Preprints.org

ChatGPT's Expanding Horizons and Transformative Impact Across Domains: A Critical Review of Capabilities, Challenges, and Future Directions

Taiwo Raphael Feyijimi [*] , John Ogbeleakhu Aliu , Ayodeji Emmanuel Oke , Douglas Aghimien

Review

ChatGPT's Expanding Horizons and Transformative Impact Across Domains: A Critical Review of Capabilities, Challenges, and Future Directions

Taiwo Raphael Feyijimi [1,*], John Ogbeleakhu Aliu [2], Ayodeji Emmanuel Oke [3] and Douglas Omoregie Aghimien [4]

[1] University of Georgia

[2] Engineering Education Transformations Institute, College of Engineering, University of Georgia, Athens, Georgia, USA

[3] Department of Quantity Surveying, Federal University of Technology Akure, Akure, Nigeria

[4] University of Johannesburg

* Correspondence: taiwo.feyijimi@uga.edu

Abstract

The advent of ChatGPT marks a significant inflection point in AI, characterized by widespread adoption and diverse perceptions ranging from "shock" in academia to "awe" in industry. Built upon the Transformer architecture, its evolution has enabled human-like text generation and complex task engagement-a capability now rapidly advancing toward autonomous Agentic AI systems that can plan, reason, and execute multi-step tasks. This paper provides a comprehensive, critical review of this expanding horizon and its impact across natural language understanding (NLU), content generation, knowledge discovery, education, and engineering. It aims to synthesize current capabilities, limitations, and ethical considerations, and to propose novel methodological and research directions for responsible global integration. This review synthesizes current literature and research findings, critically assessing performance, analyzing error rates, examining innovative techniques like Retrieval-Augmented Generation (RAG), and exploring the ethical challenges inherent in its cross-domain deployment. Generative AI demonstrates profound capabilities, pushing NLU boundaries with multimodality and enabling diverse applications. However, limitations persist in factual accuracy, bias, and explainability ("Black Box Conundrum"). Key tensions identified include the "Specialization vs. Generalization" tension in NLU, the "Quality-Scalability-Ethics Trilemma" in content generation, the "Pedagogical Adaptation Imperative" in education, and the emergence of "Human-LLM Cognitive Symbiosis" in engineering – a concept now evolving toward models of Human-Agent Orchestration as AI gains autonomy. The findings necessitate proactive adaptation across sectors, including redesigning educational pedagogy, developing AI collaboration skills, implementing robust quality control, and prioritizing ethical design and transparency. This review offers a unique cross-domain synthesis, introducing conceptual frameworks – including the Specialization vs. Generalization Tension, Quality-Scalability-Ethics Trilemma, Black Box Conundrum, Pedagogical Adaptation Imperative, Human-Agent Symbiosis, and the overarching "Ethical-Technical Co-evolution Imperative" – to illuminate complex challenges. It charts the intensification of these challenges within the emerging agentic paradigm, identifies specific research gaps, and proposes a forward-looking agenda for advancing both methodological and theoretical understanding to guide the development of responsible AI.

Keywords: ChatGPT; large language models (LLMs); agentic AI; natural language understanding (NLU); content generation; education; engineering; ethical considerations

1. Introduction: The ChatGPT Inflection Point in AI and its Applications

The advent of Chat Generative Pre-trained Transformer (ChatGPT) marks a significant inflection point in the trajectory of artificial intelligence and its pervasive influence across myriad sectors. Its rapid ascent since its public release has been characterized by widespread adoption and a mixture of acclaim and apprehension, fundamentally altering landscapes in education, healthcare, customer service, software development, and beyond (Akhtarshenas et al., 2025). Hailed as a "game changer" (Infosys Limited, 2023) and a "disruptive" technology (Murray et al., 2025), ChatGPT, built upon the Transformer architecture and the continued evolution of Large Language Models (LLMs) (Akhtarshenas et al., 2025), has demonstrated an uncanny ability to generate human-like text and engage in complex tasks. This inflection point is now rapidly evolving, as the powerful reasoning and tool-use capabilities of these models are giving rise to Agentic AI – autonomous systems that can plan, reason, and execute complex, multi-step tasks to achieve goals, representing a paradigm shift from passive tools to active participants in digital and physical environments (Xi et al., 2023).

The initial reception to ChatGPT, particularly within academic circles, has often been described as more "shock" than "awe" (Murray et al., 2025). This reaction stems largely from its immediate and palpable disruption to established educational practices, especially concerning student assessment and academic integrity (Dempere et al., 2023). The swiftness with which students adopted the tool for tasks such as coursework generation caught many institutions off-guard. This dynamic is set to intensify with the rise of AI agents, which can automate not just the writing of an assignment but the entire research and analysis process, posing even deeper challenges to traditional assessment. In contrast, industries have often focused more on the "awe" aspect, emphasizing productivity gains and novel content creation capabilities (Al Naqbi et al., 2024) – a perception now amplified by the potential of agents to automate entire workflows, not just discrete tasks (Team, 2024). This divergence in perception is context-dependent; those whose core practices are directly challenged by the technology are more likely to experience initial shock, while those who see immediate utility may be more readily impressed. This paper posits that the initial caution within academia, while potentially slowing immediate, uncritical adoption, may ultimately foster the development of more robust ethical frameworks, innovative pedagogical responses, and a deeper understanding of human-AI collaboration that will be essential for governing agentic systems.

This paper critically examines ChatGPT's multifaceted applications in natural language understanding, content generation, knowledge discovery, engineering, and education. It aims to advance theoretical understanding and propose methodological innovations, while addressing the attendant ethical considerations, research frontiers, and the implications of the emerging agentic paradigm to foster responsible and impactful global integration. The subsequent sections will dissect these specific application domains, exploring their theoretical underpinnings, recent methodological advancements, and the critical challenges that must be navigated to realize the full potential of both generative and agentic AI responsibly.

2. Advancements in Natural Language Understanding with ChatGPT: Capabilities, Innovations, and Critical Frontiers

Natural Language Understanding (NLU) lies at the heart of ChatGPT's capabilities, enabling its diverse applications. The continuous evolution of its underlying architecture and training methodologies has pushed the boundaries of what machines can comprehend and how they can interact using human language, setting the stage for more autonomous and goal-oriented AI systems.

2.1. Core NLU Architecture and Functionalities

ChatGPT's NLU prowess is built upon the Transformer architecture, a deep learning model that has revolutionized sequence-to-sequence tasks (Vaswani et al., 2017; Akhtarshenas et al., 2025). A key component of this architecture is the self-attention mechanism, which allows the model to weigh the importance of different words in an input sequence, thereby capturing long-range dependencies and contextual nuances (Infosys Limited, 2023). This enables ChatGPT to engage in complex query responses, maintain contextual understanding over extended dialogues, and generate remarkably human-like text (Akhtarshenas et al., 2025). The evolution from GPT-3.5 through GPT-4 and its variants like GPT-4o, and the newer 'o1' series, has seen significant enhancements (Hariri, 2023). Newer models boast improved contextual awareness, capable of maintaining coherence over longer conversations. Training on larger and more diverse datasets has enhanced their language understanding, leading to greater accuracy and fluency (Hariri, 2023). Advanced tokenization techniques have improved efficiency and processing speed (Hariri, 2023). Perhaps one of the most significant LLMs advancements is the integration of multimodality, allowing models like GPT-4o to process and generate content involving text, images, and audio (Hariri, 2023). These foundational NLU capabilities are crucial building blocks for the next frontier: autonomous Agentic AI systems that can act on this understanding to perform tasks.

2.2. Innovative NLU Techniques and their Impact

Beyond core architectural improvements, specific techniques are being developed to enhance ChatGPT's NLU capabilities, address its limitations, and empower it to move from passive generation to active problem-solving.

Retrieval Augmented Generation (RAG): RAG has emerged as a powerful method to mitigate knowledge gaps by dynamically incorporating external information into the generation process (Lewis et al., 2020; Yu et al., 2025). This involves vectorizing relevant documents and providing them as context, allowing the model to access up-to-date or domain-specific information not present in its training data. A notable application is in specialized engineering domains like Building Information Modeling (BIM), where RAG enabled ChatGPT-4 to significantly improve its understanding of localized Korean BIM guidelines, boosting performance by 25.7% (Yu et al., 2025). By grounding responses in verifiable external documents, RAG increases trustworthiness and is a foundational technique for reliable AI systems (Gao et al., 2024).

The Rise of Agentic AI: A paradigm-shifting innovation is the development of agentic frameworks. These frameworks empower LLMs to act as autonomous agents that can reason, create complex plans, decompose tasks, and utilize external tools (like code interpreters, APIs, or web browsers) to achieve multi-step goals (Xi et al., 2023). Unlike traditional NLU, which focuses on comprehension and response, agents use NLU to interact with an environment and execute actions, representing a move from passive text generation to proactive, goal-directed behavior (Park et al., 2023; Shinn et al., 2023).

Advancements from NLP Conferences (EMNLP/ACL 2024-2025): Recent research highlights several innovative techniques (Empirical Methods in Natural Language Processing, 2024):

- **LLMs for Data Annotation and Cleansing:** LLMs are increasingly used to automate or assist in data annotation, a traditionally labor-intensive task. For example, the Multi-News+ dataset was enhanced by using LLMs with chain-of-thought and majority voting to cleanse and classify documents, improving dataset quality for multi-document summarization tasks.
- **Factual Inconsistency Detection:** Given LLMs' propensity for hallucination, techniques to detect factual inconsistencies are crucial. Methods like FIZZ, which employ fine-grained atomic fact decomposition and alignment with source documents, offer more interpretable ways to identify inaccuracies in abstractive summaries.

- **Multimodal NLU Enhancement:** Research is exploring the integration of other modalities, such as acoustic speech information, into LLM frameworks for tasks like depression detection, indicating a move towards more holistic NLU that mirrors human multimodal.
- **"Evil Twin" Prompts:** The discovery of "evil twin" prompts, obfuscated and uninterpretable inputs that can elicit desired outputs and transfer between models, opens new avenues for understanding LLM vulnerabilities and their internal representations, posing both security risks and research opportunities.

2.3. Critical Assessment: Benchmarks, Limitations, and Human Comparison

Despite their advancements, LLMs like ChatGPT face several critical limitations:

Benchmark Performance: While versatile, general-purpose models like ChatGPT-3.5 Turbo may not always outperform specialized, fine-tuned models (e.g., BERT variants) on specific NLU benchmarks like GLUE, particularly in tasks such as paraphrase detection or semantic similarity (Akhtarshenas et al., 2025). Newer versions like GPT-4 show improvements, but the gap often remains for highly optimized task-specific models.

Inherent Limitations:

- *Misinformation and Hallucinations:* A persistent issue is the generation of plausible-sounding but incorrect or nonsensical information, often termed "hallucinations" (Akhtarshenas et al., 2025). This undermines reliability, especially in critical applications.
- *Bias:* LLMs inherit biases present in their vast training datasets, which can manifest as gender, racial, geographical, or ideological skews in their outputs (Akhtarshenas et al., 2025; Bender et al., 2021; OpenAI, 2023a). These biases can perpetuate harmful stereotypes and lead to unfair outcomes.
- *Transparency and Explainability:* The "black box" nature of LLMs makes it difficult to understand their decision-making processes or trace the origins of errors (Infosys Limited, 2023; Wolf et al., 2019; Liu et al., 2023; Mavrepis et al., 2024; Zhao et al., 2024). This lack of interpretability is a major hurdle for debugging, ensuring fairness, and building trust.
- *Contextual Understanding Limits:* While improved, LLMs can still struggle with deeply nuanced contextual understanding, complex linguistic structures (like center-embedding), rare words, sarcasm, or the subtleties of human emotion (Akhtarshenas et al., 2025; Bender et al., 2021; OpenAI, 2023a).

These risks are significantly amplified in agentic systems. An agent that hallucinates or acts on biased information can cause direct, real-world harm, making robust safety and alignment protocols paramount (Weidinger et al., 2024).

Comparison to Human NLU: Studies demonstrate that the human brain's language processing capabilities, particularly for complex syntactic structures and predictive reasoning, still surpass those of current LLMs, even when comparing against non-native human speakers (Shormani, 2024). ChatGPT, despite its sophistication, cannot be considered a complete theory of human language acquisition or processing due to fundamental differences in learning mechanisms and underlying "competence" (Lake & Baroni, 2023; Shormani, 2024).

2.4. Advancing Method and Theory in NLU through ChatGPT

The evolution of ChatGPT catalyzes new methodological and theoretical directions in NLU:

- Methodologically, a key frontier is the development of **agentic AI workflows.** These systems represent a profound shift, leveraging core NLU to interact with tools and environments to solve complex, multi-step problems (Park et al., 2023; Shinn et al., 2023).
- Techniques like RAG are foundational, providing agents with grounded, verifiable knowledge (Lewis et al., 2020; Hu & Lu, 2024; Wu et al., 2022; Yu, 2022).
- Probing methods like "evil twin" prompts (Empirical Methods in Natural Language Processing, 2024; Melamed et al., 2023; Mozes, 2024; Mozes et al., 2023; Oremus, 2023; Perez & Ribeiro, 2022; Perez et al., 2022; Xue et al., 2023) and the push towards multimodality (Hariri, 2023; OpenAI, 2024c) are creating more robust and versatile models to power these agents.

A core tension in this advancement is between **specialization and generalization.** While foundation models like ChatGPT exhibit broad capabilities (Akhtarshenas et al., 2025; OpenAI, 2023a), they are often outperformed on specific tasks by fine-tuned models (Akhtarshenas et al., 2025; Hodge Jr, 2023; Perlman, 2023; Surden, 2023). This suggests future progress lies in a sophisticated interplay between generalist models and specialized techniques. This is not merely about scaling models but designing smarter, more adaptable architectures, perfectly exemplified by agentic frameworks. In these systems, a generalist LLM acts as a central "reasoning engine" that intelligently selects and deploys specialized tools or knowledge sources (like RAG) as needed (Xi et al., 2023). Theoretically, this points to a need for new frameworks that model the trade-offs between generalization and specialization, drawing inspiration from cognitive science theories on how humans balance broad knowledge with deep, domain-specific expertise to achieve goals (Lake & Baroni, 2023).

Table 1 provides a comparative overview of different ChatGPT models and their NLU/content generation characteristics, based on available information.

Table 1. Comparative Overview of ChatGPT Models in NLU and Content Generation (Source: Authors).

Model Version	Key Architectural Features/Training Data Cutoff	Notable NLU Capabilities	Content Generation Strengths	Known Limitations	Key Benchmark Performance (Example)
ChatGPT-3.5 / 3.5-Turbo	Based on GPT-3.5, Text/Code (pre-2021/2023) (Infosys Limited, 2023; OpenAI, 2023a)	Basic text tasks, translation, conversational AI, faster responses (Infosys Limited, 2023; OpenAI, 2023a; Susskind & Susskind, 2022)	Dialogue, boilerplate tasks, initial drafts, summaries (Infosys Limited, 2023; OpenAI, 2023a; Susskind & Susskind, 2022)	Accuracy issues, bias, limited by training data cutoff, struggles with highly specialized tasks (Akhtarshenas et al., 2025; OpenAI, 2023a)	GLUE average score ~78.7% (comparable to BERT-base, lags RoBERTa-large) (Shormani, 2024). Passed Korea's BIM Expertise Exam with 65% average (Yu et al., 2025). Error rates in healthcare can be high (Ray, 2023).
ChatGPT-4	Based on GPT-4, Text/Code (pre-2023) (Sapkota et al., 2025; Achiam et al., 2023; OpenAI, 2023b)	Multimodal (text), high precision, improved reasoning, expanded context window (Hariri, 2023; Achiam et al., 2023; OpenAI, 2023b; OpenAI, 2024b)	More coherent, contextually relevant text, complex conversations, nuanced topics (Hariri, 2023; Achiam et al., 2023; OpenAI, 2023b; OpenAI, 2024b)	Still prone to hallucinations, bias; costlier; specific weaknesses in areas like local guidelines without RAG (Ray, 2023; Bender et al., 2021; Achiam et al., 2023; OpenAI,	Passed Korea's BIM Expertise Exam with 85% average (improved to 88.6% with RAG for specific categories) (Yu et al., 2025). Lower error rates in business/economics (~15-20%) compared to 3.5 (Ray, 2023).

				2023b; Nguyen et al., 2023)	
GPT-4o / GPT-4o mini	Text/Code (pre-2024) (Hariri, 2023; OpenAI, 2024c)	Multimodal (text/image/audio/video), improved contextual awareness, advanced tokenization, cost-efficiency (mini) (Hariri, 2023; OpenAI, 2024c)	Richer, more interactive responses, real-time collaboration support (Hariri, 2023; OpenAI, 2024c)	Newer models, long-term limitations still under study, but likely share core LLM challenges.	GPT-4o slightly better than 3.5-turbo and 4o-mini on research quality score estimates (correlation 0.67 with human scores using title/abstract) (Thelwall, 2024). GPT-4o mini outperforms GPT-3.5 Turbo on MMLU (82% vs 69.8%) (RedBlink, 2025).
o1-series (o1-preview, o1-mini, o1)	STEM-focused data, some general data (pre-2024/2025) (Sapkota et al., 2025; OpenAI, 2024d)	System 2 thinking, PhD-level STEM reasoning (o1-preview), fast reasoning (o1-mini), full o1 reasoning and multimodality (o1) (Sapkota et al., 2025; OpenAI, 2024d)	Analytical rigor, hypothesis generation/evaluation (biology, math, engineering) (OpenAI Help Center, n.d.; OpenAI, 2024e)	Specialized for STEM, general capabilities relative to GPT-4o may vary.	o1-mini is best performing benchmarked model on AIME 2024 and 2025 (OpenAI Help Center, n.d.; OpenAI, 2024e). Used for generating Finite Element code in geotechnical engineering (Kim et al., 2025; OpenAI, 2024f).

3. The New Epoch of Content Generation: Diverse Applications, Quality Assurance, and Ethical Imperatives

ChatGPT has inaugurated a new epoch in content generation, demonstrating remarkable versatility across an array of domains. This capability, however, is rapidly evolving from simple content creation to autonomous task execution, which introduces more complex challenges concerning quality assurance and ethical responsibility.

3.1. ChatGPT's Role in Diverse Content Generation and Task Automation

The applications of ChatGPT and its underlying models are extensive and expanding from content creation to proactive task completion:

- **Technical and Scientific Content:** In engineering, ChatGPT assists in drafting reports, generating software documentation, and producing code snippets (Neveditsin et al., 2025). Multivocal literature reviews indicate error rates for engineering tasks average around 20-30% for GPT-4 (Ray, 2023). In medicine, it is used for generating patient reports and drafting discharge summaries (Hariri, 2023), though error rates can range from 8% to 83% (Ray, 2023).

- **Marketing and SEO Content:** Marketers leverage ChatGPT for creating blog posts, ad copy, social media updates, and personalized email campaigns. It also aids in SEO by generating topic ideas and crafting meta descriptions (Fisher, 2025).

- **Legal Content:** Law firms utilize ChatGPT for drafting client correspondence, creating legal blog content, and developing marketing materials to increase efficiency (Fisher, 2025).

- **Creative Writing:** ChatGPT has shown aptitude in generating creative content such as stories, poetry, and scripts, acting as a catalyst for imaginative endeavors (Hariri, 2023; OpenAI, 2024c; Elkatmis, 2024; Niloy et al., 2024; Zhu et al., 2024).

- **Academic Content:** In academic settings, ChatGPT assists with literature reviews, drafting sections of papers, generating study materials, and creating quizzes (Alasadi & Baiz, 2023; Dwivedi et al., 2023; Isiaku et al., 2024; Michel-Villarreal et al., 2023; Preiksaitis & Rose, 2023; Wu, 2023; Cotton et al., 2024).

- **Automated Task Execution with AI Agents:** The next frontier lies in Agentic AI, where LLMs are empowered to act as autonomous agents. These agents move beyond generating content to performing complex, multi-step tasks. For example, an agent might not just write code but also debug it, or not just draft a marketing email but also execute the entire campaign by analyzing performance data and adjusting its strategy (Park et al., 2023; Xi et al., 2023). This represents a shift from a content creator to a task automator.

3.2. Methodologies for Quality Control, Coherence, and Accuracy

Ensuring the quality and reliability of AI outputs, whether static content or agentic actions, is paramount.

- **Human Oversight and Human-in-the-Loop:** This remains the most critical control measure. Expert review is essential for content where errors have severe consequences (Infosys Limited, 2023; Dwivedi et al., 2023; Susskind & Susskind, 2022). For Agentic AI, this evolves into a "human-in-the-

loop" model, where humans supervise, intervene, and approve agent actions before execution to prevent errors and ensure safety (Shneiderman, 2022).

- **Prompt Engineering:** The quality of output is highly dependent on the input prompt. Effective prompt engineering is a key skill for guiding both content generation and agent behavior (Vu et al., 2025; Arvidsson & Axell, 2023; Marvin et al., 2023; Velásquez-Henao et al., 2023; Zhou et al., 2022; Herman, 2025; Knoth et al., 2024; Pan et al., 2024).

- **Iterative Refinement:** Using feedback loops to progressively refine outputs is a common practice to improve quality for both text and agent action sequences (Kaushik et al., 2025; Hadi et al., 2023; Chen et al., 2025; Liu et al, 2023; Sivarajkumar et al., 2024; Xu et al., 2024).

- **Fact-Checking and Source Verification:** Due to the risk of hallucinations, rigorous fact-checking is essential (Akhtarshenas et al., 2025; Hodge Jr, 2023; Perlman, 2023; Surden, 2023; OpenAI, 2024a). For agents, this includes grounding their knowledge in real-time, verifiable data sources before they act.

- **Process and Tool-Use Validation:** For AI agents, quality control must extend beyond the final output to validate the entire process. This includes verifying that the agent's reasoning is sound and that it uses its tools (e.g., web browsers, APIs) correctly and safely (Shinn et al., 2023).

- **Specialized Evaluation Metrics and Tools:** Domain-specific metrics like BLURB (Akhtarshenas et al., 2025; Gu et al., 2021; Naseem et al., 2022) and tools like SelfCheckGPT (Akhtarshenas et al., 2025) are crucial for objective assessment.

- **Error Rate Analysis:** Systematic analysis of error rates provides insights into reliability and highlights areas needing improvement (Ray, 2023).

3.3. Ethical Considerations in Content and Task Automation

The power of generative AI brings forth a spectrum of ethical challenges that are amplified by the introduction of autonomy.

- **Bias:** Generated content and agentic actions can reflect and amplify societal biases from training data (Bender et al., 2021; Hariri, 2023; OpenAI, 2024c).

- **Trustworthiness and Reliability:** The probabilistic nature of LLMs means their outputs are not always factually correct or reliable, posing risks if unverified information is disseminated (Garousi, 2025; Schiller, 2024; Lee, 2024; Preiksaitis & Rose, 2023; Hariri, 2023; Wu et al., 2024; Xu et al., 2024; Xu et al., 2024).

- **Security and Misuse:** The potential for misuse is significant. Agentic AI dramatically lowers the barrier for malicious activities by enabling the automation of tasks like orchestrating large-scale phishing campaigns or propagating disinformation (Johnson & Acemoglu, 2023; OpenAI, 2023c; Veisi et al., 2025).

- **Accountability and Autonomous Action:** Agents capable of autonomous action raise profound ethical questions about accountability. Determining responsibility when an autonomous agent causes financial, social, or physical harm is a complex challenge for which legal and ethical frameworks are still nascent (Weidinger et al., 2024).

- **Social Norms and Cultural Sensitivity:** Generated content and actions must align with diverse cultural and societal expectations to avoid offense or misinterpretation (Johnson & Acemoglu, 2023; OpenAI, 2023c; Veisi et al., 2025).

- **Ethical Data Sourcing and Privacy:** Concerns persist regarding the methods used for collecting training data and the privacy of user inputs fed into ChatGPT (Daun & Brings, 2023; Marques & Bernardino, 2024; Neveditsin et al., 2025; OpenAI, 2024a).

- **Copyright and Authorship:** The generation of content raises complex questions about intellectual property rights, originality, authorship attribution, and plagiarism, especially when outputs closely resemble training data or are presented as original work (OpenAI, 2023c; Gamage et al., 2023; Hannigan et al., 2024; Jiang et al., 2024; Susnjak & McIntosh, 2024). Legal frameworks are still evolving to address these issues (Infosys Limited, 2023).

3.4. Advancing Method and Theory for Responsible Content Generation & Task Automation

To navigate this new epoch responsibly, advancements in both methodology and theory are needed. A **"Human-AI Symbiotic"** framework is proposed, where AI systems like ChatGPT handle initial drafting, information synthesis, and repetitive aspects of content creation, while human experts focus on critical review, strategic guidance, ethical filtering, creative refinement, and ensuring contextual appropriateness (Levitt & Grubaugh, 2023; U.S. Department of Education, 2023). This moves beyond simple post-editing to a more integrated collaborative process. This integrated collaboration is essential for both content generation and the governance of AI agents. Furthermore, dynamic, context-aware quality control mechanisms are crucial. Such a framework requires a paradigm shift, especially within academia, moving away from viewing generative AI as a threat that erodes essential skills. Instead, it should be seen as an opportunity to foster critical thinking and analytical capabilities (Levitt & Grubaugh, 2023). When properly integrated, this human-AI collaboration can become a powerful tool to prepare students for an increasingly AI-augmented workforce, where the ability to partner effectively with intelligent systems is a crucial skill (U.S. Department of Education, 2023). Furthermore, the development of dynamic, context-aware quality control mechanisms that adapt to the content type, intended audience, and potential risks is crucial, moving beyond static checklists or generic evaluation metrics.

The landscape of ChatGPT-driven content generation is characterized by a fundamental **"Quality-Scalability-Ethics Trilemma."** There exists an inherent tension in simultaneously achieving high-quality, nuanced content, generating this content efficiently at scale, and rigorously adhering to ethical standards such as bias mitigation, truthfulness, and intellectual property respect (Dempere et al., 2023; Gamage et al., 2023; Hannigan et al., 2024; Jiang et al., 2024; Susnjak & McIntosh, 2024; Susskind & Susskind, 2022). This inherent tension is acutely magnified with Agentic AI. Granting AI autonomy dramatically increases its scalability and potential impact, making the trade-offs with quality and ethics far more critical. Prioritizing scalability, such as rapid, fully automated content creation for mass dissemination, can lead to a degradation in content quality – resulting in generic, superficial, or inaccurate outputs – and can significantly amplify ethical risks, for example, through the widespread propagation of biased information or misinformation. Similarly, strict adherence to comprehensive ethical guidelines, including rigorous bias detection and mitigation protocols or meticulous checks for originality and factual accuracy, can inherently slow down the generation process and limit the scope of what can be feasibly automated. An autonomous system that scales rapidly without robust quality control and ethical safeguards poses unacceptable risks. Efforts to enhance quality and ethics through human review or rigorous validation often reduce the speed

and scalability that make agents so powerful (Ray, 2023; Garousi, 2025; Schiller, 2024; Nguyen et al., 2023; Lee, 2024; Preiksaitis & Rose, 2023; Wu et al., 2024; Xu et al., 2024; Xu et al., 2024).

This trilemma implies that a central challenge is developing techniques that co-optimize these dimensions or provide transparent trade-offs. This might involve multi-stage processes where different levels of quality control and ethical scrutiny are applied based on risk, or AI-assisted tools for ethical review. Theoretically, this calls for new models of "responsible generative efficiency" and "trustworthy autonomy" that can quantify, predict, and guide the balance between these competing factors, moving the field towards a more mature and accountable approach to AI content creation and task automation.

4. ChatGPT as a Catalyst for Knowledge Discovery: Methodologies, Scientific Inquiry, and Future Paradigms

ChatGPT and similar LLMs are increasingly being explored not just as information retrieval tools but as active catalysts in the process of knowledge discovery. This is evolving from simple assistance to the deployment of autonomous systems that can manage complex scientific workflows, offering new paradigms for generating novel insights.

4.1. Methodologies for Knowledge Extraction from Unstructured Data

LLMs excel at processing and interpreting unstructured text, which constitutes a massive portion of the world's data. This capability is being harnessed through increasingly sophisticated methodologies:

- **Information Extraction from Diverse Sources:** ChatGPT can parse complex documents, such as historical seedlists or health technology assessment (HTA) documents in different languages, to extract specific data points where rule-based methods falter (Dagdelen et al., 2024; Mitra et al., 2024; Yang et al., 2022; Shah et al., 2023).

- **Qualitative Data Analysis Assistance:** Researchers are exploring ChatGPT for assisting in qualitative analysis, such as generating initial codes or identifying potential themes (Chen et al., 2025; Kaushik et al., 2025; Liu et al, 2023; Sivarajkumar et al., 2024; Xu et al., 2024). However, careful prompting and validation are required, as LLMs can generate nonsensical data if not properly guided (Chen et al., 2025; Kaushik et al., 2025; Liu et al, 2023; Sivarajkumar et al., 2024; Xu et al., 2024).

- **LLMs Combined with Knowledge Graphs (KGs):** A promising methodology involves integrating LLMs with KGs. The GoAI method, for instance, uses an LLM to build and explore a KG of scientific literature to generate novel research ideas, providing a more structured approach than relying on the LLM alone (Gao et al., 2025; Pan et al., 2024).

- **Autonomous Knowledge Discovery with AI Agents:** The next methodological leap involves deploying Agentic AI to create automated knowledge discovery pipelines. These agents can be tasked with a high-level goal and then autonomously plan and execute a sequence of actions – such as searching databases, retrieving papers, extracting data, and synthesizing findings – to deliver structured knowledge with minimal human intervention (Bran et al., 2024).

- **Prompt Injection Vulnerabilities:** Research into prompt injection techniques highlights how the knowledge extraction process can be manipulated, underscoring security vulnerabilities that must be addressed for reliable knowledge discovery, especially in autonomous systems (Chang et al., 2025).

4.2. Applications in Scientific Research

ChatGPT is finding applications across the scientific research lifecycle, with agentic systems poised to integrate these functions into automated workflows:

- **Hypothesis Generation:** Models like GPT-4 can generate plausible and original scientific hypotheses, sometimes outperforming human graduate students in specific contexts (Noy & Zhang, 2023; OpenAI, 2024e; OpenAI Help Center, n.d.).

- **Literature Review Assistance:** LLMs can accelerate literature reviews by summarizing articles and identifying relevant papers and themes (Mitra et al., 2024; Dagdelen et al., 2024; Yang et al., 2022; Albadarin et al., 2024; Gabashvili, 2023; Haman & Školník, 2024; Imran & Almusharraf, 2023; Mostafapour et al., 2024; Wang et al., 2023; Waseem et al., 2023).

- **Experimental Design Support:** ChatGPT can assist in outlining experimental procedures but may require expert refinement to address oversimplifications or "loose ends" (Dai et al., 2023; Eymann et al., 2025; Fill et al., 2023; Li et al., 2023; OpenAI, 2024e).

- **Data Analysis and Interpretation:** LLMs can assist in analyzing large volumes of text data to identify patterns and emerging themes (Haltaufderheide & Ranisch, 2024; Hariri, 2023; Gabashvili, 2023; Garg et al., 2023; Li et al., 2024; Sallam, 2023; Pan et al., 2024; OpenAI, 2024c).

- **Simulating Abductive Reasoning:** LLMs can simulate abductive reasoning to infer plausible explanations or methodologies, thereby aiding research discovery (Glickman & Zhang, 2024; Huang & Chang, 2022; Bhagavatula et al., 2019; Garbuio & Lin, 2021; Magnani & Arfini, 2024; Pareschi, 2023; Xu et al., 2025).

- **Automating Research with Scientific Agents:** The culmination of these capabilities is the creation of scientific agents. These are autonomous systems designed to conduct research by integrating multiple steps. For instance, a scientific agent could be tasked with a high-level research question and then autonomously search literature, formulate a hypothesis, design and execute code for a simulated experiment, analyze the results, and draft a preliminary report, dramatically accelerating the pace of discovery (Boiko et al., 2023). OpenAI ChatGPT's and Google Gemini's Deep Research language models are good examples.

4.3. Critical Assessment of ChatGPT's Role in Advancing Research

The integration of generative AI into scientific inquiry presents a duality of prospects and challenges:

- **Acceleration and Efficiency:** AI has the potential to dramatically accelerate research by automating time-consuming tasks, allowing researchers to focus on higher-level conceptual work (Dai et al., 2023; Fill et al., 2023; Li et al., 2023; Noy & Zhang, 2023; Rice et al., 2024).

- **Accuracy and Reliability Concerns:** The propensity for hallucinations and bias is a major concern that necessitates rigorous validation of all AI-generated outputs (Bender et al., 2021; OpenAI, 2024a; Rice et al., 2024). This risk is magnified for autonomous agents, where acting on a single hallucinated fact could derail an entire research workflow.

- **The Indispensable Role of Human Expertise:** Human expertise remains crucial for critical evaluation, contextual understanding, and ensuring methodological soundness (Dai et al., 2023; Fill et al., 2023; Li et al., 2023; Noy & Zhang, 2023; Rice et al., 2024). As research becomes more automated,

the human role shifts from task execution to high-level strategic direction and critical supervision of the AI's process and outputs.

4.4. Advancing Method and Theory in AI-Augmented Knowledge Discovery

The use of AI in knowledge discovery is pushing methodological and theoretical boundaries:

- Frameworks like GoAI (Gao et al., 2025) exemplify a move toward structured methodologies that combine LLMs with KGs for more transparent idea generation.

- The concept of LLMs "simulating abductive reasoning" (Glickman & Zhang, 2024) suggests a new theoretical lens for understanding how these models contribute to scientific insight, moving beyond pattern matching toward computational reasoning.

A significant hurdle in this new paradigm is the **"Black Box Conundrum."** While tools like ChatGPT can generate novel hypotheses (Eymann et al., 2025), the internal processes remain opaque (Dempere et al., 2023). This opacity is especially problematic for science, which demands transparency and reproducibility. The shift towards autonomous scientific agents makes this conundrum more acute. For an agent's discoveries to be scientifically valid, its entire decision-making process must be transparent and verifiable. An inscrutable "reasoning" process can lead to a "crisis of explanation," undermining the principles of systematic inquiry. Consequently, advancing true knowledge discovery with AI necessitates significant progress in explainable AI (XAI) tailored for generative models (Dovesi et al., 2024). Without such advancements, AI-assisted discovery might remain a useful heuristic but will lack the demonstrable rigor required for foundational breakthroughs. This challenge calls for new theories of "computational scientific reasoning" to bridge the gap between the statistical generation of current LLMs and the logical, evidence-based reasoning that is the hallmark of the scientific method, especially as these reasoning capabilities become embodied in autonomous agents.

5. Revolutionizing Education and Training: ChatGPT's Global Impact on Pedagogy, Assessment, and Equity

The integration of ChatGPT into education has been met with a mixture of excitement and trepidation, signaling a potential revolution in pedagogy, assessment, and the pursuit of educational equity. Its capabilities offer novel ways to personalize learning and support diverse learner needs, but they also pose significant challenges to traditional paradigms, particularly concerning academic integrity and the development of critical thinking. This evolution is now accelerating toward the use of more autonomous Agentic AI systems, magnifying both the opportunities and the risks.

5.1. Applications in Education

ChatGPT's versatility has led to its application across numerous facets of the educational landscape, with agentic systems representing the next frontier.
- **Personalized Learning:** A primary application is facilitating personalized learning experiences. ChatGPT can adapt content, offer real-time feedback, and function as a virtual tutor available 24/7 (Davar et al., 2025; Li, 2025).
- **Curriculum and Lesson Planning:** Educators use ChatGPT to assist in designing courses, developing lesson plans, and visualizing theoretical concepts in practical settings (Li, 2025; Li et al., 2025).

- **Innovative Student Assessment:** ChatGPT is being explored for generating diverse assessment items and designing tasks that promote critical thinking (Davar et al., 2025). GenAI can also personalize assessments and feedback based on learner responses (Arslan et al., 2024).

- **Teaching Aids and Interactive Tools:** The technology can be harnessed to develop engaging teaching aids, virtual instructors, and interactive simulations (Davar et al., 2025).

- **Support for Diverse Learners:** ChatGPT enhances accessibility for students with disabilities and multilingual learners through translation and simplification (Chan et al., 2024).

- **Autonomous Learning Companions and Agents:** The next evolutionary step is the deployment of AI agents as personalized learning companions. These agents go beyond tutoring by autonomously managing a student's long-term learning journey. They can co-design study plans, curate resources from vast digital libraries, schedule tasks, and proactively adapt strategies based on performance, transforming the learning process into a continuous, interactive dialogue (Molenaar, 2024; Salesforce, 2025).

5.2. Impact on Critical Thinking, Academic Integrity, and Ethics

The integration of ChatGPT – a generative AI – into education brings hightened implications for cognitive skills and ethical conduct, which are amplified by agentic systems.

- **Critical Thinking:** A dichotomy exists where AI can either be used to generate thought-provoking prompts that foster analysis or, through over-reliance, erode students' ability to think deeply (Mohammed, 2025; Dempere et al., 2023). Concerns persist that students may become cognitively passive (Alghazo et al., 2025). The introduction of AI agents deepens this concern, as they could automate not just the answers but the entire process of inquiry and discovery, potentially deskilling students in research and problem-solving (Zawacki-Richter et al., 2019).
- **Academic Integrity:** The risk of plagiarism with AI-generated text is a primary concern (Dempere et al., 2023; Mohammed, 2025). With agents, this evolves from verifying authorship of text to verifying authorship of action. Strategies to uphold integrity must shift toward assessments that are inherently human-centric, such as project-based work and oral examinations (Mohammed, 2025).
- **Ethical Challenges:** Broader ethical issues include data privacy, equity, and potential biases in AI content (Dempere et al., 2023). Agentic AI introduces new dilemmas regarding student autonomy and data sovereignty. An agent managing a student's learning collects vast amounts of sensitive performance and behavioral data, raising critical questions about consent, surveillance, and how that data is used to shape a student's educational future (Prinsloo, 2020).

5.3. Global Perspectives and Educational Equity

The adoption and impact of ChatGPT in education varies significantly across global contexts, influenced by infrastructure, literacy, and policy, technological infrastructure, digital literacy, cultural norms, and institutional policies

- **Diverse International Perceptions:** Studies from regions like Pakistan and Indonesia reveal mixed student perceptions, balancing the benefits of ChatGPT as an AI assistance with concerns about its impact on deep thinking and integrity (Alghazo et al., 2025; Adiyono et al., 2025).

- **Democratization vs. Digital Divide:** ChatGPT has the potential to democratize education by providing widespread access to high-quality learning resources (Li, 2025). However, it also risks exacerbating the digital divide if access to technology, internet, and AI literacy are inequitably distributed (Chan et al., 2024). The advent of powerful, resource-intensive learning agents could

create a new, more profound equity gap between students who have access to personalized autonomous tutors and those who do not (UNESCO, 2023).

- **Cultural Context and Bias:** LLMs trained on predominantly Western datasets may perpetuate cultural biases (Dempere et al., 2023). While AI can be used to decolonize curricula, this requires careful human oversight to avoid reinforcing existing biases (Chan et al., 2024).

5.4. Advancing Educational Research, Theories, and Pedagogical Models

The advent of ChatGPT and other generative AI tools necessitates a re-evaluation of educational theories and practices.

- **Revisiting Learning Theories:** ChatGPT's capabilities challenge and offer new lenses through which to view learning theories such as constructivism (where students actively construct knowledge, potentially aided by AI tools) (Li et al., 2025) and self-determination theory (exploring AI's impact on student autonomy, competence, and relatedness) (Alghazo et al., 2025).

- **Transforming Assessment Paradigms:** Traditional assessment methods are being questioned. There is a call for innovative assessment strategies that emphasize higher-order thinking, creativity, and authentic application of knowledge, rather than tasks easily outsourced to AI (Dempere et al., 2023). This includes exploring personalized, adaptive assessments leveraging GenAI (Arslan et al., 2024).

- **Methodological Rigor in AI-in-Education Research:** There is a critical need for methodological rigor in studying AI's impact on education. Researchers must carefully define experimental treatments, establish appropriate control groups, and use valid outcome measures that genuinely reflect learning, avoiding pitfalls of earlier "media/methods" debates where technology effects were often confounded with instructional design (Weidlich & Gašević, 2025).

- **Developing New Pedagogical Models:** The situation calls for the development of new pedagogical models that constructively integrate AI. This involves training educators and students in AI literacy, prompt engineering skills, and the critical evaluation of AI-generated outputs, and designing learning experiences that leverage AI as a tool for enhancing human intellect and creativity, rather than replacing it (Kasneci et al., 2023; Zhai, 2023).

The widespread availability of ChatGPT presents a **"Pedagogical Adaptation Imperative."** The initial defensive reactions focused on mitigating threats like plagiarism are insufficient (Mohammed, 2025; Dempere et al., 2023). This "shock" to the system compels a fundamental shift toward cultivating higher-order skills like critical thinking, creativity, and metacognition (Murray et al., 2025).

This necessitates a paradigm shift where AI is not just a tool to be policed but a cognitive partner to be leveraged. The imperative is to adapt curricula to foster human-AI collaboration. This vision must now expand to include Agentic AI. The future of education will likely involve a co-evolution where educators become facilitators and strategic supervisors of AI-augmented learning experiences. Their role will shift to managing classrooms of human-agent teams, guiding students in the ethical and effective use of their personalized learning companions. Theoretically, this calls for new frameworks of "AI-Integrated Pedagogy" that conceptualize learning as a synergistic process involving human learners, human educators, and AI agents. It requires adapting theories like constructivism to account for knowledge co-constructed with an autonomous agent, and developing new models of human-AI co-regulation to ensure technology enhances, rather than undermines, human intellect and autonomy (Molenaar, 2024).

6. Engineering New Frontiers with ChatGPT: Advancing Design, Optimization, and Methodological Frameworks

ChatGPT and other LLMs are rapidly permeating various engineering disciplines, offering novel tools for design, analysis, and optimization. Their ability to process natural language, generate code, and synthesize information is creating new possibilities, with the evolution toward autonomous Agentic AI systems poised to redefine traditional engineering methodologies.

6.1. Applications in Engineering Disciplines

The application of ChatGPT in engineering is diverse and expanding from task assistance to workflow automation.

- **Software Engineering:** LLMs are used for code generation, debugging, automated code review, and documentation, with experts reporting significant time savings (Neveditsin et al., 2025; Rawat et al., 2024). LLMs can also assist in translating natural language requirements into code (Yadav et al., 2025).

- **Building Information Modeling (BIM), Architecture, and Civil Engineering:** ChatGPT is explored for semantic search, information retrieval, and task planning (Yu et al., 2025). RAG has proven effective in helping ChatGPT apply localized BIM guidelines (Yu et al., 2025).

- **Mechanical, Industrial, and General Engineering Design:** LLMs assist in idea generation, conceptual design, and formulating engineering optimization problems (Vu et al., 2025; Jiang et al., 2025).

- **Geotechnical Engineering:** ChatGPT can generate finite element analysis (FEA) code for modeling complex processes, though its effectiveness varies based on the programming library used, underscoring its role as an assistant (Kim et al., 2025).

- **Control Systems Engineering:** Studies show ChatGPT can pass undergraduate control systems courses but struggles with open-ended projects requiring deep synthesis and practical judgment (Puthumanaillam & Ornik, 2025).

- **Automated Design and Analysis with Engineering Agents:** The next frontier is the deployment of **engineering agents.** These are autonomous systems that can manage complex, multi-step engineering workflows. For example, an agent could be tasked with a high-level goal, such as designing a mechanical part, and then autonomously generate design options, use software tools to run simulations (e.g., FEA), interpret the results, and iterate on the design until specifications are met (Wang et al., 2023).

6.2. Theoretical Constructs and Novel Engineering Methodologies

The integration of ChatGPT is prompting new engineering methodologies and theoretical constructs.

- **Prompt Engineering for Optimization:** Effective problem formulation using ChatGPT relies heavily on sophisticated prompt engineering and sequential learning approaches (Vu et al., 2025)

- **Human-LLM Design Practices:** Comparative studies are yielding insights into LLM strengths (e.g., breadth of ideation) and weaknesses (e.g., design fixation), leading to recommendations for structured design processes with human oversight (Ege et al., 2025).

- **Cognitive Impact on Design Thinking:** Research is exploring how AI influences designers' cognitive processes, such as fostering thinking divergence and fluency (Jiang et al., 2025).

- **LLMs in Systems Engineering (SE):** While LLMs can generate SE artifacts, there are significant risks, including tendencies towards "premature requirements definition" and "unsubstantiated numerical estimates" (Topcu et al., 2025). These risks are magnified in autonomous agentic systems where flawed assumptions could propagate through an entire automated workflow.

- **Methodologies for Agentic Workflows:** The rise of engineering agents necessitates new methodologies for managing human-agent and agent-agent collaboration. This includes designing frameworks for task decomposition, tool selection, and process validation to ensure the reliability and safety of autonomous engineering systems (Team, 2024).

6.3. Impact on Engineer Productivity and Future Practice

The adoption of ChatGPT in engineering has clear implications for productivity and the nature of work.

- **Productivity Gains:** Studies report significant productivity increases from using LLMs for tasks like code generation and drafting (Yadav et al., 2025). The shift toward agentic AI promises to extend these gains from task assistance to end-to-end workflow automation (Rawat et al., 2024).

- **Concerns and Challenges:** Concerns exist about over-dependence on AI, which could lead to skill degradation, and anxieties about job security (Yadav et al., 2025). The need for human oversight remains critical due to potential inaccuracies and biases (Ray, 2023).

- **Preparing Future Engineers:** Engineering curricula must adapt to prepare students for workplaces where GenAI tools are prevalent. This includes teaching AI literacy, prompt engineering, and the critical evaluation of AI outputs to ensure they can effectively supervise and collaborate with AI systems (Murray et al., 2025).

6.4. Advancing Engineering Methodologies and Theoretical Frameworks

The capabilities of ChatGPT can serve as a catalyst for advancing engineering methodologies and developing new theoretical frameworks:

- **Agent-Assisted Engineering Frameworks:** There is an opportunity to develop structured frameworks that explicitly integrate AI agents at various stages of the engineering design process. These frameworks would define roles, responsibilities, and interaction protocols for human engineers and their agentic counterparts.

- **Theories of AI-Robustness in Design:** The identification of LLM failure modes (Topcu et al., 2025) can inform new theories around "AI-robustness" to predict and mitigate risks associated with using AI in critical applications.

The application of LLMs like ChatGPT in engineering is fostering a shift towards what can be conceptualized as **"Human-AI Cognitive Symbiosis."** Current evidence indicates that while LLMs can assist with a range of tasks (Yadav et al., 2025; Ege et al., 2025; Kim et al., 2025), they require significant human guidance and correction, especially in complex or safety-critical situations (Kim et al., 2025; Puthumanaillam & Ornik, 2025). Human engineers possess superior capabilities in deep contextual understanding, critical judgment, and true innovation, areas where current LLMs are limited (Ege et al., 2025).

The most effective applications arise when human engineers strategically leverage these AI systems as powerful cognitive partners. This dynamic is evolving with the advent of Agentic AI. It is less about AI replacing engineers and more about augmenting their capabilities by delegating complex workflows to

autonomous agents under human supervision. Consequently, engineering education and practice must evolve to cultivate "AI-collaboration literacy" – the skills required to effectively prompt, guide, validate, and ethically integrate the work of AI agents. The engineer's role will shift from a sole problem-solver to an orchestrator of human and AI agent collaborative systems. This necessitates new theoretical models of **"Human-Agent Symbiosis"** in engineering. Such theories would aim to elucidate how the distinct strengths of humans and AI agents can be optimally combined and how to design engineering processes that cultivate this synergy, elevating AI from mere "tools" to active, agentic partners in the engineering endeavor.

Table 2 summarizes key applications of ChatGPT in education and engineering, highlighting benefits, challenges, and novel implications.

Table 2. Key Applications of ChatGPT in Education and Engineering: Benefits, Challenges, and Novel Methodological/Theoretical Implications (Authors).

Application Area	Specific Use Cases	Documented Benefits	Key Challenges	Novel Methodological/Theoretical Implications
Education	Personalized learning, virtual tutoring (Davar et al., 2025)	Tailored content, adaptive pacing, 24/7 support, increased engagement (Davar et al., 2025)	Over-reliance, reduced critical thinking, accuracy of information, data privacy, equity of access (AlAli & Wardat, 2024)	Development of "AI-Integrated Pedagogy"; re-evaluation of constructivist and self-determination learning theories in AI contexts.
	Curriculum/Lesson Planning (Li, 2025)	Efficiency for educators, idea generation, diverse material creation (Li, 2025)	Quality of AI suggestions, maintaining teacher creativity, potential for generic content (Li, 2025)	Frameworks for AI-assisted curriculum design that balance efficiency with pedagogical soundness and teacher agency.
	Student Assessment (Chan et al., 2024)	Generation of diverse quiz/exam questions, formative feedback, personalized assessment (Li, 2025)	Academic integrity (plagiarism), difficulty assessing true understanding, fairness of AI-generated assessments (Mohammed, 2025)	New assessment paradigms focusing on higher-order skills, process over product; ethical guidelines for AI in assessment.
Engineering	Software Engineering (Code generation, debugging, QA) (Yadav et al., 2025)	Increased developer productivity, reduced coding time, improved code quality (Yadav et al., 2025)	Accuracy of generated code, over-dependence, skill degradation, security risks, bias in code (Yadav et al., 2025)	"Human-LLM Cognitive Symbiosis" models for software development; AI-collaboration literacy for engineers.
	BIM/Architecture/Civil Engineering (Info retrieval, design	Enhanced understanding of domain-specific knowledge (with RAG), task planning support (Yu et al., 2025)	Reliance on quality of RAG documents, need for domain expertise in prompt/RAG setup (Yu et al., 2025)	Methodologies for integrating LLMs with domain-specific knowledge bases (e.g., RAG) for specialized engineering tasks.

visualization) (Yu et al., 2025)			
Mechanical/Industrial Design (Ideation, prototyping, optimization) (Jiang et al., 2025)	Accelerated idea generation, exploration of diverse concepts, assistance in optimization problem formulation (Jiang et al., 2025)	Design fixation, unnecessary complexity, misinterpretation of feedback, unsubstantiated estimates (Ege et al., 2025)	"AI-Augmented Engineering Design" frameworks; theories of "AI-robustness" in design; understanding LLM impact on cognitive design processes.
Geotechnical Engineering (Finite Element Analysis code generation) (Kim et al., 2025)	Assistance in implementing numerical models, especially with high-level libraries (Kim et al., 2025)	Extensive human intervention needed for low-level programming or complex problems; requires user expertise (Kim et al., 2025)	Frameworks for human-AI collaboration in complex numerical modeling and simulation.

7. Navigating the AI Revolution: Themes, Tensions, Critical Gaps, and Future Directions

The proliferation and evolving capabilities of ChatGPT have undeniably reshaped multiple domains, yet this progress is accompanied by critical research gaps and a pressing need for a forward-looking agenda. Synthesizing findings across Natural Language Understanding (NLU), content generation, knowledge discovery, education, and engineering reveals common themes and distinct challenges that must be addressed to harness the full potential of generative AI ethically and effectively, especially as the technology advances from passive tools to autonomous Agentic AI systems.

7.1. Common Themes Across Domains

A foundational theme is the **transformative capability** of ChatGPT. Hailed as a "disruptive" technology, its ability to generate human-like text and engage in complex tasks is altering established practices. Newer models like GPT-4o and the o1 series incorporate multimodality and advanced reasoning, pointing towards future models with enhanced "System 2 thinking" (Hagendorff, 2024). This opens doors to unexplored applications with significant disruptive potential, including the deployment of autonomous agentic systems capable of complex, multi-step task execution across science, education, and engineering. However, these advancements are accompanied by **significant limitations.** These include persistent issues with factual inaccuracy ("hallucinations"), inherent biases, the **lack of transparency** ("Black Box Conundrum"), and difficulties with complex reasoning. This necessitates the **essential role of human oversight.** The concept of **Human-AI "Cognitive Symbiosis"** emerges as a crucial paradigm, which is now evolving from collaboration with passive tools to frameworks of human-agent teaming and orchestration where humans guide and supervise autonomous systems (Shneiderman, 2022). Furthermore, **pervasive ethical considerations** are interwoven throughout all domains, including bias, reliability, misuse, data privacy, and authorship. Finally, the disruptive nature of ChatGPT necessitates a **fundamental imperative for adaptation.** This involves significant changes in educational pedagogy ("Pedagogical Adaptation Imperative"), engineering design processes ("Human-Agent Cognitive Symbiosis"), and the development of new methodologies for quality control across all fields. Figure 1 provides a broad overview of the contribution of this critical review.
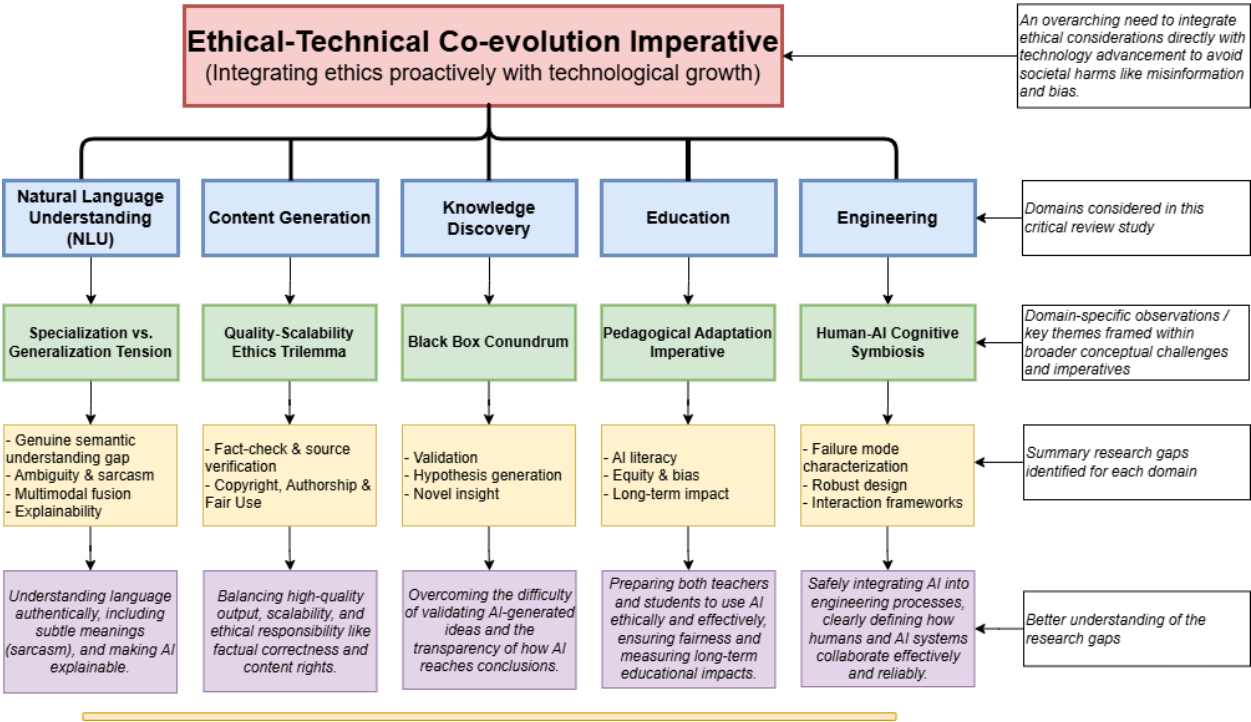
Figure 1. Overview of this study's contribution. Navigating AI's Ethical Frontier and Human-AI Synergy: Balancing Technology and Society.

7.2. Synthesis of Themes and Identification of Critical Research Gaps

The interplay of these themes creates inherent tensions and highlights critical research gaps, which are magnified by the prospect of agentic AI.

- **Natural Language Understanding (NLU):** The "Specialization vs. Generalization Tension" persists. A fundamental gap lies in discerning genuine semantic understanding versus sophisticated pattern matching (Shormani, 2024; Katzir, 2023; Baroni, 2020; Lake, 2019). This gap becomes a critical safety concern for agentic systems that must act reliably based on their understanding of commands and environmental cues. The lack of explainability hinders trust and theoretical advancement, a problem that becomes acute when an agent's reasoning cannot be audited (Achiam et al., 2023; Liu et al., 2023; Sapkota et al., 2025).

- **Content Generation:** The "Quality-Scalability-Ethics Trilemma" is a core challenge (Dempere et al., 2023; Gamage et al., 2023). With the rise of agentic AI, this trilemma intensifies, as the potential for autonomous systems to act unethically at scale poses a far greater risk than generating harmful text alone. New technical solutions and legal frameworks are urgently needed to govern the actions of these agents (Ballardini et al., 2019; Craig, 2022).

- **Knowledge Discovery:** The "Black Box Conundrum" hinders the validation of AI-generated insights (Dai et al., 2023; Noy & Zhang, 2023). When a scientific agent autonomously conducts a research workflow, the need for a transparent and reproducible "chain of reasoning" becomes paramount for scientific integrity.

- **Education:** The "Pedagogical Adaptation Imperative" demands a shift in focus to skills that complement AI. A critical gap is the lack of research on how to educate students to collaborate with and critically supervise learning agents without sacrificing their own cognitive autonomy

(Weidlich & Gašević, 2025; Kasneci et al., 2023). Ensuring equitable access to powerful learning agents is crucial to prevent a widening of educational disparities (Sabzalieva & Valentini, 2023).

- **Engineering:** The "Human-LLM Cognitive Symbiosis" must evolve into robust human-agent teaming. A major gap exists in developing validation techniques for agents in safety-critical applications and creating theoretical frameworks for trust and responsibility in these collaborative systems (Topcu et al., 2025; Miller, 2023).

Binding these is the overarching **"Ethical-Technical Co-evolution Imperative."** Technical advancements are inextricably linked with escalating ethical challenges (Bender et al., 2021). This is most evident with agentic AI, where the capacity for autonomous action demands that ethics, safety, and alignment are not afterthoughts but are embedded into the core design of the system (Weidinger et al., 2024).

7.3. Proposal of a Forward-Looking Research Agenda

To address these gaps, a concerted research agenda is proposed, with a special focus on the challenges and opportunities presented by agentic AI.

7.3.1. Methodological Advancements

- **NLU:** Develop benchmarks that assess "deep understanding" and robust reasoning, critical for safe agentic behavior.
- **Content & Action Generation:** Design adaptive quality and ethical control frameworks that are integrated directly into an agent's decision-making loop.
- **Knowledge Discovery:** Develop and validate rigorous protocols for human supervision of AI-assisted hypothesis generation and experimentation.
- **Education:** Conduct longitudinal studies on the impact of learning agents on cognitive development. Design and test AI literacy curricula focused on human-agent collaboration.
- **Engineering:** Formulate comprehensive testing and validation protocols for agents used in safety-critical design tasks and implement robust human-in-the-loop control frameworks.
- **Cross-Domain Methodologies for Agentic Systems:** A crucial priority is to develop standardized safety protocols, robust and intuitive human-in-the-loop control mechanisms, and secure "sandboxing" environments for testing the behavior of autonomous agents before deployment in real-world settings.

7.3.2. Theoretical Advancements:

- **NLU:** Formulate theories of "Explainable Generative NLU" to make agent reasoning transparent.
- **Content & Action Generation:** Develop "Ethical AI Agency Frameworks" that provide a theoretical basis for guiding the responsible actions of autonomous systems.
- **Knowledge Discovery:** Propose "Computational Creativity Theories" to explain how AI agents contribute to novel discovery.
- **Education:** Build "AI-Augmented Learning Theories" that model how students learn effectively in partnership with AI agents, exploring frameworks like "Cyborg Pedagogy."

- **Engineering:** Conceptualize "Human-Agent Symbiotic Engineering Theories" that define principles for shared cognition and distributed responsibility in human-agent teams.

- **Theories of Trustworthy Autonomy and Governance:** An overarching theoretical challenge is to develop robust theories of human-agent teaming, create computational models for agent accountability, and design governance frameworks for multi-agent ecosystems where agents interact with each other and with society (Xi et al., 2023).

This research agenda must be guided by the "Ethical-Technical Co-evolution Imperative." This implies embedding ethical design, fairness, transparency, and safety into the core R&D lifecycle of AI systems. Methodologically, this requires new ways to test for and mitigate ethical risks. Theoretically, it calls for "Co-evolutionary AI Development Frameworks" that model the interplay between technical progress and societal impact. This involves fostering "Anticipatory Governance" models for AI, where potential future impacts of widespread agent deployment are systematically explored and proactively addressed to guide innovation toward solutions that are not only powerful but also principled, equitable, and aligned with human values (Floridi & Nobre, 2024).

Table 3 outlines critical research gaps and proposes elements of a future agenda.

Table 3. Critical Research Gaps and Future Agenda for ChatGPT Research (Advancing Method and Theory).

Domain	Specific Identified Research Gap	Proposed Novel Research Question(s)	Potential Methodological Advancement	Potential Theoretical Advancement
NLU	True semantic understanding vs. mimicry; robustness to ambiguity; explainability (Sapkota et al., 2025)	How can NLU models be designed to exhibit verifiable deep understanding and provide transparent reasoning for their interpretations?	Development of "Deep Understanding Benchmarks"; new XAI techniques for generative NLU.	Theories of "Explainable Generative NLU"; models of computational semantics beyond statistical co-occurrence.
Content Generation	Ensuring factual accuracy; dynamic quality control; IP & copyright (Dempere et al., 2023)	What adaptive mechanisms can ensure real-time quality and ethical compliance in AI content generation across diverse contexts?	Adaptive, context-aware QA frameworks; blockchain or other technologies for provenance tracking.	"Ethical AI Content Frameworks"; theories of "Responsible Generative Efficiency."
Knowledge Discovery	Validating AI-generated hypotheses; moving from info extraction to insight; ethical AI in	How can LLMs be integrated into the scientific method to reliably generate and validate novel, theoretically	Rigorous validation protocols for AI-discovered knowledge; hybrid LLM-KG-	"Computational Creativity Theories" for scientific discovery; models of AI-

	science (Rice et al., 2024)	grounded hypotheses?	Experimental methodologies.	assisted abductive reasoning.
Education	Longitudinal impact on learning & critical thinking; AI literacy curricula; equity & bias in EdAI (Dempere et al., 2023); K-12 & special education gaps (Dimeli & Kostas, 2025)	What pedagogical frameworks optimize human-AI collaboration for deep learning and critical skill development across diverse learners and contexts?	Longitudinal mixed-methods studies; co-design of AI literacy programs with educators and students; comparative studies in underrepresented educational settings.	"AI-Augmented Learning Theories"; frameworks for "Cyborg Pedagogy"; theories of ethical AI integration in diverse educational systems.
Engineering	LLMs in safety-critical tasks; understanding LLM failure modes in complex design (Topcu et al., 2025); human-LLM collaboration frameworks (Empirical Methods in Natural Language Processing, 2024); NL to code/design beyond software (Yadav et al., 2025)	How can engineering design and optimization processes be re-theorized to effectively and safely incorporate LLM cognitive capabilities?	Protocols for LLM validation in complex simulations; frameworks for human-in-the-loop control for safety-critical engineering AI.	"Human-AI Symbiotic Engineering Design Theories"; theories of "AI-Robustness" in engineering systems.

7.4. Practical Implications for Method, Theory, and Practice

The synthesis of these themes and the proposed research agenda have profound implications, particularly as the field moves from passive language generation toward autonomous Agentic AI.

- **Method:** The identified limitations necessitate new methodological approaches. This includes developing robust validation protocols for both generated content and agentic actions, advancing techniques like Retrieval-Augmented Generation (RAG) to ground agent knowledge (Lewis et al., 2020), and establishing prompt engineering as a core skill for effective human-agent interaction. Crucially, new methods are needed for designing, testing, and ensuring the safety and reliability of complex, multi-step agentic workflows (Team, 2024).

- **Theory:** The challenges and emergent interactions demand new theoretical frameworks. These include theories for Explainable NLU, Responsible Generative Efficiency, and AI-assisted Abductive Reasoning. In education and engineering, this means developing AI-Augmented Learning Theories and Human-Agent Symbiotic Engineering Theories. These frameworks are the essential theoretical underpinnings for building trustworthy and beneficial AI agents. Overarching this is the need for Co-evolutionary AI Development Frameworks that model the interplay between technical and ethical progress, which is paramount for guiding agentic systems (Floridi & Nobre, 2024).

- **Practice:** The practical implications are vast, requiring significant adaptation. This includes revising educational pedagogy to focus on skills like critical thinking and AI literacy, training professionals in human-agent teaming (HAT) (Seeber et al., 2020), implementing rigorous quality assurance for AI outputs, and prioritizing ethical design and bias mitigation. The shift in practice is from using AI as a tool to leveraging it as a cognitive partner; this partnership is evolving into one where humans provide strategic oversight and ethical judgment for increasingly autonomous AI agents (Shneiderman, 2022).

This integrated view underscores that the future of AI is not merely a technical challenge but a complex interplay of technological advancement, ethical considerations, and societal adaptation.

8. Limitation of this Critical Review Study

This study is subject to several limitations. Primarily, as a review, its conclusions are drawn from the synthesis of existing literature, research findings, and conference proceedings rather than from original empirical research. This reliance on secondary sources means the study's scope and depth are constrained by the quality and recency of available publications. Furthermore, the focus is specifically on ChatGPT and similar large language models (LLMs), which may not fully represent the nuances or challenges associated with other classes of AI models. A significant constraint is the dynamic nature of the field; the insights and identified gaps, while current at the time of writing, are susceptible to becoming quickly outdated as AI capabilities evolve at an unprecedented pace. Finally, the review acknowledges the inherent difficulty in objectively assessing "deep understanding," as distinguishing between genuine semantic comprehension and sophisticated statistical pattern matching in LLMs remains a fundamental research challenge.

9. Conclusions

The journey through ChatGPT's applications reveals a consistent theme: while these tools can automate, augment, and accelerate many tasks, they are not panaceas. In NLU, the tension between generalization and specialization persists. In content generation, the pursuit of quality, scalability, and ethics forms a complex trilemma. For knowledge discovery, the "black box" nature of LLM reasoning poses a conundrum for scientific rigor. In education, the "pedagogical adaptation imperative" calls for a fundamental rethinking of teaching and learning. Similarly, in engineering, the concept of "human-LLM cognitive symbiosis" suggests a future of human-AI collaboration. All these domain-specific challenges converge and intensify with the advent of Agentic AI, which embodies the next frontier of capability, complexity, and risk.

The contributions of this critical review lie in framing these observations within broader conceptual challenges. Identifying the "Specialization vs. Generalization Tension," the "Quality-Scalability-Ethics Trilemma," the "Black Box Conundrum," the "Pedagogical Adaptation Imperative," and the evolution of "Human-LLM Cognitive Symbiosis" into "Human-Agent Orchestration" provides a structured way to understand the trajectory of ChatGPT and similar other generative AI tools. These frameworks underscore that advancing method and theory is not merely a technical

pursuit but one deeply intertwined with the ethical and practical challenges of steering increasingly autonomous systems.

Harnessing the benefits of AI requires an unwavering commitment to ethical development, deployment, and continuous critical evaluation. This involves addressing biases, ensuring transparency, safeguarding privacy, and proactively considering societal impacts, especially for autonomous agents. The path forward is not one of unbridled technological determinism but of thoughtful, human-centered innovation (Shneiderman, 2022). As AI systems continue their rapid evolution toward greater autonomy, the global community must engage in ongoing interdisciplinary dialogue. The research agenda proposed herein offers starting points for such endeavors. By embracing critical innovation and a profound sense of responsibility, it is possible to navigate the complexities of the AI revolution, mitigating its risks while leveraging its immense potential to advance knowledge, enhance human capabilities, and contribute to global well-being. The journey is complex, but with diligent inquiry and ethical stewardship, the expanding horizons of agentic AI can indeed lead to a more informed, efficient, and equitable future.

References

1. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

2. Adiyono, A., Al Matari, A. S., & Dalimarta, F. F. (2025). Analysis of Student Perceptions of the Use of ChatGPT as a Learning Media: A Case Study in Higher Education in the Era of AI-Based Education. Journal of Education and Teaching (JET), 6(2), 306-324.

3. Akhtarshenas, A., Dini, A., & Ayoobi, N. (2025). ChatGPT or A Silent Everywhere Helper: A Survey of Large Language Models. arXiv preprint arXiv:2503.17403.

4. Al Naqbi, H., Bahroun, Z., & Ahmed, V. (2024). Enhancing work productivity through generative artificial intelligence: A comprehensive literature review. Sustainability, 16(3), 1166.

5. Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. Discover Education, 3(1), 60.

6. Alghazo, R., Fatima, G., Malik, M., Abdelhamid, S. E., Jahanzaib, M., & Raza, A. (2025). Exploring ChatGPT's Role in Higher Education: Perspectives from Pakistani University Students on Academic Integrity and Ethical Challenges. Education Sciences, 15(2).

7. Arslan, B., Lehman, B., Tenison, C., Sparks, J. R., López, A. A., Gu, L., & Zapata-Rivera, D. (2024). Opportunities and challenges of using generative AI to personalize educational assessment. Frontiers in Artificial Intelligence, 7, 1460651.

8. Arvidsson, S., & Axell, J. (2023). Prompt engineering guidelines for LLMs in Requirements Engineering.

9. Atchley, P., Pannell, H., Wofford, K., Hopkins, M., & Atchley, R. A. (2024). Human and AI collaboration in the higher education environment: Opportunities and concerns. Cognitive Research: Principles and Implications, 9(1), 20. https://doi.org/10.1186/s41235-024-00547-9

10. Ballardini, R. M., He, K., & Roos, T. (2019). AI-generated content: authorship and inventorship in the age of artificial intelligence. In Online Distribution of Content in the EU (pp. 117-135). Edward Elgar Publishing.

11. Baroni, M. (2020). Linguistic generalization and compositionality in modern artificial neural networks. Philosophical Transactions of the Royal Society B, 375(1791), 20190307.

12. Belzner, L., Gabor, T., & Wirsing, M. (2023, October). Large language model assisted software engineering: prospects, challenges, and a case study. In International Conference on Bridging the Gap between AI and Reality (pp. 355-374). Cham: Springer Nature Switzerland.

13. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).

14. Bhagavatula, C., Bras, R. L., Malaviya, C., Sakaguchi, K., Holtzman, A., Rashkin, H., ... & Choi, Y. (2019). Abductive commonsense reasoning. arXiv preprint arXiv:1908.05739.

15. Boiko, D. A., MacKnight, R., & Gomes, G. (2023). Emergent autonomous scientific research capabilities of large language models. arXiv. https://doi.org/10.48550/arXiv.2304.05332

16. Bran, A., Cox, S. R., & Schilter, P. (2024). ChemCrow: Augmenting large-language models with a tool-set for chemistry. arXiv. https://doi.org/10.48550/arXiv.2304.05376

17. Chan, R. Y., Sharma, S., & Bista, K. (Eds.). (2024). ChatGPT and Global Higher Education: Using Artificial Intelligence in Teaching and Learning. STAR Scholars Press.

18. Chang, X., Dai, G., Di, H., & Ye, H. (2025). Breaking the Prompt Wall (I): A Real-World Case Study of Attacking ChatGPT via Lightweight Prompt Injection. arXiv preprint arXiv:2504.16125.

19. Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2025). Unleashing the potential of prompt engineering for large language models. Patterns.

20. Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228-239.

21. Craig, C. J. (2022). The AI-copyright challenge: Tech-neutrality, authorship, and the public interest. In Research handbook on intellectual property and artificial intelligence (pp. 134-155). Edward Elgar Publishing.

22. Dagdelen, J., Dunn, A., Lee, S., Walker, N., Rosen, A. S., Ceder, G., ... & Jain, A. (2024). Structured information extraction from scientific text with large language models. Nature Communications, 15(1), 1418.

23. Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y. S., Gašević, D., & Chen, G. (2023, July). Can large language models provide feedback to students? A case study on ChatGPT. In 2023 IEEE international conference on advanced learning technologies (ICALT) (pp. 323-325). IEEE.

24. Daun, M., & Brings, J. (2023, June). How ChatGPT will change software engineering education. In Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1 (pp. 110-116).

25. Davar, N. F., Dewan, M. A. A., & Zhang, X. (2025). AI chatbots in education: challenges and opportunities. Information, 16(3), 235.

26. Dempere, J., Modugu, K., Hesham, A., & Ramasamy, L. K. (2023, September). The impact of ChatGPT on higher education. In Frontiers in Education (Vol. 8, p. 1206936). Frontiers Media SA.

27. Dimeli, M., & Kostas, A. (2025). The Role of ChatGPT in Education: Applications, Challenges: Insights From a Systematic Review. Journal of Information Technology Education: Research, 24, 2.

28. Dovesi, D., Malandri, L., Mercorio, F., & Mezzanzanica, M. (2024). A survey on explainable AI for Big Data. Journal of Big Data, 11(1), 6. https://doi.org/10.1186/s40537-023-00863-9

29. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). Opinion Paper:"So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International journal of information management, 71, 102642.

30. Ege, D. N., Øvrebø, H. H., Stubberud, V., Berg, M. F., Elverum, C., Steinert, M., & Vestad, H. (2025). ChatGPT as an inventor: Eliciting the strengths and weaknesses of current large language models against humans in engineering design. AI EDAM, 39, e6.

31. Elkatmis, M. (2024). ChatGPT and Creative Writing: Experiences of Master's Students in Enhancing. International Journal of Contemporary Educational Research, 11(3), 321-336.

32. Empirical Methods in Natural Language Processing. (2024). The 2024 Conference on Empirical Methods in Natural Language Processing. https://aclanthology.org/events/emnlp-2024/

33. Eymann, V., Lachmann, T., & Czernochowski, D. (2025). When ChatGPT Writes Your Research Proposal: Scientific Creativity in the Age of Generative AI. Journal of Intelligence, 13(5), 55.

34. Fill, H. G., Fettke, P., & Köpke, J. (2023). Conceptual modeling and large language models: impressions from first experiments with ChatGPT. Enterprise Modelling and Information Systems Architectures (EMISAJ), 18, 1-15.

35. Fisher, J. (2025, May). ChatGPT for Legal Marketing: 6 Ways to Unlock the Power of AI. AI-CASEpeer. https://www.casepeer.com/blog/chatgpt-for-legal-marketing/

36. Floridi, L., & Nobre, C. (2024). Artificial intelligence, and the new challenges of anticipatory governance. Ethics and Information Technology, 26(2), 24. https://doi.org/10.1007/s10676-024-09761-y

37. Gabashvili, I. S. (2023). The impact and applications of ChatGPT: a systematic review of literature reviews. arXiv preprint arXiv:2305.18086.

38. Gamage, K. A., Dehideniya, S. C., Xu, Z., & Tang, X. (2023). ChatGPT and higher education assessments: More opportunities than concerns?. Journal of Applied Learning and Teaching, 6(2), 358-369.

39. Gao, R., Yu, D., Gao, B., Hua, H., Hui, Z., Gao, J., & Yin, C. (2025). Legal regulation of AI-assisted academic writing: challenges, frameworks, and pathways. Frontiers in Artificial Intelligence, 8, 1546064.

40. Gao, X., Zhang, Z., Xie, M., Liu, T., & Fu, Y. (2025). Graph of AI Ideas: Leveraging Knowledge Graphs and LLMs for AI Research Idea Generation. arXiv preprint arXiv:2503.08549.

41. Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., & Wang, H. (2024). Retrieval-augmented generation for large language models: A survey. ACM Computing Surveys, 57(5), 1–46. https://doi.org/10.1145/3667794

42. Garbuio, M., & Lin, N. (2021). Innovative idea generation in problem finding: Abductive reasoning, cognitive impediments, and the promise of artificial intelligence. Journal of Product Innovation Management, 38(6), 701-725.

43. Garg, R. K., Urs, V. L., Agarwal, A. A., Chaudhary, S. K., Paliwal, V., & Kar, S. K. (2023). Exploring the role of ChatGPT in patient care (diagnosis and treatment) and medical research: A systematic review. Health Promotion Perspectives, 13(3), 183.

44. Garousi, V. (2025). Why you shouldn't fully trust ChatGPT: A synthesis of this AI tool's error rates across disciplines and the software engineering lifecycle. arXiv preprint arXiv:2504.18858.

45. Glickman, M., & Zhang, Y. (2024). AI and generative AI for research discovery and summarization. arXiv preprint arXiv:2401.06795.

46. Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., ... & Poon, H. (2021). Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare (HEALTH), 3(1), 1-23.

47. Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., ... & Mirjalili, S. (2023). Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. Authorea Preprints, 1, 1-26.

48. Hagendorff, T. (2024). A virtue ethics-based framework for the corporate ethics of AI. AI and Ethics, 4(2), 653–666. https://doi.org/10.1007/s43681-023-00327-0

49. Haltufderheide, J., & Ranisch, R. (2024). ChatGPT and the future of academic publishing: A perspective. The American Journal of Bioethics, 24(2), 4-11.

50. Haman, M., & Školník, M. (2024). Using ChatGPT for scientific literature review: a case study. IASL, 1(1), 1-13.

51. Hannigan, T. R., McCarthy, I. P., & Spicer, A. (2024). Beware of botshit: How to manage the epistemic risks of generative chatbots. Business Horizons, 67(5), 471-486.

52. Hariri, W. (2023). Unlocking the potential of ChatGPT: A comprehensive exploration of its applications, advantages, limitations, and future directions in natural language processing. arXiv preprint arXiv:2304.02017.

53. Herman, S. (2025, February). The art of the prompt: AI prompts for every creative. Adobe. https://www.adobe.com/creativecloud/ai/discover/ai-prompts.html

54. Hodge Jr, S. D. (2023). Revolutionizing Justice: Unleashing the Power of Artificial Intelligence. SMU Sci. & Tech. L. Rev., 26, 217.

55. Hu, Y., & Lu, Y. (2024). Rag and rau: A survey on retrieval-augmented language model in natural language processing. arXiv preprint arXiv:2404.19543.

56. Huang, J., & Chang, K. C. C. (2022). Towards reasoning in large language models: A survey. arXiv preprint arXiv:2212.10403.

57. Hupkes, D., Dankers, V., Mul, M., & Bruni, E. (2020). Compositionality decomposed: How do neural networks generalise?. Journal of Artificial Intelligence Research, 67, 757-795.

58. Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT in facilitating the process of literature review. Available at SSRN 4404768.

59. Infosys Limited. (2023). A perspective on ChatGPT, Its Impact and Limitations. https://www.infosys.com/techcompass/documents/perspective-chatgpt-impact-limitations.pdf

60. Isiaku, L., Muhammad, A. S., Kefas, H. I., & Ukaegbu, F. C. (2024). Enhancing technological sustainability in academia: leveraging ChatGPT for teaching, learning and evaluation. Quality Education for All, 1(1), 385-416.

61. Jiang, C., Huang, R., & Shen, T. (2025). Generative AI-Enabled Conceptualization: Charting ChatGPT's Impacts on Sustainable Service Design Thinking With Network-Based Cognitive Maps. Journal of Computing and Information Science in Engineering, 25(2).

62. Jiang, Y., Hao, J., Fauss, M., & Li, C. (2024). Detecting ChatGPT-generated essays in a large-scale writing assessment: Is there a bias against non-native English speakers?. Computers & Education, 217, 105070.

63. Johnson, S., & Acemoglu, D. (2023). Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity. Hachette UK.

64. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, 103, 102274.

65. Katzir, R. (2023). Why large language models are poor theories of human linguistic cognition. A reply to Piantadosi (2023). Manuscript. Tel Aviv University. url: https://lingbuzz. net/lingbuzz/007190.

66. Kaushik, A., Yadav, S., Browne, A., Lillis, D., Williams, D., Donnell, J. M., ... & Arora, M. (2025). Exploring the Impact of Generative Artificial Intelligence in Education: A Thematic Analysis. arXiv preprint arXiv:2501.10134.

67. Keysers, D., Schärli, N., Scales, N., Buisman, H., Furrer, D., Kashubin, S., ... & Bousquet, O. (2019). Measuring compositional generalization: A comprehensive method on realistic data. arXiv preprint arXiv:1912.09713.

68. Kim, T., Yun, T. S., & Suh, H. S. (2025). Can ChatGPT implement finite element models for geotechnical engineering applications?. International Journal for Numerical and Analytical Methods in Geomechanics.

69. Knoth, C. H., Kieslich, K., Fraumann, G., & Gfrereis, P. (2024). Opportunities and limits of using large language models in evidence synthesis: a descriptive case study. Systematic Reviews, 13(1), 1-13.

70. Lake, B. M. (2019). Compositional generalization through meta sequence-to-sequence learning. Advances in neural information processing systems, 32.

71. Lake, B., & Baroni, M. (2018, July). Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. In International conference on machine learning (pp. 2873-2882). PMLR.

72. Lake, B. M., & Baroni, M. (2023). Human-like systematic generalization through a meta-learning neural network. Nature, 623(7985), 115–121. https://doi.org/10.1038/s41586-023-06668-3

73.   Lee, H. (2024). The rise of ChatGPT: Exploring its potential in medical education. Anatomical sciences education, 17(5), 926-931.

74.   Levitt, G., & Grubaugh, S. (2023). Artificial intelligence and the paradigm shift: Reshaping education to equip students for future careers. The International Journal of Social Sciences and Humanities Invention, 10(6), 7931-7941. https://doi.org/10.18535/ijsshi/v10i06.02

75.   Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems, 33, 9459-9474.

76.   Li, M. (2025). The impact of ChatGPT on teaching and learning in higher education: challenges, opportunities, and future scope. Encyclopedia of Information Science and Technology, Sixth Edition, 1-20.

77.   Li, R., Liang, P., Wang, Y., Cai, Y., Sun, W., & Li, Z. (2025). Unveiling the Role of ChatGPT in Software Development: Insights from Developer-ChatGPT Interactions on GitHub. arXiv preprint arXiv:2505.03901.

78.   Liu, Y., Deng, G., Xu, Z., Li, Y., Zheng, Y., Zhang, Y., ... & Liu, Y. (2023). Jailbreaking chatgpt via prompt engineering: An empirical study. arXiv preprint arXiv:2305.13860.

79.   Liu, Y., Kong, W., & Merve, K. (2025). ChatGPT applications in academic writing: a review of potential, limitations, and ethical challenges. Arquivos Brasileiros de Oftalmologia, 88(3), e2024-0269.

80.   Liu, Y., Yao, Y., Ton, J. F., Zhang, X., Guo, R., Cheng, H., ... & Li, H. (2023). Trustworthy llms: a survey and guideline for evaluating large language models' alignment. arXiv preprint arXiv:2308.05374. https://arxiv.org/abs/2308.05374

81.   Magnani, L., & Arfini, S. (2024). Model-based abductive cognition: What thought experiments teach us. Logic Journal of the IGPL, jzae096.

82.   Marques, N., Silva, R. R., & Bernardino, J. (2024). Using chatgpt in software requirements engineering: A comprehensive review. Future Internet, 16(6), 180.

83.   Marvin, G., Hellen, N., Jjingo, D., & Nakatumba-Nabende, J. (2023, June). Prompt engineering in large language models. In International conference on data intelligence and cognitive informatics (pp. 387-402). Singapore: Springer Nature Singapore.

84.   Mavrepis, P., Makridis, G., Fatouros, G., Koukos, V., Separdani, M. M., & Kyriazis, D. (2024). XAI for all: Can large language models simplify explainable AI?. arXiv preprint arXiv:2401.13110.

85.   Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies. U.S. Department of Education.

86.   Melamed, R., McCabe, L. H., Wakhare, T., Kim, Y., Huang, H. H., & Boix-Adsera, E. (2023). Prompts have evil twins. arXiv preprint arXiv:2311.07064.

87.   Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. Education Sciences, 13(9), 856.

88.   Miller, D. (2023). Exploring the impact of artificial intelligence language model ChatGPT on the user experience. International Journal of Technology Innovation and Management (IJTIM), 3(1), 1-8.

89.   Mitra, M., de Vos, M. G., Cortinovis, N., & Ometto, D. (2024, September). Generative AI for Research Data Processing: Lessons Learnt From Three Use Cases. In 2024 IEEE 20th International Conference on e-Science (e-Science) (pp. 1-10). IEEE.

90.   Mohammed, A. (2025, March). Navigating the AI revolution: Safeguarding academic integrity and ethical considerations in the age of innovation. BERA. https://www.bera.ac.uk/blog/navigating-the-ai-revolution-safeguarding-academic-integrity-and-ethical-considerations-in-the-age-of-innovation

91. Molenaar, I. (2024). Human-AI co-regulation: A new focal point for the science of learning. Npj Science of Learning, 9(1), 29. https://doi.org/10.1038/s41539-024-00222-3

92. Mostafapour, M., Asoodar, M., & Asoodar, M. (2024). Advantages and disadvantages of using ChatGPT for academic literature review. Cogent Engineering, 11(1), 2315147.

93. Mozes, M. A. J. (2024). Understanding and Guarding against Natural Language Adversarial Examples (Doctoral dissertation, UCL (University College London)).

94. Mozes, M., He, X., Kleinberg, B., & Griffin, L. D. (2023). Use of llms for illicit purposes: Threats, prevention measures, and vulnerabilities. arXiv preprint arXiv:2308.12833.

95. Murray, M., Maclachlan, R., Flockhart, G. M., Adams, R., Magueijo, V., Goodfellow, M., ... & Lauro, V. (2025). A 'snapshot'of engineering practitioners views of ChatGPT-informing pedagogy in higher education. European Journal of Engineering Education, 1-26.

96. Naseem, U., Dunn, A. G., Khushi, M., & Kim, J. (2022). Benchmarking for biomedical natural language processing tasks with a domain specific ALBERT. BMC bioinformatics, 23(1), 144.

97. Naveed, J. (2025). Optimized Code Generation in BIM with Retrieval-Augmented LLMs.

98. Neveditsin, N., Lingras, P., & Mago, V. (2025). Clinical insights: A comprehensive review of language models in medicine. PLOS Digital Health, 4(5), e0000800.

99. Nguyen, M. N., Nguyen Thanh, B., Vo, D. T. H., Pham Thi Thu, T., Thai, H., & Ha Xuan, S. (2023). Evaluating the Efficacy of Generative Artificial Intelligence in Grading: Insights from Authentic Assessments in Economics. Available at SSRN 4648790.

100. Niloy, A. C., Akter, S., Sultana, N., Sultana, J., & Rahman, S. I. U. (2024). Is Chatgpt a menace for creative writing ability? An experiment. Journal of computer assisted learning, 40(2), 919-930.

101. Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187-192.

102. OpenAI. (2023a). GPT-3.5 Turbo. https://openai.com/index/gpt-3-5-turbo-fine-tuning-and-api-updates/

103. OpenAI. (2023b). GPT-4 Technical Report. https://openai.com/index/gpt-4-research/

104. OpenAI. (2023c). Safety & alignment. https://openai.com/safety/

105. OpenAI. (2024a). ChatGPT FAQ. https://help.openai.com/en/collections/3742473-chatgpt

106. OpenAI. (2024c). Hello GPT-4o. https://openai.com/index/hello-gpt-4o/

107. OpenAI. (2024d). Introducing o1: Our next step in AI research. https://openai.com/o1/

108. OpenAI. (2024e). o1-mini: Our best performing model on AIME. https://openai.com/o1/

109. OpenAI. (2024f). o1-preview: Advanced reasoning in STEM. https://openai.com/index/introducing-openai-o1-preview/

110. OpenAI Help Center. (n.d.). What is the ChatGPT model selector? Retrieved June 11, 2025, from https://help.openai.com/en/articles/7864572-what-is-the-chatgpt-model-selector

111. Oremus, W. (2023). The clever trick that turns ChatGPT into its evil twin. The Washington Post. URL https://www.washingtonpost.com/technology/2023/02/14/chatgpt-dan-jailbreak/.

112. Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying large language models and knowledge graphs: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14(1), e1518. https://doi.org/10.1002/widm.1518

113. Pareschi, R. (2023). Abductive reasoning with the GPT-4 language model: Case studies from criminal investigation, medical practice, scientific research. Sistemi intelligenti, 35(2), 435-444.

114. Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3586183.3606763

115. Perlman, A. (2023). The implications of ChatGPT for legal services and society. Mich. Tech. L. Rev., 30, 1.

116. Perez, E., Huang, S., Song, F., Cai, T., Ring, R., Aslanides, J., ... & Irving, G. (2022). Red teaming language models with language models. arXiv preprint arXiv:2202.03286.

117. Perez, F., & Ribeiro, I. (2022). Ignore previous prompt: Attack techniques for language models. arXiv preprint arXiv:2211.09527.

118. Preiksaitis, C., & Rose, C. (2023). Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. JMIR medical education, 9, e48785.

119. Prinsloo, P. (2020). Data frontiers and frontiers of power in higher education: A view of the an/archaeology of data. Teaching in Higher Education, 25(4), 394–412. https://doi.org/10.1080/13562517.2020.1758223

120. Puthumanaillam, G., & Ornik, M. (2025). The Lazy Student's Dream: ChatGPT Passing an Engineering Course on Its Own. arXiv preprint arXiv:2503.05760.

121. Rawat, A. S., Fazzini, M., George, T., Gokulan, R., Maddila, C., & Arrieta, A. (2024). A new era of software development: A survey on the impact of large language models. ACM Computing Surveys, 57(5), 1–40. https://doi.org/10.1145/3667793

122. Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems, 3, 121-154.

123. RedBlink. (2025, April). Llama 4 vs ChatGPT: Comprehensive AI Models Comparison 2025. https://redblink.com/llama-4-vs-chatgpt/

124. Reich, J. (2020). Failure to disrupt: Why technology alone can't transform education. Harvard University Press.

125. Rice, S., Crouse, S. R., Winter, S. R., & Rice, C. (2024). The advantages and limitations of using ChatGPT to enhance technological research. Technology in Society, 76, 102426.

126. Sabzalieva, E., & Valentini, A. (2023). ChatGPT and artificial intelligence in higher education: Quick start guide.

127. Salesforce. (2025, June 23). AI agents in education: Benefits & use cases. Salesforce. https://www.salesforce.com/education/artificial-intelligence/ai-agents-in-education/

128. Sallam, M. (2023, March). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In Healthcare (Vol. 11, No. 6, p. 887). MDPI.

129. Sapkota, R., Raza, S., & Karkee, M. (2025). Comprehensive analysis of transparency and accessibility of chatgpt, deepseek, and other sota large language models. arXiv preprint arXiv:2502.18505.

130. Schiller, C. A. (2024). The human factor in detecting errors of large language models: A systematic literature review and future research directions. arXiv preprint arXiv:2403.09743.

131. Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elbanna, A., ... & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. Information & Management, 57(2), 103174. https://doi.org/10.1016/j.im.2019.103174

132. Shah, N., Jain, S., Lauth, J., Mou, Y., Bartsch, M., Wang, Y., ... & Luo, Y. (2023). Can large language models reason about medical conversation? arXiv preprint arXiv:2305.00412.

133. Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. arXiv. https://doi.org/10.48550/arXiv.2303.11366

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

135. Shormani, M. Q. (2024). Non-native speakers of English or ChatGPT: Who thinks better? arXiv preprint arXiv:2412.00457.

136. Sivarajkumar, S., Kelley, M., Samolyk-Mazzanti, A., Visweswaran, S., & Wang, Y. (2024). An empirical evaluation of prompting strategies for large language models in zero-shot clinical natural language processing: algorithm development and validation study. JMIR Medical Informatics, 12, e55318.

137. Surden, H. (2023). ChatGPT, AI large language models, and law. Fordham L. Rev., 92, 1941.

138. Susnjak, T., & McIntosh, J. (2024). Academic integrity in the age of ChatGPT. Change: The Magazine of Higher Learning, 56(1), 21-27.

139. Susskind, R., & Susskind, D. (2022). The future of the professions: How technology will transform the work of human experts. Oxford University Press.

140. Team, A. (2024). The agentic design pattern: A new paradigm for building AI systems. Andreessen Horowitz. https://a16z.com/the-agentic-design-pattern-a-new-paradigm-for-building-ai-systems/

141. Thelwall, M. (2024). Evaluating research quality with large language models: an analysis of ChatGPT's effectiveness with different settings and inputs. Journal of Data and Information Science, 241218-241218.

142. Topcu, T. G., Husain, M., Ofsa, M., & Wach, P. (2025). Trust at Your Own Peril: A Mixed Methods Exploration of the Ability of Large Language Models to Generate Expert-Like Systems Engineering Artifacts and a Characterization of Failure Modes. Systems Engineering.

143. U.S. Department of Education, Office of Educational Technology. (2023). Artificial intelligence and the future of teaching and learning: Insights and recommendations. https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf

144. UNESCO. (2023). Guidance for generative AI in education and research. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

145. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems 30 (pp. 5998–6008). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

146. Veisi, O., Bahrami, S., Englert, R., & Müller, C. (2025). AI Ethics and Social Norms: Exploring ChatGPT's Capabilities From What to How. arXiv preprint arXiv:2504.18044.

147. Velásquez-Henao, J. D., Franco-Cardona, C. J., & Cadavid-Higuita, L. (2023). Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering. Dyna, 90(SPE230), 9-17.

148. Vu, N. G. H., Wang, K., & Wang, G. G. (2025). Effective prompting with ChatGPT for problem formulation in engineering optimization. Engineering Optimization, 1-18.

149. Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., Fan, L., & Anandkumar, A. (2023). Voyager: An open-ended embodied agent with large language models. arXiv. https://doi.org/10.48550/arXiv.2305.16291

150. Waseem, F., Al-Ghamdi, D., Al-Ghamdi, A., & Ahmad, I. (2023). Unlocking the potential of ChatGPT in requirements engineering: a study of benefits and challenges. Arabian Journal for Science and Engineering, 1-15.

151. Weidlich, J., & Gašević, D. (2025). ChatGPT in education: An effect in search of a cause. PsyArXiv Preprints.

152. Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Laura, L., & Gabriel, I. (2024). An overarching risk analysis and management framework for frontier AI. arXiv. https://doi.org/10.48550/arXiv.2405.02111

153. Wiedemer, T., Mayilvahanan, P., Bethge, M., & Brendel, W. (2023). Compositional generalization from first principles. Advances in Neural Information Processing Systems, 36, 6941-6960.

154. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2019). Huggingface's transformers: State-of-the-art natural language processing. arXiv preprint arXiv:1910.03771. https://arxiv.org/abs/1910.03771

155. Wu, T., He, S., Liu, J., Sun, Y., Liu, K., Han, T. X., ... & Zhao, J. (2024). A brief overview of the dark side of AI: The case of ChatGPT. Communications of the ACM.

156. Wu, Y., Zhao, Y., Hu, B., Minervini, P., Stenetorp, P., & Riedel, S. (2022). An efficient memory-augmented transformer for knowledge-intensive nlp tasks. arXiv preprint arXiv:2210.16773.

157. Xi, Z., Chen, W., Guo, X., He, H., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., & Wang, R. (2023). The rise and potential of large language model based agents: A survey. arXiv. https://doi.org/10.48550/arXiv.2309.07864

158. Xue, J., Zheng, M., Hua, T., Shen, Y., Liu, Y., Bölöni, L., & Lou, Q. (2023). Trojllm: A black-box trojan prompt attack on large language models. Advances in Neural Information Processing Systems, 36, 65665-65677.

159. Yadav, S., Qureshi, A. M., Kaushik, A., Sharma, S., Loughran, R., Kazhuparambil, S., ... & Lillis, D. (2025). From Idea to Implementation: Evaluating the Influence of Large Language Models in Software Development--An Opinion Paper. arXiv preprint arXiv:2503.07450.

160. Yang, X., Chen, A., PourNejatian, N., Shin, H. C., Smith, K. E., Parisien, C., ... & Wu, Y. (2022). Gatortron: A large clinical language model to unlock patient information from unstructured electronic health records. arXiv preprint arXiv:2203.03540.

161. Yu, W. (2022, July). Retrieval-augmented generation across heterogeneous knowledge. In Proceedings of the 2022 conference of the North American chapter of the association for computational linguistics: human language technologies: student research workshop (pp. 52-58).

162. Yu, Y., Kim, S., Lee, W., & Koo, B. (2025). Evaluating ChatGPT on Korea's BIM Expertise Exam and improving its performance through RAG. Journal of Computational Design and Engineering, 12(4), 94-120.

163. Zakir, M. H., Bashir, S., Nisar, K., Ibrahim, S., Khan, N., & Khan, S. H. (2024). Navigating the Legal Labyrinth: Establishing Copyright Frameworks for AI-Generated Content. Remittances Review, 9(1), 2515-2532.

164. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? International Journal of Educational Technology in Higher Education, 16(1), 39. https://doi.org/10.1186/s41239-019-0171-0

165. Zhai, X. (2023). ChatGPT for next generation science learning. Available at SSRN 4331313.

166. Zhao, H., Chen, H., Yang, F., Liu, N., Deng, H., Cai, H., ... & Du, M. (2024). Explainability for large language models: A survey. ACM Transactions on Intelligent Systems and Technology, 15(2), 1-38.

167. Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., & Ba, J. (2022, November). Large language models are human-level prompt engineers. In The Eleventh International Conference on Learning Representations.

168. Zhu, D., Chen, J., Shen, X., Li, X., & Elhoseiny, M. (2024). A survey on multimodal large language models. arXiv. https://doi.org/10.48550/arXiv.2311.13282

169. Zhu, S., Wang, Z., Zhuang, Y., Jiang, Y., Guo, M., Zhang, X., & Gao, Z. (2024). Exploring the impact of ChatGPT on art creation and collaboration: Benefits, challenges and ethical implications. Telematics and Informatics Reports, 14, 100138.