Generative Al Literacy: A Comprehensive Framework for Literacy and Responsible Use

CHENGZHI ZHANG, Georgia Institute of Technology, USA BRIAN MAGERKO, Georgia Institute of Technology, USA

After the release of several AI literacy guidelines, the rapid rise and widespread adoption of generative AI-such as ChatGPT, Dall·E, and Deepseek-have transformed our lives. Unlike traditional AI algorithms (e.g., convolutional neural networks, semantic networks, classifiers) captured in existing AI literacy frameworks, generative AI exhibits distinct and more nuanced characteristics. However, a lack of robust generative AI literacy is hindering individuals' ability to evaluate critically and use these models effectively and responsibly. To address this gap, we propose a set of guidelines with 12 items for generative AI literacy, organized into four key aspects: (1) Guidelines for Generative AI Tool Selection and Prompting, (2) Guidelines for Understanding Interaction with Generative AI, (3) Guidelines for Understanding Interaction with Generative AI, and (4) Guidelines for High-Level Understanding of Generative AI. These guidelines aim to support schools, companies, educators, and organizations in developing frameworks that empower their members—such as students, employees, and stakeholders—to use generative AI in an efficient, ethical, and informed way.

CCS Concepts: • Computing methodologies \rightarrow Artificial intelligence; • Social and professional topics \rightarrow Computing literacy; • General and reference \rightarrow Surveys and overviews.

Additional Key Words and Phrases: Generative AI, AI Literacy, Education, Guideline

ACM Reference Format:

1 INTRODUCTION

Since the release of the "Five Big Ideas in AI" by the AI4K12 organization in 2019 [108], followed by AI literacy competency frameworks proposed by Long & Magerko in 2020 [63] and Ng et al. in 2021 [77], the field of artificial intelligence has undergone significant transformation. A key development has been the rise of generative AI, which has gained widespread popularity in recent years. Tools like ChatGPT ¹ and Dall-E ² gained worldwide popularity in early 2023, sparking a surge in applications such as Copilot ³ and Deepseek ⁴. Unlike traditional AI algorithms like decision trees, semantic networks, and convolutional neural networks (CNNs) captured in existing widely accepted AI literacy guidelines, generative AI exhibits distinct and more specialized characteristics. For instance, it has

Authors' addresses: Chengzhi Zhang, czhang694@gatech.edu, Georgia Institute of Technology, Atlanta, Georgia, USA; Brian Magerko, magerko@gatech.edu, Georgia Institute of Technology, Atlanta, Georgia, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

©~2025 Association for Computing Machinery.

Manuscript submitted to ACM

¹ChatGPT: https://chatgpt.com/

²Dall-E https://openai.com/index/dall-e-3/

³Copilot: https://copilot.microsoft.com/

⁴Deepseek: https://www.deepseek.com/

demonstrated advanced multilingual comprehension capabilities [1] and achieved human-level performance on cognitive aptitude tests-including General Language Understanding Evaluation (GLUE) and Stanford Question Answering Dataset (SQuAD) [94]. However, these impressive human/superhuman capabilities often lead users to overestimate the intelligence of generative AI systems, a phenomenon known as "Eliza effect" [34]. This term refers to the tendency to attribute greater intelligence to responsive computer programs than they actually possess [104]. Such overconfidence in generative AI's abilities has resulted in widespread misconceptions about its limitations [51, 55]. These misunderstandings can have far-reaching consequences across various domains. For instance, they can erode professionalism in the workplace [75], compromise integrity in academic settings such as scientific writing [2], and even result in fatal outcomes for both adults [120] and minors [8].

The unique characteristics of generative AI-differentiating from general AI models—and the potential harms it can cause due to public misconceptions highlight the need for specialized AI literacy guidelines tailored to generative AI. Unlike AI algorithms like decision trees that discriminate and classify (e.g., spam filters) and convolutional neural networks (CNNs) that produce deterministic results (e.g., speech-to-text generators), mainstream large-pretrained generative models often rely on advanced transformer architectures [114]. These architectures introduce a high degree of non-determinism and variability in their outputs. However, existing AI literacy guidelines fail to address these specificities, as they primarily focus on non-generative AI systems.

Current literature on "generative AI literacy" is limited but growing. Examples include twelve defining competencies for generative AI literacy [6], four generative AI literacy dimensions including "knowledge", "application", "evaluation" and "ethics" [88], The Generative AI Literacy Assessment Test (GLAT) [49], and workshops aimed at educating students on the use of generative AI [102]. While the twelve competencies outlined in [6] provide a high-level framework (e.g., "3. Knowledge of the capabilities and limitations of generative AI tools"), our work consolidates these ideas into actionable guidelines for policymakers and educators to design effective learning interventions. Our framework focuses on four key aspects of using generative models: (1) Guidelines for Generative AI Tool Selection and Prompting, (2) Guidelines for Understanding Interaction with Generative AI, and (4) Guidelines for High-Level Understanding of Generative AI. By providing practical guidelines for understanding and interacting with generative AI, this work aims to bridge the gap between its rapid technological advancements and the public's ability to use it responsibly, critically, and effectively.

2 DEFINING GENERATIVE AI LITERACY

What is generative AI. Generative AI refers to computational techniques that are capable of generating seemingly new, meaningful content such as text, images, or audio–from training data [36]. Similarly, Muller et al. describe generative AI systems as those that produce novel and creative content, including images, texts, music, video, code, and other forms of design [74]. Beyond these definitions, David Foster's book provides a more comprehensive visualization of the history of generative AI, categorizing its development and highlighting representative products up to early 2023 [38]. Foster identifies six main categories of generative AI: (1) Variational Autoencoder (VAE) (e.g. [95]), (2) Generative Adversarial Network (GAN) [41], (3) Autoregressive/Transformer [114], (4) Normalizing Flow, (5) Energy-Based/Diffusion Model and (6) Multimodal Model (e.g. DALL·E ⁵). After the birth of AI systems in the 1950s [44], generative AI like VAE and GAN models gained prominence in the 2010s, particularly in image generation tasks. However, the introduction

⁵Dall·E: https://openai.com/index/dall-e-3/

of the Transformer architecture in 2017 [114] marked a new turning point, leading to the development of ChatGPT (GPT stands for "Generative Pre-training Transformer") and sparking a global surge in generative AI applications.

To define generative AI literacy guidelines, we conducted an exploratory review of existing literature. Given the rapidly evolving nature of generative AI, we anticipate that future researchers will expand upon this work to develop more refined guidelines. This framework serves as an initial step toward developing more comprehensive resources for educators, policymakers, and researchers to design learning interventions, such as curricula, workshops, and other tools, aimed at improving public understanding of generative AI. We define generative AI literacy as:

A set of guidelines for designing learning interventions aimed at enhancing users' understanding of generative AI, enabling them to interact with it effectively, responsibly, and critically. These guidelines are intended to assist educators, schools, companies, and organizations in developing frameworks that empower their members, such as students, employees, and stakeholders, to use generative AI in ethical and informed ways.

The most widely utilized form of generative AI is the generative language model, commonly referred to as large language models (LLMs) or large, pre-trained language models (PLMs) [71]. While generative AI broadly includes applications such as video generation [57], image generation [61], design generation [122], and code generation [103]. Our guidelines primarily focus on text-to-text generative models. These models are the most widely used and studied, though many of the guidelines can also apply to other types of generative AI systems.

3 METHOD

While systematic literature reviews are commonly used to identify practices in well-established research fields, generative AI literacy is a nascent research area with limited coverage in existing literature. To address this gap, we employed a scoping study to systematically gather and develop a comprehensive set of guidelines. The scoping study is "an approach to reviewing the literature which to date has received little attention in the research methods literature" [7], and it has previously been adopted in defining AI literacy competency [63]. Differentiating from a systematic review that aims at "searching for particular study designs," the goal of scoping studies is to have "in-depth and broad results" [7], ensuring a comprehensive coverage of existing literature.

In this active searching process, guided by our definition of generative AI literacy, we formulated two key selection criteria to direct our search: (1) Is this knowledge essential for generative AI users with a non-technical background? (2) Will including this in the guideline help learners interact with generative AI more effectively, ethically, and responsibly?

In the search process, we began by querying the ACM library, the Springer library, and Google Scholar using the keywords such as "generative AI literacy", "generative AI characteristics", and "generative artificial intelligence literacy". However, the preliminary search only yielded a limited set of literature. Thus, we expanded our search to include both peer-reviewed articles and grey literature (e.g., government reports, policy documents, and working papers) while incorporating additional keywords including "generative AI", "ChatGPT", "LLM", "foundational models", "guideline", and "use wisely". To enhance the comprehensiveness of our search, we incorporated two AI-empowered academic search tools—Semantic Scholar ⁶ and Elicit ⁷. These tools enabled us to identify literature that was semantically relevant but did not necessarily include those keywords in their titles. For example, the book *Generative Artificial Intelligence: What Everyone Needs to Know*, which provides guidelines for generative AI users, was discovered through semantic search,

⁶Semantic Scholar: https://www.semanticscholar.org/

⁷Elicit: https://elicit.com/

Table 1. Literature reviewed by venue type

Venue Type	
Conference papers	19
Journal papers	26
Books	4
Other grey literature	43

and it did not appear in keyword-based queries [52]. Additionally, we incorporated snowball sampling, a technique recommended in scoping study processes [7]—by examining the bibliographies of relevant studies to identify additional sources [91]. This approach allowed us to gather a diverse range of materials, including peer-reviewed, generative AI use guidelines from accredited university libraries [47, 59, 81, 82], international non-profit organizations like UN-ESCO [110] and Counsel for European Union [21], and public-facing course from globally recognized leader in AI [76]. This initial search and subsequent refinement resulted in a draft set of guidelines, with each item supported by at least one relevant literature.

Following the initial search, the first author conducted an exploratory reading process that involved reviewing the abstracts and skimming the contents of the collected literature. For relevant sections, the first author conducted a thorough reading and labeled each literature according to the draft guideline items. Throughout this process, the first author consulted the second author for professional input and advice. This involves adding, consolidating, or removing draft guideline items. Additionally, they searched for relevant literature to support and refine the guidelines until reaching a literature saturation. In total, we reviewed 92 literature and developed a set of 12 guidelines, with each literature supporting at least one guideline item. The Table 1 listed the categories of literature reviewed. For consistency and to emphasize the educational focus of the guidelines, we use the term "learners" throughout the paper to refer to users of generative AI.

4 GUIDELINES FOR GENERATIVE AI TOOL SELECTION AND PROMPTING

• G1: Learners need to carefully determine whether they should use, and which generative AI tool to use. (Supporting References: [22, 47, 94, 110])

Learners need to determine whether the scenario is appropriate for using generative AI at the forefront. The United Nations Educational, Scientific and Cultural Organization (UNESCO) provided a flow chart to help determine the safe use of ChatGPT [110]: if users can not verify the accuracy of the output and correctness is critical, then it is unsafe to use it. Universities have varying stances on using generative AI. Some are embracing it as an educational tool and offering courses like "Facilitating Learning and Research with Generative AI" [82] while others consider submitting AI-generated content in whole or in part generated with the use of generative as cheating [112]. Learners should consult the institution's policy and evaluate the context before using generative AI. Different generative AI tools have distinct capacities. Reference sheets listing key features, such as advantages, disadvantages, privacy policy, and trained data of key generative AI tools can help learners select the proper tool for their best advantage [47, 94]. Learners should consider institutional policies, the use contexts, and tool performance to decide whether they should use generative AI and which generative AI tool to use.

• G2: Learners need to develop safe and effective prompting skills, ensuring user privacy while interacting with efficiency. (Supporting References: [6, 50, 58, 59, 76, 81, 82, 102, 110])

How people prompt generative AI. Existing research with users prompting LLMs reveals the users' prompting characteristics: (1) opportunistic but not systematic, (2) users have a tendency to over-generalize the responses, (3) users hold the inappropriate mental model derived from human-to-human instructional experience [124]. For generative image AI, similar research also witnessed users' inappropriate mental models, including expecting generative image AI to interpret prompts like humans, thus leading to failed iterations [65]. Apart from these two studies with adults, a study with children further witnessed that children interact with chatbots as if they were interacting with people (e.g., asking "Are you a boy or girl?") [11]. These all revealed the misconceptions and inappropriate prompting techniques people hold towards generative AI, highlighting the need for adequate prompting skills.

Effective prompting skills involve prompting generative models, like chatbots, with questions they are capable of answering. A well-accepted framework is the CLEAR prompting guideline [62], which stands for "Concise", "Logical", "Explicit", "Adaptive" and "Reflective". The guideline emphasizes not only crafting a single well-structured prompt but also iteratively refining it based on the model's output. The CLEAR guideline has been endorsed by university libraries [58, 59]. In the context of text-to-image generative models, researchers have developed specialized guidelines [60]. Some products (e.g., Dall·E) have released their product-specific guidelines [25]. Prompting text-to-image generators requires additional expertise, such as adjusting the parameters and seeds to achieve the desired results [121]. Beyond effectiveness, safety is a critical consideration. Librarians advise against inputting personal data [58] and recommend privacy-protected AI tools [84]. Those measures can help safeguard data privacy and increase effectiveness.

5 GUIDELINES FOR UNDERSTANDING INTERACTION WITH GENERATIVE AI

• G3: Generative AI has a limited context window, so users need to re-frame or repeat context as needed. (Supporting References: [48, 49, 79, 124])

Generative AI's limited context window is a significant challenge for large language models (LLMs) [48]. This limitation causes AI to forget earlier inputs during long conversations and "struggle to maintain continuity within long dialogs" [79]. Researchers are actively working to extend these context windows [29, 48], but this issue persists. Empirical studies show that users who are unaware of this constraint often experience frustration [124]. This is a constraint highlighted in the Generative AI Literacy Assessment Test (GLAT) [49]. To avoid frustration, users must recognize the differences between interacting with generative AI and humans. This includes re-framing and repeating the context as needed in order to maintain continuity in conversations.

• G4: Generative AI lacks social cognition but can demonstrate fake empathy and theory of mind. (Supporting References: [24, 79, 97, 116, 126])

LLMs generate word sequences by deploying probabilistic methods to predict the next word. While these models can simulate human-like responses, they lack the capacity for social cognition, including theory of mind (ToM). Even when techniques are employed to make LLMs appear as though they possess ToM [126], this remains a superficial imitation rather than a genuine understanding.

As defined by Premack, theory of mind refers to the ability to attribute mental states—such as beliefs, intentions, and emotions—to oneself and others, and to use this understanding to predict behavior [92]. While Google's LaMDA reportedly passed a famous ToM assessment administered by Blaise Agüera y Arcas [79], this result remains controversial. An up-to-date pre-print with existing literature concluded that although new LLMs demonstrated advanced ToM

Manuscript submitted to ACM

abilities, they still rely on spurious correlations instead of solid understanding [97]. This highlights a lack of genuine ToM abilities.

LLMs also fail to achieve shared mental state construction (SSM), a concept described in the *International Encyclopedia of the Social & Behavioral Sciences* as the ability of team members to use shared knowledge to predict task needs and anticipate others' actions, enabling adaptive behavior [100]. Generative AI lacks the capacity for participatory sense-making, which involves the dynamic generation and transformation of meaning through interaction [27], showing a more general lack of understanding of social cognitive processes. Despite these shortcomings, people often turn to AI for on-demand social validation [109], which can reinforce people's existing biases and trap users in "bias bubbles." Moreover, LLMs can demonstrate fake empathy and increase the vulnerability of people interacting with LLMs, as it creates an illusion of connection and understanding [24]. These limitations highlight LLMs' broader inability to engage in genuine social cognitive processes and warn users against projecting wrong mental models after witnessing the human behaviors they demonstrate [89].

6 GUIDELINES FOR UNDERSTANDING GENERATIVE AI'S OUTPUT

• G5: Generative AI can output content that is harmful to people, so learners should treat the outputs with caution and skepticism. (Supporting References: [39, 80, 81, 84, 110])

Despite the multilingual capabilities and superhuman fact-retrieval abilities demonstrated by large language models (LLMs), generative AI outputs are not infallible. It can produce harmful, inappropriate, deceptive, and misleading content. For instance, there have been reports of teenagers and adults using chatbots-powered by generative AI-that generated dangerous instructions, such as those leading to electric shocks [10] or misleading advice with severe ethical implications, including instructions that could result in harming their guardians [80]. Additionally, generative AI has led teenagers into emotionally and sexually abusive relationships, and caused fatal consequences for both minors [18] and adults [106].

To mitigate these risks, some companies have adopted red-teaming practices before launching their generative AI models, and shared related tools with developers building generative AI applications on their company platform [15, 70]. While these measures help assess and reduce the likelihood of generative AI producing harmful content, they cannot entirely eliminate the risks. Even with such safeguards in place, learners must remain vigilant about the potential for harmful outputs.

• G6: Generative AI can output misinformation and disinformation, so learners should critically consume the results for correctness, especially in high-stakes contexts. (Supporting References: [49, 83, 90, 107, 111, 113, 115])

Differentiating from harmful contents—that emphasize non-false but damaging contents that can directly harm individuals or groups—misinformation, and disinformation focus on "falsified facts" which are verifiable if users take the time to validate their accuracy. This aligns with longstanding calls in academic literature for critical information literacy [14] and critical media literacy [54], which emphasize equipping users with the skills to interrogate the authenticity of content.

In some literature, it is called the hallucination effect [28]. However, this term is misleading. According to the American Psychology Association (APA), "hallucination" means "a false sensory perception that has a compelling sense of reality despite the **absence** of an external stimulus." This deviates from the "seemingly true" but false information, "factually incorrect" and "fabricated information" that learners encounter with AI-generated content. Instead, the terms Manuscript submitted to ACM

"misinformation" and "disinformation" are more appropriate. The APA defines misinformation as "false or inaccurate information—getting the facts wrong," and disinformation as "false information which is deliberately intended to mislead—intentionally misstating the facts" [4]. As a widely used generative AI tool, ChatGPT has gained notoriety for generating fabricated academic citations [2], misleading narratives [105], and misinformation like bogus case law [75]. These all pose significant challenges to people who use ChatGPT as a "search engine" without knowing those true limitations. Learners should distinguish between low-stakes scenarios (e.g., using generative AI for inspiration or drafting informal emails) and high-stakes scenarios (e.g., providing legal clauses for lawsuits). In high-stakes situations, regular fact-checking is essential to mitigate risks. While AI-generated misinformation is not a new issue, the ease, speed, and credibility with which generative AI can produce such content are unprecedented [125]. Due to the speed, researchers have raised concerns that large language models (LLMs) could pollute our information ecosystems [67, 101], exacerbating the spread of false information.

• G7: Generative AI can lack results explainability, so learners need to cross-check the information. (Supporting References: [32, 49, 99, 119])

The lack of explainability in large language models (LLMs), rooted in the complex mathematical nature of deep neural networks, renders their decision-making processes a "black box" to learners. As a result, learners must fact-check important information generated by these systems. While tools like Deepseek (with web-search feature), as well as academic tools such as Sourcely ⁸ and Elicit ⁹, can provide sources for information, many generative models remain opaque and lack transparency in their outputs.

Explainable AI (XAI) system aims to address this issue by making AI behaviors more intelligible to humans through detailed explanations [43]. XAI is seen as critical for verifying information and providing traceable, investigable sources. Existing XAI approaches include technically visualizing neural networks—the foundational building blocks of generative AI [123], and socially constructing situated XAI systems [32] with automated explanations [33]. However, these methods still fall short of the true "explainability" found in human-to-human interactions, pushing the responsibility on users to fact-check AI-generated results. Even if with systems that can provide evidence and explainability, learners still need to check the linkage and the rationale between the provided source and the generative AI's output, ensuring the provided explanations support the claims. **To ensure best accuracy, users must compare AI outputs with ground truth information.** Specifically, university libraries recommend strategies like "lateral reading" [119], which involves cross-referencing multiple sources to verify validity [96]. This approach can help mitigate the risks associated with relying on opaque AI systems.

• G8: The innate bias in the generation dataset that can result in biased results, so learners need to critically evaluate output for unmitigated bias. (Supporting References: [12, 13, 19, 37, 62, 81, 85, 93, 112])

Bender et al. provide a detailed analysis of the origins of bias in AI systems from a computational linguistics perspective. They highlight that bias stems from the datasets used to train these models, which often include biased information from internet sources and a lack of media coverage for less-digitalized communities [12], etc. As a result, ChatGPT and similar models can exhibit gender and racial bias [87], cultural and linguistic bias [94], and bias against individuals with disabilities [40], etc. These biases prevent the models from providing fair and balanced feedback. While bias in text generation can often be implicit, bias with text-to-image generators is explicit. For example, prompting an image

⁸Sourcely: https://www.sourcely.net/

⁹Elicit: https://elicit.com/

AI with "Black African doctors providing care for white suffering children" has produced children who are invariably black while doctors are white [31]. Learners should be aware of the bias and be critical of the generated results.

7 GUIDELINES FOR HIGH-LEVEL UNDERSTANDING OF GENERATIVE AI

• G9: Digital and non-digital content encountered in our daily lives can possibly be AI-generated. Learners need to critically evaluate how genuine media content is. (Supporting References: [16, 21, 45, 73, 125])

Existing generative AI literacy guidelines [13] emphasize the ability to distinguish AI-generated content as a key competency. However, some argue that humans struggle to reliably identify such content [16, 72], and it is becoming an increasingly more difficult task. Therefore, we propose that users should focus more on **being aware of the existence of AI-generated content and being critical of what they see** rather than striving to distinguish that content. This is particularly relevant in higher education, where AI-generated content detection tools are used despite their often being unreliable and prone to bias. As a result, many institutions discourage the faculty and staff from using those detectors [58, 81]. The rise of generative AI has also heightened the risks of fraud, scams, and impersonation crimes through techniques like deepfaking [16, 45]. These techniques make it easier for malicious actors to deceive individuals and organizations, underscoring the need for vigilance regarding the source of the content we consume.

While some organizations, including academic institutions [81] and companies [70] have taken steps to transparently disclose the use of generative AI, such responsible AI practices are not yet widespread. In everyday scenarios, such as customer service, the role of generative AI often goes unacknowledged. It highlights the importance of users remaining cautious and critical of the genuineness of content encountered in the wild. This also undermines the importance of critical information and media literacy highlighted in G6.

• G10: Learners need to understand how generative models "know" (and "don't know") concepts compared to human knowing. (Supporting References: [21, 66, 73, 79])

Generative AI models are capable of "knowing" concepts and information, but not necessarily in the same way and depth that humans do. LLMs' being able to predict the word sequence does not necessarily mean they can engage in thinking and reasoning [66, 118]. Knowledge can be grounded in a cognitive system in referential, sensorimotor, relational, communicative, and epistemic ways [73]. But generative AI models—like LLMs—are typically trained with a humongous dataset on the internet, thus only grounded in linguistic and symbolic ways. The phenomenon has been captured in the famous "Chinese room argument" [20], which presents that they "do not have the ability to process and understand the meaning in the way humans do" [21]. Generative AI for images operates based on visual data formats, but none of the existing generative AI to date has sensorimotor grounding in tactiles and smells (e.g., LLMs can provide step-by-step instructions on how to play violin, but they do not understand the true meaning of "press the string against the fingerboard"). Understanding how a generative AI model "knows" things can be a critical skill in influencing how one chooses, interacts with, and critically consumes these technologies.

• G11: Learners need to know there are hidden human labor, environmental, and broader profound social impacts of generative AI. (Supporting References: [9, 23, 26, 46, 50, 56, 76, 110, 117])

Generative AI models demand an immense amount of human labor behind the scenes, far beyond just computational power. A vast support system is required to maintain and update these models, highlighting that their operation is not magical but deeply reliant on human effort. As Crawford's work *Atlas of AI* reveals, this hidden labor extends to technologies like Amazon Echo, alongside significant environmental costs [23]. Beyond these concerns, researchers have raised alarms about the potential for AI to reduce aesthetic diversity, as its inherent biases shape what it deems Manuscript submitted to ACM

"beautiful" [67]. Moreover, the broader societal implications of generative AI are profound. It can undermine critical thinking [53, 56], infringe copyright [46], homogenize collective creativity [5, 30], hold fake promise of creativity [17], and exacerbate educational inequalities [50]. Generative AI is also reshaping the skills needed in the workplace [35, 69]. These issues underscore the importance of understanding the wider social impacts of AI. Learners should be aware of these consequences as they engage with and evaluate the role of generative AI in society.

• G12: The capabilities of generative AI evolve rapidly. Learners need to understand its up-to-date capabilities and limitations. (Supporting References: [42, 68, 78, 86])

The rapid evolution of generative AI means capabilities once deemed impossible can shift dramatically in short timeframes—a phenomenon reminiscent of the pace of technological change in the 1990s. While historical parallels offer context, today's advancements are uniquely driven by frameworks like Moore's Law, which continues to underpin exponential growth in computing power and algorithmic efficiency [98]. We have also witnessed the inference cost performing at GPT-3.5's level dropped 280 times within two years [68]. Moreover, studies demonstrate measurable progress: GPT-4 significantly outperformed GPT-3.5 in financial literacy tests [78]. Late 2024 marked a turning point as models like ChatGPT and Bard gained web-search functionality [42, 86], enabling them to have up-to-date information. This is a leap not yet universal across AI systems. Looking ahead, OpenAI's 2025 announcement of its artificial general intelligence (AGI) vision underscores the accelerating and unpredictable trajectory of generative AI [64]. To navigate the uncertainty, learners should update their knowledge of AI's capabilities and limitations, as they are often temporary, not absolute.

8 DISCUSSION

A review of AI literacy constructs over the past five years [3] reveals that AI literacy guidelines are continually evolving, as are the learning interventions designed by educators. With the unprecedented rise of generative AI, however, it exhibits characteristics that differ significantly from those depicted in existing AI literacy frameworks. This discrepancy highlights the need for updated guidelines that address the unique learning challenges of generative AI. As we search for current generative AI policies in university guidelines [59, 81, 82, 112], and government documents [21, 110], we found these efforts do not fully cover the scope of broader learning objectives and considerations. This gap underscores the need for a more comprehensive framework to guide learners in developing generative AI literacy. As we track the literature we reviewed forming these 12 guidelines, we also witnessed a disproportionate coverage of each guideline item in reviewed literature, with guideline items G2: safe and effective prompting, G6: misinformation and disinformation, G8: innate bias with generation results, and G11: broader social impact, covered in more literature than other guideline items. It further underscores the need for a holistic and comprehensive framework to define generative AI literacy. In the guideline, we intentionally omitted understanding the working mechanisms of generative AI, as such technical information is unlikely to enhance everyday learners' ability to interact with generative AI.

Our work aims to construct a holistic guideline that serves as a starting point for fostering future discussions and learning interventions on generative AI literacy. Unlike the more established AI literacy competencies [63, 77], generative AI is a relatively new and rapidly evolving field [38], making it challenging to create an all-encompassing framework. In the meantime, we need to incorporate timely grey literature to complement the lengthy academic peerreview process. This approach allowed us to capture the rapidly evolving characteristics of generative AI. Nevertheless, we hope our guidelines will provide a more solid foundation for addressing the learning objectives of generative AI users. While there have been notable efforts in defining generative AI literacy [6], developing generative AI literacy

assessment tests [49], and designing workshops [102], there is still a need for more comprehensive and robust learning materials. Future learning interventions could incorporate design considerations [63] to help learners interact with generative AI effectively, responsibly, and critically. Given the fast-paced evolution of generative AI, learning interventions need to be updated with the latest challenges that learners face.

REFERENCES

- Kabir Ahuja, Harshita Diddee, Rishav Hada, Millicent Ochieng, Krithika Ramesh, Prachi Jain, Akshay Nambi, Tanuja Ganu, Sameer Segal, Maxamed Axmed, et al. 2023. Mega: Multilingual evaluation of generative ai. arXiv preprint arXiv:2303.12528 (2023).
- [2] Hussam Alkaissi and Samy I McFarlane. 2023. Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus 15, 2 (2023).
- [3] Omaima Almatrafi, Aditya Johri, and Hyuna Lee. 2024. A Systematic Review of AI Literacy Conceptualization, Constructs, and Implementation and Assessment Efforts (2019-2023). Computers and Education Open (2024), 100173.
- [4] American Psychological Association. 2023. Misinformation and Disinformation. https://www.apa.org/topics/journalism-facts/misinformation-disinformation. https://www.apa.org/topics/journalism-facts/misinformation-disinformation.
- [5] Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization effects of large language models on human creative ideation. In Proceedings of the 16th conference on creativity & cognition. 413–425.
- [6] Ravinithesh Annapureddy, Alessandro Fornaroli, and Daniel Gatica-Perez. 2024. Generative AI literacy: Twelve defining competencies. Digital Government: Research and Practice (2024).
- [7] Hilary Arksey and Lisa O'Malley. 2005. Scoping studies: towards a methodological framework. International journal of social research methodology 8. 1 (2005), 19–32.
- [8] Imane El Atillah. 2023. Ai chatbot blamed for "encouraging" young father to take his own life. https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-.
- [9] Alex Barrett and Austin Pack. 2023. Not quite eye to AI: student and teacher perspectives on the use of generative artificial intelligence in the writing process. International Journal of Educational Technology in Higher Education 20, 1 (2023), 59.
- [10] BBC News. 2021. Alexa tells 10-year-old girl to touch live plug with penny. https://www.bbc.com/news/technology-59810383 Accessed: 2025-01-25
- [11] Yasmine Belghith, Atefeh Mahdavi Goloujeh, Brian Magerko, Duri Long, Tom Mcklin, and Jessica Roberts. 2024. Testing, Socializing, Exploring: Characterizing Middle Schoolers' Approaches to and Conceptions of ChatGPT. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–17.
- [12] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. 610–623.
- [13] Aras Bozkurt. 2024. Why generative AI literacy, why now and why it matters in the educational landscape?: Kings, queens and GenAI dragons. 283–290 pages.
- [14] Anna Cristina Brisola and Andréa Doyle. 2019. Critical information literacy as a path to resist "fake news": Understanding disinformation as the root problem. Open Information Science 3. 1 (2019), 274–286.
- [15] Jonas Bundschuh, M Greta Ruppert, and Yvonne Späck-Leigsnering. 2023. Pyrit: A finite element based field simulation software written in Python. COMPEL-The international journal for computation and mathematics in electrical and electronic engineering 42, 5 (2023), 1007–1020.
- [16] Anupama Chadha, Vaibhav Kumar, Sonu Kashyap, and Mayank Gupta. 2021. Deepfake: an overview. In Proceedings of second international conference on computing, communications, and cyber-security: IC4S 2020. Springer, 557–566.
- [17] Tuhin Chakrabarty, Philippe Laban, Divyansh Agarwal, Smaranda Muresan, and Chien-Sheng Wu. 2024. Art or artifice? large language models and the false promise of creativity. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–34.
- [18] CNN. 2024. 'There are no guardrails.' This mom believes an AI chatbot is responsible for her son's suicide. https://www.cnn.com/2024/10/30/tech/teen-suicide-character-ai-lawsuit/index.html Accessed: 2025-01-25.
- [19] Mark Coeckelbergh. 2020. AI ethics. Mit Press.
- [20] David Cole. 2004. The Chinese room argument. (2004).
- [21] Counsel for European Union. 2023. ChatGPT in the Public Sector: Overhyped or Overlooked? https://www.consilium.europa.eu/media/63818/art-paper-chatgpt-in-the-public-sector-overhyped-or-overlooked-24-april-2023_ext.pdf. Accessed: 2025-01-10.
- [22] Andrew Cox. 2024. Algorithmic literacy, AI literacy and responsible generative AI literacy. Journal of Web Librarianship 18, 3 (2024), 93-110.
- [23] Kate Crawford. 2021. The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.
- [24] Andrea Cuadra, Maria Wang, Lynn Andrea Stein, Malte F Jung, Nicola Dell, Deborah Estrin, and James A Landay. 2024. The illusion of empathy? notes on displays of emotion in human-computer interaction. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–18.
- [25] Dallery.Gallery. 2025. The DALL-E 2 Prompt Book. https://dallery.gallery/the-dalle-2-prompt-book/ Accessed: Jan. 30, 2025.
- [26] Anh Dang and Hui Wang. 2024. Ethical use of generative AI for writing practices: Addressing linguistically diverse students in US Universities' AI statements. Journal of Second Language Writing 66 (2024), 101157.

[27] Hanne De Jaegher and Ezequiel Di Paolo. 2007. Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the cognitive sciences* 6 (2007), 485–507.

- [28] Matias del Campo and Neil Leach. 2022. Machine Hallucinations: Architecture and Artificial Intelligence. John Wiley & Sons.
- [29] Yiran Ding, Li Lyna Zhang, Chengruidong Zhang, Yuanyuan Xu, Ning Shang, Jiahang Xu, Fan Yang, and Mao Yang. 2024. LongRoPE: extending LLM context window beyond 2 million tokens. In Proceedings of the 41st International Conference on Machine Learning (Vienna, Austria) (ICML'24). IMLR.org, Article 442, 14 pages.
- [30] Anil R Doshi and Oliver P Hauser. 2024. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances* 10, 28 (2024), eadn5290.
- [31] Carmen Drahl. 2023. AI was asked to create images of Black African docs treating white kids. How'd it go? {https://www.npr.org/sections/goatsandsoda/2023/10/06/1201840678/ai-was-asked-to-create-images-of-black-african-docs-treating-white-kids-howd-it-}, urldate = 2023-10-06.
- [32] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. [n. d.]. Expanding Explainability: Towards Social Transparency in AI systems. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2021) (CHI '21). Association for Computing Machinery, 1–19. https://doi.org/10.1145/3411764.3445188
- [33] Upol Ehsan, Pradyumna Tambwekar, Larry Chan, Brent Harrison, and Mark O Riedl. 2019. Automated rationale generation: a technique for explainable AI and its effects on human perceptions. In *Proceedings of the 24th international conference on intelligent user interfaces*. 263–274.
- [34] Clemens Eisenmann, Jakub Mlynář, Jason Turowetz, and Anne W Rawls. 2024. "Machine Down": making sense of human-computer interaction—Garfinkel's research on ELIZA and LYRIC from 1967 to 1969 and its contemporary relevance. AI & SOCIETY 39, 6 (2024), 2715–2733.
- [35] Ziv Epstein, Aaron Hertzmann, Investigators of Human Creativity, Memo Akten, Hany Farid, Jessica Fjeld, Morgan R Frank, Matthew Groh, Laura Herman, Neil Leach, et al. 2023. Art and the science of generative AI. Science 380, 6650 (2023), 1110–1111.
- [36] Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch, and Patrick Zschech. 2024. Generative ai. Business & Information Systems Engineering 66, 1 (2024), 111–126.
- [37] Carnegie Endowment for International Peace. 2025. The World According to Generative Artificial Intelligence. https://carnegieendowment.org/research/2025/01/the-world-according-to-generative-artificial-intelligence?lang=en. Accessed: 2025-01-10.
- [38] David Foster. 2023. Generative deep learning. "O'Reilly Media, Inc.".
- [39] The Fulcrum. 2023. AI is Fabricating Misinformation: A Call for AI Literacy in the Classroom. https://thefulcrum.us/ai-literacy-misinformation
- [40] Kate Glazko, Yusuf Mohammed, Ben Kosa, Venkatesh Potluri, and Jennifer Mankoff. 2024. Identifying and Improving Disability Bias in GPT-Based Resume Screening. In The 2024 ACM Conference on Fairness, Accountability, and Transparency. 687–700.
- [41] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. Advances in neural information processing systems 27 (2014).
- $[42]\ Google.\ 2024.\ Bard\ and\ Google\ AI\ Search\ Updates.\ https://blog.google/technology/ai/bard-google-ai-search-updates/.\ Accessed:\ 2025-04-26.$
- [43] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. 2019. XAI—Explainable artificial intelligence. Science robotics 4, 37 (2019), eaay7120.
- [44] Michael Haenlein and Andreas Kaplan. 2019. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. California management review 61, 4 (2019), 5–14.
- [45] Asri Gresmelian Eurike Hailtik and Wiwik Afifah. 2023. Criminal responsibility of artificial intelligence committing deepfake crimes in Indonesia. Asian Journal of Social and Humanities 2, 4 (2023), 776–795.
- [46] Melissa Heikkilä. 2022. This artist is dominating AI-generated art. And he's not happy about it. MIT Technology Review 125, 6 (2022), 9-10.
- [47] Hong Kong University of Science and Technology (HKUST) Library. 2024. Generative AI in Education. https://libguides.hkust.edu.hk/ai-literacy Accessed: 2025-01-10.
- [48] Ali Reza Ibrahimzada. 2024. Program Decomposition and Translation with Static Analysis. In Proceedings of the 2024 IEEE/ACM 46th International Conference on Software Engineering: Companion Proceedings. 453–455.
- [49] Yueqiao Jin, Roberto Martinez-Maldonado, Dragan Gašević, and Lixiang Yan. 2024. GLAT: The Generative AI Literacy Assessment Test. arXiv preprint arXiv:2411.00283 (2024).
- [50] Yueqiao Jin, Lixiang Yan, Vanessa Echeverria, Dragan Gašević, and Roberto Martinez-Maldonado. 2025. Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. Computers and Education: Artificial Intelligence 8 (2025), 100348.
- [51] Aditya Johri, Ashish Hingle, and Johannes Schleiss. 2024. Misconceptions, Pragmatism, and Value Tensions: Evaluating Students' Understanding and Perception of Generative AI for Education. arXiv preprint arXiv:2410.22289 (2024).
- [52] Jerry Kaplan. 2024. Generative Artificial Intelligence: What Everyone Needs to Know. Oxford University Press.
- [53] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences 103 (2023), 102274.
- [54] Douglas Kellner and Jeff Share. 2005. Toward critical media literacy: Core concepts, debates, organizations, and policy. Discourse: studies in the cultural politics of education 26, 3 (2005), 369–386.
- [55] Celeste Kidd and Abeba Birhane. 2023. How AI can distort human beliefs. Science 380, 6651 (2023), 1222-1223.

[56] Hao-Ping Hank Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. (2025).

- [57] Yitong Li, Martin Min, Dinghan Shen, David Carlson, and Lawrence Carin. 2018. Video generation from text. In *Proceedings of the AAAI conference* on artificial intelligence. Vol. 32.
- [58] Texas Tech University Libraries. 2023. AI Literacy and Pedagogy in the Age of Generative AI. https://guides.library.ttu.edu/ai-literacy
- [59] Harvard Library. 2024. Quick Guide to Harvard Library. https://guides.library.harvard.edu/c.php?g=1330621&p=10034696 Accessed: 2025-01-10.
- [60] Vivian Liu and Lydia B Chilton. 2022. Design guidelines for prompt engineering text-to-image generative models. In Proceedings of the 2022 CHI conference on human factors in computing systems. 1–23.
- [61] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal image generation for news illustration. In Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. 1–17.
- [62] Leo S Lo. 2023. The CLEAR path: A framework for enhancing information literacy through prompt engineering. The Journal of Academic Librarianship 49, 4 (2023), 102720.
- [63] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI conference on human factors in computing systems. 1–16.
- [64] Time Magazine. 2025. How OpenAI's Sam Altman Is Thinking About AGI and Superintelligence in 2025. https://time.com/7205596/sam-altman-superintelligence-agi/. Time (2025). Accessed: 2025-02-19.
- [65] Atefeh Mahdavi Goloujeh, Anne Sullivan, and Brian Magerko. 2024. Is It AI or Is It Me? Understanding Users' Prompt Journey with Text-to-Image Generative AI Tools. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–13.
- [66] Kyle Mahowald, Anna A Ivanova, Idan A Blank, Nancy Kanwisher, Joshua B Tenenbaum, and Evelina Fedorenko. 2024. Dissociating language and thought in large language models. Trends in Cognitive Sciences (2024).
- [67] Lev Manovich. 2018. AI aesthetics. Strelka press Moscow.
- [68] Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, Tobi Walsh, Armin Hamrah, Lapo Santarlasci, Julia Betts Lotufo, Alexandra Rome, Andrew Shi, and Sukrut Oak. 2025. The AI Index 2025 Annual Report. https://hai.stanford.edu/ai-index/2025-ai-index-report. Accessed: 2025-01-10.
- [69] McKinsey & Company. 2025. The state of AI: How organizations are rewiring to capture value. Web https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai Accessed: 2025-04-25.
- [70] Microsoft. 2024. Responsible AI Transparency Report. https://www.microsoft.com/en-us/corporate-responsibility/responsible-ai-transparency-report Accessed: 2025-01-25.
- [71] Bonan Min, Hayley Ross, Elior Sulem, Amir Pouran Ben Veyseh, Thien Huu Nguyen, Oscar Sainz, Eneko Agirre, Ilana Heintz, and Dan Roth. 2023.
 Recent advances in natural language processing via large pre-trained language models: A survey. Comput. Surveys 56, 2 (2023), 1–40.
- [72] Modern Language Association and Conference on College Composition and Communication. 2024. Student Guide to AI Literacy. https://aiandwriting.hcommons.org/student-guide-to-ai-literacy/. Accessed: 2025-01-10.
- [73] Dimitri Coelho Mollo and Raphaël Millière. 2023. The vector grounding problem. arXiv preprint arXiv:2304.01481 (2023).
- [74] Michael Muller, Anna Kantosalo, Mary Lou Maher, Charles Patrick Martin, and Greg Walsh. 2024. GenAICHI 2024: Generative AI and HCI at CHI 2024. In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. 1–7.
- [75] LARRY NEUMEISTER. 2023. Lawyers blame ChatGPT for tricking them into citing bogus case law. https://apnews.com/article/artificial-intelligence-chatgpt-courts-e15023d7e6fdf4f099aa122437dbb59b
- [76] Andrew Ng. 2023. Generative AI for Everyone. https://www.coursera.org/learn/generative-ai-for-everyone#modules
- [77] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. 2021. Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence 2 (2021), 100041.
- [78] Paweł Niszczota and Sami Abbas. 2023. GPT has become financially literate: Insights from financial literacy tests of GPT and a preliminary test of how people use it as a source of advice. Finance Research Letters 58 (2023), 104333.
- [79] David A. Noever and Samantha E. Miller Noever. 2023. Large Language Models and the Reverse Turing Test. Neural Computation 35, 3 (2023), 309–342. https://doi.org/10.1162/neco_a_01563
- [80] NPR. 2024. Lawsuit: A chatbot hinted a kid should kill his parents over screen time limits https://www.npr.org/2024/12/10/nx-s1-5222574/kids-character-ai-lawsuit
- [81] University of British Columbia. 2024. DRAFT Principles and Guidelines for Generative Artificial Intelligence (GenAI) in Teaching and Learning. https://it.ubc.ca/sites/it.ubc.ca/files/Guidelines-GenAI-TL_V2.pdf
- $[82] \ \ The University of Hong Kong. 2023. \ Al Literacy for Education (Student). \ \ https://learning.hku.hk/catalog/course/ai-literacy-for-education-student/literacy-for-education-s$
- [83] University of Rochester. 2023. Generative AI Use in Education. https://www.rochester.edu/provost/gen-ai-education/
- [84] Wellington Institute of Technology. 2023. AI | Artificial Intelligence: AI Introduction. https://whitireia.libguides.com/AI Accessed: 2025-01-10.
- [85] Jesutofunmi A Omiye, Jenna C Lester, Simon Spichak, Veronica Rotemberg, and Roxana Daneshjou. 2023. Large language models propagate race-based medicine. NPJ Digital Medicine 6, 1 (2023), 195.
- [86] OpenAI. 2024. Introducing ChatGPT Search. https://openai.com/index/introducing-chatgpt-search/. Accessed: Accessed: 2025-04-26.

[87] Anaelia Ovalle. 2023. Decoding The Digital Fuk\'u: Deciphering Colonial Legacies to Critically Assess ChatGPT in Dominican Education. arXiv preprint arXiv:2310.17533 (2023).

- [88] Xianghan O'Dea, Davy Tsz Kit Ng, Mike O'Dea, and Viacheslav Shkuratskyy. 2024. Factors affecting university students' generative AI literacy: Evidence and evaluation in the UK and Hong Kong contexts. Policy Futures in Education (2024), 14782103241287401.
- [89] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*. 1–22.
- [90] Peter S Park, Simon Goldstein, Aidan O'Gara, Michael Chen, and Dan Hendrycks. 2024. AI deception: A survey of examples, risks, and potential solutions. Patterns 5, 5 (2024).
- [91] Charlie Parker, Sam Scott, and Alistair Geddes. 2019. Snowball sampling. SAGE research methods foundations (2019).
- [92] David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? Behavioral and brain sciences 1, 4 (1978), 515-526.
- [93] ProPublica. 2016. Machine Bias. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
- [94] Partha Pratim Ray. 2023. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems (2023).
- [95] Ali Razavi, Aaron Van den Oord, and Oriol Vinyals. 2019. Generating diverse high-fidelity images with vq-vae-2. Advances in neural information processing systems 32 (2019).
- [96] Salt Lake Community College. 2023. AI, ChatGPT, and the Library. https://libguides.slcc.edu/ChatGPT/InformationLiteracy Accessed: 2025-04-26.
- [97] Karahan Sarıtaş, Kıvanç Tezören, and Yavuz Durmazkeser. 2025. A Systematic Review on the Evaluation of Large Language Models in Theory of Mind Tasks. arXiv preprint arXiv:2502.08796 (2025).
- [98] Robert R Schaller. 1997. Moore's law: past, present and future. IEEE spectrum 34, 6 (1997), 52-59.
- [99] Johannes Schneider. 2024. Explainable generative ai (genxai): A survey, conceptualization, and research agenda. Artificial Intelligence Review 57, 11 (2024), 289.
- [100] Neil J Smelser, Paul B Baltes, et al. 2001. International encyclopedia of the social & behavioral sciences. Vol. 11. Elsevier Amsterdam.
- [101] Adam Sobieszek and Tadeusz Price. 2022. Playing games with AIs: the limits of GPT-3 and similar large language models. *Minds and Machines* 32, 2 (2022), 341–364.
- [102] Miriam Sullivan, Michael McAuley, Danielle Degiorgio, and Paul McLaughlan. 2024. Improving students' generative AI literacy: A single workshop can improve confidence and understanding. Journal of Applied Learning and Teaching 7, 2 (2024).
- [103] Jiao Sun, Q Vera Liao, Michael Muller, Mayank Agarwal, Stephanie Houde, Kartik Talamadupula, and Justin D Weisz. 2022. Investigating explainability of generative AI for code through scenario-based design. In Proceedings of the 27th International Conference on Intelligent User Interfaces.
 212–228
- [104] Lawrence Switzky. 2020. ELIZA effects: Pygmalion and the early development of artificial intelligence. Shaw 40, 1 (2020), 50–68.
- [105] Stuart A. Thompson Tiffany Hsu. 2023. Disinformation Researchers Raise Alarms About A.I. Chatbots. https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html Accessed: 2025-01-10.
- [106] The Brussels Times. 2023. Belgian man dies by suicide following exchanges with chatbot. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with chatbot. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with chatbot. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with chatbot. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatbot. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatbot. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatbot. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatbot. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatbot. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatbot. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatbot. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatbot. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatbot. https://www.brusselstimes.com/asselstimes-with-chatbot. https://www.brusselstimes-with-chatbot. https://www.brusselstimes-wi

[111] Northwestern University. 2024. The Hallucination Problem: A Feature, Not a Bug. https://casmi.northwestern.edu/news/articles/2024/the-hallucination-problem-a-feature-not-a-bug.ht

- [107] The New York Times. 2024. See How Easily A.I. Chatbots Can Be Taught to Spew Disinformation. https://www.nytimes.com/interactive/2024/05/19/technology/biased-ai-chatbots.html Accessed: 2025-02-10.
- [108] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. Envisioning AI for K-12: What should every child know about AI?. In Proceedings of the AAAI conference on artificial intelligence, Vol. 33. 9795–9799.
- $[109]\,$ Sherry Turkle. 2024. Who Do We Become When We Talk to Machines? (2024).
- [110] UNESCO. 2023. ChatGPT and Artificial Intelligence in Higher Education: Quick Start Guide. https://unesdoc.unesco.org/ark:/48223/pf0000385146 Accessed: 2025-01-10.
- Accessed: 2025-01-10.
 [112] The George Washington University. 2023. Guidelines for Using Generative Artificial Intelligence at the George Washington University.
- 112] The George Washington University. 2023. Guidelines for Using Generative Artificial Intelligence at the George Washington University. https://provost.gwu.edu/sites/g/files/zaxdzs5926/files/2023-04/generative-artificial-intelligence-guidelines-april-2023.pdf
- [113] Irvine Libraries University of California. 2023. AI in Research Guide. https://guides.lib.uci.edu/research_ai/home
- [114] A Vaswani. 2017. Attention is all you need. Advances in Neural Information Processing Systems (2017).
- [115] William H Walters and Esther Isabelle Wilder. 2023. Fabrication and errors in the bibliographic citations generated by ChatGPT. Scientific Reports 13, 1 (2023), 14045.
- [116] Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. 2021. Towards mutual theory of mind in human-ai interaction: How language reflects what students perceive about a virtual teaching assistant. In Proceedings of the 2021 CHI conference on human factors in computing systems. 1–14.
- [117] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, et al. 2022. Taxonomy of risks posed by language models. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency. 214–229.

[118] Peter West, Ximing Lu, Nouha Dziri, Faeze Brahman, Linjie Li, Jena D Hwang, Liwei Jiang, Jillian Fisher, Abhilasha Ravichander, Khyathi Chandu, et al. 2023. THE GENERATIVE AI PARADOX: "What It Can Create, It May Not Understand". In *The Twelfth International Conference on Learning Representations*.

- [119] Sam Wineburg and Sarah McGrew. 2017. Lateral reading: Reading less and learning more when evaluating digital information. (2017).
- [120] Chloe Xiang. 2023. 'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says. https://www.vice.com/en/article/wxjxgx/openai-and-microsoft-sued-for-dollar3-billion-over-alleged-chatgpt-privacy-violations. Accessed: 2025-01-10.
- [121] Katherine Xu, Lingzhi Zhang, and Jianbo Shi. 2024. Good Seed Makes a Good Crop: Discovering Secret Seeds in Text-to-Image Diffusion Models. arXiv preprint arXiv:2405.14828 (2024).
- [122] Han Yan, Haijun Zhang, Linlin Liu, Dongliang Zhou, Xiaofei Xu, Zhao Zhang, and Shuicheng Yan. 2022. Toward intelligent design: An ai-based fashion designer using generative adversarial networks aided by sketch and rendering generators. IEEE Transactions on Multimedia 25 (2022), 2323–2338.
- [123] Litao Yan, Elena L Glassman, and Tianyi Zhang. 2021. Visualizing examples of deep neural networks at scale. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [124] JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [125] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic lies: Understanding ai-generated misinformation and evaluating algorithmic and human solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.
- [126] Pei Zhou, Aman Madaan, Srividya Pranavi Potharaju, Aditya Gupta, Kevin R McKee, Ari Holtzman, Jay Pujara, Xiang Ren, Swaroop Mishra, Aida Nematzadeh, et al. 2023. How FaR Are Large Language Models From Agents with Theory-of-Mind? arXiv preprint arXiv:2310.03051 (2023).