REVIEW OF ARTIFICIAL INTELLIGENCE IN EDUCATION

DOI: https://doi.org/10.37497/rev.artif.intell.educ.v7ii.1377

PERSPECTIVE RESEARCH

READING WITH, NOT THROUGH, AI: A HUMAN-CENTERED FRAMEWORK FOR MULTIMODAL LITERATURE EDUCATION

Lendo com a IA, e não apenas por meio dela: uma estrutura centrada no ser humano para o ensino multimodal da literatura

Milan Mašát
Palacký University Olomouc (Czech Republic).
Email: milan.masat@upol.cz

ABSTRACT | Background: Multimodal generative artificial intelligence (GenAI) can process written language, images, typography, layout, and interface-like elements, creating opportunities for literature education but also new epistemic risks. Vision-capable models may combine accurate recognition, plausible inference, and invented visual detail in a single fluent interpretation. The central challenge is not access to multimodal information, but disciplined separation of observation, inference, interpretation, and human judgment. **Objective:** This perspective paper develops Multimodal Evidence-Bounded Interpretation (MEBI), a human-centered framework designed to preserve close reading, interpretive plurality, evidential accountability, and student authorship when multimodal GenAI is used with literary texts. **Methods:** The study uses an evidence-informed conceptual synthesis and a worked multimodal case. Recent peer-reviewed research on GenAI-supported learning, student agency, literary analysis, multimodal AI, and visual hallucination is integrated with established scholarship on multimodality and graphic narrative and with human-centered AI guidance. The framework is demonstrated through selected scan locations from Nora Dåsnes's Czech graphic narrative Na kočičí svědomí (2024). **Results:** MEBI comprises five design propositions—human-first encounter, modal traceability, observation–inference separation, preserved interpretive plurality, and human adjudication and authorship—and a five-phase sequence: Encounter, Describe, Hypothesize, Audit, and Adjudicate & Author. Five evidence tags track verbal, image, sequential-spatial, typographic-graphic, and digital-interface evidence. A worked case, classroom sequence, evidence ledger, and assessment heuristic operationalize the framework. **Conclusion:** MEBI positions AI as a generator of descriptions, hypotheses, counter-readings, and questions rather than as interpretive authority. It offers a researchable design for multimodal literature education while requiring empirical validation across genres, languages, age groups, and educational contexts.

Keywords | generative AI; multimodal AI; literature education; AI literacy; close reading; visual hallucination

Received: 25 June 2026
Revised: 11 August 2026
Accepted: 13 August 2026
e-ISSN: 2965-4688

Corresponding Author: Milan Mašát – Email: milan.masat@upol.cz

How to cite this article: Mašát, M. (2026). Reading With, Not Through, AI: A Human-Centered Framework for Multimodal Literature Education. *Review of Artificial Intelligence in Education, 7*(i), e01377. https://doi.org/10.37497/rev.artif.intell.educ.v7ii.1377

RESUMO | Contexto: A inteligência artificial generativa multimodal (GenAI) pode processar linguagem escrita, imagens, tipografia, layout e elementos semelhantes a interfaces, criando possibilidades para o ensino de literatura e, simultaneamente, novos riscos epistêmicos. Modelos com capacidade visual podem combinar reconhecimento correto, inferência plausível e detalhes visuais inventados em uma interpretação fluente. **Objetivo:** Este artigo de perspectiva desenvolve a Multimodal Evidence-Bounded Interpretation (MEBI), uma estrutura centrada no ser humano destinada a preservar a leitura atenta, a pluralidade interpretativa, a responsabilidade evidencial e a autoria do estudante no uso de GenAI multimodal com textos literários. **Métodos:** O estudo combina uma síntese conceitual orientada por evidências com um caso multimodal demonstrativo. Pesquisas revisadas por pares sobre aprendizagem apoiada por GenAI, agência estudantil, análise literária, IA multimodal e alucinação visual são integradas a estudos consolidados sobre multimodalidade e narrativa gráfica e a orientações de IA centrada no ser humano. A estrutura é demonstrada com locais selecionados do escaneamento da narrativa gráfica tcheca Na kočičí svědomí (2024), de Nora Dåsnes. **Resultados:** A MEBI reúne cinco proposições de design e uma sequência de cinco fases: Encontro, Descrição, Hipótese, Auditoria e Julgamento & Autoria. Cinco etiquetas de evidência rastreiam evidências verbais, imagéticas, sequencial-espaciais, tipográfico-gráficas e de interface digital. Um caso demonstrativo, uma sequência didática, um registro de evidências e uma rubrica heurística operacionalizam a proposta. **Conclusão:** A MEBI posiciona a IA como geradora de descrições, hipóteses, contraleituras e perguntas, e não como autoridade interpretativa. A proposta permanece conceitual e requer validação empírica entre gêneros, línguas, faixas etárias e contextos educacionais.

Palavras-chave | inteligência artificial generativa; IA multimodal; ensino de literatura; literacia em IA; leitura atenta; alucinação visual

INTRODUCTION

Generative artificial intelligence (GenAI) has moved rapidly from text-only conversational systems toward multimodal models that can process combinations of written language, images, document layouts, screenshots, and other visual inputs. For education, this shift matters because the object being learned is often already multimodal. In literature education, the issue is particularly acute in graphic narratives, picture books, illustrated diaries, comics, and digitally inflected fiction, where meaning is distributed across wording, image, sequence, typography, page architecture, and conventions borrowed from messaging or social-media interfaces. A model that can "see" a page can therefore appear unusually well suited to the task. The same capability, however, changes the epistemic risk. An unsupported textual answer can be checked against words on a page; an unsupported multimodal answer may blend accurate recognition, plausible inference, and invented visual detail into one fluent interpretation.

Current educational evidence does not justify either a blanket rejection or an unqualified adoption of GenAI. A systematic review and meta-analysis of experimental studies identified 69 articles on ChatGPT-supported learning and found an expanding, heterogeneous evidence base rather than a single context-independent effect (Deng et al., 2025). Large field and laboratory studies likewise indicate that the design of assistance matters. Bastani et al. (2025), in a randomized field experiment with nearly one thousand secondary-school mathematics students, found that access to a standard GPT-4-style interface improved practice performance but was followed by poorer unassisted test performance relative to a control condition; safeguards designed around teacher input substantially reduced this negative learning effect. Darvishi et al. (2024), studying 1,625 students across ten courses, similarly found that students tended to rely on AI assistance

rather than fully internalize what the assistance was doing. Fan et al. (2025) reported a related divergence in a randomized writing study: the ChatGPT-supported group improved its essay scores, while knowledge gain and transfer did not differ significantly. These findings support a distinction that is central to the present article: better AI-assisted task performance cannot be assumed to demonstrate durable learning (Yan et al., 2025).

The distinction is especially important for literature. Literary competence cannot be reduced to producing a coherent summary or a plausible thematic statement. It involves attending to textual and formal detail, tolerating ambiguity, comparing interpretations, coordinating claims with evidence, and assuming responsibility for a reading. Recent literature-specific research shows genuine possibilities for GenAI but also a need for deliberate boundaries. Bellot et al. (2025) used AI-generated alternative endings to George Orwell's Animal Farm as objects of comparison in undergraduate literature teaching, demonstrating the value of treating AI output as material to be examined rather than as an answer key. Chan and Wong (2025), using Bloom's Taxonomy to evaluate GPT-4 literary analysis across three texts, found strengths in accuracy and relevance alongside limitations in depth and the subtleties of humanistic inquiry. Mašát et al. (2026) responded to this problem with the TASU routine (Text → AI → Student → Teacher), an evidence-first model intended to preserve close reading, interpretive reasoning, and student authorship.

Multimodal GenAI creates a further step in this problem. Screen-vision systems can use visual context during learning activities, but the available empirical evidence remains sparse and task-dependent. Torres Vico and Fernández-Herrero (2026) found the clearest differences between multimodal and comparison conditions in organization-related activities rather than across all learning tasks. More broadly, multimodal model research documents systematic visual hallucination and image-context reasoning failures (Guan et al., 2024; Huang et al., 2024). In a classroom, therefore, the issue is not simply whether a model can identify panels, facial expressions, speech balloons, or interface elements. The didactic question is whether students can distinguish what the model has actually grounded in the artifact from what it has inferred, generalized, or fabricated.

This article proposes Multimodal Evidence-Bounded Interpretation (MEBI) as a human-centered framework for AI-supported literature education. Evidence-bounded does not mean that interpretation is reduced to literal description. Literary interpretation necessarily exceeds description. The boundary concerns accountability: every interpretive move should make visible what evidence it draws on, what inferential step it takes, what alternative reading remains plausible, and who finally owns the judgment. MEBI therefore positions multimodal GenAI as a generator of descriptions, hypotheses, counter-readings, and questions, not as an interpretive authority.

The framework is aligned with human-centered AI governance. UNESCO's guidance is explicitly anchored in an approach that promotes "human agency, inclusion, equity, gender equality, and cultural and linguistic diversity" (Miao & Holmes, 2023, section "A human-centred approach"). Its AI Competency Framework for Teachers organizes 15 competencies across human-centred mindset, ethics of AI, AI foundations and applications, AI pedagogy, and AI for professional learning (Miao & Cukurova, 2024). In the European context, Article 4 of the AI Act requires providers and deployers to take measures supporting AI literacy; the provision was amended in July 2026, and current implementation details are reflected in European Commission guidance (European Commission,

n.d.; European Parliament & Council of the European Union, 2024). For literature education, these governance principles are most useful when translated into observable classroom practices rather than treated as abstract policy language.

Accordingly, the study addresses three research questions:

- RQ1. Which epistemic and pedagogical risks become salient when multimodal GenAI is used to interpret graphic and visually complex literary narratives?
- RQ2. Which evidence and agency principles can reduce these risks without eliminating the productive use of AI for comparison, questioning, and interpretive plurality?
- RQ3. How can these principles be operationalized in a worked literature-education case using an excerpt from Nora Dåsnes's Na kočičí svědomí?

The article first synthesizes evidence on GenAI-supported learning, student agency, literature education, multimodality, and visual hallucination. It then defines five design propositions and translates them into the MEBI workflow. Finally, the framework is demonstrated on a supplied scan excerpt from Dåsnes's graphic narrative. The worked case is methodological rather than experimental: it illustrates how an evidence-bounded lesson can be designed and audited, but it does not claim effects on student achievement, motivation, or reading competence.

LITERATURE REVIEW

GenAI in education: benefit is conditional on design

The most defensible conclusion from the current educational literature is conditional rather than celebratory or prohibitive. GenAI can improve access to explanations, feedback, ideation, and task completion, but learning effects depend on task type, learner characteristics, timing, instructional framing, and the amount of cognitive work retained by the learner. Deng et al. (2025) synthesized experimental studies across educational levels and disciplines and found substantial heterogeneity in implementation. Sanz-Tejeda et al. (2026), reviewing 136 open-access studies of GenAI-assisted academic reading and writing in higher education, likewise described both efficiency and support benefits and recurring concerns about overreliance, authorship, academic integrity, and the quality of cognitive engagement.

The strongest causal warning comes from studies that remove AI after supported practice. Bastani et al. (2025) showed that unguarded assistance can create a performance–learning trade-off: students may solve more during supported practice while acquiring less independent capability. This does not mean that GenAI is intrinsically harmful. In the same study, a tutor constrained to provide pedagogically designed hints substantially mitigated the negative effect. The result is directly relevant to literature education because it shifts attention from the question "Is AI allowed?" to the more actionable question "What kinds of assistance preserve the intellectual work the assignment is meant to develop?"

Yan et al. (2025) formalize the same problem by distinguishing immediate performance gains from learning. Durable learning depends on psychological processes such as metacognition, self-efficacy, cognitive effort, and retrieval that an assistance system can either support or bypass. Fan et al. (2025) provide empirical support for this distinction in academic writing: AI-supported students improved products without corresponding gains in knowledge or transfer. Their term "metacognitive laziness" is useful not as a label for students, but as a design warning: a system that removes monitoring, planning, evaluation, and revision from the learner may improve the artifact while weakening the learning process that the artifact is supposed to evidence.

This design perspective is consistent with research on agency. Darvishi et al. (2024) found that students relied on AI feedback and did not fully preserve performance when assistance was removed; self-regulation strategies partly compensated for the loss. Zhai et al. (2024), in a systematic review of over-reliance on AI dialogue systems, identified risks to analytical thinking, critical evaluation, and decision-making when students accept AI-generated outputs without adequate scrutiny. These findings make a strong case for visible verification routines. In literature education, verification cannot mean fact-checking alone. It must include checking whether a claim is warranted by the formal and semantic evidence of the literary artifact.

GenAI in literary analysis and literature education

Research that directly examines GenAI in literature teaching remains small compared with the broader AI-in-education literature, but it is now sufficiently developed to identify productive roles. Bellot et al. (2025) framed AI-generated alternative endings as comparative material in an undergraduate Animal Farm activity. The pedagogical value lies not in treating the generated ending as a legitimate continuation by default but in asking students to analyze narrative coherence, characterization, historical context, stylistic fidelity, and authenticity. In this role, GenAI externalizes an alternative that students can criticize.

Chan and Wong (2025) evaluated ChatGPT-4 literary analyses across Bloom's cognitive levels using rubrics and judgments from four university teachers. Their findings support a structured partnership model: AI can provide relevant and often accurate analytical material, yet it has difficulty with depth, creativity, and the subtleties of humanistic interpretation. The implication is not that human interpretation is mysteriously inaccessible to machines in every case. It is that fluency and analytical structure are insufficient evidence of interpretive adequacy. A response must be judged in relation to the primary work, the question asked, the conventions of literary argument, and the possibility of reasonable disagreement.

Mašát et al. (2026) provide a literature-specific policy and practice response by insisting that students encounter the text before AI and that AI use remain auditable through evidence ledgers, warrants, and rejection logs. Their TASU model protects the sequence Text → AI → Student → Teacher so that AI does not become the first or final reader. MEBI extends this logic to artifacts in which "the text" is not exclusively verbal. In a graphic narrative, evidence may be carried by gaze direction, panel scale, color or line, the juxtaposition of speech and image, page turns, handwriting,

interface conventions, or the gap between what a character writes and what the image shows. An evidence-first framework must therefore specify how evidence is located across modes.

Multimodal meaning in graphic narratives

Multimodality is not an AI invention. Social-semiotic accounts have long treated communication as the orchestration of multiple modes rather than as language plus optional decoration (Kress, 2010). In visual narratives for younger readers, image, wording, framing, gaze, distance, salience, and page composition jointly organize interpersonal and narrative meaning (Painter et al., 2013). Comics theory likewise emphasizes that meaning emerges not only within panels but through sequence and the cognitive relation established across panels and gutters (Cohn, 2013; Mikkonen, 2017).

These traditions matter because they prevent a common technological simplification: assuming that multimodal AI merely adds image recognition to textual interpretation. A graphic narrative is not a bag of objects plus dialogue. The sequence in which information becomes available, the spatial relation between panels, the visual rhythm of a page, the distinction between diegetic writing and narrator commentary, and the simulation of digital interfaces can all affect interpretation. Any classroom use of multimodal AI therefore requires a unit of analysis richer than "What is in the picture?"

Istrate (2026) makes a related point in multimodal pedagogy: easier production of multimodal materials does not remove the need for pedagogical expertise. In his formulation, "more powerful tools risk producing more elaborate noise" (Introduction). For literature education, the analogous risk is elaborate interpretive noise: a fluent model may supply a psychologically plausible explanation that is weakly grounded in the page, or may collapse productive ambiguity into a single confident reading. The pedagogical problem is thus one of epistemic calibration, not only technical capability.

Multimodal AI, visual hallucination, and evidential uncertainty

Multimodal language models can integrate visual and verbal information, but benchmark research demonstrates that this integration is fallible. Huang et al. (2024) constructed 1,200 visual-hallucination instances across eight modes and showed that major multimodal systems can invent incorrect image details. Guan et al. (2024) developed HallusionBench to test entangled language hallucination and visual illusion, showing persistent difficulty in image-context reasoning even among advanced models. These studies do not measure literary interpretation, and it would be inappropriate to transfer their benchmark scores directly to a classroom. Their relevance is methodological: visual input does not eliminate hallucination; it introduces an additional layer that must be audited.

Educational research on screen-vision models reinforces the need for contextual caution. Torres Vico and Fernández-Herrero (2026) found task-dependent benefits in higher education, with the clearest differences in organization-related activities and with students also reporting technical and ethical concerns. The study does not establish that multimodal AI improves literary

reading. Instead, it shows why the mode of AI access must be treated as a pedagogical variable rather than as a neutral upgrade.

The challenge is intensified by the interpretive status of literary evidence. A model may accurately recognize that a character is drawn with lowered eyes; "the character is ashamed" is an inference. It may correctly identify message bubbles; "the sender is manipulating the recipient" is a higher-order inference. It may recognize a sequence of panels; "the page represents social isolation" is an interpretation. None of these inferences is automatically invalid, but each requires a different degree of warrant. If those levels are not made explicit, multimodal AI can compress observation, inference, and interpretation into a single declarative sentence.

Human-centered AI and AI literacy as interpretive practice

A human-centered approach does not require minimal automation. It requires that automation remain compatible with meaningful human control, responsibility, and capability development (Shneiderman, 2022). UNESCO applies this principle to education by placing human agency and ethical judgment at the center of GenAI governance (Miao & Holmes, 2023) and by framing teacher competence in terms that extend beyond technical tool use (Miao & Cukurova, 2024). These documents are particularly compatible with literature education because interpretive judgment is not a residual step after automation; it is one of the core learning goals.

AI literacy is therefore treated here as an epistemic practice. A student is not AI-literate merely because the student can formulate a sophisticated prompt. In an interpretive task, literacy includes knowing what the system can access, distinguishing generated language from source evidence, recognizing plausible hallucination, comparing outputs, documenting uncertainty, and deciding when a claim should be rejected. This formulation also avoids a model-specific curriculum. Model interfaces and performance will change faster than literature curricula; evidence-checking, source accountability, and human authorship are more durable educational principles.

On this basis, the literature review yields five design propositions rather than statistical hypotheses. The article is a perspective study and does not test causal relationships; the propositions specify conditions to be examined in future empirical research.

- P1 — Human-first encounter. Students should make an initial encounter with the primary literary artifact before receiving AI interpretation, so that the model does not define the first horizon of attention.
- P2 — Modal traceability. AI-supported claims should be traceable to identifiable verbal, visual, spatial/sequential, typographic/graphic, or interface evidence.
- P3 — Observation–inference separation. The workflow should explicitly distinguish what is present in the artifact from what is inferred or interpreted from it.
- P4 — Preserved plurality. AI should be used to generate competing readings and counterevidence rather than to converge prematurely on one "correct" interpretation.
- P5 — Human adjudication and authorship. The student should make the final interpretive judgment, justify acceptance or rejection of AI contributions, and author the assessed response.

METHODOLOGY

Design

The study uses an evidence-informed conceptual synthesis with a worked multimodal case. It is not a systematic review, meta-analysis, model benchmark, classroom intervention, or study of human participants. The purpose of the evidence synthesis is to derive a transparent set of design principles from current empirical and policy literature and then to test their internal coherence against a concrete multimodal literary artifact.

The evidence scan was updated through 9 August 2026. Sources were identified through targeted searches of peer-reviewed publisher platforms, DOI records, scholarly discovery interfaces, and official policy websites. Search combinations included terms equivalent to: generative AI AND education; ChatGPT AND learning; generative AI AND student agency; generative AI AND literary analysis; AI AND literature education; multimodal AI AND higher education; visual hallucination AND multimodal language model; and AI literacy AND education. Priority was given to peer-reviewed empirical studies, systematic reviews or meta-analyses, recent perspective papers that explicitly synthesize evidence, and official UNESCO and European Union documents. For the rapidly changing GenAI literature, the main temporal window was 2023–2026. Earlier sources were retained only where they provided established conceptual tools for multimodality, visual narrative, comics, or human-centered design.

Sources were excluded from the evidential argument when they were vendor marketing materials, nontraceable commentary, retracted work, or preprints for which a peer-reviewed version was available. Claims about current systems were phrased conservatively because model capabilities and product interfaces change rapidly. DOI-bearing publisher records or official institutional pages were used to verify bibliographic information wherever available.

Worked-case material

The worked case uses a 14-page PDF scan supplied for the present analysis from the Czech edition of Nora Dåsnes's Na kočičí svědomí, translated by Jitka Jindřišková and published by Albatros in 2024 (Dåsnes, 2024). The copyright page identifies the Norwegian original as Ti kniver i hjertet, first published by Aschehoug in 2020. Because the scan images do not consistently display printed book pagination and some PDF pages contain photographed spreads, the present study does not invent page numbers. Instead, the PDF images are labeled S1–S14 in scan order. S1 is the cover and S14 is the copyright/back-matter image; the analytical material is concentrated in S2–S13.

The excerpt is suitable for the worked case because meaning is visibly distributed across multiple semiotic resources. The scan includes sequential comic panels, speech balloons, handwritten diary-like comments, drawings embedded in notes, typographic emphasis, color-coded message-like exchanges, profile-card or social-media-like layouts, and scenes in which domestic or peer interaction is mediated by digital devices. These features create a realistic test for a framework that requires evidence to be located across modes rather than only quoted as prose.

No copyrighted page image is reproduced in this article. The worked case refers to scan locations and describes only those formal and narrative features necessary for the methodological demonstration.

Analytic procedure

The worked case proceeded in four stages. First, each scan image was inspected for meaning-bearing resources and assigned one or more classroom evidence tags. Second, potential AI tasks were classified according to whether they asked for description, inference, interpretation, comparison, or evaluation. Third, foreseeable error modes were mapped to the educational evidence reviewed above, especially overreliance, performance–learning divergence, visual hallucination, and premature interpretive closure. Fourth, these mappings were converted into a classroom workflow that requires each AI-supported claim to pass through a human evidence audit before it can enter a final interpretation.

For classroom use, MEBI operationalizes five evidence tags. They are heuristic categories rather than an ontological claim that modes are always separable:

- [V] Verbal evidence: printed dialogue, narration, handwritten words, captions, and other legible language.
- [I] Image evidence: depicted actions, gesture, gaze, facial expression, objects, framing, and pictorial relations.
- [S] Sequential-spatial evidence: panel order, juxtaposition, scale, salience, page placement, and relations across panels or spreads.
- [T] Typographic-graphic evidence: lettering, emphatic marks, diagrammatic elements, color coding, icons, emoji-like signs, and handwritten style.
- [D] Digital-interface evidence: simulated messaging, profile cards, platform conventions, reactions, or other interface-like forms represented inside the narrative.

The criterion for a successful audit is not that every interpretation becomes certain. Rather, the audit should make uncertainty inspectable: students can show where evidence is strong, where an inference is plausible but contestable, and where an AI claim is unsupported.

RESULTS

The Multimodal Evidence-Bounded Interpretation framework

The principal result of the conceptual synthesis is the Multimodal Evidence-Bounded Interpretation (MEBI) framework. MEBI comprises three coordinated layers: five design propositions (P1-P5), five instructional phases, and five modal evidence tags. The phase sequence is Encounter → Describe → Hypothesize → Audit → Adjudicate & Author. The propositions govern when and how AI may enter the interpretive process, while the evidence tags make claims traceable across verbal,

image, sequential-spatial, typographic-graphic, and digital-interface resources. AI is deliberately absent from Encounter and cannot own the final assessed interpretation. Figure 1 summarizes this integrated architecture.

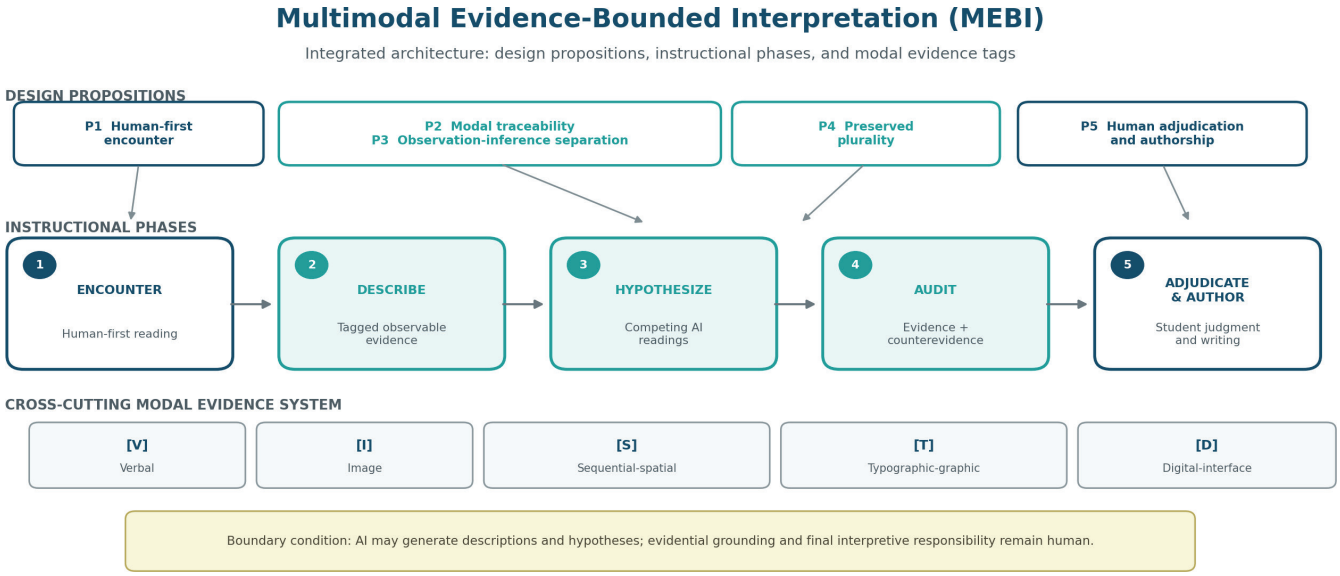

Figure 1. Integrated architecture of Multimodal Evidence-Bounded Interpretation (MEBI). The five-phase sequence is governed by five design propositions (P1-P5) and a cross-cutting modal evidence system [V], [I], [S], [T], [D]. AI is excluded from Encounter, may support Describe and Hypothesize, is audited by the student in Audit, and cannot author the final assessed interpretation.

Across the sequence, the evidence tags function as a shared vocabulary for traceability rather than as mutually exclusive ontological categories; a single interpretive claim may depend on several tags. P2-P4 therefore cut across more than one instructional phase, whereas P1 and P5 define the human boundary conditions at the beginning and end of the workflow.

Phase 1: Encounter — establish a human first reading

Students first inspect the literary artifact without generative assistance. The output is intentionally modest: three to five observations, one question, and one provisional interpretation. The aim is not to produce a superior "human" reading before AI appears. It is to ensure that the learner has an independent record of attention against which later AI suggestions can be compared. This phase operationalizes P1 and responds to evidence that extensive assistance can displace self-regulation or create dependence (Darvishi et al., 2024; Fan et al., 2025).

Phase 2: Describe — constrain AI to evidence before meaning

AI is then asked to describe the selected page or spread while separating directly observable features from uncertain recognition. A suitable prompt is: Describe only what is visibly or verbally present. Tag every observation [V], [I], [S], [T], or [D]. If an element is unclear, say that it is unclear. Do not infer motives, relationships, emotions, identities, or themes in this step.

This constraint does not make the model reliable. It makes its errors easier to see. Students compare the output with the artifact and mark omitted, misrecognized, or invented details. The result is an evidence ledger rather than an interpretation.

Phase 3: Hypothesize — use AI for plural, contestable readings

Only after the evidence ledger is available does AI generate interpretive hypotheses. The system is asked for at least two readings that are genuinely distinct and for evidence that would support or weaken each one. A prompt can require: Offer two competing interpretations of the scene. For each, list the evidence tags on which the reading depends, state one counterargument, and identify any claim that cannot be verified from the supplied page alone.

This phase operationalizes P4. It treats AI's generative breadth as a resource for comparison while avoiding the authority effect created by a single polished answer. Bellot et al. (2025) provide a literature-specific precedent for this contrastive use of generated material.

Phase 4: Audit — coordinate claims, evidence, and counterevidence

Students now evaluate each AI claim against the primary artifact. Every claim is classified as supported, plausible but underdetermined, unsupported, or contradicted. The student records at least one warrant explaining why the evidence supports the claim and, where possible, one piece of counterevidence. Visual or interface claims receive the same scrutiny as verbal claims. If the model states that a character is angry, for example, the student must distinguish between observable features such as posture or lettering and the higher-order emotional label.

The audit is the core of MEBI because it converts AI literacy into disciplinary literacy. Verification is not an external "check the chatbot" activity. It is the very work of literary argument: locating evidence, evaluating relevance, and reasoning from form to meaning.

Phase 5: Adjudicate & Author — return judgment to the student

The student writes the final interpretation in their own words and documents which AI contributions were accepted, modified, or rejected. The final text should cite scan locations or ordinary page numbers where available and should preserve uncertainty when the evidence does not justify certainty. A short rejection log is pedagogically valuable because it makes non-use visible;

learning can be demonstrated not only by adopting a useful suggestion but also by identifying a seductive, unsupported one.

Teacher assessment focuses on the quality of evidence coordination and judgment, not on whether the student's conclusion matches the model's. This final phase operationalizes P5 and keeps AI from becoming either an unacknowledged co-author or a hidden grader of its own interpretation.

Evidence-to-design mapping

Table 1. Contemporary evidence and its implication for MEBI

Evidence source	Evidence type and finding relevant to this study	MEBI design implication
Deng et al. (2025)	Systematic review/meta-analysis of 69 experimental articles; effects vary across contexts and implementations.	Avoid claims that "AI improves learning" in general; specify task and design conditions.
Bastani et al. (2025)	Large randomized field experiment; unguarded GPT improved supported performance but reduced subsequent unassisted performance, while pedagogical safeguards mitigated harm.	Preserve productive learner effort and build guardrails into the interaction.
Darvishi et al. (2024)	Randomized experiment with 1,625 students across ten courses; students tended to rely on AI assistance.	Include an assistance-removal logic: first read and final judgment remain human.
Fan et al. (2025)	Randomized writing study with 117 university students; product improvement did not translate into significant knowledge gain or transfer.	Do not equate polished interpretation with learned interpretive competence.
Bellot et al. (2025)	Literature-teaching study using AI-generated alternative endings as analytical material.	Use AI output as an object of comparison and critique.
Chan & Wong (2025)	Teacher-rated literary analysis showed potential alongside limits in depth and humanistic subtlety.	Require human evaluation of analytical depth and evidence.
Torres Vico & Fernández-Herrero (2026)	Quasi-experimental study of screen-vision multimodal AI; benefits were task-dependent.	Treat multimodality as a pedagogical variable, not as an automatic improvement.
Huang et al. (2024); Guan et al. (2024)	Benchmarks document visual hallucination and image-context reasoning failures.	Audit visual claims and separate recognition from interpretation.
Miao & Holmes (2023); Miao & Cukurova (2024)	Human-centered guidance emphasizes agency, ethics, and teacher competence.	Keep accountable human judgment at the beginning and end of the workflow.

Table 1 is not a meta-analytic synthesis and the studies are not directly comparable. Its purpose is to make the chain from evidence to design choice explicit. The framework avoids importing effect sizes from mathematics, writing, or screen-based tasks into literature education. Instead, it uses those findings to identify failure modes that a literature-specific design should address.

A modal evidence ledger for graphic narratives

MEBI requires a simple representation for evidence that can be used by students and teachers. Table 2 illustrates the ledger structure.

Table 2. Modal evidence ledger

Tag	What the learner records	Example of an admissible observation	Example of a claim requiring further inference
[V]	Exact or paraphrased wording with location	A speech balloon contains a short reply; a handwritten note changes size or emphasis.	The speaker is lying.
[I]	Depicted action, gesture, gaze, object, framing	A character looks away; a phone is visible; figures are separated in the frame.	The character feels rejected.
[S]	Relation across panels/page space	A close-up follows a group scene; one panel occupies markedly more space.	The page proves a permanent change in identity.
[T]	Graphic/typographic treatment	A phrase is enlarged; color separates two strands of annotation; icons punctuate a note.	The color necessarily symbolizes one fixed emotion.
[D]	Simulated digital-interface feature	Message bubbles, profile-style cards, or reaction icons are represented.	The displayed interface gives a complete and objective account of the relationship.

The right-hand column is deliberately not labeled "wrong." Literary interpretation depends on inference. MEBI instead asks students to identify when they have crossed from evidence to inference and to supply a warrant proportionate to that move.

Worked case: Na kočičí svědomí

The supplied scan illustrates why multimodal traceability is needed. The sequence cannot be adequately described as a set of prose passages accompanied by pictures. Different scan images activate different configurations of mode, and the evidential demands change accordingly.

S2 foregrounds a forest setting, dialogue, and a visually enlarged speech balloon. A multimodal model might correctly recognize the figures, landscape, and speech. It could also infer embarrassment, exclusion, or relationship tension. Under MEBI, those emotional labels cannot enter the evidence ledger as observations. Students would first tag visible posture, spatial separation, wording, and the relative graphic salience of the balloon; only then could they debate what these details suggest.

S3 shifts toward handwritten reflection and sketch-like annotation. Here typography and the embodied appearance of writing are part of the narrative voice. A text-only transcription would preserve lexical content while losing scale, line, color, and the relation between writing and drawings. A multimodal model may be useful precisely because it can notice these visual features, but it may also normalize them into prose and thereby erase the formal difference that students are meant to interpret. MEBI therefore requires separate [V] and [T] entries before a claim about tone or self-presentation is accepted.

S5–S6 contain diagrammatic self-analysis: lists, arrows, "detective"-like visual motifs, icons, and caricatural drawings are integrated with the handwritten narrative. These pages are epistemically interesting because the character's own attempt to infer motives is represented graphically. An AI system can easily produce a second-order psychological explanation—an interpretation of a character who is already interpreting herself. The evidence audit should therefore ask two distinct questions: What does the page show the character claiming? and What does the page invite the

reader to conclude about that claim? Collapsing these levels would erase irony, self-correction, or narrative distance.

S7 represents message-like exchanges in colored bubbles alongside drawings and reflective handwriting. The digital-interface form is not transparent evidence of the social world; it is a represented communication environment inside the narrative. A model that summarizes "what happened" may merge message content with assumptions about the relationship between senders. MEBI tags the bubbles as [D] and their wording as [V], requiring any claim about intention or relational power to be separately warranted.

S8 adopts a profile-card or social-media-like visual grammar. Interface conventions can strongly cue contemporary readers, but they can also prompt AI systems to import generic knowledge about real social platforms. The audit should therefore prohibit claims about features, metrics, or platform functions not visibly present in the artifact. This is a direct application of visual-hallucination evidence: prior knowledge of what an interface "usually" contains is not evidence of what the page actually contains.

S9–S11 return to sequential interpersonal scenes, including domestic interactions and device-mediated communication. Here [S] becomes particularly important. Meaning depends on what is shown before and after a close-up, who occupies shared or separate panels, and how spoken language interacts with the pictured response. A model-generated character analysis should be tested against the whole local sequence rather than one decontextualized facial expression.

S12–S13 broaden the social and outdoor action and combine group composition with note-like elements. The shift itself can become evidence if the learner can locate it across the sequence. MEBI therefore allows a larger-scale interpretive claim—such as movement from private reflection toward shared action—only when the student can connect multiple scan locations and acknowledge counterevidence.

Table 3 turns these observations into teachable prompts.

Table 3. Worked-case tasks and evidence audits

Scan location	AI-supported task	Foreseeable failure mode	Required student audit
S3	Describe the relation between handwriting, sketches, and verbal reflection.	AI converts graphic form into a prose summary and treats tone as self-evident.	Separate [V] from [T]; identify which tonal claim is inferred.
S5–S6	Generate two readings of the diagrammatic self-analysis.	AI presents a psychological diagnosis or one definitive motive.	Distinguish the character's represented inference from the reader's inference; reject diagnostic language not warranted by the text.
S7	Explain how message-like bubbles contribute to the scene.	AI invents sender intention or off-page context.	Mark [D] and [V] evidence; list what remains unknown.
S8	Compare the profile-like cards with conventional narration.	AI imports functions from real platforms that are not depicted.	Verify every interface claim visually; delete unsupported platform assumptions.
S9–S11	Propose a character-state interpretation across the sequence.	AI overweights one facial expression or dialogue line.	Require at least three pieces of evidence across two modes and one counterexample.
S12–S13	Formulate a claim about change across the excerpt.	AI narrates a linear development that the local excerpt cannot establish.	Use multiple scan locations and qualify the scope of the claim to "within this excerpt."

A 60–90 minute classroom sequence

MEBI can be implemented without turning the lesson into a lesson about prompting. A practical sequence is:

1. Human close reading (10–15 minutes). Students inspect a selected spread or scan location, create five tagged observations, and write one provisional claim without AI.
2. Constrained AI description (10 minutes). The class submits only the selected, legally and institutionally permitted excerpt or uses a teacher-prepared description if image upload is inappropriate. Students compare the model's tagged description with their own ledger.
3. Error and uncertainty audit (10–15 minutes). Students mark model statements as supported, uncertain, unsupported, or contradicted and record at least one visual or verbal correction.
4. Counter-reading generation (10–15 minutes). AI is asked for two incompatible interpretations with evidence and counterevidence. Students choose neither, one, or parts of both.
5. Student-authored interpretation (15–25 minutes). Students write a paragraph or short response grounded in the primary artifact and attach a brief AI-use note listing one accepted, one modified, and one rejected AI contribution.
6. Teacher/peer challenge (optional 10 minutes). A peer or teacher identifies the weakest warrant and asks the student to strengthen, qualify, or withdraw it.

The sequence intentionally includes points at which AI is not used. This is not a symbolic gesture. It creates evidence that the student can notice, reason, and decide without continuous assistance.

Assessment criteria

Because MEBI is proposed rather than psychometrically validated, the following rubric is a design heuristic for classroom piloting, not a validated measurement instrument.

Table 4. Proposed assessment rubric for an MEBI task

Criterion	Weight	High-quality evidence of learning
Primary-artifact grounding	30%	Claims are tied to accurately located verbal and/or visual evidence; no invented details.
Multimodal integration	20%	The response explains relations across at least two relevant modes rather than listing them separately.
Inference discipline	20%	Observation, inference, and interpretation are distinguishable; uncertainty is appropriately qualified.
Counter-reading and evaluation	15%	The student considers a plausible alternative or counterevidence and explains why the final reading is preferable or remains open.
Human authorship and AI audit	15%	The final wording is student-authored and the process note shows critical acceptance, modification, or rejection of AI output.

DISCUSSION

From text-first AI use to multimodal traceability

The first contribution of MEBI is to extend evidence-first literature pedagogy from textual sequence to multimodal traceability. TASU (Mašát et al., 2026) establishes the crucial temporal rule that the text comes before AI and that student and teacher judgment follow. MEBI retains that architecture but expands what counts as locatable evidence. In a graphic narrative, a page citation alone may be insufficient if a claim depends on the relation between lettering and image, between panels, or between a simulated interface and surrounding narration. The modal evidence ledger makes the warrant inspectable without pretending that the modes are independent in actual reading.

This extension is important because multimodal AI can create a false sense of evidential completeness. When a model can process an image, students may assume that it has "read the whole page." Research on visual hallucination shows why this assumption is unsafe (Guan et al., 2024; Huang et al., 2024). More fundamentally, visual access does not tell the learner which aspects of the page the model used, misread, or silently ignored. MEBI therefore treats traceability as a pedagogical requirement even when the model is technically multimodal.

AI as a contrastive reader rather than an authority

The second contribution is a functional reframing of AI in literary analysis. Bellot et al. (2025) demonstrate the value of AI-generated alternatives when they become objects of analysis. MEBI generalizes this contrastive function: the system can supply a second description, a rival hypothesis, a counterargument, or a deliberately different reading. Its value comes from creating material against which students exercise judgment.

This position also accommodates Chan and Wong's (2025) finding that GenAI can perform competently across many levels of literary analysis while still showing limitations in depth. The educational objective should not be to preserve student agency by artificially restricting AI to trivial tasks. A strong model can be given genuinely demanding tasks, provided the student must evaluate the output against the artifact and retain authorship of the final reasoning. Human-centered design is compatible with powerful automation when control and accountability remain meaningful (Shneiderman, 2022).

Performance is not the learning outcome

A recurring risk in AI-supported humanities teaching is evaluating the final text while ignoring how it was produced. This is precisely the performance–learning problem identified outside literature. Bastani et al. (2025) show that students can perform better with AI while learning less for subsequent unassisted work. Fan et al. (2025) similarly separate improved writing products from knowledge gain and transfer, and Yan et al. (2025) argue that the field must distinguish observable task success from durable changes in knowledge and skill.

For literature education, this distinction has a direct assessment consequence. A polished AI-assisted interpretation may be a poor indicator of the student's ability to read. MEBI therefore relocates assessment toward process artifacts that are lightweight but epistemically informative: the first-reading note, the modal ledger, a correction of AI error, a warrant, a counter-reading, and a rejection log. These do not eliminate opportunities for misuse, but they align assessment more closely with the reasoning the course intends to develop.

The framework also explains why blanket "use AI to improve your essay" assignments are educationally underspecified. Such instructions optimize the product without specifying what thinking must remain with the learner. In contrast, MEBI allocates functions: AI may widen hypotheses and challenge a reading; the learner must locate, discriminate, warrant, and decide.

Preserving ambiguity as a disciplinary objective

Literary ambiguity deserves special treatment because it is often misclassified as a problem to be solved. In many educational AI applications, uncertainty is undesirable: a factual answer should be correct, an equation should be solved, and a misconception should be corrected. Literature often requires a different epistemic stance. Multiple readings can remain viable when they are differently warranted, and a responsible interpretation may end by qualifying rather than resolving uncertainty.

Generative models are optimized to continue discourse coherently and often respond to requests for "the meaning" by producing a unified thematic account. That fluency can create premature closure. P4 therefore requires counter-readings and counterevidence. The point is not to produce relativism in which every reading is equally acceptable. On the contrary, evidence-bounded plurality makes evaluation stricter: students must compare how well different claims account for the artifact.

The Dåsnes worked case illustrates this need. Handwritten self-commentary, diagrammatic reflection, social-media-like forms, and sequential images can represent different levels of awareness and voice. A psychologically neat explanation may erase the tension between what a character says, depicts, performs, or retrospectively narrates. Requiring two readings and an evidence audit protects the formal complexity that makes the work pedagogically valuable.

AI literacy as disciplinary epistemic vigilance

UNESCO's human-centered framing and teacher competency model are often discussed at the level of policy, but literature teaching offers a concrete site for operationalization. Human agency becomes observable when students make a first reading before AI, reject unsupported model claims, and own the final interpretation. Ethics becomes observable when copyrighted materials and personal data are handled appropriately, when AI use is disclosed, and when models are not treated as invisible co-authors. AI foundations become relevant when students understand that visual access does not guarantee visual truth. AI pedagogy becomes the design of tasks in which assistance augments rather than substitutes for reading.

This approach also makes AI literacy less dependent on ephemeral prompt techniques. Prompting matters, but a student who can elicit an impressive answer without knowing how to audit it is not fully literate in an educational sense. The durable competence is epistemic vigilance: knowing when an output requires checking, what source can check it, and what uncertainty remains after checking. Zhai et al. (2024) show why this matters for analytical and critical thinking; Huang et al. (2024) and Guan et al. (2024) extend the issue to visual claims.

The European AI governance context strengthens the institutional relevance of such competencies. After the July 2026 amendments to Article 4 of the AI Act, providers and deployers remain required to take measures supporting AI literacy, while national market-surveillance authorities began supervising and enforcing the provision from 2 August 2026 (European Commission, n.d.). A literature department does not need to turn this into legal instruction. It can respond by defining transparent, auditable classroom routines for AI use and by documenting which interpretive functions remain human responsibilities.

Practical governance for institutions and teachers

MEBI implies several low-cost governance choices. First, institutions should distinguish permitted assistance from assessed responsibility. A student may be permitted to use AI for alternative interpretations while still being individually responsible for source evidence and final wording. Second, assignment instructions should specify whether visual materials may be uploaded to third-party systems. Copyright, licensing, privacy, and institutional data policies apply before pedagogical convenience. Instructors can avoid unnecessary uploads by using public-domain works, licensed excerpts, institutionally approved tools, or teacher-prepared descriptions where appropriate.

Third, model outputs should not be used as an invisible standard for grading literary interpretations. An AI-generated "reference answer" can reproduce the same closure and hallucination problems the framework is designed to resist. AI may assist teachers with question generation or comparative examples, but teacher judgment should remain accountable and the criteria disclosed to students.

Fourth, access and inclusion must be considered at the task-design level. A student should not receive a lower grade because they lack a paid multimodal model. The core learning sequence should have a low-tech path: teacher-provided AI outputs can be audited by the entire class, or pairs can work from the same archived output. This is compatible with human-centered and equity-oriented AI governance (Miao & Holmes, 2023).

Relationship to multimodal pedagogy

Istrate (2026) argues that GenAI makes multimodal production easier while increasing the need for pedagogical design. MEBI applies the same principle to multimodal reception. The fact that a system can describe a page does not mean that it has designed a useful reading experience, and the fact that it can produce an interpretation does not mean that the learner has interpreted. The

teacher's role shifts from guarding access to designing epistemic friction: requiring the learner to stop, locate evidence, compare claims, and justify decisions.

This also avoids a simplistic "human versus AI" framing. The human contribution is not valuable merely because it is human; students can also overgeneralize, misread images, or settle too quickly on a theme. AI can expose those weaknesses by generating a counter-reading or noticing a formal feature a student missed. MEBI is therefore reciprocal: students audit AI, but AI can also be used to stress-test student interpretations. The asymmetry is responsibility. The system does not bear educational accountability for the judgment; the student and teacher do.

Limitations and Research Agenda

MEBI should be interpreted as a conceptual design proposal rather than an empirically validated intervention. Six limitations delimit its present claims. First, the framework has not yet been tested with students in a controlled or naturalistic classroom study; no claim of effectiveness, learning gain, or transfer is therefore warranted. Second, the worked case draws on one excerpt from the Czech translation of a Norwegian graphic narrative, so transfer across languages, cultural contexts, age groups, and curricula remains untested. Third, MEBI is currently operationalized most fully for graphic narratives; its applicability to picture books, illustrated novels, manga, digital fiction, poetry with visual form, and other multimodal literary genres should be demonstrated rather than assumed. Fourth, the proposed assessment rubric is a design heuristic, not a psychometrically validated instrument; its weighting, inter-rater reliability, construct validity, and sensitivity to change require empirical evaluation. Fifth, scalability remains uncertain. The five evidence tags and audit routines may impose additional instructional and cognitive load, particularly for teachers or students unfamiliar with semiotic terminology, multimodal analysis, or GenAI auditing, and implementation studies should examine training requirements and time costs. Sixth, the evidential basis combines studies from different disciplines and educational levels, while multimodal model capabilities are changing rapidly. Cross-domain findings are therefore used to identify plausible failure modes, not to infer literature-specific effect sizes, and model-specific error rates should not be treated as stable. The present study also does not estimate differential effects associated with prior literary competence, AI literacy, language background, disability, or access.

Future research should therefore proceed along three complementary lines: classroom effectiveness, cross-genre and cross-context transfer, and implementation feasibility. A feasible study would compare at least three conditions: human-only close reading, unstructured multimodal AI assistance, and MEBI-structured assistance. Outcomes should be separated into assisted performance and unassisted transfer. Measures could include evidence accuracy, unsupported-inference rate, quality of warrants, interpretive plurality, delayed transfer to a new graphic narrative, and student agency or self-regulation. Process data should capture which AI suggestions students accept, modify, and reject. Such a design would directly test whether evidence-bounded multimodal assistance avoids the performance–learning divergence documented in other domains.

Replication should then examine picture books, illustrated novels, manga, born-digital fiction, and other multimodal literary forms across languages, age groups, curricular settings, and levels of

teacher experience. Such studies should report not only student outcomes but also implementation time, teacher usability, training requirements, and the extent to which the modal evidence tags require adaptation for particular genres or learner populations.

A second research direction is model auditing. The same literary pages could be submitted, where legally permitted, to several multimodal systems using standardized prompts. Researchers could code recognition errors, invented details, unsupported psychological inferences, flattening of visual form into prose, and sensitivity to counterevidence. This would create a literature-specific benchmark without confusing benchmark performance with pedagogy. A third direction is teacher education: preservice and practicing teachers could be studied as they design MEBI tasks, revealing which forms of AI literacy are most difficult to translate into classroom instructions.

CONCLUSION

Multimodal generative AI creates a genuine opportunity for literature education because it can interact with the same mixtures of word, image, sequence, and interface that characterize many contemporary literary works. The opportunity is also an epistemic risk. Visual access can make AI output seem better grounded than it is, while fluent interpretation can hide the transition from observation to inference. Current evidence across education shows that AI-supported performance and learning are not equivalent, that students can become reliant on assistance, and that pedagogical guardrails matter (Bastani et al., 2025; Darvishi et al., 2024; Fan et al., 2025; Yan et al., 2025). Literature-specific research indicates that AI can be productive when its outputs are treated as comparative objects rather than authoritative readings (Bellot et al., 2025; Chan & Wong, 2025; Mašát et al., 2026).

MEBI translates these findings into five principles: human-first encounter, modal traceability, observation–inference separation, preserved interpretive plurality, and human adjudication and authorship. Its five-phase sequence—Encounter → Describe → Hypothesize → Audit → Adjudicate & Author—keeps AI available for demanding intellectual work while retaining the disciplinary responsibilities of close reading and evidence-based argument.

The worked case from Dåsnes's Na kočičí svědomí demonstrates why this matters. Handwriting, sequential panels, diagrammatic self-commentary, digital-message forms, and social-media-like layouts cannot be reduced to extracted prose without loss. A responsible AI-supported reading must therefore show not only what an interpretation says but where its evidence resides and how the inferential move is made.

The framework should be understood as a researchable design proposal, not as evidence of effectiveness. Its value is falsifiable and practical: future studies can test whether students using MEBI make fewer unsupported claims, produce stronger warrants, preserve more independent judgment, and transfer interpretive skills better than students using unstructured AI assistance. Until such evidence is available, the most defensible principle for multimodal AI in literature education is not "AI first" or "AI never," but AI bounded by the artifact and answerable to the reader.

Funding: This work was supported by the project HumanAID: AI zaměřená na člověka pro udržitelnou a adaptabilní společnost (reg. no. CZ.02.01.01/00/23_025/0008691).

Ethics statement: The study did not involve human participants, personal data, or an experimental educational intervention. The worked case analyzes a published literary work and does not reproduce copyrighted page images.

Data availability: No new dataset was generated. The worked case is based on the cited published edition of Nora Dåsnes's Na kočičí svědomí and the scan excerpt supplied for methodological analysis.

Author contributions: Milan Mašát: Conceptualization; methodology; investigation; writing—original draft; writing—review and editing.

Artificial Intelligence (AI) disclosure: OpenAI's ChatGPT was used during manuscript preparation to assist with structural drafting, language editing, and organization of evidence. All direct quotations, bibliographic metadata, legal and policy claims, and substantive source-dependent statements were checked by the author against the cited publications, DOI records, official institutional sources, or the primary literary work. AI was not listed as an author and did not make final interpretive, methodological, or scholarly judgments. Responsibility for the integrity, originality, accuracy, and ethical compliance of the manuscript remains with the human author.

REFERENCES

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

Bellot, A. R., Gutiérrez-Colón Plana, M., & Baran, K. A. (2025). Redefining literature education: The role of ChatGPT in undergraduate courses. International Journal of Artificial Intelligence in Education, 35, 3185–3201. https://doi.org/10.1007/s40593-025-00497-3

Chan, C. K. Y., & Wong, H. Y. H. (2025). Assessing the ability of generative AI in English literary analysis through Bloom's Taxonomy. Discover Computing, 28, 127. https://doi.org/10.1007/s10791-025-09572-8

Cohn, N. (2013). The visual language of comics: Introduction to the structure and cognition of sequential images. Bloomsbury Academic.

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. Computers & Education, 210, 104967. https://doi.org/10.1016/j.compedu.2023.104967

Dåsnes, N. (2024). Na kočičí svědomí (J. Jindřišková, Trans.). Albatros. (Original work published 2020 as Ti kniver i hjertet, Aschehoug.)

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, 105224. https://doi.org/10.1016/j.compedu.2024.105224

European Commission. (n.d.). AI literacy – Questions & answers. Shaping Europe's digital future. Retrieved August 9, 2026, from https://digital-strategy.ec.europa.eu/en/faqs/ai-literacy-questions-answers

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology, 56(2), 489–530. https://doi.org/10.1111/bjet.13544

Guan, T., Liu, F., Wu, X., Xian, R., Li, Z., Liu, X., Wang, X., Chen, L., Huang, F., Yacoob, Y., Manocha, D., & Zhou, T. (2024). HallusionBench: An advanced diagnostic suite for entangled language hallucination and visual illusion in large vision-language models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 14375–14385). https://doi.org/10.1109/CVPR52733.2024.01363

Huang, W., Liu, H., Guo, M., & Gong, N. Z. (2024). Visual hallucinations of multi-modal large language models. In Findings of the Association for Computational Linguistics: ACL 2024 (pp. 9614–9631). https://doi.org/10.18653/v1/2024.findings-acl.573

Istrate, O. (2026). Generative AI and multimodal pedagogy: Implications for teaching, learning, and assessment. Frontiers in Education, 11, 1828953. https://doi.org/10.3389/feduc.2026.1828953

Kress, G. (2010). Multimodality: A social semiotic approach to contemporary communication. Routledge.

Mašát, M., Kopecký, K., Krejčí, V., & Voráč, D. (2026). Anchored to the text, owned by the student: A policy & practice review for generative AI in literature education. Frontiers in Education, 11, 1805617. https://doi.org/10.3389/feduc.2026.1805617

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

Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Mikkonen, K. (2017). The narratology of comic art. Routledge.

Painter, C., Martin, J. R., & Unsworth, L. (2013). Reading visual narratives: Image analysis of children's picture books. Equinox.

Sanz-Tejeda, A., Domínguez-Oller, J. C., Baldaquí-Escandell, J. M., Gómez-Díaz, R., & García-Rodríguez, A. (2026). The impact of generative AI on academic reading and writing: A synthesis of recent evidence (2023–2025). Frontiers in Education, 10, 1711718. https://doi.org/10.3389/feduc.2025.1711718

Shneiderman, B. (2022). Human-centered AI. Oxford University Press.

Torres Vico, M., & Fernández-Herrero, J. (2026). Impact of large multimodal language models (MLMs) with screen-vision in higher education: A quasi-experimental study. Technology, Knowledge and Learning. Advance online publication. https://doi.org/10.1007/s10758-026-10006-7

Yan, L., Greiff, S., Lodge, J. M., & Gašević, D. (2025). Distinguishing performance gains from learning when using generative AI. Nature Reviews Psychology, 4, 435–436. https://doi.org/10.1038/s44159-025-00467-5

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments, 11, 28. https://doi.org/10.1186/s40561-024-00316-7