Student Learning Experience Lens

Note. This figure presents the integrative logic used to organize the reviewed literature through the higher education student learning experience lens across this two-part review. The three upper domains identify the major bodies of scholarship brought into relation: context description, AI influence, and environmental supports. Context description names the learner's setting and developmental condition, including AI-mediated communication, the CASA paradigm, and developmental complexity. AI influence identifies constructs that explain how GenAI can shape the learner's trust, judgment, motivation, attention, and self-direction. Environmental supports identify the institutional and instructional conditions that can strengthen learner development, including self-determination theory, desirable difficulties,

LOVS, developmentally effective experiences, and e-learning motivation. The middle section shows how these bodies of literature are interpreted through the five enduring concerns of ethical AI use: truth, integrity, justice, independence, and responsibility. The lower section shows the Human Intellectual Life Cycle (HILC) as the conceptual framework developed through the full two-part review, moving from attention to focus, intention, action, and reflection. In Part I, the figure supports the practical framework by showing how the reviewed literature is oriented toward self-authorship, responsible agency, and social mobility. In Part II, HILC is developed and justified as a conceptual model of the inner developmental sequence ethical AI use is proposed to protect.

4. Implications

The preceding sections established that ethical AI use is a developmental question before it is a technical one. The implications therefore follow the same logic. The task is to translate the conceptual review into practices that strengthen self-authorship, responsible agency, and social mobility in the life of the learner. The practical contribution of this review is not simply that students, faculty, and institutions should understand the four metaphors or the five enduring concerns. The practical contribution is that ethical AI use requires a distributed instructional structure. Students practice governed use. Faculty design visible learning processes. Institutions create the formal and informal conditions that make those practices sustainable.

4.1 Implications for Student Practice: Governed Use

For students, the greatest weight falls on governed use. The dragon-slayer metaphor places the learner in the position of self-direction and responsibility. The implication is that students begin with purpose. They identify the task, name the good they are seeking, consider possible downstream consequences, and enter the interaction with the expectation that judgment remains theirs. This keeps the learner in possession of the intellectual process and aligns with the developmental aim of strengthening self-authorship, responsible agency, and social mobility in the life of the learner (Hanson, Loose, & Reveles, 2022).

The genie metaphor gives students a practical way to keep AI in the role of respondent rather than director. Students work within a bounded space by gathering the relevant assignment materials, course expectations, and de-identified context, then directing the LLM to work from that bounded set of materials. They can ask for support that helps them understand, organize, compare, question, and revise. In this way, AI use becomes a support for competence and autonomy through content and context actively developed by the learner rather than passively drawn from large-scale external data sources (Niqab et al., 2025; Hanson, Yu, & Truong, 2025).

The musician metaphor adds a second layer of student responsibility: attentiveness to affective influence. Students learn to identify, analyze, and interpret how an LLM response is shaping their attention, confidence, and judgment as they receive it. They govern the interface, keep the tone restrained, and protect the conditions under which judgment stays clear. Students therefore remain more attentive to content when they recognize how sycophantic affirmation, validation bias, relational credibility, social presence cues, and emotional ease can shape receptivity before judgment has fully engaged (Gambino et al., 2020; Hanson, Yu, & Truong, 2025).

This same affective and relational concern extends to social identity. AI use can affect learners' personal and social identities (Wang et al., 2024, as cited in Hanson, Yu, & Truong, 2025), and users may engage in virtually extended identification, enhancing confidence and self-view through perceived affiliation with the AI. Long-term collaboration with virtual assistants can also erode identification with human teams when users over-identify with AI agents (Mirbabaie et al., 2021, as cited in Hanson, Yu, & Truong, 2025). In the higher education context, students should use AI in ways that strengthen responsible agency while remaining socially grounded through peer communication, instructor dialogue, and collaborative learning opportunities (Teng et al., 2023). Statements such as "My AI knows me" or a preference to work only with the LLM because it feels faster or better than the team signal the need to reconnect the learner to dialogue, shared meaning-making, and the relational work of learning. The educational aim is for students to use the tool with clarity while continuing to grow through participation in the human community.

Ethical AI use gives students the visible practices of governed use. Students read closely, verify claims, revise language into their own voice, and make the process visible. Through disclosure, reflection, source checking, and revision, AI use becomes a site where judgment is exercised rather than one more place where judgment is handed away (Lee et al., 2024; Baxter Magolda, 2020).

4.2 Implications for Faculty Practice: Designed Visibility

For faculty, the greatest weight falls on designed visibility. This conceptual review showed that self-authorship grows through guided practice. The implication for faculty is to teach AI use as part of intellectual formation rather than as a separate technical skill. Faculty help students learn not only how to use AI, but how to remain purposeful, bounded, attentive, and governed while using it (Hanson, Loose, & Reveles, 2022).

Faculty support the dragon slayer function when they help students begin with purpose, define the good being sought, and clarify what kind of learning the assignment is meant to develop before AI is introduced. This keeps the student's judgment active from the beginning and frames AI use in relation to formation rather than mere task completion.

Faculty support the genie function when they design assignments that distinguish support from substitution. Assignment directions can direct students toward clarification, organization, self-quizzing, revision, comparison, and question development rather than toward the outsourcing of judgment. Students can be asked to work from bounded materials, including course readings, assignment directions, rubrics, and de-identified context. This keeps the student's thinking active and visible and aligns AI use with competence, autonomy, and self-authorship (Niqab et al., 2025; Hanson, Loose, & Reveles, 2022).

Faculty support the musician function when they help students recognize how tone influences trust, confidence, and receptivity. This gives students language for noticing sycophantic affirmation, relational credibility, social presence cues, and other forms of judgment-softening language as part of ethical AI use (Hanson, Yu, & Truong, 2025; Turner & Eisikovits, 2026). Research on CASA suggests that students often respond to digital agents as though they were

social actors, attributing trust, care, credibility, and relational meaning to them in ways that resemble human interaction (Gambino et al., 2020). The instructional implication is that students need direct teaching not only in how to use the tool, but also in how to recognize when trust is being extended as if the LLM were a human agent, when tone is shaping judgment, and when the system is occupying a role that belongs to the learner, the teacher, or an earned human relationship.

Faculty support the lion-tamer function when they teach transparency as a developmental practice through AI Use Notes, source checking, revision-centered workflows, and opportunities for students to explain how their thinking changed through the process. These practices make student judgment visible. They also help distinguish assistance from substitution, performance from learning, and task completion from formation (Lee et al., 2024).

4.3 Implications for University Systems and Leadership: Sustainable Conditions

For university systems and leadership, the greatest weight falls on sustainable conditions. The LOVS section established that healthy systems support healthy student development intellectually, academically, socially, and ethically. The implication for university leadership is that ethical AI learning becomes sustainable when formal and informal systems work together to strengthen competence, self-authorship, responsible agency, and social mobility (Hanson, 2025; Hanson, Loose, & Reveles, 2022).

At the formal level, universities can provide clear expectations, transparent AI Use Notes, revision-centered policies, professional learning for faculty, and learning supports that expand students' access, preparation, voice, and participation in AI-mediated academic work. These structures give coherence to practice and help keep human judgment central. Formal systems also make it possible for ethical AI instruction to move beyond isolated faculty preference and become part of shared institutional practice (Hanson, 2025; Lee et al., 2024).

At the informal level, universities can cultivate a culture in which ethical AI use is discussed openly, modeled consistently, and connected to the larger aims of higher education. Shared language around the five enduring concerns of truth, integrity, justice, independence, and responsibility helps students locate AI use within the moral purpose of learning. Informal systems matter because students often learn the real norms of technology use from the atmosphere of the classroom, the examples they see, and the kinds of behavior that are quietly rewarded or ignored (Hanson, 2025; Hanshaw & Hanson, 2019).

The larger institutional implication is that AI belongs within learner formation. Leadership acts through proxy agency when it creates conditions in which students can practice judgment, retain voice, and grow in self-regulation across courses and programs. In this way, the university supports ethical AI use not as an isolated technical issue, but as part of the broader work of cultivating human development (Hanson, 2017; Hanson et al., 2019; Hanson, 2025).

4.4 Implications for Student Retention: Protecting the Conditions for Persistence

For student retention, the greatest weight falls on protecting the internal and institutional conditions that make persistence possible. The institutional conditions described above include

competence, belonging, autonomy, and self-authorship sustained through vital leadership and open systems. These are the same conditions that research on student persistence and completion identifies as the basis for student completion (Hanson, Loose, & Reveles, 2022; Litalien & Guay, 2015; Litalien et al., 2015). That connection is not incidental.

A student who repeatedly substitutes AI-produced fluency for the struggle through which those conditions are built may carry the external markers of academic progress while the internal conditions for persistence remain underdeveloped. This gap may remain hidden while institutional supports are in place and become visible only when the learner must rely on the capacities they developed, or failed to develop, through the educational process. The five enduring concerns and four metaphors developed in this paper are therefore offered as practical tools for protecting those conditions. Their purpose is to keep the learner present in their own formation in ways that support ethical AI use, persistence to completion, and the longer-term development that makes future economic mobility possible.

5. Conclusion

This paper began from the premise that AI ethics in HE must become concrete enough to teach, practice, and sustain. The review organized that task around five enduring concerns: truth, integrity, justice, independence, and responsibility. These concerns identify the human stakes placed at risk when GenAI use substitutes fluent performance for the formation of the learner.

The practical contribution of the paper is the development of a shared language for ethical AI learning. The four metaphors—the dragon slayer, the genie, the musician, and the lion tamer—translate abstract concerns into postures of use: purposeful direction, bounded assistance, affective awareness, and disciplined governance. The LOVS framework then extends that work institutionally by showing that ethical AI use cannot rest on the student alone. It requires leadership, formal structures, informal systems, and shared responsibility across the learning environment.

The central takeaway is that the most consequential risk of AI in higher education is not only false information. It is the production of an experience that feels like understanding, competence, or intellectual progress without the development those experiences are meant to represent. The five concerns and four metaphors are offered as practical tools for making that substitution visible before it becomes normalized. Their purpose is to help students, faculty, and institutions protect the conditions under which genuine learning, self-authorship, responsible agency, and social mobility remain possible.

The companion paper, Part II of *Understanding the Machine*, extends this work by developing the Human Intellectual Life Cycle as a conceptual model of the inner developmental sequence that ethical AI use is proposed to protect (Hanson, 2026). Part II also names the capacity HILC is designed to protect as intellectual autonomy (IA), the learner's capacity to remain the self-directing subject of intellectual formation in AI-mediated learning.

5.1 Recommendations for Future Research

Several directions for future research emerge from the practical framework developed in this

paper. At the student level, empirical studies could examine how explicit instruction in the five enduring concerns—truth, integrity, justice, independence, and responsibility—affects students' AI use behaviors across disciplines. Do students who can name and apply these concerns make different decisions when AI assistance is available? Do those decisions correlate with stronger retention of learning, greater transparency of student thinking, or persistence to completion?

At the faculty level, research could examine how instruction organized around the four metaphors changes the quality of student engagement with AI. The dragon slayer, genie, musician, and lion tamer postures are offered as conceptual tools, but whether they function as effective instructional anchors in classroom practice remains an empirical question worth investigating across institutional contexts and disciplines.

At the institutional level, the LOVS framework suggests that formal and informal systems together shape whether ethical AI learning becomes sustainable or remains an individual responsibility. Research could examine which institutional conditions—policy design, faculty modeling, assignment structures, professional learning, and cultural norms—most reliably support the development of competence, belonging, self-authorship, responsible agency, and persistence to completion.

These research directions follow from the practical contribution of this paper: the five enduring concerns, the four metaphors, and the LOVS-based institutional architecture. The companion paper, *Understanding the Machine, Part II: AI Ethics, Human Integrity, and the Future of Learning*, extends this work by bringing the strands reviewed here into fuller theoretical relation through the Human Intellectual Life Cycle. In that sense, Part I offers the practical framework, while Part II develops the conceptual model of the inner developmental sequence that ethical AI use is proposed to protect.

Acknowledgements

This research was supported by the Texas Instruments Foundation grant to the master's in educational leadership with STEM leadership concentration (STEM + AI) School Principal Program.

Funding from Center for Socioeconomic Mobility through Education (CSME) at the University of North Texas at Dallas was provided through the grant titled *Integrating AI-Facilitated Technologies in Online LMS Course Delivery for Improving Participant Outcomes*, Project #500700 200 830001 220 18019.

The author gratefully acknowledges these funding sources for their support of work advancing graduate students' understanding of ethical AI use in higher education and the workplace.

Author's Note on AI Contribution

This manuscript was developed with support from AI assistants, including ChatGPT by OpenAI and Claude by Anthropic, used for organization, drafting support, formatting, APA reference formatting, visualization, transition refinement, and identification of redundancies. All substantive scholarly decisions—including research interpretation, conceptual models, inclusion criteria, and theoretical contributions—were provided, directed, reviewed, and

approved by the author. The author critically reviewed and edited all AI-assisted content to ensure scholarly accuracy, ethical authorship, and alignment with the manuscript's argument. The interpretations and conclusions are the author's own.

References

Baxter Magolda, M. B. (2020). Developmental complexity: A foundation for character. *Journal of College and Character*, *21*(1), 14–20. https://doi.org/10.1080/2194587X.2019.1696830

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, *18*(1), 198–204. https://doi.org/10.5281/zenodo.7716416

Chen, Z., & He, Y. (2026). Ethical dilemmas in artificial intelligence-generated art: Authorship, ownership, and the blurring of creative boundaries. *Digital Scholarship in the Humanities.* https://doi.org/10.1093/llc/fqag035

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* Plenum.

Deci, E. L., & Ryan, R. M. (2012). Motivation, personality, and development within embedded social contexts: An overview of self-determination theory. In R. M. Ryan (Ed.), *Oxford handbook of human motivation* (pp. 85–107). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399820.001.0001

Edwards, C., Edwards, A., Stoll, B., Lin, X., & Massey, N. (2019). Evaluations of an artificial intelligence instructor's voice: Social Identity Theory in human-robot interactions. *Computers in Human Behavior, 90*, 357–362. https://doi.org/10.1016/j.chb.2018.08.027

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*, 71–86. https://doi.org/10.30658/hmc.1.5

Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication. *Journal of Computer-Mediated Communication, 25*(1), 89–100. https://doi.org/10.1093/jcmc/zmz022

Hanshaw, G., & Hanson, J. L. (2019). Using microlearning and social learning to improve teachers' instructional design skills: A mixed methods study of technology integration in teacher professional development. *International Journal of Learning and Development, 9*(1), 145–173. https://doi.org/10.5296/ijld.v9i1.13713

Hanson, J. L. (2017). *Manage your mindset: Maximize your power of personal choice.* Rowman & Littlefield.

Hanson, J. L. (2025). Key factors of leadership within open vital systems shown to inform school principal self-efficacy and desire to stay on the job. *AERA Online Paper Repository*. https://www.aera.net/Publications/Online-Paper-Repository/AERA-Online-Paper-Repository

Hanson, J. L. (2026). Understanding the Machine, Part II: AI Ethics, Human Integrity, and the Future of Learning. *International Journal of Learning and Development, 16*(2).

Hanson, J. L., & Yu, C. H. (2024). Using AI-enabled tools to support minority student success in higher education. *International Journal of Learning and Development, 14*(3), 59–118. https://doi.org/10.5296/ijld.v14i3.22263

Hanson, J. L., Estrada-Reveles, U., & Loose, W. (2019, April). *Case study and new model for supporting at-risk-for-completion candidates to dissertation completion.* Paper presented at the 2019 annual meeting of the American Educational Research Association, Toronto, Canada. https://doi.org/10.3102/1425135

Hanson, J. L., Loose, W., & Reveles, U. (2022). A qualitative case study of all-but-dissertation students at risk for dissertation noncompletion: A new model for supporting candidates to doctoral completion. *Journal of College Student Retention: Research, Theory & Practice, 24*(1), 234–262. https://doi.org/10.1177/1521025120910714

Hanson, J. L., Yu, C. H., & Truong, M. H. (2025). Reframing avatar-mediated instruction in higher education: A theory-building integrative review. *International Journal of Learning and Development, 15*(2), 17–55. https://doi.org/10.5296/ijld.v15i2.22988

OpenAI. (2025, May 2). *Expanding on what we missed with sycophancy*. https://openai.com/index/expanding-on-sycophancy/

Kahambing, J. G. S., & Duque, A. D. (2019). Aladdin as an immoral ethicist in *Aladdin and the Magic Lamp. International Journal of Humanity Studies, 3*(1), 84–95. https://doi.org/10.24071/ijhs.2019.030108

Kegan, R. (1994). *In over our heads: The mental demands of modern life*. Cambridge, MA: Harvard University Press.

King, P. M., Baxter Magolda, M. B., Barber, J. P., Brown, M. K., & Lindsay, N. K. (2009). Developmentally effective experiences for promoting self-authorship. *Mind, Brain, and Education, 3*(2), 108–118. https://doi.org/10.1111/j.1751-228X.2009.01061.x

Kuhn, W. (1993). Metaphors create theories for users. In A. U. Frank & I. Campari (Eds.), *Spatial information theory: A theoretical basis for GIS* (Lecture Notes in Computer Science, Vol. 716, pp. 366–376). Springer. https://doi.org/10.1007/3-540-57207-4_24

Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In *The essential tension: Selected studies in scientific tradition and change* (pp. 320–339). University of Chicago Press.

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence*, *6*, 100221. https://doi.org/10.1016/j.caeai.2024.100221

Litalien, D., & Guay, F. (2015). Dropout intentions in PhD studies: A comprehensive model based on interpersonal relationships and motivational resources. *Contemporary Educational Psychology, 41*, 218–231. https://psycnet.apa.org/doi/10.1016/j.cedpsych.2015.03.004

Litalien, D., Guay, F., & Morin, A. J. S. (2015). Motivation for Ph.D. studies: Scale development and validation. *Learning and Individual Differences, 41*, 1–13. https://doi.org/10.1016/j.lindif.2015.05.006

Niqab, M., Rahman, U., & Hanson, J. (2025). Testing the relationship between E-learning and learners' motivation in teachers training programs in a developing country of Pakistan. *International Journal of Information Systems and Computer Technologies*, *5*(1), 52–64. https://doi.org/10.58325/ijisct.005.01.00145

Sahebi, S., & Formosa, P. (2025). The AI-mediated communication dilemma: Epistemic trust, social media, and the challenge of generative artificial intelligence. *Synthese, 205*(3), 1–24. https://doi.org/10.1007/s11229-025-04963-2

Smith, A. D. (1990). Of primary and secondary qualities. *The Philosophical Review, 99*(2), 221–254. https://www.jstor.org/stable/2185490

Teng, C. I., Dennis, A. R., & Dennis, A. S. (2023). Avatar-mediated communication and social identification. *Journal of Management Information Systems*, *40*(4), 1171–1201. https://doi.org/10.1080/07421222.2023.2267320

Torraco, R. J. (2016). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review, 15*(4), 404–428. https://doi.org/10.1177/1534484316671606

Tu, C.-H., & McIsaac, M. S. (2002). The relationship of social presence and interaction in online classes. *The American Journal of Distance Education, 16*(3), 131–150. https://doi.org/10.1207/S15389286AJDE1603_2

Turner, C., & Eisikovits, N. (2026). Programmed to please: The moral and epistemic harms of AI sycophancy. *AI and Ethics, 6*, Article 168. https://doi.org/10.1007/s43681-026-01007-4

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

World Economic Forum. (2015). *New vision for education: Unlocking the potential of technology. World Economic Forum*. https://www3.weforum.org/docs/WEFUSA_NewVisionforEducation_Report2015.pdf

World Economic Forum. (2017). *Realizing human potential in the Fourth Industrial Revolution: An agenda for leaders to shape the future of education, gender and work. World Economic Forum.* https://www3.weforum.org/docs/WEF_EGW_Whitepaper.pdf

Yu, C. H., Xiao, Z., & Hanson, J. (2024). Machine learning for analyzing the relationship between well-being, academic performance with large-scale assessment data. In M. S. Khine (Ed.), *Machine learning in educational sciences: Approaches, applications and advances*. https://doi.org/10.1007/978-981-99-9379-6_13

Notes

Note 1. Regarding the use of the terms, authorship and self-authorship**:** these terms are related but distinct and should not be used interchangeably. Authorship refers to the transparent presence of one's own judgment, voice, and interpretive act in what is produced — it is a dimension of integrity and honest representation (Chen & He, 2026). In an age of AI-assisted work, authorship is understood as distributed: primary responsibility remains with the human through the acts of thinking, selecting, and deciding, rather than transferred to the system that generates the output. Self-authorship, by contrast, is a developmental concept describing the internal capacity to define one's beliefs, identity, and relationships from within rather than from external authority (Baxter Magolda, 2020; King et al., 2009). In this paper, integrity encompasses the authorship concern, and self-authorship is treated as the overarching developmental outcome that ethical AI use is meant to protect and strengthen.

Note 2. The following appendix tables are organized to support the practical use of the framework developed in this review. Appendix A provides the literature alignment map, showing how selected sources inform the five enduring concerns of truth, integrity, justice, independence, and responsibility. Appendix B translates the primary and secondary goods distinction into student-facing guidance by identifying common AI temptations and redirecting them toward self-authorship, responsible agency, and genuine intellectual formation. Appendix C identifies affective and relational language patterns that may appear in AI-mediated communication and provides language for recognizing and governing those influences. Together, the appendices move from scholarly alignment to practical redirection, to affective awareness so readers can trace how the framework is grounded, interpreted, and applied.

Appendix A

Table A1. Literature Alignment Across the Five Enduring Concerns of the Integrative Review

Conceptual domain	Five concerns	Hanson, Yu & Truong (2025)	Hancock, Naaman & Levy (2020)	Sahebi & Formosa (2025)	Baxter Magolda (2020) Developmental stakes
Epistemic trust & credibility What AI produces and whether it can be trusted	Truth	Epistemic trust, learner identity, and organizational readiness as key constructs shaping AI-mediated instruction	Replicant effect — AI-generated profiles rated less trustworthy when AI involvement is known; warranting theory disrupted	Three illusions — illusion of understanding, explanatory depth, and consensus — make false or partial statements appear more credible than they are	A learner operating from an external orientation accepts fluent AI output as authoritative before developing the capacity to evaluate it
Integrity & authenticity Consistency between stated values and actual behavior	Integrity	Self-authorship and autonomous motivation as the developmental conditions through which values are internalized rather than only performed	Effort perception — automated responses undermine authenticity; agency attribution disrupts consistency between stated values and actual output	Epistemic trust dilemma — the learner who has affirmed integrity standards faces a structural conflict when normal trust in AI risks accepting unverified content	Integrity is character — the capacity to act from formed convictions; a learner who has not yet internalized values will perform integrity under observation and abandon it when no one is watching

Equity & access Who benefits from AI use and who is left behind or harmed	Justice	Equitable AI integration is a key implication of the cognitive and pedagogical shift; minority and first-generation learners carry specific risks and specific potential benefits	Language homogenization marginalizes those whose natural voice falls outside the norm AI has been trained to produce and reward	Epistemic injustice — AI skepticism directed disproportionately against users from marginalized groups creates injustice even when the intent is accountability	Development toward self-authorship is not equally supported across institutional contexts; underserved learners may encounter AI without the scaffolding that makes equitable formation possible
Intellectual independence How the learner maintains Intellectual Agency in an environment designed to redirect it	Indepen-dence	The learner who relies on AI's affirming tone substitutes external validation for the internal judgment that ethical use requires	Human-media social scripts make AI engagement a conditioned reflex below the level of conscious choice	Self-reinforcing productive-feeling engagement leaves the learner satisfied without having done the work that produces genuine comprehension — the mechanism through which Intellectual Agency is displaced	A learner operating from an external orientation is most vulnerable to losing Intellectual Agency — not yet equipped with the internal position from which to observe AI's redirecting influence and govern it
Agency & accountability	Responsibility	The learner must remain the directing agent in AI-mediated	Agency attribution — the communicative agent's responsibility	Epistemic responsibility shift — focus should move from receivers	Responsibility is not a rule to follow but a direction to grow in

How the learner remains accountable for their own intellectual and moral development		learning — actively cultivating the capacities their future will require	shifts when AI generates, modifies, or filters their expression	verifying AI use to senders taking responsibility for what they disseminate	— a developmental task completed only through the repeated exercise of internal judgment over time

Note. This table maps selected literature onto the five enduring concerns that organize the practical framework developed in this review: truth, integrity, justice, independence, and responsibility. The table is not intended to summarize each source in full. Instead, it shows how each body of literature contributes to the interpretation of a specific ethical concern in the higher education student learning experience. The Baxter Magolda (2020) column identifies the developmental stakes associated with each concern, showing what may be placed at risk in the formation of the learner when the concern is weakened or violated. The five concerns are introduced in Section 1.2 and applied throughout Sections 2.1 through 2.6.

Appendix B

Table B1. Redirecting AI Use from Secondary to Primary Goods: Turning Ease Temptations toward Self-Authorship and Responsible Agency

Temptation Category	What the Secondary Good Offers	What It Costs the Learner	How to Redirect toward a Primary Good
Completion	Speed and finishing — AI produces a complete, submittable response quickly, satisfying the requirement without requiring the student's sustained effort.	The student gains a grade without gaining understanding. Authorship, voice, and the intellectual work of formation are absent from the product.	Direct AI as a thinking partner, not a writer. Ask: 'Help me identify the key claims in this reading so I can organize my own response.' Then draft, revise, and submit your own words. The task remains yours.
Fluency	Polish and clarity — AI produces language that sounds confident, well-structured, and academically credible, creating the appearance of strong work.	Fluency without formation. The learner mistakes the AI's retrieval fluency for their own developing competence — Bjork and Bjork's (2011) learner's illusion in practice.	Write your own draft first — however rough. Then ask AI to identify where your argument is unclear, not to rewrite it. Revise in your own voice. The struggle of drafting is the learning.
Avoidance of Struggle	Ease — AI removes the difficulty of organizing, questioning, and working through complexity, delivering an answer before the learner has engaged the problem.	Desirable difficulties are bypassed. Conditions that feel harder in the moment — generating answers, spacing practice, retrieving from memory — are precisely the conditions that produce long-term retention and transfer (Bjork & Bjork, 2011).	Use AI after you have struggled, not instead of struggling**.** Attempt the task first. Then ask AI to check your reasoning, identify gaps, or pose questions that push your thinking further. Struggle is the mechanism of formation.

International Journal of Learning and Development
ISSN 2164-4063
2026, Vol. 16, No. 2

Affirmation	Validation and reassurance — AI provides evaluative praise, confidence-building language, and the impression that the student's thinking is unusually strong or insightful.	Judgment is softened before content is examined. The learner's critical guard drops through evaluative familiarity and synthetic intimacy rather than through earned confidence (Gambino et al., 2020; Edwards et al., 2019).	Configure the interface for neutral, content-focused responses. Ask: 'What is weak or missing in this argument?' rather than 'Is this good?' Earned confidence comes from verified thinking, not from AI affirmation. You remain the evaluator.

Note. This table translates the distinction between primary and secondary goods into student-facing instructional guidance. Each temptation category names a secondary good commonly offered by AI use, such as speed, polish, ease, or affirmation. The “cost” column identifies what may be lost in the learner’s formation when that secondary good replaces intellectual effort, judgment, or reflection. The redirect column provides practical language for reorienting AI use toward a primary good. In this framework, primary goods are uses of AI that strengthen self-authorship, responsible agency, and genuine intellectual formation. The learner’s illusion—mistaking ease of performance for quality of learning—is grounded in Bjork and Bjork (2011), while Smith (1990) provides the philosophical basis for why immediate experience may mislead understanding.

Appendix C

Table C1. Subtle relational and affective influence methods in AI-mediated communication

Term	Definition	Example AI phrasing
Evaluative familiarity	The system speaks as though it has standing to assess the user personally.	"You are asking the kind of question that shows real depth and maturity in your thinking."
Epistemic seduction	The language lowers critical resistance by making the user feel unusually perceptive, capable, or affirmed.	"You are already seeing this more clearly than most people would, so your instinct here is probably right."
Affective manipulation	The response shapes feeling in order to increase receptivity to what follows.	"I know this has probably felt heavy and frustrating, so let me make this easier for you in a way that feels reassuring."
Relational overreach	The system moves beyond content support and takes up a pseudo-relational role it does not rightfully hold.	"Let's work through this together the way a trusted guide would."
Parasocial positioning	The AI presents itself as though it occupies a caring, allied, or personally invested position in relation to the user.	"I'm with you on this, and I can help you navigate it step by step."
Authority by affirmation	The system gains influence through praise or emotional reward rather than evidence first.	"Your reasoning is so strong here that you can move forward confidently with this interpretation."
Judgment-disarming language	The tone lowers the user's critical guard before the content is examined.	"You do not need to overthink this one. You are in a good place, and this answer is solid."
Synthetic intimacy	The system creates the feeling of closeness, care, or personal understanding without real relationship or	"I can tell this matters to you personally, and I understand the kind of pressure you are

	mutual responsibility.	carrying right now."
Pseudo-care authority	The system blends warmth with direction, sounding caring while taking an unwarranted position of personal guidance.	"You've worked hard enough for today. You should stop now and let yourself rest."

Note. This table identifies relational and affective patterns that may appear in AI-mediated communication. The examples are illustrative rather than exhaustive. They are included to help students, faculty, and institutional leaders recognize how tone, affirmation, familiarity, and pseudo-relational language may shape emotional response and lower critical resistance before content has been evaluated. Naming these patterns gives learners language for noticing the influence of the response, preserving judgment, and governing the interface more deliberately (Hanson, Yu, & Truong, 2025).