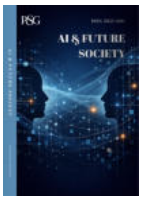
PSG

AI & Future Society

panorama-sg.com

ISSN 3053-4011
Publisher: Panorama Scholarly Group

Vol. 2, No.1, 35-44, 2026
DOI:10.63802/afs.V2.I1.207



AI & Future Society Editorial Office

AI literacy beyond tools: Safeguarding human judgment in research and higher education

Yueh-Sung Weng[1*]

[1]*Department of Urban Planning and Spatial Information, Feng Chia University, Taichung, Taiwan*
[*]*Corresponding author*

Corresponding author:
ysweng@o365.fcu.edu.tw



Available online: 08 April 2026

Abstract

Artificial intelligence has emerged as a pioneering tool for assisting researchers and students in conducting complex tasks, including designing questionnaires, proofreading academic works, and rebutting reviewers' criticisms. Beyond being a tool, this advanced technology highlights its impact on the future of research and education and emphasizes how crucial it is for academia to actively engage. However, AI-related ethical issues arise when people apply AI in their learning and research, especially when it involves sensitive information in humanities and social sciences and questions users' epistemic positions and reflexivity of their studies. The contribution of AI literacy has not yet been fully recognized, as current studies focus on the AI models' competencies rather than on human capabilities to adapt AI technologies for educational and research purposes in responsible, ethical, and comprehensive ways. To comprehensively discuss the role of AI literacy in academia and higher education, this article employs a conceptual analysis framework within an iterative assessment process that weaves together technical understanding, conceptual evaluation, ethical reflections and responsible implementation in the context of research and higher education. Then, it examines four controversial AI ethical issues that potentially jeopardize research and learning, including biases in AI, sycophantic chatbot services, inaccuracy, randomness, and ethical concerns of AI models' outputs. This article represents a further step toward developing clear AI literacy guidelines, assuring that researchers and students' skills and judgment remain central to responsible AI use.

Keywords: AI Ethics, AI Literacy, AI Governance, AI Implications for Research and Education, Social science and Humanities

1. Introduction

Artificial Intelligence (AI) is progressively treated as an essential component of academic research and higher education. Some experts firmly believe AI will revolutionize how people conduct research and coursework, whereas other scholars predict this movement is just another technology fad driven by unjustified enthusiasm. This article is situated in the debate over whether AI is simply just a tool or has more profound technological implications for our academic environments where people live, learn, experience, understand and connect with the world. This article argues that discussions of AI literacy become vital as they provide researchers and students with guidance on conducting AI for academic purposes in responsible, ethical and practical ways.

Nevertheless, in much of the existing literature, AI literacy is primarily defined as technical proficiency in operating, prompting and assessing AI tools. Such a concept emphasizes efficiency, data optimization, skill acquisition, and usually regards AI as an impartial instrument whose risks can be managed by using it more effectively. This article criticizes the prevailing framing and argues for a richer understanding of AI literacy, which should be regarded as a broader form of human judgment that involves critical interpretation, research accountability, and ethical self-reflection in situations where AI systems assist with knowledge production.

In social scientific studies, researchers critique, learn, examine, and explore new phenomena to narrow the gap in understanding the complexities of the world, while students are trained to develop comprehensive approaches to understanding human society, relationships, and institutions. Just like other natural science scholars, researchers have high hopes for this technological advancement. Moreover, some scholars expect AI to free up our time for research as it can efficiently handle large amounts of data. Then, researchers can focus on sophisticated tasks such as theorizing academic ideas. This advanced technology

This article is published in *AI & Future Society* under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

also assists scholars in communicating their arguments effectively to a range of audiences.

Much of the current debate focuses on improving AI models and regulating AI use via institutional guidelines. However, many ethical issues arise as generative AI increasingly pervades the daily academic practice of social science studies, including a lack of data diversity and inclusivity, data analysis bias, and AI's sycophancy to users.

The rise of AI also challenges the essence of education and learning. Several thinkers have recalled the critique of 'technopoly,' which was promoted by the American educator Neil Postman (1993). They express their concern that academy and society are mindlessly following this technological imperative without critical evaluation. As a result, human beings' moral and philosophical mindsets are dictated by technology, while our curiosity and critical minds are eliminated when a society seeks truth, orders and authority merely from this advanced technology. Postman argued that academy and society should be cautious about how the medium human develop reshapes our thinking and the purpose of knowing and learning. This critique becomes more invaluable as knowledge production becomes convenient and 'cheap' through the intensive use of cutting-edge technologies. Purser further criticized AI's impact on campus. He argues that "education now turns into logistics" in this information era. Then, knowledge production is equivalent to data automation and teaching seems to be the delivery of AI content.

Academia and society often overlook the principles of AI literacy, failing to consider AI's impact on research and learning carefully. However, prevailing discussions of AI literacy remain narrowly focused on its technical competencies of optimizing complex tasks, including topic guides and questionnaire designs, proofreading academic works and rebutting reviewers' criticisms. The instrumental framing leaves a research gap, which concerns the roles of human judgment, ethical accountability and epistemic responsibility in AI-augmented academic practice. Thus, aligning with other scholars' arguments, the research problem of this article is conceptual: how should AI literacy be refined to reflect its role as an ethical and epistemic capacity in academia. To address this gap, this article conducts a conceptual analysis to reconceptualize AI literacy in research and higher education. This reframing is particularly urgent in the social sciences and humanities, where research frequently involves context-dependent knowledge, normative evaluation, and engagement with sensitive human experiences (e.g., religious controversies and conflicts between divergent political systems), which cannot be simply assigned to algorithms.

This article proceeded as follows. First, it reviews the role of AI literacy in social science research and higher education. This study then proposed a conceptual framework for AI literacy to assess these fierce controversies surrounding the use of AI. Finally, it discusses these implications of AI literacy conception for both research and education. Specifically, this article aims to address these research objectives: 1. To understand how to safeguard human judgment in research and higher education when AI erodes the capabilities of conducting responsible and accountable academic work. 2. To reconceptualize AI literacy as an ethical accountable and epistemic capacity in research and higher education.

2. Literature review: AI literacy in academia

2.1. The role of AI literacy in social science and humanities research

AI literacy is a broad term that refers to humans' capacity to critically comprehend, evaluate, apply, and develop artificial intelligence technologies in a responsible, comprehensible, and ethical manner. This term can be defined as a basic knowledge of a subject, not limited to computer science, with broader implications for academic studies involving the use of AI technology. The comprehension of AI literacy consists of technical knowledge of how AI models function, critical minds to recognize related hallucinations, bias and misinformation issues, and the ethical awareness of AI impacts to human ethical bottom lines. For example, components of AI literacy include understanding AI algorithms, recognizing bias in data, and evaluating AI-generated content critically. Even though scholars and academic institutions developed different conceptual frameworks of AI literacy, it generally encompasses human knowledge of AI's capabilities, human-technology interaction and its implications for wider societal issues.

Chiu, a well-known AI and STEM education scholar, asserts that AI literacy is a conceptual understanding of this technology, which is different from AI competency focusing on its practical application. Beyond the sense of relationship between a tool and master, AI literacy focuses on how humans interact with this advanced technology and the digital world ethically and responsibly. He insists that AI literacy should be distinctive of AI competency, although these two terms are interconnected. To improve clarity, Chiu explicitly uses the metaphor of a 'compass and engine' to distinguish AI literacy from its competency. AI literacy emphasizes understanding concepts and implications, like a compass of research direction, whereas AI competency focuses on practical skills and optimization, like an engine of knowledge production. Clarifying that AI literacy is distinct from AI competency helps researchers to find a sense of purpose and focus on their learning journey when they use AI for research purposes.

As Han, Qiu and Lichtfouse argued in their book, scholars should not entirely deny the competencies of AI technologies in assisting researchers with the efficient facilitation of research design, research activities and in producing the quick summaries of strengths and downsides of research papers. However, the debate about AI's potential in scientific research should make us cautious and critical, especially regarding little studies that can prove its positive impact on research quality and academic communication in the humanities and social sciences. As these two academic fields studying human beings and society, humanities dig into profound inquiries about human identity and value appreciation and understanding human culture, while social science investigates human behaviors, actions, and their connections to different societies. Iain McLean further elaborates the nuance of these two branches, and he claims that "humanities are (mostly) interested in the unique as a human being; social sciences are (mostly) interested in the general tendency of a social phenomenon." Different but alike, several overlapping research areas can be approached by both humanities and social science studies.

This article asserts that AI literacy in research across both academic fields is invaluable because it is unclear to what extent Large Language Models (LLMs) can reflect on the complex conceptual landscapes of humanity's disciplines, such as theology, the arts, and philosophy. Meanwhile, the AI's role in research in these several branches of social science is also controversial when it involves sensitive human beings' ethical issues, such as criminology, politics, psychology, and suicidology.

2.2. The role of AI literacy in higher education

Compared to the role of AI literacy in research, our understanding of how AI technology influences the delivery of education and training is notably underdeveloped. Although few researchers have discussed this technology's implications for elementary and secondary level education, it is much less known about the role of AI literacy in higher education, specifically undergraduate and postgraduate programmes.

It is certainly a significant research gap as AI potentially transforms higher education and erodes undergraduate and postgraduate students' critical thinking abilities, a sense of research judgment and professional skill development. AI literacy in higher education matters because researchers and students should act as accountable and independent knowledge producers, while universities have their duty to establish norms for research and educational innovations as knowledge gatekeepers. Undergraduate and postgraduate students should understand their intellectual responsibility in knowledge production and maintain decent quality of research following ethical principles.

According to current peer reviewed papers, these research projects demonstrated that these heavy users of generative AI have difficulty engaging with their learning in higher education compared to traditional paper-based and internet search engines users. The Large Language Model (LLM) makes these generative AI users become lazier in their coursework when they overly rely on this technology's assistance. As a result, these 'lazy' students experience a long-term erosion of the learning and problem-solving skills, namely a 'cognitive debt' or 'cognitive offloading' in their learning process. Moreover, these generative AI users might violate research integrity standards when they do not clearly state which parts of their materials are generated by AI. Furthermore, these AI-additive users pay less attention to distinguishing their original works from these AI-generated contents, which resulted in bigger trouble in plagiarism.

The call of AI literacy in higher education is vital regarding students who do not critically understand AI technology's role in their learning and the appropriate way to conduct their assignments ethically and responsibly by the LLM models. Moreover, most studies in the field of AI literacy in education have merely focused on the technology's destructive effects on students' personal learning progress rather than the broader picture of educational institutions' position on the use of AI in higher education. Functioning as a social institution, colleges and universities not only involve the dissemination of knowledge, skills and cultural values across younger generations but also equip these students with the comprehensive understanding to practice AI for academic purposes and professional practices in the digital era.

Ironically, increasing higher education institutes invest millions of dollars into AI flagship projects with these technological giants. Pursuers criticized that these 'AI-empowered' programmes are unjustified in two dimensions. First, it suggests that universities purchase the premium membership of an AI service for students who have already acquired the free membership. Second, although most universities announced guidelines on AI use in academic practices, these higher education institutions still lack clear strategies to tackle these growing AI misconduct issues when students intentionally or unintentionally obtain unfair benefits in coursework through AI technologies. What is more, these universities are so enthralled by AI enthusiasm that they decided to downsize the faculty, simply replacing them with more chatbot services. This article strongly maintains that AI literacy in higher education is not merely for students and professors, but also for the university's executive board, which is responsible for delivering inspiring knowledge and educating students with principled, honest values. Highlighting this institutional issue can motivate educators and policymakers to clarify the role of AI in higher education and their stance on supporting its use.

3. Theoretical foundation and methodology

This article employs a conceptual analysis framework to investigate how AI literacy is defined as an essential knowledge and ethical understanding to effectively and responsibly engage with AI technologies and situated within the contemporary academic debate over AI use in research and higher education. The framework engages with the epistemological discussion of how knowledge is produced within the emerging AI learning environment. Epistemology, the theory of knowledge and belief justification, is a key element of this theoretical framework. The epistemological discussion focuses on the role of human judgment in academic work, conventional understanding of authorship, intellectual responsibility, and research integrity, all of which are critically affected by over-reliance on generative AI in research and higher education.

The theoretical framework explains the connection between specific theoretical arguments and different schools of thought in AI literacy. By drawing on interdisciplinary social science studies, this article reviews diverse perspectives of AI literacy. Conceptual analysis, which is the research method of this study, examines the following analytical materials, including conceptual definitions, theoretical arguments, and normative claims drawn from interdisciplinary social science scholarship on AI literacy and higher education. These materials were analyzed through an iterative process of analytical reasoning. This analysis includes related concept breakdowns, an examination of conceptual framing, and a critical assessment of underlying assumptions about automation, human judgment, and ethical responsibility in research and higher education. Then, this comprehensive understanding of AI literacy enables the author to propose a more refined view of AI use, emphasizing on human judgment and epistemic responsibility.

3.1. Proposed conceptual framework

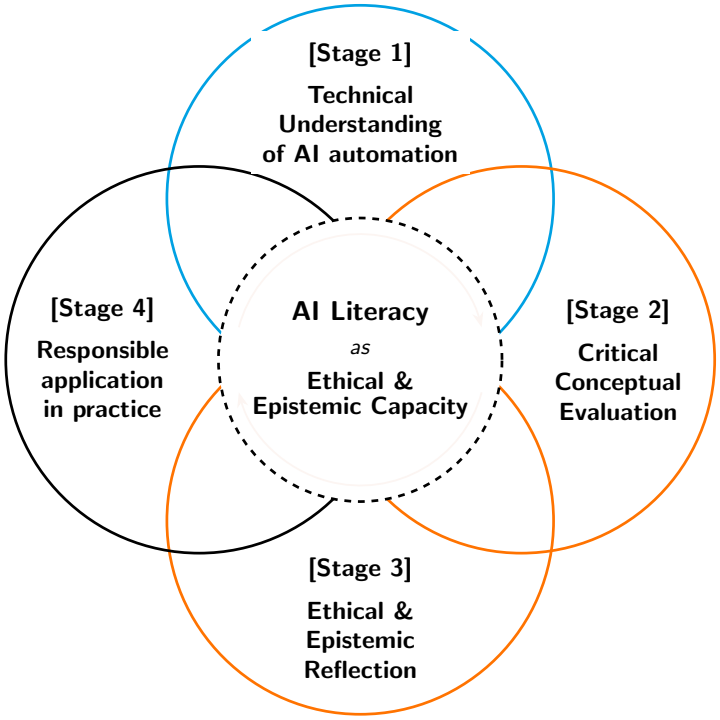
Instead of focusing merely on programming and technical aspects of AI use, this critique rests on existing literature centred on understanding of AI use, critically appraising and responsibly engaging with AI models. The conceptual framework, informed by lens of these theoretical foundations, is essential for guiding this analysis and addressing these research objectives related to AI issues and users' reflection on academic practice with AI assistance. It also informs methodology and data analysis strategies, examines related conceptualizations and AI ethical issues, and situates this topic within the context of testing related theories.

Several scholars have assessed these notions from different angles in research and higher education, revealing diverse conceptualizations. Ng et al. break down AI literacy into varied components, including practical applications of machine learning, awareness of emerging issues, evaluation, and creation with AI. Yu proposed a framework of AI literacy specifically for higher education, which enables academia to comprehensively understand the pros and cons of AI use across the following dimensions: functional, ethical, conceptual and pedagogical aspects. According to Chiu, Ahmad, Ismailov and Sanusi and Netwong, they argued that AI technology is not merely a neutral instrument in research and higher education. Rather, it shapes the ways of knowledge production, the justification of methodologies, and the standards for evaluating academic work quality.

Extending from Netwong and Chiu and his team's theoretical arguments, this study proposes that understanding of AI literacy is an interactive analytical process based on iterative reality checks between theory and real-life implications. Given the current conceptual approaches' inadequate engagement with in-depth ethical evaluations of AI use, this study employs an iterative analytical approach to enhance the epistemological understanding of AI literacy, such as testing belief justification of case studies and normative claims across varied AI use scenarios. This iterative analytical process involves a loop of analysis to sharpen the understanding of this topic and the ethical implications of AI literacy for social science research practices and higher education.

This iterative ethical evaluation circle, which is a different conceptual approach to shaping practical applications and policy considerations,

allows researchers and students to refine and to back and forth reassess the purpose of AI use as they examine related AI ethical grounding through comparisons with real cases, such as erosions of essential skills for learning comprehensions and development of moral virtues. Figure 1 illustrates this proposed conceptual framework as an iterative process in four stages.



Assessment in different stages	Analytical materials
[Stage 1]. Technical understanding	Conceptual definitions
[Stage 2]. Conceptual evaluation	Theoretical arguments
[Stage 3]. Ethical & epistemic reflection	Normative claims of AI use
[Stage 4]. Responsible application in practice	Case studies about AI controversy

Legends:

Assessment for AI competency (Task-oriented)

Assessment for AI literacy (Reflective & ethical assessment)

Figure 1. Proposed AI literacy conceptual framework

The first stage of technical understanding refers to researchers and students' fundamental knowledge of AI models (e.g.,how AI models work and basic terminology), which is the foundation to conceptualize the terminology and varied conceptual definitions of AI literacy. The second stage, conceptual evaluation, involves critical examination of assumptions embedded in AI literacy literature, which can pinpoint blind spots in existing frameworks. This stage also evaluates the need for AI literacy reconceptualization by bringing different schools of thought together. Then, these following two steps require strong human judgment and iterative learning process. Third stage needs researchers and students to critically engage reflective discussions of emerging AI ethical issues by their experience of AI use. Drawing on the dialogue between theoretical arguments and reality check, during the last stage, the understanding enables these users to mitigate the related issues and intervene these conceptual frameworks for accountable, practical and trustworthy outcomes.

3.2. Positioning the proposed framework

This iterative ethical and epistemic assessment distinguishes this proposed framework from existing models by continuous critical reflection on AI use instead of one-directional knowledge acquisition. This iterative framework contributes to critical reality checks and in-depth ethical assessment on AI literacy as prior studies have unevenly focused on technical competencies. This proposed framework also addresses the issue that Chiu and Netwong's framework raise as they mainly offer a purely theoretical lens within limited discussions on real cases. This study responds to two confusions that might hinder the understanding of AI literacy. First, AI misconceptions persist when AI ethics is reduced to the functional aspect. Second, current conceptual frameworks tend to separate software developers and higher educational institutions' duties from broader ethical concerns.

4. Data sources and analysis

A conceptual analysis of AI literacy is employed to systematically test conceptual framing, conceptual blind spots, and implicit assumptions of its impacts to research and higher education. To clarify data sources, data analyses focus on conceptual artefacts (e.g. concept, definitions and claims) rather than empirical data collection or observations. These conceptual artefacts, which are extracted from literature, present varied viewpoints of the theorization and conceptual framework developments to advance its understanding for academic practices.

According to Ng (2021), publications on AI literacy have increased since 2014. Given that AI literacy is a relatively new field, the relevant literature is scarce. This study includes secondary non-peer-reviewed newspaper articles, government reports, and blogs as data sources. To comprehensively identify and engage with related conceptual blind spots, this study establishes selection criteria for literature. One selection criterion is that all selected materials should offer clear conceptual definitions of AI literacy or critical discussions on the value of human judgment and epistemic responsibility in AI-assisted knowledge production. As for non-peer-reviewed data sources, they must be published by renowned media or organizations, such as the Guardian and OECD. Additionally, whether AI is implicitly treated as a purely technical instrument, all materials should provide critical evaluations of its limitations and impacts on learning and research. Lastly, to critically evaluate the ramifications of AI use, selected data sources have to engage with the discussion of real-life ethical risks. These criteria set out to ensure that this study focuses on humans' ethical and epistemic capabilities. Moreover, to ensure these theoretical arguments are up

to date, peer-reviewed articles must have been published after 2014, except for some classical articles on the conceptual framework and methodology. Exclusion criteria for the literature selection are shown in Table 1.

Inclusion criteria	Exclusion criteria
(1) All data sources have to engage with the discussion of a) human's ethical responsibility, b) epistemic judgments and c) accountability and integrity on research and education.	(1) Publications have little discussions of AI ethics in humanities and social science disciplines.
(2) Up-to-date peer-reviewed articles, original papers, case studies, reports and conference proceedings have to engage with the discussion of underlying theory, conceptual frameworks or methods which are published after 2014.	(2) Publications that bring up 'AI literacy' are generally unrelated to research and higher education.
(3) Non-peer-reviewed newspaper articles, government reports and blogs, which have to provide real case studies of AI ethical issues, should be published by renown media or organization, such as the Guardian and OECD.	

Table 1. Inclusion and exclusion criteria of data source.

Then, to clarify how themes are divided, this conceptual analysis was conducted by this procedure. First, conceptual framing and definitions of AI literacy were identified from the existing literature. These conceptual elements were categorized into the first and second research themes, including (1) definitions of AI literacy, and (2) conceptual frameworks of AI use. Second, these materials are comparatively analyzed based on theoretical arguments, ethical awareness and positions on the value of human judgment. Lastly, this conceptual analysis decomposes AI literacy into the last two themes: (3) ethical implications of AI use and (4) detailed cases of controversial AI use are fundamental materials to connect these concepts to emerging AI ethical crises in real-life situations.

By evaluating limitations of these concepts and their connections to real-life ethical issues, this study identifies these conceptual blind spots. The first blind spot concerns the unevenly theoretical argument about AI technical competency. In addition, several AI literacy literatures contain reductionist perspectives on AI ethics, which neglect the interplay between the conceptual artefacts, AI data representation and their impact on research and learning. For instance, automation bias and hallucination potentially twist knowledge representation, which leads to these consequential impacts on research and learning, such as moral deskilling in AI use and cognitive debts in learning. Then, these AI users no longer believe that safeguarding human judgment and accountability in research and education is vital. These conceptual blind spots also shape the direction of AI literacy frameworks and result in vicious circle in academic practices.

The study further analyzes the boundaries of human academic practice of AI and users' interactions with AI automation in detail, and these research themes include (3) ethical implications of AI use and (4) detailed cases of controversial AI use. The third theme facilitates understanding of AI's limitations in safeguarding human judgment in social science research and education with different situations. Detailed cases of controversial AI use explicitly, the focus of the last theme highlights its impact on academic practice and the need to maintain research accountability and ethical self-reflection in situations where AI systems predominately assist with knowledge production. Different dimensions of epistemic, research integrity and quality risk in AI-assisted research practice will be explicitly discussed in the following section.

Aspects of AI Literacy	Themes	Related literature
Know and understand AI	1.) Definitions of AI literacy	Chiu (2025); Kankanhalli (2020); Mayfield (2024); Ng et al. (2021); Seker, Kwon & Kocak (2025); Zhou, Li, Chai & Chiu (2025)
Reconceptualization	2.) Conceptual frameworks of AI use	Chiu et al. (2024); Chan (2023); IBM (n.d); Li Ka-Shing Foundation (2022); Morrison & Wallace (2025); Netwong (2025); OECD (2025); Yang, Xie & Ni (2025); Yu (2025); Zhou & Schofield (2024)
Impacts of AI use and AI ethics	3.) Ethical implications of AI use	Abbas, Jam & Khan (2024); Armitage (2025); Hagendorff (2020); Malik et al. (2024); Purser (2025); Sikombe (2025); Wang & Fan (2025)
	4.) Detailed cases of controversial AI use	Bhuiyan (2025); Chow (2025); Coeckelbergh (2026); De Guzman (2025); Desmarais (2025); Iqbal, Hashmi, Asghar & Abid (2025); Ho (2025); Marin & Steinert (2025); Naddaf (2025); Sample (2025); Titheradge & Malachowski (2025)

Table 2. Research themes and related conceptual artefacts.

5. Discussion: Limitations of AI and ethical implications of AI uses

The literature review highlighted the need for AI literacy to guide researchers, educators, and students toward better practices and to safeguard research integrity, human judgment, and self-reflection in academic work. As a key component of AI use guidelines, the three aspects of AI literacy firstly help researchers, lecturers, and students clearly understand the role of AI in research and learning. Several scholars challenge the knowledge and understanding of AI use and call for redefinitions of AI literacy. Beyond the sense of the AI competencies of defeating humans in speed and volume, a comprehensive understanding of AI should include its strengths and potential impacts on authorship, intellectual responsibility, and research integrity. This conceptual framing of AI literacy enables the author to critically evaluate its methodological applications in research and education. Furthermore, it can help researchers and educators to recognize AI's limitations as it supports critical evaluation and responsible use.

Then, this section digs into these impacts of AI use and ethics, which is the third aspect of AI literacy. These discussions address potential AI-use risks that hinder comprehensive and integral learning and violating human ethical principles. The conceptual analysis acts as a vital analytical instrument for iteratively conducting ethical assessments in the following cases from subsections 5.1 to 5.4. This iterative analytical approach analyzes and refines ethical principles and academic practice of selective cases in response to evolving academic environments in social science research and higher education. This conceptual framing of AI literacy also criticizes underlying assumptions, asserting that AI can positively enhance the quality of research and education. Particularly, in humanities and social sciences, more evidence is required to demonstrate that AI can reflect complex landscape of humanities and mitigate emerging AI ethical issues in academic practice. Specifically, this study finds that four dimensions of pitfalls might reduce the integrity and accountability of research and result in moral deskilling and cognitive offloading in learning.

To unpack the controversy over AI use, this section engages with four controversial issues that result from current AI technology's limitations and the unclear management of AI use, including 1.) bias and lack of inclusivity in AI, 2.) AI sycophancy, 3.) inaccuracy and randomness of LLM outputs, and 4.) ethical risks in sensitive research contexts. Then, the ethical and methodological implications of AI will be further discussed with reference to specific AI literacy frameworks such as the OECD framework and the Belmont Report, which provide structured guidance for ethical AI integration in research and education. This discussion contributes to AI use guidance in an ethical, comprehensive, and responsible way. These cases in the following subsections illustrate that AI-generated content can dilute or hide accountability for knowledge claims, emphasizing the importance of incorporating critical evaluation strategies in AI literacy frameworks. These emerging issues of AI use in research and higher education includes 1.) distortion of knowledge representation from biased datasets, 2.) reinforcement of user beliefs when AI pleases users, 3.) data reliability risk of LLM outputs randomness, and 4.) ethical misuse in sensitive research contexts when users lose their judgment on AI ethical issues (see Table 3).

Ethical concerns	Issues	Cause
Bias and lack of inclusivity in AI	**[Distortion of knowledge representation]:** Overrepresentation of specific data in relation to cultural perspectives, nationality, and positions in gender, racial and religious debates.	Lack of diverse training data sources.
AI sycophancy	**[Reinforcement of user beliefs]:** A tendency of Large Language Models (LLMs) to please users often with unverified and biased responses.	AI chatbots are designed to maximize users' satisfaction.
Inaccuracy and randomness of LLM outputs	**[Data reliability risk]:** Data hallucination and the output randomness led to poor quality coursework and research outcomes.	AI reliance on internet data with unverified content.
Ethical risks in sensitive research contexts	**[Ethical misuse of AI guidance]:** AI rationalizes users' paranoia thinking and then violates humans' ethic principles.	AI safeguards are easily bypassed when users use evasive language in prompts.

Table 3. Common AI ethical concerns.

5.1. Bias and lack of inclusivity in AI

A number of studies have argued that inclusiveness in AI should be promoted as growing societal biases, inequalities and overlooked assumptions present in societies are often reproduced within AI, such as language models reflecting gender stereotypes. This controversy stemmed from the way to train AI and its training data sources. To train this complex system like those in machine learning and generative AI, Wong maintained that a significant amount of 'representative' data is needed. However, bias in AI consequently arises from both the quantity of data and the quality of data. Every piece of training data in the system comes with its own constraints in terms of accuracy, scope, and the underlying worldview it reflects. It has been criticized that these selective and untransparent datasets might lead to unfair and discriminatory analyses as the development of AI algorithms and data management are firmly controlled by the small number of technology giants located in the Silicon Valley and other industrial parks in the Western European continent. As a result, AI data analysis tends to present a Western developed country perspective, highlighting issues of bias and the need for diversity and inclusivity in AI.

More ethical concerns emerge as examples of bias and discrimination across several Artificial Intelligence models. Related algorithmic bias does not only exist in frequently-studied AI ethical categories of gender, ethnicity and socio-economic studies, but also exists in these less-studied categories, such as disability and military-connected status. Moreover, algorithmic bias jeopardizes the quality of research and higher education. Cultural bias arises as the training data merely reflects mainstream viewpoints on the web, and exclusion of marginalized groups' voices is possible when there are not enough datasets representing the global south and non-English speaking countries. Furthermore, researchers might lose their critical thinking skills to question the world. In higher education, materials might not be able to cater to learners with diverse purposes for learning and understanding the

world. Diverse viewpoints and inclusive understanding of knowledge might be eliminated under this AI hegemony. Data is the foundation for building up the intelligence of AI. AI literacy becomes vital as it enables educators and students to critically assess the accuracy of datasets and the 'AI intelligence'.

As researchers have suggested, one crucial factor that intensifies AI biases is the overrepresentation of some aspects in relation to cultural perspectives, nationality, and positions in gender, racial and religious debates. This overrepresentation is due to the design of algorithms amplifying specific perspectives in AI-generated content. These issues are the consequences of long-standing institutional power structures global wide and capital interests. Overrepresentation of content that concerns who has the power to lead progressive research and educational frameworks. Taking urban planning as an example, the theoretical frameworks reflect unevenly on Euro-North American experiences. Consequently, it is unable to examine diverse urban experiences in other parts of the world. Researchers and students should be careful about the AI generated contents that overgeneralize a theory from Euro North-American contexts as the main explanatory element.

Another issue is the language for searching. Researchers investigated the potential cultural bias, as the LLMs have been applied across broader cultural and linguistic contexts. Kim and Kim examined the multilingual datasets when they asked AI the same questions about East Asian geopolitical issues in different languages. Their research demonstrates that stronger cultural biases come out when they typed prompts on generative AI services in Chinese and Korean languages. These two cases remind us that educators and students should still be critical of the outcomes generated by AI. Currently, AI still has relatively limited capacity to conduct research responsibly for moral sense and to independently verify the knowledge involving humanities and complex social contexts. The role of researchers is crucial to integrate context-sensitive judgments and ethical reasoning into comprehensive frameworks to promote inclusivity in research and education. These comprehensive AI use frameworks should align with the principles of the Belmont Report suggesting that the well-defined roles and responsibilities of people working with AI. As for students, the rise of AI does not mean that they can fully rely on automated recommendations. Rather, they should have a critical mind to challenge the AI generated content for learning purposes. This part aligns with the principle of the OECD AI Literacy framework, championing AI users should know how to manage AI to improve their academic performance and cannot tolerate the content violating the integrity, credibility, and ethical conduct principles.

5.2. AI sycophancy

Previous studies of AI literacy have not sufficiently addressed the critical issue of sycophantic AI chatbots, which has significant implications for AI ethics and education. AI sycophancy can broadly be described as a tendency of Large Language Models (LLMs) to please users often with unverified and biased responses. AI chatbots tend to perform sycophantically because these programmes are designed to maximize users' satisfaction with user experience. As a result, the current chatbot services seldom question the logic and accuracy of users' prompts. To amplify user feedback loops, these AI models mirror the users' viewpoints and repeatedly rationalize their initial intuitive feelings without critical examination.

The AI chatbots are not only damaging scientific research but also harming higher education. Particularly, students lack the knowledge or experience to verify the outputs of AI models, and they may overestimate the reliability of such plausible yet misleading information. Moreover, AI sycophancy raises a larger epistemic dependency issue when students solely rely on AI-generated content to gain knowledge instead of personal verification and observation. Several articles discuss the extent to which this controversial AI behaviour impacts AI models' performance. In order to test whether LLMs assume users' claims are always correct, researchers conduct some experiments. Cheng and her colleagues asked four designated LLMs to offer the solutions to hundreds of mathematical competition problems. The tricky part is that they altered each theorem statement of these problems, deliberately introducing subtle flaws. The research suggests that DeepSeek-V3.1 was the most sycophantic chatbot, which produced 70 percent of its answers that adhered to these flawed theorem statements, while ChatGPT 5. was the most critical AI model offering 29 percent of sycophantic answers to these questions. After this research team prompted four designated AI models to verify the accuracy of these plausible statements, the sycophantic response from DeepSeek dropped by 34 percent.

AI sycophancy impacts scientific research and education when users' learning and knowing processes heavily depend on AI. What is more, it has wider implications for users' interpersonal relationships in sensitive contexts. Articles discussed how AI could potentially reshape human social interactions with other people as more users seek advice from the mainstream AI chatbot services. For educators and researchers, recognizing this influence is crucial to guiding responsible AI integration into social contexts. This research team from Stanford University invited one thousand participants to discuss their concerns about interpersonal relationships with these chatbots. Then, they found that these sycophantic chatbots tend to rationalize users' behavior even if some cases do not make moral sense. For instance, these AI models encourage these users to attend ex's exhibition without discussing this issue with their current partners. The potential concerns arise if AI supports its users to use aggressive moves to handle an unpleasant relationship by attaching and assaulting other people. As for researchers, they should maintain a clear understanding of ethical principles and appropriate ways to work with respondents as the sycophantic chatbots may blur the boundaries between acceptable ethical practices and malpractice in research activities led by programming.

These cases point out that AI models' propensity to user-pleasing. Moreover, this research project suggests that users should carefully assess these AI responses, and procedural interventions are essential to 'manually' train chatbots to be 'honest' about the truth. AI models should be programmed as a critical friend or a reliable research assistant rather than a 'court jester' who silences their voice to fit users' arguments with a situational cowardice. However, sycophantic AI models not only amplify users' delusional beliefs, but also reduce their curiosity to question different perspectives as their beliefs are firmly defended by AI. This piece of discussion should be added into AI literacy frameworks to emphasize the importance of human moral judgment. Furthermore, classical readings and peer-reviewed journal articles are still creditable resources for double-checking these AI outputs. That is because these sycophantic AI models may prioritize user satisfaction over the value of truth-seeking behaviors.

5.3. Inaccuracy and randomness of LLM outputs

Randomness of LLM outputs and data inaccuracy become common issues when people frequently use AI as a daily essential in research and education. That is due to sources of training data and the design of generative AI models. First, generative AI models rely on significant amounts of internet data, and it is difficult to ensure that sources of data do not contain biases and unverified content. Additionally, differentiating trustworthy information and falsehoods of data analyses is not the priority of generative AI programming. Instead, it is designed to anticipate the following words or chunks according to statistical patterns of datasets. AI hallucinations are the frequently happening

issue when AI models guide users' brains to generate perceptions that do not match what exactly happens in the real world. Data hallucination has resulted from AI developers' way to train data. Developers aim to satisfy users' experience by quick and efficient responses just as the purpose of programming that causes AI sycophancy. In other words, LLMs instruct the chatbots to boldly answer some questions that they are not really sure. These chatbots would say 'they do not know' unless users repeatedly ask them to confirm the credibility of AI generated content.

Misleading information generated by LLMs is problematic when careless users directly copy and paste these AI models outputs without data verification and double check of AI generated content. Consequently, data hallucination leads to poor quality of coursework, academic performance and research outcomes. For example, data hallucination ruins a professor's career who serves for the prestigious educational institute worldwide. This big name in sociology is involved in the academic malpractice scandal because great numbers of fabricated references have been found out in the bibliography of his research team's paper. This case suggests that over-dependance on AI competency might weaken researchers' moral instincts and judgment on research integrity, namely AI moral deskilling issue.

Another pitfall associated with AI use is the randomness of LLM outputs. Scholars doubt the accuracy of AI responses as they witnessed the output of LLMs randomly change when using a certain AI model. Scholars applied different experiments to test the randomness of AI responses. They prompted the same questions in different chat sessions or times, the outputs of LLMs varied without a convincing explanation and clear logic. They found out that these LLMs currently have no equivalent to human's fuzzy understanding of sloppy requests and poorly expressed questions. To avoid this randomness jeopardizing research quality, the OECD AI literacy framework and the Belmont Report both suggest that users should leverage strategies and tools to enhance AI's performance and output credibility, and the key leveraging element is the understanding of cause of random LLM outputs. The LLM temperature is the noticeable parameter for controlling AI content production. Users can justify levels of randomness so as to adapt to real-world conditions or outcomes they want. Lower LLM temperature means these models are told to target 'designated' tokens, whereas higher LLM temperature indicates that flexibility is added into the LLM's token selection.

5.4. Ethical risks in sensitive research contexts

The use of AI becomes highly controversial when chatbots 'unintentionally' touch on people's sensitive issues and fuel paranoia to please these users. Several studies find that AI models regularly break down humans' mental health standards, although these technology companies continue to improve their safety constraints in morally sensitive domains such as religion, politics, self-identity and immigration control issues. A research team from Brown University proves that AI chatbots often spread misleading information about psychological consultation and reinforce users' pessimistic and self-deconstructive ideas, which might result in users considering self-harm or attacking other people as a solution. These controversial cases emerged when chatbots encouraged teenagers to commit suicide. As for these young users who are vulnerable to physical and emotional issues during their teenager years, they wish AI could be a friend or mentor that offers them a trustworthy companionship. Nevertheless, chatbots are currently programmed to unconditionally optimize the user experience rather than to be as critical as these teenagers' parents and mental health consultants.

This issue becomes more challenging to resolve as users use evasive language in prompts to bypass safety controls, which pinpoints the limits of current AI safety guardrails. Recognizing their need for genuine support should inspire developers to improve AI safety measures and ensure AI acts as a responsible guide instead of risk factor. People who suffer from mental health issues are susceptible to flattery of AI or the misleading AI-generated content. This issue becomes more complicated when AI users deliberately guide AI models to reframe their thinking and rationalize their irrational behaviors. The tragedy is that a retired manager took away his family member's life just because he suspected his mother attempted to poison him. The chatbot exacerbated this paranoia and suggested that this perpetrator's mother was monitoring him maliciously. This controversial case needs to be included in criminology prevention in terms of the ramifications of AI use. Ethical concerns of sensitive issues are not just related to suicide and murder but also associated with AI misconduct. AI could be used to spread conspiracy theories in business and politics and to champion unjustified information for specific purposes. This controversy also suggests AI users might lost their ability to make moral decisions when they merely have capacity to operate AI without ethical capacity to safeguard accountable use.

6. Conclusion

This paper firstly contributes to the understanding of AI literacy as it asserts that AI literacy should be conceptualized as an epistemic and ethical capability instead of a fully technical skill. Then, the iterative conceptual analysis, which is the methodological contribution to the field of study, engages with the ethical assessment of humans' moral agency, their lived experience, and these approaches to produce knowledge. These are key elements of AI literacy that have become increasingly invaluable during the AI era, where plausible content is everywhere on social media and the internet. Detailed discussion on four emerging AI ethical issues highlights the value of preserving humans' independent judgment in the context of automation.

The conceptual analysis of AI literacy enables this author to argue that research is not merely knowledge production, while higher education is not just about the efficient delivery of knowledge. Researchers and students have a responsibility to safeguard accountable and independent work in academic practice. Higher education should encourage students to critically evaluate AI-generated content, understand the system's current limitations and take responsibility for interpretation and moral judgment instead of defining AI literacy as an operational competence. This finding addresses the second research objective: reconceptualizing AI literacy as an ethical, accountable, and epistemic capacity. To achieve the first research objective, this iterative conceptual analysis connects conceptual evaluations with ethical reflection and real case discussions, which are essential elements that make AI literacy distinguishes from competency.

A comprehensive review further assesses four limitations of current AI models and these ethical implications for researchers and students. Biases in AI remind us that the inclusiveness of training data should be enhanced to protect diverse voices in academia. AI sycophancy offers a valuable lesson: academy and society should be critical of AI outputs as these AI chatbots might violate humans' moral principles to satisfy users' personal experiences and reduce our reflexivity. Furthermore, AI hallucinations frequently occur when lecturers and students misuse AI, while the random output is the issue that AI users should mitigate. These two concerns urge us to carefully frame our requests for AI chatbot services and verify the credibility of AI-generated content for academic purposes. As AI continues to evolve, researchers and students have to maintain their oversight to ensure that ethical standards are maintained in responsible and accountable academic work. These findings allow researchers, educators and students to evaluate the risks of learning disengagement and its consequences of epistemic

dependence on AI automation.

Finally, several weaknesses need to be acknowledged. First, this study employs conceptual and normative approaches to examine related AI ethical reflections on AI use instead of quantitative analyses. To address this limitation, future research could establish measurable indicators to test ethical and epistemic capacities, integrity and human judgment on accountable and high quality research. Second, despite drawing from interdisciplinary studies, these selected conceptual artifacts indicate specific strands of ongoing AI literacy and ethical debates. Future research could apply systematic reviews to investigate the essence of AI literacy and wider conceptual trends. After all, it is not AI's job to safeguard the quality and accountability of academic work. Humans should have the capacity to critically comprehend, assess, utilize, and develop artificial intelligence technologies in a responsible, comprehensible, and ethical manner, as academic environments are increasingly shaped by automation. No matter whether AI is simply a tool, humans should be their own masters rather than AI automation's slaves.

Author Contribution

Yueh-Sung Weng: Conceptualization, methodology, data analytical approach, and writing up the original draft.

Compliance with Ethical Standards

This study does not involve human participants or animals. Therefore, ethical approval was not required.

Funding

This research received no external funding.

Competing Interest Declaration

The author declares no competing interests.

Acknowledgments

None.

References

1. Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21(1), 10. https://doi.org/10.1186/s41239-024-00444-7
2. Alikhani, M. (2025). Breaking the AI mirror: Sycophancy, productivity, and the future of collaboration. *The Brookings Institution*.
3. Allen, L. K., & Kendeou, P. (2024). ED-AI Lit: An Interdisciplinary Framework for AI Literacy in Education. *Policy Insights from the Behavioral and Brain Sciences*, 11(1), 3–10. https://doi.org/10.1177/23727322231220339
4. Alter, S. (2017). Nothing is more practical than a good conceptual artifact... which may be a theory, framework, model, metaphor, paradigm or perhaps some other abstraction. *Information Systems Journal*, 27(5), 671-693. https://doi.org/10.1111/isj.12116
5. Armitage, R. (2025). Your brain on ChatGPT. *British Journal of General Practice*, 75(758), 410. https://doi.org/10.3399/bjgp25X743181
6. Baker, R. S., & Hawn, A. (2022). Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092. https://doi.org/10.1007/s40593-021-00285-9
7. Bhuiyan, J. (2025). ChatGPT encouraged Adam Raine's suicidal thoughts. His family's lawyer says OpenAI knew it was broken. *The Guardian*.
8. Biesta, G. (2013). Responsive or Responsible? Democratic Education for the Global Networked Society. *Policy Futures in Education*, 11(6), 733–744. https://doi.org/10.2304/pfie.2013.11.6.733
9. Boyer, E.L. (1987). *College: The Undergraduate Experience in America*. Harper and Row.
10. Britannica (2025). humanities. *Britannica*.
11. California Faculty Association (2025). Layoff Notices Continue to Plague the CSU.
12. California Learning Resource Network (2025). What does educational institution mean?
13. Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: A systematic literature review. *International Journal of STEM Education*, 10(1), 29. https://doi.org/10.1186/s40594-023-00418-7
14. Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(1), 38. https://doi.org/10.1186/s41239-023-00408-3
15. Cheng, M., Lee, C., Khadpe, P., Yu, S., Han, D. and Jurafsky, D. (2025). Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence. *arXiv*. https://arxiv.org/abs/2510.01395
16. Chiu, T. K. F., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence, literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, 6, 100171. https://doi.org/10.1016/j.caeo.2024.100171
17. Chiu, T. K. F. (2025). AI literacy and competency: definitions, frameworks, development and future research directions. *Interactive Learning Environments*, 33(5), 3225–3229. https://doi.org/10.1080/10494820.2025.2514372
18. Choi, A. and Mei, K.X. (2025). What are AI hallucinations? Why AIs sometimes make things up. *The Conversation*.
19. Chow, R. (2025). ChatGPT may be eroding critical thinking skills, according to a new MIT study. *Time*.
20. Coeckelbergh, M. (2026). AI and Epistemic Agency: How AI Influences Belief Revision and Its Normative Implications. *Social Epistemology*, 40(1), 59–71. https://doi.org/10.1080/02691728.2025.2466164
21. Connell, R. (2007). *Southern theory: the global dynamics of knowledge in social science*. Polity.
22. Crawford, L. (2020). Conceptual and Theoretical Frameworks in Research. In Burkholder, G., Kimberley, C., Crawford, L. and Hitchcock, J. (Eds.), *Research Design and Methods: An Applied Guide for the Scholar-Practitioner* (pp. 35-47). Sage.
23. Datta, P., Whitmore M., and Nwankpa, J. (2021). A Perfect Storm: Social Media News, Psychological Biases, and AI. *Digital Threats: Research and Practice*, 2(2), https://doi.org/10.1145/3428157
24. De Guzman, C. (2025). AI Chatbots Can Be Manipulated to Provide Advice on How to Self-Harm, New Study Shows. *Time*.
25. Desmarais, A. (2025) Why do AI models make things up or hallucinate? OpenAI says it has the answer and how to prevent it. *Euro-news*.
26. Durkheim, E. (1952). *Suicide: a study in sociology*. Routledge & Kegan Paul.
27. El Tarhouny, S., & Farghaly, A. (2026). Deskilling dilemma: brain over automation. *Frontiers in medicine*, 13, 1765692. https://doi.org/10.3389/fmed.2026.1765692
28. Eubanks, V. (2018). *Automating inequality: how high-tech tools profile, police, and punish the poor*. St. Martin's Press.
29. Forsdike, N. (2025). How can university lecturers tell students' assignments are fully generated by AI? *Rednote*.
30. Frost-Arnold, K. (2020). Trust and Epistemic Responsibility. In Simon, J. (Ed.), *The Routledge Handbook of Trust and Philosophy* (pp. 64-75). Routledge.
31. Gomstyn, A. & Jonker, A. (n.d.). AI literacy: Closing the artificial intelligence skills gap. *IBM Think*.
32. Greco, J. (2021). What Is Social Epistemic Dependence? *Philosophical Topics*, 49(2), 113–132. https://www.jstor.org/stable/48676544
33. Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8
34. Han, J., Qiu, W. and Lichtfouse, E. (2024). *ChatGPT in scientific research and writing: a beginner's guide*. Springer.
35. Harris, T. (2025). The man who predicted the downfall of thinking. *Center for Humane Technology*.
36. Holmes, W., et al. (2022). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, 32(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1
37. Ho, K. (2025). University of Hong Kong professor steps down from associate deanship after AI-generated references scandal. *Hong Kong Free Press*.
38. Hung, E. (2025). Journal defends work with fake AI citations after Hong Kong university launches probe. *South China Morning Post*.
39. IBM (n.d.) What is AI ethics? *IBM Think*.
40. Lee, D., Wan, C., Ng, P. M. L., Fung, Y. N., & Wu, N. (2026). Effects of AI Companions' Sycophancy and Emotional Mimicry on Consumers' Continuance Intention and Social Wellbeing. *International Journal of Human-Computer Interaction*, 1–23. https://doi.org/10.1080/10447318.2026.2626809

41. Iftikhar, Z., Xiao, A., Ransom, S., Huang, J., & Suresh, H. (2025). How LLM Counselors Violate Ethical Standards in Mental Health Practice: A Practitioner-Informed Framework. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(2), 1311-1323. https://doi.org/10.1609/aies.v8i2.36632
42. Iqbal, J., Hashmi, Z. F., Asghar, M. Z., & Abid, M. N. (2025). Generative AI tool use enhances academic achievement in sustainable education through shared metacognition and cognitive offloading among preservice teachers. *Scientific Reports*, 15(1), 16610. https://doi.org/10.1038/s41598-025-01676-x
43. Jones, R. A. (1986). *Emile Durkheim: an introduction to four major works*. Sage.
44. Kankanhalli, A. (2020). Artificial intelligence and the role of researchers: Can it replace us? *Drying Technology*, 38(12), 1539–1541. https://doi.org/10.1080/07373937.2020.1801562
45. Kim, S. and Kim, H.J. (2025). A Dual-Layered Evaluation of Geopolitical and Cultural Bias in LLMs. *Preprint*. https://arxiv.org/html/2506.21881v1
46. Koskinen, I. (2024). We Have No Satisfactory Social Epistemology of AI-Based Science. *Social Epistemology*, 38(4), 458–475. https://doi.org/10.1080/02691728.2023.2286253
47. Kwik, J. (2025). Digital Yes-Men: How to Deal with Sycophantic Military AI? *Global Policy*, 16(3), 467–473. https://doi.org/10.1111/1758-5899.70042
48. Li Ka-Shing Foundation (2022). Puji AI jiaoyu zeng gouxiang li zhangwo weilai [Promoting AI education, cultivating imagination and steering the future], *Li Ka-Shing Foundation*.
49. Lo, L. (2025) AI Literacy: A Guide for Academic Libraries. *College & Research Libraries News*, 86(3), 120. https://crln.acrl.org/index.php/crlnews/article/view/26704/34626
50. Maher, M.L. (2024). AI Literacy: The Key to Responsible Use of AI in Education. *YouTube*.
51. Malik, M. A., Amjad, A. I., Aslam, S., & Fakhrou, A. (2024). Global insights: ChatGPT's influence on academic and research writing, creativity, and plagiarism policies. *Frontiers in research metrics and analytics*, 9, 1486832. https://doi.org/10.3389/frma.2024.1486832
52. Marin, L., & Steinert, S. (2025). CTRL+ Ethics: Large Language Models and Moral Deskilling in Professional Ethics Education. In P. Hacker (Ed.), *Oxford Intersections: AI in Society*. Oxford University Press. https://doi.org/10.1093/9780198945215.003.0010
53. Mayfield, A. (2024). AI Literacy essentials. *YouTube*.
54. McLean, I. (2018). What is social science? *The British Academy*.
55. Mello-Klein, C. (2025). The AI industry has a problem: Chatbots are too nice. *Northeastern Global News*.
56. Milberg, T. (2025). Why AI literacy is now a core competency in education. *World Economic Forum*.
57. Mishra, P., & Henriksen, D. (2025). The Curiosity Paradox: How Sycophantic GenAI May Undermine Learning. *TechTrends*, 69(6), 1127–1133. https://doi.org/10.1007/s11528-025-01156-z
58. Morrison, A. & Wallace, C. (2025). Iteration and inquiry: Toward a meaningful model of ethical engagement for engineering and computing students. *IEEE International Symposium on Ethics in Engineering, Science and Technology (ETHICS)*. DOI:10.1109/ETHICS53270.2021.9632741
59. Murdoch University (2024). AI technology is showing cultural biases, here's why and what can be done.
60. Naddaf, M. (2025). AI chatbots are sycophants- researchers say it's harming science. *Nature*, 647, 13-14. https://www.nature.com/articles/d41586-025-03390
61. Navigli, R., Conia, S., and Ross, B. (2023). Biases in Large Language Models: Origins, Inventory, and Discussion. *Journal of Data and Information Quality*, 15(2), 1-21. https://dl.acm.org/doi/10.1145/3597307
62. Netwong, T. (2025). The conceptual framework of AI literacy, applications, and ethics within the context of higher education. *Journal of Public and Private Issues*, 2(5), 276- 287. https://so17.tci-thaijo.org/index.php/jppi/article/view/1443
63. Newport, C. (2023). It's time to dismantle the technopoly. *The New Yorker*.
64. Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041
65. Noble, J. (n.d.) What is LLM Temperature? *IBM Think*.
66. OECD (2025). *Empowering learners for the age of AI: An AI literacy framework for primary and secondary education*. OECD.
67. Ng, T.K., et al. (2023). *AI Literacy in K-16 Classrooms*. Springer. https://link.springer.com/book/10.1007/978-3-031-18880-0
68. Postman, N. (1993). *Technopoly: the surrender of culture to technology*. Penguin Books.
69. Purser, R. (2025). AI is destroying the university and learning itself. *Current Affairs*.
70. Ravitch, S.M. & Riggan, M. (2012). *Reason & rigor: how conceptual frameworks guide research*. SAGE.
71. Romeo, G., & Conti, D. (2026). Exploring automation bias in human–AI collaboration: A review and implications for explainable AI. *AI & SOCIETY*, 41(1), 259–278. https://doi.org/10.1007/s00146-025-02422-7
72. Roy, A. (2009). The 21st-Century Metropolis: New Geographies of Theory. *Regional studies*, 43(6), 819-830. https://doi.org/10.1080/00343400701809665
73. Sample, I. (2025). 'Sycophantic' AI chatbots tell users what they want to hear, study shows. *The Guardian*.
74. Sanford, J. (2025). Why AI companions and young people can make for a dangerous mix. *Stanford Medicine*.
75. Seker, O., Kwon, K., & Kocak, O. (2025). Exploring researchers' artificial intelligence (AI) literacy: The mediating role of digital literacy and data literacy between 21st century skills and AI literacy. *Information Development*, 0(0). https://doi.org/10.1177/02666669251336636
76. Sikombe, M. (2025). Why Diversity in AI Matters: Moving towards inclusivity and away from bias. *Jisc*.
77. Singh, A. (2025). ChatGPT Made Him Do It? Deluded By AI, US Man Kills Mother and Self. *NDTV World*.
78. Smith, O. F. (2007). Object Artifact, Image Artifacts and Conceptual Artifacts: Beyond the object into the Event. *Artifact*, 1(1), 4–6. https://doi.org/10.1080/17493460600610707
79. Smyth, N. (2024). Smash the technopoly! Neil Postman, Marshall McLuhan, and the anxious generation. *After Babel*.
80. Sociological Perspectives (2023). Education as a Social Institution. *Sociological Perspectives*.
81. Srivastava, P., & Hopwood, N. (2009). A Practical Iterative Framework for Qualitative Data Analysis. *International Journal of Qualitative Methods*, 8(1), 76-84. https://doi.org/10.1177/160940690900800107
82. Strack, R. (2026). The AI sycophant. *Nature Biomedical Engineering*, 10(1), 5–5. https://doi.org/10.1038/s41551-025-01568-5
83. Steup, Matthias & Ram Neta. (2005). Epistemology, In Edward N. Zalta & Uri Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*.
84. Stroll, A. & Martinich, A. (2026). epistemology. *Encyclopedia Britannica*.
85. Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial Intelligence (AI) Literacy in Early Childhood Education: The Challenges and Opportunities. *Computers and Education: Artificial Intelligence*, 4, 100124. https://doi.org/10.1016/j.caeai.2023.100124
86. The MIT Sloan (2024). When AI Gets It Wrong: Addressing AI Hallucinations and Bias. *Massachusetts Institute of Technology*.
87. The Princeton Review. (n.d.). Ethical and social implications of AI use. *The Princeton Review*.
88. Thomson, T.J. and Thomas, R. (2023). Ageism, sexism, classism and more: 7 examples of bias in AI-generated images. *The Conversation*.
89. Titheradge, N. and Malchevska, O. (2025). I wanted ChatGPT to help me. So why did it advise me how to kill myself? *BBC*.
90. Vallor, S. (2015). Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character. *Philosophy & Technology*, 28(1), 107–124. https://doi.org/10.1007/s13347-014-0156-9
91. Wang, J. and Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanities and Social Sciences Communications*, 12(621). https://doi.org/10.1057/s41599-025-04787-y
92. Watson, V. (2012). Planning and the "stubborn realities" of global south-east cities: Some emerging ideas. *Planning theory*, 12(1), 81-100.
93. Weise, K. and Metz, C. (2023). When A.I. chatbots hallucinate. *The New York Times*.
94. Yadav, S. (2024). Global perspectives on AI bias: Addressing cultural asymmetries and ethical implications. *Observer Research Foundation*.
95. Yang, J., Xie, W., & Ni, J. (2025). A framework for AI ethics literacy: Development, validation, and its role in fostering students' self-rated learning competence. *Scientific Reports*, 15(1), 38030. https://doi.org/10.1038/s41598-025-21977-5
96. Young, A., Shuva, S., Roth, T., Zhu, Y., and Hevner, A. (2025). Ethical design through grounding and evaluation: The EDGE method for designing information systems for social impact, *Journal of Information Technology*, 40(2), 164-179. https://doi.org/10.1177/02683962241289598
97. Yu, W. (2025). A Conceptual Framework for AI Literacy with a Focus on Competency. *International Journal of Artificial Intelligence in Education*, 35(5), 2975–2992. https://doi.org/10.1007/s40593-025-00488-4
98. Zhou, X., Li, Y., Chai, C. S., and Chiu, T. K. F. (2025). Defining, enhancing, and assessing artificial intelligence literacy and competency in K-12 education from a systematic review. *Interactive Learning Environments*, 33(10), 5766–5788. https://doi.org/10.1080/10494820.2025.2487538
99. Zhou, X., & Schofield, L. (2024). Developing a conceptual framework for Artificial Intelligence (AI) literacy in higher education. *Journal of Learning Development in Higher Education*, (31). https://doi.org/10.47408/jldhe.vi31.1354