Artificial Intelligence Education Studies
Volume 1, Issue 2, 2025, Page 48-62
https://doi.org/10.6914/aiese.010204

Artificial Intelligence Education Studies

Addressing "Hallucinations" in AI-Generated Content: Strategies for Developing Student Fact-Checking and Information Evaluation Skills

Yingzhe LI

Hong Kong Center of Social Sciences Email: lisophiahk@gmail.com https://orcid.org/0009-0008-3326-6397

Abstract

The integration of Artificial Intelligence (AI) into educational settings presents transformative opportunities alongside significant challenges, notably the phenomenon of AI "hallucinations"—the generation of plausible yet fabricated or inaccurate information. This paper addresses the critical need to equip students with robust fact-checking and information evaluation skills to navigate these AI-generated outputs. It begins by deconstructing AI hallucinations, examining their definitions, technical underpinnings, and profound implications for student learning and epistemic trust. Foundational literacies, including critical thinking, information literacy, and digital citizenship, are explored as essential prerequisites for effectively engaging with AI. The core of the paper details a range of pedagogical strategies, such as explicit instruction in verification techniques like lateral reading and the SIFT method, the cultivation of critical engagement through healthy skepticism, and the application of Socratic pedagogy to interrogate AI claims. Furthermore, it considers various approaches to assessing student competencies in this evolving information ecosystem, from adapting established critical thinking tests to developing authentic, performance-based tasks and utilizing pre-post intervention designs. The discussion highlights implementation challenges, ethical dimensions, and crucial future research trajectories. Ultimately, this paper argues for a proactive, multi-faceted approach to fostering the skills necessary for students to become discerning consumers and responsible users of AI-generated content, ensuring they can harness AI's benefits while mitigating its risks.

Keywords artificial intelligence in education; AI hallucinations; critical thinking; fact-checking; information literacy

1 Introduction

1.1 The Proliferation of AI in Educational Environments

Artificial Intelligence (AI) is rapidly reshaping numerous sectors, with education being no exception[1]. The integration of AI tools, including generative AI, chatbots, intelligent tutoring systems, and adaptive learning platforms, is becoming increasingly prevalent across diverse educational levels, from K-12 to higher education, and spanning various academic disciplines[1]. These technologies promise to revolutionize traditional teaching and learning paradigms by offering personalized learning experiences, enhancing student engagement, and potentially improving academic outcomes[1]. AI-driven tools can analyze vast amounts of student data to identify learning gaps, tailor interventions, provide real-time feedback, and customize learning pathways to meet individual student needs and paces[1]. Studies suggest that such personalized environments can foster improved self-efficacy and a more positive attitude toward education[1].

However, this increasing reliance on AI in education is not without its complexities. While the potential benefits are significant, the effective and ethical integration of AI requires careful consideration of its limitations and the

© The Authors 2025. Article History Received: April 19, 2025; Accpted: May 10, 2025; Published: June 30, 2025. ***Artificial Intelligence Education Studies***, ISSN 3079-8086 (print), ISSN 3079-8094 (online), a Quarterly, founded on 2025, published by Creative Publishing Co., Limited. Email: wtocom@gmail.com, https://ai-es.org, https://cpcl.hk.

new challenges it introduces. The allure of AI's capabilities, particularly in automating tasks and providing instant information, necessitates a deeper understanding of how these tools function and the nature of the content they produce.

One of the most compelling aspects of AI in education is its capacity for personalization. AI algorithms can adapt content delivery and support based on individual learning patterns and knowledge levels, moving beyond a one-size-fits-all model to foster a more inclusive and potentially effective learning environment[2]. Yet, this very strength in personalization carries an inherent risk. If the AI generates inaccurate or misleading information—a phenomenon known as "hallucination"—this misinformation can also be personalized. Such tailored falsehoods, presented with the same confidence as factual information, might be more readily accepted by students, especially if they perceive the AI as understanding their specific needs or tailoring content uniquely for them. This could lead to the formation of deeply ingrained, individualized misconceptions that are arguably more challenging to identify and correct than errors encountered in more generalized, traditional educational resources. The potential for AI to deliver personalized misinformation underscores the heightened need for students to possess strong critical evaluation skills.

1.2 The Emergent Challenge of AI-Generated "Hallucinations"

A significant impediment to the reliable and uncritical use of AI in educational contexts is the phenomenon of AI-generated "hallucinations." An AI hallucination occurs when an AI system produces fabricated, nonsensical, or inaccurate information, yet presents this information with a high degree of confidence, making it difficult for users, especially students, to discern its reliability[3]. These outputs can appear plausible and contextually coherent, further complicating the task of verification[4]. The confident presentation of false information poses a direct threat to the integrity of the learning process, as students may unknowingly incorporate inaccuracies into their knowledge base.

The problem is exacerbated by the nature of many AI tools, particularly Large Language Models (LLMs), which are designed to generate human-like text based on patterns learned from vast datasets. These systems do not "understand" information in the human sense but rather predict sequences of words, making them susceptible to generating content that is syntactically correct and fluent but semantically flawed or entirely untrue[4]. This emergent challenge requires a proactive response from the education sector.

Frequent encounters with AI hallucinations, if not adeptly managed through the development of student evaluation skills, also carry the risk of eroding trust in AI technologies. If students and educators repeatedly find AI outputs to be unreliable or misleading, their confidence in these tools could diminish significantly. This lack of trust might lead to a reluctance to utilize AI even for tasks where it could offer genuine educational benefits, thereby hindering the positive transformative potential that AI holds for education. Developing students' abilities to critically assess AI-generated content is therefore not merely about mitigating a risk but also about fostering an environment where AI can be integrated effectively and responsibly.

1.3 Thesis: The Imperative for Cultivating Student Fact-Checking and Information Evaluation Skills

The prevalence of AI hallucinations and their potential impact on learning and epistemic trust underscore the urgent need for a pedagogical shift. It is no longer sufficient for students to be passive consumers of information; they must become active, critical evaluators of all information sources, including those generated by AI. This paper argues that the rise of sophisticated AI tools necessitates a concerted effort to cultivate robust fact-checking, information evaluation, and critical thinking skills in students. Equipping students with these competencies is essential not only for maintaining academic integrity and fostering responsible digital citizenship but also for developing sound epistemic foundations—their understanding of how knowledge is constructed, validated, and applied. The ability to critically dissect and verify AI-generated content is paramount in an era where the lines between authentic and artificial information are increasingly blurred.

1.4 Paper Overview

This paper will systematically address the challenge of AI hallucinations in education. Section 2 will deconstruct AI "hallucinations," offering definitions, exploring the technical mechanisms behind their occurrence, and discussing their educational implications. Section 3 will examine the foundational literacies—critical thinking, information literacy, and digital citizenship—that underpin students' ability to navigate the complexities of AI-generated content.

Section 4 will propose and detail specific pedagogical strategies designed to foster fact-checking and information evaluation capabilities, including explicit instruction in verification techniques and the use of Socratic pedagogy. Section 5 will explore methods for assessing these crucial student competencies in an AI-infused information ecosystem. Section 6 will discuss broader challenges, ethical considerations, and future research trajectories in this domain. Finally, Section 7 will conclude by reaffirming the necessity of these skills and calling for collaborative action to prepare students for the realities of an AI-driven world.

2 Deconstructing AI "Hallucinations": Definitions, Mechanisms, and Educational Implications

2.1 Defining AI "Hallucinations"

AI "hallucinations" are broadly defined as instances where an AI system, particularly a generative AI model, produces information that is fabricated, nonsensical, or factually inaccurate, yet presents it as if it were correct and reliable[3]. These outputs are often convincing and contextually coherent with the prompt or ongoing dialogue but lack a factual basis or are independent of the user's input in a way that leads to falsehoods[4]. The confident delivery of such misinformation makes it particularly challenging for users, especially students who may lack deep domain knowledge or advanced critical evaluation skills, to identify the inaccuracies[3].

Researchers have further classified hallucinations. For instance, one distinction is made between intrinsic hallucinations, where the AI's output directly contradicts the source content it was supposedly based on (or its own conversational history), and extrinsic hallucinations, where the output introduces information whose accuracy cannot be verified against the provided source content or conversational history, effectively being new, unverified, and potentially fictional content[4]. Intrinsic hallucinations point to a misinterpretation or misrepresentation of available information, while extrinsic hallucinations often involve the generation of entirely new, albeit plausible-sounding, details.

2.2 Considering "Confabulation" as a More Apt Descriptor

While the term "hallucination" is widely used, some researchers argue that "confabulation" might be a more technologically accurate descriptor for the phenomenon observed in AI[4]. In clinical psychology, "hallucination" refers to a sensory experience without an external stimulus, often associated with certain neurological or psychiatric conditions, implying a conscious perceptual process[4]. Generative AI, however, lacks consciousness, subjective experience, or awareness. Therefore, attributing "hallucinations" to AI can be misleading, anthropomorphizing the technology by suggesting a human-like cognitive error.

"Confabulation," in a clinical context, involves the brain creating fabricated or distorted memories, influenced by existing knowledge, experiences, or contextual cues, often to fill in gaps in memory without the conscious intent to deceive[4]. This aligns more closely with how LLMs operate: they reconstruct information based on patterns and relationships in their vast training data, influenced by the immediate prompt (contextual cues). When these reconstructions lead to inaccuracies or fabrications, it is more akin to a system "confabulating" plausible-sounding outputs based on its learned data and probabilistic nature, rather than "perceiving" something that isn't there[4]. Adopting the term "confabulation" could help frame the issue more accurately as a limitation of the technology's information processing rather than a flaw in a non-existent AI consciousness.

This distinction has pedagogical importance. If students understand AI errors as "confabulations" rooted in the way these systems are trained and how they generate responses, they may be less inclined to view AI as an infallible oracle. This encourages a shift from merely teaching students to identify errors in AI output to fostering a more fundamental understanding of how AI systems work, including the critical influence of their training data and algorithmic processes. Such an understanding promotes a deeper, more structural critique of AI's capabilities and inherent limitations.

2.3 Technical Underpinnings: Probabilistic Models and Training Data Issues

AI hallucinations, or confabulations, are not random occurrences but are intrinsically linked to the fundamental architecture and training of LLMs. These models are probabilistic; they generate text by predicting the most likely next word (or token) in a sequence, based on the patterns learned from the massive datasets they were trained on[4]. This probabilistic nature, while enabling fluent and coherent text generation, is a primary source of hallucinations.

Several factors related to training data contribute significantly to the likelihood of hallucinations. First, Frequency and Accessibility of Patterns: Patterns, expressions, or concepts that appear frequently in the training data can be easily and quickly accessed by the model during response generation. This can lead the AI to output such information even if it is not contextually accurate or relevant to the user's prompt, simply because it is a high-probability sequence[4]. Second, Conflicting Information: Large training datasets inevitably contain conflicting, inconsistent, or contradictory information. These internal tensions within the data can cause the AI to generate responses that are illogical or self-contradictory[4]. Third, Outdated, Incomplete, or False Information: The knowledge cutoff date of an AI model means its training data may be outdated. Furthermore, if the training data itself contains inaccuracies, biases, or incomplete information, the AI is likely to reproduce and perpetuate these flaws in its outputs[4]. The sheer volume and often unstructured nature of the text corpora used for training make it exceedingly difficult to curate perfectly accurate and unbiased datasets. Fourth, Algorithmic Bias: Beyond the data itself, the algorithms used to train and operate these models can introduce or amplify biases. Assumptions made in the design of these algorithms can lead to skewed or systematically inaccurate results, contributing to what might appear as hallucinations but are, in fact, reflections of inherent biases[4].

In essence, the large, often uncurated text corpora, the stochastic (probabilistic) behavior of the models, and the inherent weaknesses in current language modeling techniques collectively contribute to the generation of hallucinatory content[4]. Recognizing these technical origins is crucial for developing informed educational strategies. For instance, the fact that hallucinations often stem from issues like outdated or biased training data provides concrete "teachable moments." When an AI produces an outdated fact or a biased statement, educators can guide students to question why the AI might have generated that output, prompting discussions about data quality, data bias, the importance of curated datasets, and the recency of information—all fundamental components of data literacy that extend well beyond AI interaction.

2.4 Impact on Student Learning and Epistemic Trust

The implications of AI hallucinations for student learning are multifaceted and potentially detrimental. When AI confidently presents misinformation, students, particularly those who are less experienced or knowledgeable in a specific subject area, may readily accept these falsehoods as facts[3]. This can lead to the assimilation of incorrect concepts, the development of flawed mental models, and difficulties in subsequent learning that builds upon these erroneous foundations.

Moreover, an uncritical reliance on AI tools without adequate verification habits can inadvertently diminish students' own critical thinking skills. If AI is perceived as a source of quick and easy answers, students may become less inclined to engage in the effortful cognitive processes of analysis, evaluation, and independent problem-solving[1]. This "cognitive offloading" can hinder the development of the very skills needed to identify and counteract AI hallucinations.

Beyond the immediate impact on knowledge acquisition and skill development, AI hallucinations can have broader implications for students' epistemic trust—their understanding of knowledge, the credibility of different information sources, and their confidence in their own ability to discern truth. Repeated exposure to unreliable AI outputs, especially if not accompanied by strategies for critical evaluation, could lead to confusion, cynicism, or an unhealthy skepticism towards all information sources. Conversely, it could also foster an over-reliance on flawed AI, undermining trust in more traditional, curated sources of knowledge.

The creative potential of AI, even through its "hallucinations"[4], presents a pedagogical paradox. Students must learn to differentiate between contexts where AI's generative, sometimes confabulatory, capabilities might be harnessed for creative exploration versus contexts that demand stringent factual accuracy. This requires a sophisticated level of discernment—a metacognitive skill to assess when AI's imaginative errors are a feature to be played with versus a bug to be rigorously fact-checked. This nuanced understanding is more complex than simply labeling all hallucinations as inherently negative.

3 Foundational Literacies: Critical Thinking, Information Literacy, and Digital Citizenship in the Age of AI

3.1 The Enduring Role of Critical Thinking in Evaluating AI Outputs

Critical thinking, broadly understood as the ability to engage in reflective and independent thinking, is paramount in the age of AI. The Delphi Report's consensus definition describes critical thinking as "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment was based"[7]. This definition highlights the active, intentional nature of critical thought.

AI tools, despite their sophistication, are not substitutes for human critical thinking; rather, their proliferation makes such thinking even more indispensable[1]. Students must employ critical thinking skills to sift through AI-generated information, question its assertions, and assess its validity. Core critical thinking skills, such as those measured by instruments like the California Critical Thinking Skills Test (CCTST)—including analysis, interpretation, inference, evaluation, and explanation[8]—and the Watson-Glaser Critical Thinking Appraisal (WGCTA)—which assesses inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments[10]—are directly applicable to dissecting AI claims. For example, analysis helps break down complex AI responses, interpretation aids in understanding their meaning and potential implications, inference allows for drawing logical conclusions about their plausibility, and evaluation is crucial for judging their credibility and the strength of any implicit arguments.

Beyond these cognitive skills, the development of critical thinking dispositions is equally vital. Attributes such as inquisitiveness (the curiosity to question), truth-seeking (the desire for accuracy), open-mindedness (willingness to consider different perspectives and the possibility of AI error), analyticity (prizing the application of reason and evidence), systematicity (being organized and focused in inquiry), and confidence in reasoning are crucial for motivating students to actively engage with and scrutinize AI-generated content rather than passively accepting it[7].

However, a significant concern arises from the potential for "cognitive offloading"[6]. If students become overly reliant on AI for tasks that traditionally help build critical thinking—such as summarizing complex texts, drafting arguments, or solving problems—they may miss crucial opportunities for cognitive development. This could lead to an atrophy of the very skills needed to critically evaluate the AI's outputs, including its hallucinations. This creates a potential vicious cycle: reduced critical thinking leads to greater reliance on AI, which further diminishes critical thinking. Educational approaches must therefore emphasize the strategic use of AI as a tool to augment human intellect and support critical inquiry, rather than as a replacement for fundamental cognitive effort.

3.2 Information Literacy as a Prerequisite for Navigating AI-Generated Content

Information literacy encompasses the set of abilities requiring individuals to "recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information"[13]. In an environment where AI can instantly generate vast amounts of information, these skills are more critical than ever. Information literacy provides the framework for students to approach AI-generated content with a discerning eye.

Teaching information literacy with a fact-checking curriculum can significantly help students evaluate digital information, including AI outputs, more accurately[14]. Key components of information literacy, such as evaluating source credibility (e.g., assessing the author, purpose, objectivity, accuracy, and currency of information)[15], are directly transferable to the assessment of AI-generated content. While the "source" is now an algorithm rather than a human author or traditional publisher, the underlying principles of evaluation remain relevant, albeit requiring adaptation.

Traditional information literacy frameworks often focus on evaluating human authors, their credentials, affiliations, and the reputation of publishing venues. When the "author" is an AI, the focus of evaluation must necessarily shift. Instead of asking "Who is the author?", students need to learn to ask questions like: "What is known about the AI model that generated this information (e.g., its developer, its training data, its known limitations or biases)?" "What was the prompt given to the AI, and how might that have influenced the output?" "What are the processes by which this AI constructs answers?" This implies an evolution of information literacy pedagogy to incorporate elements of "algorithmic literacy"—an understanding of how algorithms, including those in AI, curate, generate, and sometimes distort information. The ability to critically assess the AI itself as a source, or at least as a conduit of information, becomes a central tenet of information literacy in the AI era.

3.3 Developing Digital Literacy for Responsible AI Engagement

Digital literacy refers to an individual's ability to find, evaluate, create, and communicate information through digital platforms and technologies[17]. In the context of AI, digital literacy extends beyond general computer skills to include a critical understanding of AI tools themselves. This involves recognizing how these tools operate (at least at a conceptual level), their potential for embedding and perpetuating bias, their susceptibility to manipulation, and the ethical considerations surrounding their use[17].

Students need to be educated not just as critical consumers of AI-generated content but also as responsible and ethical users of AI technologies[17]. This includes understanding how their own interactions with AI (e.g., the prompts they use) can influence the outputs they receive and recognizing their responsibility in how they use and disseminate AI-generated information. The ease with which AI can produce shareable content means that students can inadvertently become vectors for the spread of misinformation if that content contains hallucinations. Therefore, a crucial component of digital citizenship in the age of AI is the ethical imperative to rigorously verify AI-generated content before sharing or relying on it, directly linking digital literacy to fact-checking skills and the "check before you share" principle[18]. This emphasizes the student's role not merely as a passive recipient but as an active agent within the digital information ecosystem, accountable for the veracity of the information they propagate.

4 Pedagogical Strategies for Fostering Fact-Checking and Information Evaluation Capabilities

4.1 Explicit Instruction in Verification Techniques

4.1.1 Lateral Reading: Principles and Application

Lateral reading is a powerful strategy employed by professional fact-checkers. Instead of staying on a single webpage or analyzing a source in isolation (vertical reading), lateral readers open multiple browser tabs to investigate the original source or claim by consulting other trusted websites and resources[19]. This approach allows for a quick assessment of credibility by seeing what the broader consensus or other reputable entities say about the information or its originator. Research by the Stanford History Education Group (SHEG) has demonstrated the remarkable effectiveness of lateral reading; professional fact-checkers using this technique were significantly more successful and faster at identifying unreliable websites and misinformation compared to Stanford undergraduates and even history PhDs, who often engaged in less effective vertical reading strategies[20].

To apply lateral reading to AI-generated claims, students can be taught to: first, identify a specific, verifiable claim made by an AI; second, open new browser tabs; third, search for that claim on reputable news sites, academic databases (like Google Scholar or specific subject databases), or established fact-checking websites; and fourth, look for corroborating or conflicting information from multiple independent sources. The iCivics lesson "Artificially Speaking: AI Chatbot Claims" provides a practical example of applying lateral reading to investigate claims produced by AI chatbots, encouraging students to verify chatbot outputs externally[17].

4.1.2 The SIFT Method (Stop, Investigate, Find, Trace)

The SIFT method, developed by Mike Caulfield, offers a memorable four-step heuristic for rapidly evaluating online information, which is highly applicable to AI-generated content[14]. The four moves are: 1. Stop: This initial step emphasizes the importance of pausing before deeply engaging with or, critically, sharing information encountered online, including AI outputs[19]. Students should ask themselves: Who (or what AI model) is behind this information? What might be its purpose or inherent biases? When was this information generated or last updated? Can I implicitly trust this output? Taking a moment to reflect can prevent immediate acceptance of plausible-sounding falsehoods. 2. Investigate the Source: This involves quickly researching the AI tool or platform that generated the content[19]. For an AI model, this might mean searching for information about its developers, its training data (if available), its known capabilities, limitations, and documented biases. A quick search like "<AI tool name> Wikipedia" or "<AI tool name> limitations" can yield valuable context. 3. Find Better (or Trusted) Coverage: Instead of deeply analyzing the AI's initial output in isolation, this step encourages students to look for trusted, alternative coverage of the claims made by the AI[19]. This is where lateral reading comes into play. Students should cross-check the information with reputable news organizations, expert websites, academic sources, or official reports to see if the AI's claims are corroborated or contradicted. The focus is on the claim itself, seeking verification from reliable external sources. 4. Trace Claims, Quotes, and Media to the Original Context: AI models sometimes present information, quotes, or even descriptions of media out of their original context, or they might fabricate sources altogether[19]. This step involves teaching students to try to find the origin of specific pieces of information an AI might present. This could include checking the dates of cited events or data, looking for the primary source of a quotation, or using reverse image search tools (like Google Images or TinEye) to verify the origin and context of any images the AI might describe or generate. Information taken out of context can be highly misleading.

4.1.3 Source Credibility Assessment Frameworks (Adapting for AI)

Traditional source evaluation frameworks, such as CRAAP (Currency, Relevance, Authority, Accuracy, Purpose) or the 5 W's (Who, What, When, Where, Why)[16], provide useful criteria for assessing information. While designed for human-authored sources, these frameworks can be adapted for evaluating AI-generated content[15]. For example, for Authority, one might ask: Who or what developed the AI model? What is known about its training data and potential biases? For Accuracy, one might ask: Can the claims be verified against trusted external sources? Does the AI cite sources, and are they real and accurately represented? For Purpose, one might ask: What was the AI designed to do? Is it optimized for factual accuracy, creativity, or conversation? Are there commercial interests influencing its design or output? For Currency, one might ask: How up-to-date is the AI's training data? Is it capable of accessing real-time information? For Objectivity/Bias, one might ask: Does the AI exhibit any noticeable slant or bias in its language or the information it presents/omits? While vertical reading (digging deep into the source itself)[13] has its place, particularly for understanding the nuances of a well-vetted document, for initial verification of unknown or potentially unreliable AI outputs, lateral reading should be the primary strategy.

4.1.4 Leveraging Fact-Checking Tools and Platforms

Students should be made aware of and taught how to use established fact-checking resources. Websites like Snopes, PolitiFact, and FactCheck.org investigate common misconceptions, political claims, and viral stories, and can be invaluable for verifying specific factual assertions that an AI might generate[21]. Tools like TinEye or Google Reverse Image Search can help verify the authenticity and origin of images[21], which is increasingly important as AI image generation becomes more sophisticated. Furthermore, resources like AllSides, which show news coverage from different political perspectives, can help students understand potential biases if an AI is summarizing or drawing information from news sources that have a particular slant[21].

4.2 Cultivating Critical Engagement with AI-Generated Content

4.2.1 Promoting Healthy Skepticism and Verification Habits

Educators should encourage students to view AI as a "partner, not a replacement"[3]. This perspective emphasizes that AI outputs require human oversight, critical assessment, and verification. A healthy skepticism means not taking AI-generated information at face value, especially if it seems too perfect, overly simplistic, aligns too closely with pre-existing biases, or is designed to evoke a strong emotional response[13]. The goal is to instill a default habit of questioning and seeking corroboration.

4.2.2 Analyzing AI Outputs for Bias, Coherence, and Evidentiary Support

Students should be taught to actively look for signs of bias in AI-generated content. This could manifest in the language used, the perspectives prioritized, or the information omitted[17]. They should also learn to assess the internal logical coherence of AI responses: Do the different parts of the response make sense together? Are there contradictions? Crucially, students should be encouraged to demand, or seek externally, evidentiary support for claims made by AI. If an AI makes a factual assertion, what is the basis for that claim? Can it be substantiated by reliable sources?

4.3 Employing Socratic Pedagogy and Inquiry-Based Learning

4.3.1 Utilizing Socratic Dialogue to Interrogate AI Claims

The Socratic method, characterized by a shared dialogue between teacher and students driven by the teacher's continual probing questions, aims to explore underlying beliefs, assumptions, and the foundations of knowledge[22].

This method is not about providing answers but about stimulating critical thinking and self-reflection[24]. It can be powerfully applied to AI-generated content by prompting students to move beyond surface-level acceptance and deeply interrogate the information presented by AI[25]. Instead of the teacher directly stating whether an AI's claim is true or false, they can pose Socratic questions such as: "The AI has stated X. What makes you think that is accurate (or inaccurate)?" "What are the potential assumptions the AI might be operating under to produce this response?" (Even if AI doesn't have "assumptions" like humans, this question can lead to discussions about its training data and algorithms.) "What evidence would you need to find to confidently verify (or refute) this AI's claim?" "If this AI's statement is true, what are the logical implications? If it's false, what are the consequences of believing it?" "Are there alternative explanations or perspectives the AI might have missed? Why do you think it focused on this particular aspect?" While Socratic chatbots are being developed to facilitate such dialogues[24], the emphasis here is on human-led Socratic inquiry directed at the outputs of general AI tools. This process encourages students to articulate their reasoning, uncover their own biases, and develop a more nuanced understanding of the complexities involved in evaluating AI-generated information. Applying Socratic questioning in this way can help students grapple with the idea that AI responses, while not based on human-like assumptions, are profoundly shaped by the "assumptions" embedded within their algorithms and the patterns prevalent in their training data (e.g., an AI might infer causality from mere correlation if that pattern is strong in its data). This fosters a deeper critique than simple fact-checking.

4.3.2 Designing Activities for Critical AI Tool Usage

Active learning experiences are key. Educators can design activities where students use AI tools for specific tasks (e.g., generating a summary, drafting an argument, researching a topic) and then engage in structured self-evaluation or peer-evaluation of the AI's output[6]. This evaluation should focus explicitly on identifying potential hallucinations, assessing accuracy, checking for biases, and verifying any sources provided by the AI. For example, students could be asked to "Promote Active Engagement With Scientific Data" by using AI to generate data or interpretations, then having students critically analyze, test hypotheses against it, and identify patterns themselves rather than accepting the AI's conclusions[6]. Another activity is to "Use AI to Facilitate Scientific Argumentation" by encouraging students to use AI as a tool to gather initial evidence for a debate, but with the explicit follow-up task of rigorously fact-checking every piece of AI-provided information and identifying its original source[6]. Comparing outputs from different AI tools on the same prompt, or from the same tool using varied prompting strategies, can also be an illuminating exercise. This helps students understand the variability in AI responses, the influence of prompting, and the inconsistent nature of hallucinations.

The process of fact-checking AI hallucinations can sometimes be frustrating or disorienting for students, especially when outputs are subtly misleading or confidently incorrect. Pedagogical strategies should therefore implicitly aim to build students' emotional resilience and persistence. Creating a classroom environment where the iterative process of investigation, including initial "failures" to quickly verify or moments of confusion, is seen as a normal and valuable part of learning can encourage students to persevere rather than giving up or defaulting to blind acceptance.

Furthermore, while many strategies focus on debunking existing AI misinformation, there is significant value in prebunking. This involves inoculating students against future AI hallucinations by proactively teaching them about the common types of errors AI models are prone to make (e.g., fabricating citations, misinterpreting nuanced prompts, exhibiting known biases due to training data, generating plausible but nonsensical explanations). This understanding, grounded in the technical underpinnings of AI (as discussed in Section 2.3), can help students anticipate and be more readily skeptical of these specific error categories when they encounter them, making subsequent debunking efforts more efficient and targeted.

5 Assessing Student Competencies in an AI-Infused Information Ecosystem

5.1 Challenges and Considerations for Assessment

Assessing nuanced skills like the critical evaluation of AI-generated content presents considerable challenges. These competencies extend beyond simple knowledge recall and involve complex cognitive processes, critical dispositions, and practical application in dynamic digital environments. Traditional assessment methods, such as multiple-choice tests, often fall short in capturing the process-oriented nature of fact-checking and information evaluation[26]. Concerns have been raised about whether standardized critical thinking tests, while useful, fully measure the practical application of these skills in novel contexts like interacting with AI[26].

Table 1: Pedagogical Strategies for Combating AI Hallucinations

Strategy Name	**Core Principle**	**Key Steps/Application to AI-Generated Content**	**Strengths in Addressing Hallucinations**	**Potential Implementation Challenges/Considerations**
Lateral Reading	Investigate a source/claim by consulting multiple external, trusted sources before deeply engaging with the original.	1. Identify AI claim. 2. Open new tabs. 3. Search claim on reputable sites (news, academic, fact-checkers). 4. Compare findings.	Quickly assesses credibility by leveraging external expertise; effective against sophisticated misinformation.	Requires access to diverse online resources; students may need guidance in selecting "trusted" external sources.
SIFT Method (Stop, Investigate, Find, Trace)	A four-step heuristic for rapid evaluation of online information.	Stop: Pause, question initial AI output. Investigate: Research the AI tool itself (developer, limitations). Find: Seek trusted coverage of the AI's claims. Trace: Attempt to find original context for AI's specific data points/quotes.	Provides a structured, memorable process; applicable to various online content, including AI.	"Investigating the source" for AI can be complex; tracing claims from opaque AI models can be difficult.
Adapted Source Credibility Frameworks	Apply traditional evaluation criteria (Currency, Relevance, Authority, Accuracy, Purpose) with modifications for AI.	Evaluate AI based on its developer, training data, known biases (Authority); verifiability of claims (Accuracy); recency of data (Currency); intended function (Purpose).	Leverages existing student knowledge of source evaluation; provides a comprehensive checklist.	Criteria need careful adaptation; "Authority" of an AI is non-traditional and harder to define.
Leveraging Fact-Checking Tools	Utilize established fact-checking websites and tools to verify specific AI claims.	Teach students to use sites like Snopes, PolitiFact, reverse image search tools (TinEye) to check assertions or media presented by AI.	Provides access to expert-verified information; can quickly debunk common falsehoods AI might repeat.	AI may generate novel hallucinations not yet covered by fact-checkers; relies on third-party sites.
Promoting Healthy Skepticism	Cultivate a mindset of questioning AI outputs rather than passive acceptance.	Encourage viewing AI as a fallible tool requiring human oversight; question overly perfect or emotionally charged AI responses.	Fosters an essential critical disposition; reduces gullibility towards confidently presented misinformation.	Can be overdone, leading to excessive cynicism if not balanced with AI's utility.
Socratic Questioning of AI Outputs	Use probing questions to encourage deep critical examination of AI claims and underlying assumptions.	Teacher-led or peer discussions asking "Why did AI say that?", "What evidence supports/refutes this?", "What's missing?".	Develops deeper analytical skills; helps uncover implicit biases or logical flaws in AI responses; promotes active learning.	Requires skilled facilitation; can be time-consuming.
Prebunking AI Error Types	Proactively teach students about common categories of AI hallucinations based on technical understanding.	Explain typical AI flaws (e.g., source fabrication, data-driven biases, misinterpretation of nuance) before students heavily use AI.	Inoculates against future deception; makes students more alert to specific vulnerabilities of AI.	Requires educators to understand AI limitations; AI evolution may introduce new error types.

A key consideration is the need for assessments that reflect authentic interactions with AI tools and their outputs. The way a student responds to a static, decontextualized question about critical thinking may differ significantly from how they actually behave when confronted with a plausible but inaccurate AI response in a real-world or simulated scenario. Furthermore, the rapid evolution of AI means that assessment tasks must be adaptable and focus on transferable principles rather than tool-specific knowledge that may quickly become outdated.

5.2 Adapting Critical Thinking Assessments (e.g., CCTST, Watson-Glaser)

Established standardized tests for critical thinking can play a role in assessing foundational skills relevant to evaluating AI-generated content. The California Critical Thinking Skills Test (CCTST) is designed to measure core reasoning skills such as analysis, interpretation, inference, evaluation, explanation, induction, and deduction[8]. These are all cognitive skills vital for dissecting and judging the veracity of AI outputs. The Watson-Glaser Critical Thinking Appraisal (WGCTA) assesses a similar set of skills, focusing on inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments[12]. Recognizing an AI's unstated "assumptions" (stemming from its training data) or evaluating the strength of an argument it constructs are directly applicable tasks.

These tests can provide valuable baseline data on students' general critical thinking abilities, which are prerequisites for effectively dealing with AI hallucinations. However, to more directly assess skills related to AI, consideration could be given to how items on such tests might be adapted, or how new sections or complementary instruments could be developed. This might involve creating scenarios specifically involving AI-generated content, where students are asked to apply their analytical and evaluative skills to identify potential hallucinations, biases, or logical flaws within that context.

5.3 Developing Performance-Based Tasks and Authentic Assessments

Given the limitations of traditional tests, performance-based assessments are increasingly advocated for measuring complex skills[11]. These tasks require students to actively do something with AI-generated content, demonstrating their verification and evaluation processes in a more authentic context. Examples of such tasks include: AI Output Analysis, where students are provided with an AI-generated text that contains subtle hallucinations and are tasked with identifying these issues; AI-Assisted Research Report with Verification Log, where students use AI for information gathering but must maintain and submit a detailed verification log; and a Simulated Fact-Checking Challenge, where students participate in a timed challenge to determine the veracity of AI-generated statements.

A crucial aspect of these assessments is evaluating the process of verification, not just the final "product" or whether the student correctly identified an error. This requires looking at the methods students employed: Did they attempt lateral reading? Did they consult multiple, diverse sources? Did they recognize potential AI biases or limitations? Did they articulate their reasoning clearly? This often necessitates observational components (e.g., think-aloud protocols where students verbalize their thought process while performing the task) or reflective components (e.g., a written reflection on their verification strategy and challenges encountered). Rubrics for such tasks should include criteria for both the accuracy of the evaluation and the soundness of the verification process.

A significant challenge in designing such assessments lies in balancing authenticity with controllability. Using live AI tools in assessments offers high ecological validity, but their outputs can be unpredictable, making standardized assessment difficult. Conversely, using pre-prepared, simulated AI outputs offers high controllability but may lack the full authenticity of interacting with a live, dynamic AI. A balanced approach, perhaps involving a mix of controlled simulations and more open-ended tasks with live AI tools, might be the most effective strategy.

5.4 Utilizing Pre-Post Intervention Designs for Program Evaluation

To measure the effectiveness of specific pedagogical interventions, pre-post test designs are a valuable research methodology[31]. In this design, students' skills or knowledge are assessed before the intervention (pre-test), the pedagogical strategy is implemented, and then the same or a parallel assessment is administered afterward (post-test). The difference in scores can provide evidence of the intervention's impact. Key considerations for designing effective pre-post tests in this context include[32]: having clear learning objectives, using appropriate question types, ensuring authenticity and relevance, avoiding bias, and maintaining conciseness.

Pre-test data is not only useful for baseline measurement but can also be a powerful tool to inform and enhance instruction[33]. Sharing anonymized, aggregated pre-test responses with students can highlight common misconceptions or skill gaps, making subsequent instruction more relevant and engaging[33]. In educational research, particularly in real-world classroom settings where true randomization is often not feasible, statistical techniques like Analysis of Covariance (ANCOVA) can be employed. ANCOVA allows researchers to statistically control for pre-existing differences between groups when analyzing post-test scores, thereby providing a more accurate estimate of the intervention's effect[35].

Beyond formal assessments, fostering students' ability to self-assess their confidence in AI-generated information and their own evaluation skills is critical for developing independent, lifelong learners. Critical thinking inherently involves self-regulation[7]. Integrating metacognitive prompts into learning activities can support this. For example, after a student uses an AI tool, they could be asked: "How confident are you in the accuracy of this AI's response on a scale of 1-5? What specific factors contribute to your level of confidence? What steps did you take to verify this information?" Such reflective practices help students become more aware of their own evaluation processes and the ongoing need for critical vigilance.

6 Discussion: Challenges, Ethical Dimensions, and Future Research Trajectories

6.1 Overcoming Implementation Hurdles

Successfully integrating strategies to combat AI hallucinations into educational practice faces several significant obstacles. One critical barrier is Teacher Preparedness and Professional Development. Research indicates that many

Table 2: Assessment Approaches for Fact-Checking and Information Evaluation Skills in the Context of AI

Assessment Type	**Description & Example Application**	**Specific Skills Assessed (re: AI Hallucinations & Evaluation Process)**	**Advantages**	**Limitations/Considerations for AI Context**
Adapted Standardized CT Tests	Using existing critical thinking tests, potentially with new items/sections tailored to AI scenarios. Ex: Scenario where AI gives conflicting info, student identifies logical fallacies or needed verification.	General CT skills: analysis, inference, evaluation, assumption recognition, deduction, interpretation. Application to AI-specific examples.	Well-validated for general CT; provides standardized scores; can measure foundational skills.	May not capture nuanced AI interaction skills; adapting for AI requires careful design and validation. Generality might miss AI-specific process skills.[26]
Performance-Based Tasks	Students actively engage with (real or simulated) AI content and demonstrate verification processes. Ex: Given AI text with hallucinations, students identify, verify, correct, and explain.	Application of verification techniques (lateral reading, SIFT), source evaluation, bias detection, identification of hallucinations, articulation of reasoning.	High authenticity if using real AI; assesses application of skills in context; can evaluate process and product.	Real AI output is variable, making standardization hard; simulated AI may lack full authenticity; can be time-consuming to score.
Pre-Post Tests for Interventions	Assessing skills before and after a specific pedagogical intervention aimed at improving AI evaluation abilities. Ex: Test on identifying types of AI bias before and after a module on algorithmic literacy.	Measures change in specific targeted skills (e.g., ability to use SIFT, knowledge of AI limitations, confidence in evaluation).	Useful for evaluating program/intervention effectiveness; pre-test data can inform instruction.	Requires careful design to ensure validity and reliability; learning effects from pre-test possible; attribution of change solely to intervention can be challenging.
Reflective Journals / Self-Assessments	Students reflect on their process of using AI, their confidence in AI outputs, and their verification strategies. Ex: After an AI-assisted task, students write about challenges in verifying AI info.	Metacognition, self-regulation, awareness of own biases, ability to articulate evaluation process and confidence levels.	Promotes deeper learning and metacognitive awareness; provides qualitative insights into student thinking.	Subjective; may be difficult to standardize for grading; relies on student honesty and articulateness.
Think-Aloud Protocols	Students verbalize their thoughts while performing an AI evaluation task. Ex: Student talks through their steps and reasoning while fact-checking an AI's historical claim.	Real-time demonstration of thought processes, decision-making during verification, use of strategies.	Provides rich qualitative data on cognitive processes; reveals difficulties and strategies used.	Time-intensive to administer and analyze; observer effect possible; may not suit all learners.

educators feel ill-equipped to teach about AI or use it effectively in their classrooms[36]. Surveys reveal that a lack of AI training and support is a primary concern among educators[37]. Many teacher preparation programs are lagging, often mentioning AI only in the context of student plagiarism rather than as a transformative tool with inherent risks like hallucinations[36]. Effective, ongoing professional development is therefore essential.

Another challenge is Curricular Integration. Integrating these new literacies into an already crowded curriculum requires a coordinated effort. The rapid evolution of AI technology presents a "moving target" problem for curriculum development[1]. Strategies taught today might be less effective tomorrow. This implies a need to focus on teaching adaptable, principle-based critical thinking and verification methods that are applicable to any information source.

Finally, Access and Equity remains a persistent issue. Ensuring equitable access to AI tools, internet connectivity, and educational resources is paramount[2]. Without addressing these disparities, efforts to develop AI literacy could inadvertently widen existing achievement gaps[37].

6.2 Ethical Dimensions and Fostering Responsible AI Literacy

The phenomenon of AI hallucinations brings several ethical dimensions to the forefront. Fostering responsible AI literacy means educating students not only about how to use AI and critically evaluate its outputs but also about the broader societal, ethical, and moral considerations surrounding AI[1]. This includes discussions on Algorithmic Bias, where biases in training data can lead AI to produce skewed or unfair outputs[1]; Privacy, regarding the implications of using AI tools that collect student data; Intellectual Property and the ownership of AI-generated content; and The

Ethics of AI in Assessment, including the potential for bias in AI grading tools.

A specific ethical dilemma for educators is the act of "using AI to teach about AI's flaws." If an AI tool is used in a lesson about hallucinations, there's an inherent risk that the tool itself might hallucinate in an unhelpful or confusing way, derailing the learning objective. This implies that educators need to be highly discerning in selecting appropriate AI tools for pedagogical purposes and must carefully scaffold activities.

6.3 Identifying Research Gaps and Proposing Future Studies

While the need for developing student skills to address AI hallucinations is recognized, there are numerous areas where further research is essential. There is a pressing need for more empirical research on the Effectiveness of Pedagogical Strategies to determine which methods best improve students' ability to evaluate AI hallucinations[1]. Longitudinal Studies are needed to track the development and retention of these critical evaluation skills over time[28]. Research into the Cognitive Processes students use when evaluating AI content would also be valuable. Further investigation is needed into Teacher Preparedness and Impact, connecting different levels of teacher training to student outcomes[36]. Research should also explore Culturally Responsive Pedagogy, recognizing that students from different backgrounds may perceive AI differently. There is also a need for Assessment Tool Development to create and validate reliable and authentic tools for measuring these skills. Finally, more research is needed on the Impact of AI on Critical Thinking Development to understand the nuanced effects of AI use on students' general cognitive skills.

While the focus of this paper is largely on developing individual student skills, a broader implication for future consideration is the fostering of a classroom or institutional culture that values and practices collective verification and critical discourse around AI-generated content. Moving beyond individual responsibility to a "community of verification" could distribute the cognitive load and create a more resilient learning environment.

7 Conclusion

7.1 Reaffirming the Necessity of Equipping Students for the AI Era

The rapid integration of Artificial Intelligence into all facets of life, particularly education, heralds an era of unprecedented opportunities and complex challenges. Among these challenges, the phenomenon of AI-generated "hallucinations"—the confident presentation of fabricated or inaccurate information—stands out as a critical concern demanding immediate and sustained pedagogical attention. As this paper has argued, the capacity of AI to produce plausible yet erroneous content necessitates a fundamental shift in how students are taught to engage with information. It is no longer sufficient to impart knowledge; we must cultivate the critical faculties that enable students to discern truth from falsehood, accuracy from error, and authentic insight from sophisticated confabulation.

The development of robust fact-checking, information evaluation, and critical thinking skills is not merely an academic exercise. These competencies are foundational for informed decision-making, academic integrity, responsible digital citizenship, and effective lifelong learning in an increasingly AI-driven world. Failure to equip students with these skills risks creating a generation vulnerable to misinformation, unable to fully harness the benefits of AI, and potentially hindered in their cognitive development due to over-reliance on fallible automated systems. The strategies outlined—from explicit instruction in verification techniques like lateral reading and the SIFT method to the cultivation of critical engagement through Socratic inquiry and the adaptation of source credibility frameworks—offer pathways to empower students as discerning consumers and critical users of AI.

7.2 A Call for Collaborative Action among Educators, Policymakers, and Researchers

Addressing the challenge posed by AI hallucinations effectively requires a concerted and collaborative effort from all stakeholders in the educational ecosystem. Educators are at the front line and need access to high-quality professional development, evidence-based pedagogical resources, and supportive institutional environments[36]. Teacher education programs must urgently revise their curricula to prepare new teachers for the realities of AI in the classroom[36].

Policymakers at local, state, and national levels have a crucial role in developing guidelines and standards for the ethical and effective use of AI in education, ensuring equitable access, and supporting the necessary infrastructure and training initiatives[37]. This includes fostering the development of clear AI policies within educational institutions.

Researchers must continue to investigate the impact of AI on learning, the efficacy of different pedagogical interventions, the cognitive processes involved in evaluating AI-generated content, and the evolving nature of AI hallucinations themselves. This research will provide the evidence base needed to refine strategies and inform policy. AI developers also bear a responsibility to work towards mitigating hallucinations, increasing transparency in how their models operate, and collaborating with the education sector to create tools that are both powerful and safe for learning environments.

The co-evolution of AI and education is an ongoing process. It calls for proactive adaptation, continuous learning, and a commitment to critical inquiry from all involved. By embracing these principles, the educational community can strive to ensure that AI serves as a true catalyst for enhanced learning and human intellectual advancement, rather than an inadvertent source of confusion and misinformation. The goal is not to fear AI, but to foster a generation capable of navigating its complexities with wisdom, discernment, and ethical responsibility.

To Cite This Article Yingzhe LI. (2025). Addressing "Hallucinations" in AI-Generated Content: Strategies for Developing Student Fact-Checking and Information Evaluation Skills. *Artificial Intelligence Education Studies*, *1*(2), 48-62. https://doi.org/10.6914/aiese.010204

Reference

[1] Alshater, M. M. (2025). The impact of artificial intelligence (AI) on students' academic performance, engagement, and well-being in higher education: A systematic review. *Education Sciences*, *15*(3), 343. Retrieved June 12, 2025, from https://www.mdpi.com/2227-7102/15/3/343

[2] Singh, H., & Singh, A. (2024). AI in education: A review of personalized learning and educational technology. *GSC Advanced Research and Reviews*, *21*(2), 329 338. Retrieved June 12, 2025, from https://gsconlinepress.com/journals/gscarr/content/ai-education-review-personalized-learning-and-educational-technology

[3] York College of Pennsylvania. (n.d.). *Hallucinations*. Artificial Intelligence in Research - LibGuides. Retrieved June 12, 2025, from https://library.ycp.edu/c.php?g=1387233&p=10262462

[4] Al-Khaifi, A., Singh, Y., & De, A. (2025). Is artificial intelligence hallucinating? *Journal of Clinical and Experimental Hepatology*, *15*(3), 513 515. Retrieved June 12, 2025, from https://pmc.ncbi.nlm.nih.gov/articles/PMC11681264/

[5] National Science Teaching Association. (2023, October 25). *To think or not to think: The impact of AI on critical-thinking skills*. NSTA Blog. Retrieved June 12, 2025, from https://www.nsta.org/blog/think-or-not-think-impact-ai-critical-thinking-skills

[6] National Science Teaching Association. (2023, October 25). *To think or not to think: The impact of AI on critical-thinking skills*. NSTA Blog. Retrieved June 12, 2025, from https://www.nsta.org/blog/think-or-not-think-impact-ai-critical-thinking-skills

[7] O'Sullivan, P. S., Blevins-Stephens, W. L., Smith, F. M., & Vaughan-Wrobel, B. (1993). *The "California Critical Thinking Skills Test" and the National League for Nursing accreditation requirement in critical thinking*. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA. Retrieved June 12, 2025, from https://scispace.com/pdf/the-california-critical-thinking-skills-test-and-the-3nbwktj0mu.pdf

[8] Insight Assessment. (n.d.). *California Critical Thinking Skills Test*. Retrieved June 12, 2025, from https://insightassessment.com/iaproduct/california-critical-thinking-skills-test/

[9] Austin Peay State University. (2021). *CCTST score explanation*. Retrieved June 12, 2025, from https://www.apsu.edu/institutional-effectiveness/quality-assurance/senior-exam/cctst_score_explanation_2021.pdf

[10] Pearson TalentLens UK. (n.d.). *Watson Glaser Test*. Retrieved June 12, 2025, from https://www.talentlens.com/recruitment/assessments/watson-glaser-critical-thinking-appraisal.html

[11] Monitask. (n.d.). *What is critical thinking assessment?*. Retrieved June 12, 2025, from https://www.monitask.com/en/business-glossary/critical-thinking-assessment

[12] ADEA. (n.d.). *CTS tools for assessment*. Retrieved June 12, 2025, from https://www.adea.org/home/ADEAevents/teaching-resources/critical-thinking-skills-toolbox/pages/cts-tools-for-assessment

[13] Weimer, M. (2020, January 22). *Teaching information literacy in an age of misinformation*. Faculty Focus. Retrieved June 12, 2025, from https://www.facultyfocus.com/articles/effective-classroom-management/teaching-information-literacy-in-an-age-of-misinformation/

[14] Caulfield, M. (2019). The art of the real: fact checking as information literacy instruction. *Communication in Information Literacy*, *13*(2), 296-305. Retrieved June 12, 2025, from https://www.researchgate.net/publication/338521131_The_art_of_the_real_fact_checking_as_information_literacy_instruction

[15] FIU Libraries. (n.d.). *Evaluating information*. Information Literacy Toolbox. Retrieved June 12, 2025, from https://library.fiu.edu/infolit/evaluate
[16] Relay Graduate School of Education. (n.d.). *Evaluating sources*. Information Literacy – LibGuides. Retrieved June 12, 2025, from https://relay.libguides.com/information_literacy/evaluating_sources
[17] iCivics. (n.d.). *How to teach critical thinking in the digital age*. Retrieved June 12, 2025, from https://vision.icivics.org/how-to-teach-critical-thinking-in-the-digital-age/
[18] Baker, P. (2024, May 22). *5 critical thinking skills for digital literacy: Spotting misinformation & manipulation*. Killer Innovations. Retrieved June 12, 2025, from https://killerinnovations.com/5-critical-thinking-skills-for-digital-literacy-spotting-misinformation-manipulation/
[19] Minnesota State University, Mankato. (n.d.). *Lateral reading & SIFT*. Source Evaluation – Library Class and Information Guides. Retrieved June 12, 2025, from https://libguides.mnsu.edu/sourcecredibility/lateralreading
[20] The Chicago School. (n.d.). *Lateral reading: Be a pro fact checker*. Navigating Misinformation and Disinformation. Retrieved June 12, 2025, from https://library.thechicagoschool.edu/c.php?g=1425760&p=10608575
[21] Penn State University Libraries. (n.d.). *Resources for fact-checking*. News Literacy – Library Guides. Retrieved June 12, 2025, from https://guides.libraries.psu.edu/c.php?g=620262&p=4319365
[22] The Institute for Learning and Teaching. (n.d.). *The Socratic method: Fostering critical thinking*. Colorado State University. Retrieved June 12, 2025, from https://tilt.colostate.edu/the-socratic-method/
[23] Saint Leo University. (2022, October 14). *The Socratic method of teaching: What it is, its benefits, and how to use it*. Saint Leo University Blog. Retrieved June 12, 2025, from https://www.saintleo.edu/about/stories/blog/socratic-method-teaching-what-it-its-benefits-and-examples
[24] Dillenbourg, P., & Jermann, P. (2024). Chatbots & Socrates: Dialogues in learning. *ERCIM News*, (136). Retrieved June 12, 2025, from https://ercim-news.ercim.eu/en136/special/chatbots-socrates-dialogues-in-learning
[25] University of Connecticut, Center for Excellence in Teaching and Learning. (n.d.). *Critical thinking and other higher-order thinking skills*. Retrieved June 12, 2025, from https://cetl.uconn.edu/resources/design-your-course/teaching-and-learning-techniques/critical-thinking-and-other-higher-order-thinking-skills/
[26] Gadanidis, G. (2019). What's wrong with the California Critical Thinking Skills Test? In *Popular Culture and the Formidable Teacher*. eCampusOntario Pressbooks. Retrieved June 12, 2025, from https://ecampusontario.pressbooks.pub/popsterscrrar2019/chapter/whats-wrong-with-the-california-critical-thinking-skills-test-ct-testing-and-accountability/
[27] EdInstruments. (n.d.). *California Critical Thinking Skills Test (CCTST)*. Retrieved June 12, 2025, from https://edinstruments.org/instruments/california-critical-thinking-skills-test-cctst
[28] Nuutinen, S. (2006). *Evaluation of IS curriculum design: A pilot study using the California Critical Thinking Skills Test*. University of Jyväskylä. Retrieved June 12, 2025, from https://jyx.jyu.fi/bitstreams/707cf722-ce8a-46a0-8566-2c4a1ba94fbd/download
[29] Saregar, A., Pitaloka, P. A., & Susilowati, N. E. (2023). Assessment of critical thinking ability in science learning using Watson-Glaser Critical Thinking Appraisal (WGCTA). *JINAV: Journal of Innovation and Learning*, *3*(3), 223–231. Retrieved June 12, 2025, from https://jinav.org/index.php/soshum/article/view/2599
[30] Pearson. (n.d.). *Watson-Glaser Critical Thinking Appraisal*. Retrieved June 12, 2025, from https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/efficacy-and-research/reports/Watson-Glaser_One_Page_Summary.pdf
[31] Ranganathan, P., & Aggarwal, R. (2019). Study designs: Part 4 Interventional studies. *Perspectives in Clinical Research*, *10*(3), 138–140. Retrieved June 12, 2025, from https://pmc.ncbi.nlm.nih.gov/articles/PMC6647894/
[32] EMpower. (n.d.). *Guidance for pre- and post-test design*. Retrieved June 12, 2025, from https://empowerweb.org/assets/uploads/tools-resources/461/5._guidance_for_pre-_and_post-test_design.pdf
[33] Walker, A. K., & Latham, B. D. (2020). Maximizing one-shot impact: Using pre-test responses in the information literacy classroom. *Southeastern Librarian*, *64*(1), Article 6. Retrieved June 12, 2025, from https://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=1461&context=seln
[34] Ranganathan, P., & Aggarwal, R. (2019). Study designs: Part 5 Interventional studies (II). *Perspectives in Clinical Research*, *10*(4), 189–191. Retrieved June 12, 2025, from https://pmc.ncbi.nlm.nih.gov/articles/PMC6801992/
[35] Olejnik, S. F., & Covert, R. W. (2003). The use of hierarchical ANCOVA in curriculum studies. *Journal of Modern Applied Statistical Methods*, *2*(1), 60–70. Retrieved June 12, 2025, from https://jmasm.com/index.php/jmasm/article/view/239/237

[36] Center on Reinventing Public Education. (2024, May 15). *New research finds schools of education fail to prepare teachers to use AI*. Retrieved June 12, 2025, from https://crpe.org/new-research-finds-schools-of-education-fail-to-prepare-teachers-to-use-ai/

[37] SMART Technologies. (2024, August 28). *New survey reveals major gaps on the future of AI*. District Administration. Retrieved June 12, 2025, from https://districtadministration.com/opinion/artificial-intelligence-survey-reveals-major-gaps-in-k12/

Editor *Changkui LI wtocom@gmail.com*