technologies

MDPI

Review

From Google Gemini to OpenAI Q* (Q-Star): A Survey on Reshaping the Generative Artificial Intelligence (AI) Research Landscape

Timothy R. McIntosh [1,2,*], Teo Susnjak [3], Tong Liu [3], Paul Watters [4], Dan Xu [5], Dongwei Liu [6] and Malka N. Halgamuge [1]

[1] Accounting, Information Systems & Supply Chain, RMIT University, Melbourne, VIC 3000, Australia
[2] Cyberoo Pty Ltd., 81-83 Campbell Street, Surry Hills, Sydney, NSW 2010, Australia
[3] School of Mathematical and Computational Sciences, Massey University, Auckland 0632, New Zealand
[4] Cyberstronomy Pty Ltd., Melbourne, VIC 3000, Australia
[5] Australia and New Zealand Banking Group Limited, 833 Collins St., Melbourne, VIC 3008, Australia
[6] Coles Group, 800 Toorak Rd., Hawthorn East, Melbourne, VIC 3123, Australia
* Correspondence: timothy.mcintosh@rmit.edu.au

Abstract: This comprehensive survey explored the evolving landscape of generative Artificial Intelligence (AI), with a specific focus on the recent technological breakthroughs and the gathering advancements toward possible Artificial General Intelligence (AGI). It critically examined the current state and future trajectory of generative AI, exploring how innovations in developing actionable and multimodal AI agents with the ability scale their "thinking" in solving complex reasoning tasks are reshaping research priorities and applications across various domains, while the survey also offers an impact analysis on the generative AI research taxonomy. This work has assessed the computational challenges, scalability, and real-world implications of these technologies while highlighting their potential in driving significant progress in fields like healthcare, finance, and education. Our study also addressed the emerging academic challenges posed by the proliferation of both AI-themed and AI-generated preprints, examining their impact on the peer-review process and scholarly communication. The study highlighted the importance of incorporating ethical and human-centric methods in AI development, ensuring alignment with societal norms and welfare, and outlined a strategy for future AI research that focuses on a balanced and conscientious use of generative AI as its capabilities continue to scale.

Keywords: AI ethics; artificial general intelligence (AGI); artificial intelligence (AI); Gemini; generative AI; mixture of experts (MoE); multimodality; Q* (Q-star); test-time compute; agentic AI; research impact analysis

Academic Editor: Valeri Mladenov

Received: 28 October 2024
Revised: 23 January 2025
Accepted: 23 January 2025
Published: 30 January 2025

Citation: McIntosh, T.R.; Susnjak, T.; Liu, T.; Watters, P.; Xu, D.; Liu, D.; Halgamuge, M.N. From Google Gemini to OpenAI Q* (Q-Star): A Survey on Reshaping the Generative Artificial Intelligence (AI) Research Landscape. *Technologies* **2025**, *13*, 51. https://doi.org/10.3390/technologies13020051

1. Introduction

Artificial Intelligence (AI) has advanced dramatically since Alan Turing's "Imitation Game" [1], which first posed questions about machine cognition. Early computational theories [2,3] and the arrival of neural networks and machine learning [4–6] laid the groundwork for the deep learning era. These historical milestones set the stage for the next wave of progress, culminating in the creation of breakthrough Large Language Models (LLMs) such as OpenAI's ChatGPT [7], Google's Gemini [8] and Anthropic's Claude [9], which have spurred fresh debates on AI consciousness and the consequences of advanced automation [10–12]. The most recent models extend their earlier capabilities beyond text-based interactions by handling inputs like images and audio while also demonstrating

reasoning abilities that encompass multistep tasks, advanced mathematics, and code generation, therein repeatedly breaking benchmark performances and requiring more challenging ones to be devised [13].

Building on these strides, the field has seen further developments in both model architectures and operational strategies. Agentic AI equips LLMs with the ability to perform sophisticated tasks by leveraging external tools and APIs without constant human oversight [14]. Meanwhile, test-time compute or inference-time scaling increases efficiency by dynamically allocating computational resources at inferences based on the complexity of the query, as illustrated by OpenAI's o-series models [7]. Prompting strategies, including Chain-of-Thought [15] and subsequent extensions like Tree-of-Thought [16] and Forest-of-Thought [17], provide technological foundations that enable the newest LLMs to follow a structured sequence of reasoning steps. These techniques support decision making via the generation long internal reasoning chains from which suitable solutions can be sampled and derived step by step, enabling a sophisticated multistage analysis of the possible solution space [18].

In parallel, Mixture of Experts (MoE) architectures have attracted attention for their capacity to distribute different computational tasks among specialized submodels [19]. This concept operates alongside the push toward omni-modal intelligence, where AI is expected to seamlessly process and integrate information across multiple sensory modalities. These approaches reflect a broader focus on versatile and adaptive systems that can handle a wide range of applications. The recently announced 'thinking models' released by OpenAI and Google showcase previously unseen model capabilities in planning and reasoning-intensive tasks [20] that potentially point toward a future where AI systems may approximate general intelligence [21]. It is these particular and most consequential advancements that have produced models capable of deliberation, which have been linked to OpenAI's technological breakthroughs speculated to be termed initially Q* and subsequently Strawberry [22].

Against this backdrop, our study investigates the transformative effect of modern generative AI innovations, including MoE architectures, Agentic AI, omni-modal frameworks and elastic compute strategies during inference. In particular, we focus on the most notable capabilities that the recent LLMs have acquired, namely, complex reasoning and signs of planning, which we will refer to as *Q*-capabilities* in the remainder of the manuscript. We examine how these developments are shaping research priorities, practical applications, and societal interactions, as well as the ethical, societal, and academic questions they raise. By presenting key insights and lessons learned, our findings aim to offer a roadmap for navigating this fast-evolving technological frontier.

1.1. Changing AI Research Popularity

As the field of LLMs continues to evolve, exemplified by recent breakthroughs, a multitude of studies have surfaced with the aim of charting future research paths, which have varied from identifying emerging trends to highlighting areas poised for swift progress. The dichotomy of established methods and early adoption is evident, with "hot topics" in LLM research increasingly shifting toward multimodal and omni-modal capabilities, as well as conversation-driven learning. The propagation of preprints has expedited knowledge sharing, but it also brings the risk of reduced academic scrutiny. Issues like inherent biases, noted by Retraction Watch, along with concerns about the misuse of generative AI, plagiarism, and forgery, present substantial hurdles [23–25]. The academic world, therefore, stands at an intersection, necessitating a unified drive to refine research directions in light of the fast-paced evolution of the field, which appears to be partly traced through the changing popularity of various research keywords over time. The release of commercial and powerful open-source generative models has been influential. As depicted in Figure 1,

the rise and fall of certain keywords appear to have correlated with significant industry milestones, such as the release of the "Transformer" model in 2017 [26], the GPT model in 2018 [27], and the commercial ChatGPT-3.5 in December 2022. For instance, the spike in searches related to "Deep Learning" coincides with the breakthroughs in neural network applications, while the interest in "Natural Language Processing" surges as LLMs increasingly redefine what is possible in language understanding and generation. The enduring attention to "Ethics/Ethical" in AI research, despite some fluctuations, reflects the continuous and deep-rooted concern for the moral dimensions of AI, underscoring that ethical considerations are not merely a reactionary measure, but an integral and persistent dialogue within the AI discussion [28].

It is academically intriguing to postulate whether these trends signify a causal relationship, where technological advancements drive research focus, or if the burgeoning research itself propels technological development. This paper also explores the profound societal and economic impacts of AI advancements. We examine how AI technologies are reshaping various industries, altering employment landscapes, and influencing socio-economic structures. This analysis highlights both the opportunities and challenges posed by AI in the modern world, emphasizing its role in driving innovation and economic growth, while also considering the ethical implications and potential for societal disruption. Future studies could yield more definitive insights, yet the synchronous interplay between innovation and academic curiosity remains a hallmark of AI's progress.

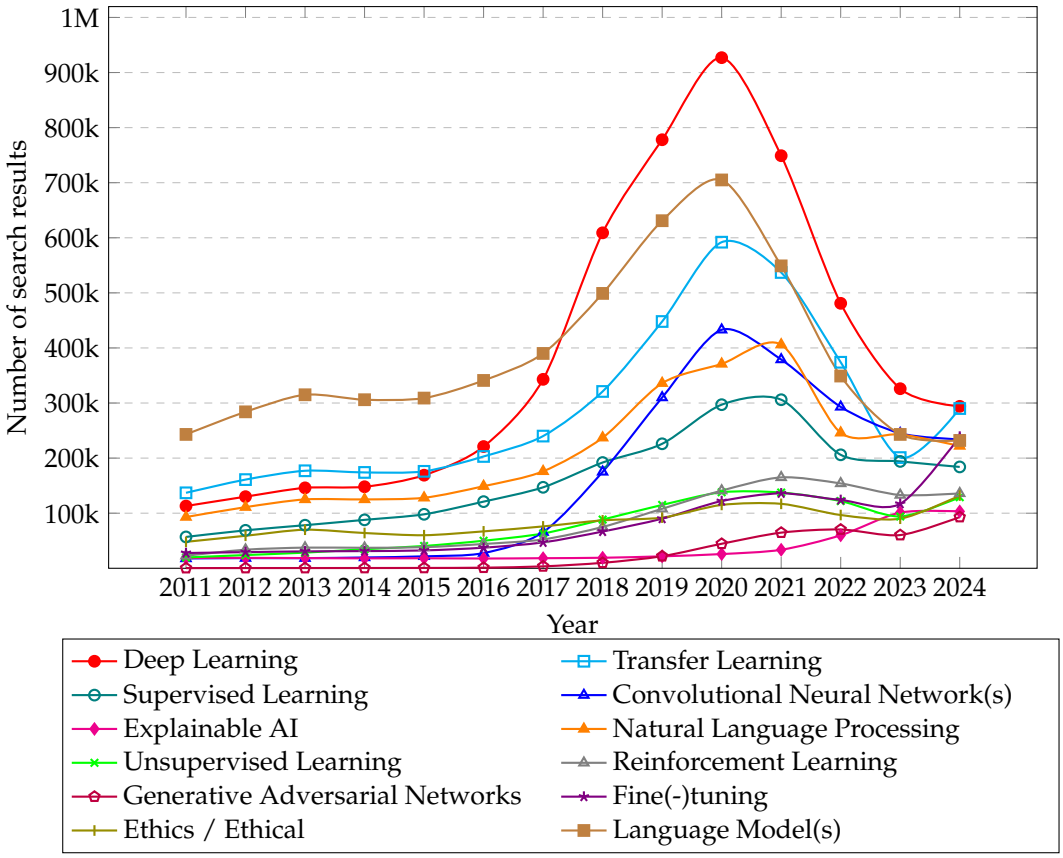

Figure 1. Number of search results on Google Scholar with different keywords by year (as collected in January 2025). (The legend entries correspond to the keywords used in the search query, which was constructed as "(AI OR artificial OR (machine learning) OR (neural network) OR computer OR software) AND ([specific keyword])").

Meanwhile, the exponential increase in the number of preprints posted on arXiv under the Computer Science > Artificial Intelligence (cs.AI) category, as illustrated in Figure 2, appears to signify a paradigm shift in research dissemination within the AI community. While the rapid distribution of findings enables swift knowledge exchange, it also raises concerns regarding the validation of information. The surge in preprints may lead to the propagation of unvalidated or biased information, as these studies do not undergo the rigorous scrutiny and potential retraction typical of peer-reviewed publications [29,30]. This trend underlines the need for careful consideration and critique in the academic community, especially given the potential for such unvetted studies to be cited and their findings propagated.

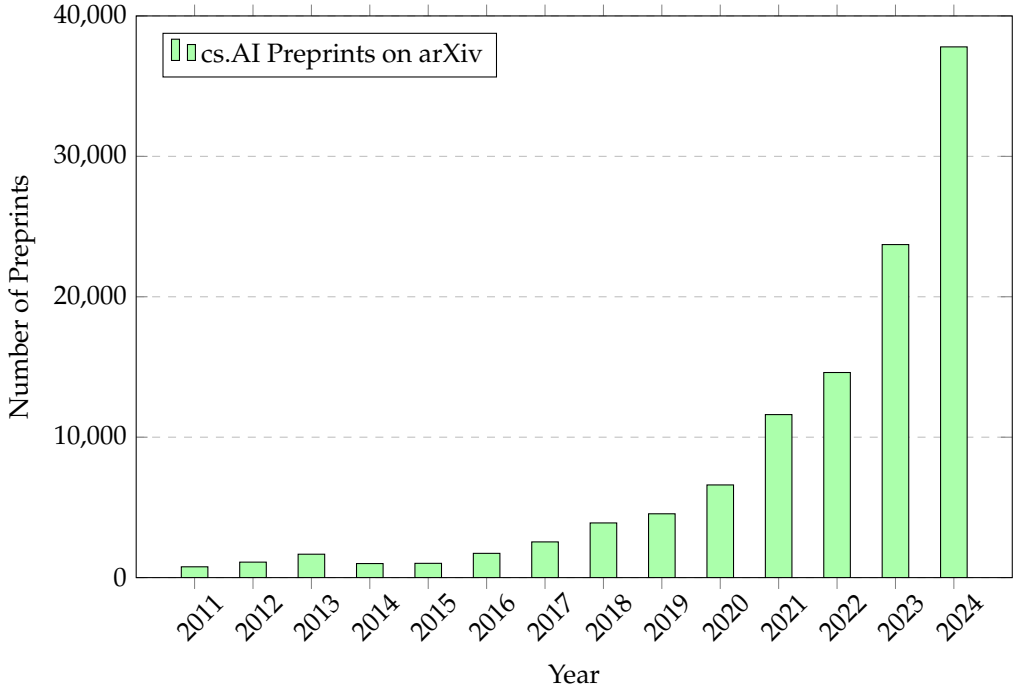

Figure 2. Annual number of preprints posted under the cs.AI category on arXiv.org.

1.2. Objectives

The impetus for this investigation has been the technological leaps forward demonstrated by the frontier LLMs both from OpenAI and Google, which have demonstrated the ability to "think" for extended durations before they respond. These models exhibit reasoning patterns akin to human problem solving by iteratively refining their problem-solving strategies, correcting errors, simplifying challenging steps, and adapting their approaches when current methods are insufficient. Notably, they exhibit advanced capabilities in complex reasoning, demonstrating their ability to decompose intricate problems into manageable subtasks and at least, emulate planning behaviors that were previously thought to be beyond the capabilities of existing LLM architectures [31,32]. Consequently, these noteworthy achievements call for an investigation into the current trends in generative AI research, which is critical for appraising the potential for obsolescence or insignificance in extant research themes while concurrently delving into burgeoning prospects within the rapidly transforming LLM panorama. This inquiry is reminiscent of the obsolete nature of encryption-centric or file-entropy-based ransomware detection methodologies, which have been eclipsed by the transition of ransomware collectives toward data theft strategies utilizing varied attack vectors, relegating contemporary studies on crypto-ransomware to the status of latecomers [33,34]. Advances in AI are anticipated to not only enhance the complex reasoning capabilities of LLMs and facilitate their integration into decision-making

systems and IT infrastructure but also in the areas of omni-modality and robotics. LLM advancements to date have already heralded the obsolescence of conventional, statistics-driven natural language processing techniques in many domains [35], and others are have also been disrupted. Nonetheless, the perennial imperative for AI to align with human ethics and values persists as a fundamental tenet [36–38], and the conjectural Q-Star initiative offers an unprecedented opportunity to instigate discourse on how such advancements might reconfigure the LLM research topography. Our research methodology involved a structured literature search using key terms like 'Large Language Models' and 'Generative AI'. To that end, this paper examines the potential impact of emerging technologies on research paths in AI, highlighting their role in opening new areas of exploration. It identifies several emerging research domains—MoE, multimodality, Agentic AI, adaptive models, and Artificial General Intelligence (AGI)—as pivotal in transforming generative AI research. The study presents a survey-style analysis to outline a research roadmap, focusing on both current and emerging trends in the field.

1.3. Survey Methodology and Scope

We use the term "survey" in the context of a "narrative literature review" or "scoping review" rather than a sociological study involving questionnaires. Our objective is to map current progress in generative AI, identify emerging research directions, and assess their relevance to academic and industrial domains.

Database Selection and Search Strategy: To capture the multidisciplinary nature of LLMs and generative AI, we consulted six major digital libraries commonly used in computer science and AI-related fields: IEEE Xplore, Scopus, ACM Digital Library, ScienceDirect, Web of Science, and ProQuest Central. We designed our queries by combining keywords such as "Generative AI", "Large Language Models", "Mixture of Experts", "Gemini", "Q-Star", and "Artificial General Intelligence". Given the rapid advancements since the publication of the *Transformer* model in 2017, our search primarily focused on literature from 2017 to 2024.

Inclusion and Exclusion Criteria: We screened articles based on the following criteria:

- Relevance to Generative AI Topics: We included works that significantly discuss LLMs, MoE architectures, multimodal learning, or AGI-oriented methods.
- Peer-Reviewed or Preprint: Priority was given to journal and conference publications; however, we also considered selective preprints (e.g., arXiv) if they substantially contributed new findings.
- Domain Focus: Papers centered on purely sociological, philosophical, or economic aspects without explicit generative AI components were excluded.

While we did not employ a strict systematic review protocol (e.g., PRISMA), our multi-database approach helped ensure breadth and diversity in the sources included. Following initial screening, duplicates and clearly out-of-scope items were removed.

Quality Assessment and Synthesis: A pool of ∼600 candidate papers emerged from the initial queries. Through iterative reading and relevance checking, the final set was narrowed to 362 key references. In synthesizing the literature, we did the following:

1. We organized papers by thematic clusters (e.g., Mixture-of-Experts Advances, Multimodal Extensions, or AGI Research Outlook).
2. We cross-compared conflicting or convergent results to highlight consensus or outstanding debates.
3. We derived a conceptual taxonomy to illustrate how new developments (e.g., Llama 2, Q-Star, etc.) potentially reshape the research landscape.

The resulting analysis formed the basis of our proposed classification framework, discussions regarding emerging directions, and reflection on ethical and technical challenges.

Limitations of Our Review: Our methodology did not involve a formal systematic review process nor explicit grading of evidence quality (e.g., GRADE). Additionally, with the volume of AI-related research expanding rapidly, certain recent arXiv papers might not be included if they appeared too late in our review cycle. Nevertheless, the review offers a comprehensive cross-section of key trends, drawing from prominent works and capturing the major research trajectories in generative AI.

The major contributions of this study are as follows:

1. Detailed examination of the evolving landscape in generative AI, emphasizing the advancements and innovations in their complex reasoning capabilities and their wide-ranging implications within the AI domain.
2. Analysis of the transformative effect of advanced generative AI systems on academic research, exploring how these developments are altering research methodologies, setting new trends, and potentially leading to the obsolescence of traditional approaches.
3. Thorough assessment of the ethical, societal, and technical challenges arising from the integration of generative AI in academia, underscoring the crucial need for aligning these technologies with ethical norms, ensuring data privacy, and developing comprehensive governance frameworks.

The rest of this paper is organized as follows: Section 2 explores the historical development of generative AI. Section 3 presents a taxonomy of current generative AI research. Section 4 explores the MoE model architecture, its innovative features, and its impact on transformer-based language models. Section 5 discusses the enabling underlying architectures and reported capabilities of the latest frontier models, which embody those of the speculated Q* (In this paper, the term Q* serves as a symbolic representation that encompasses the recognized complex reasoning and early-stage planning advancements seen in frontier LLMs.) project. Section 6 discusses the projected capabilities of AGI. Section 7 examines the impact of recent advancements on the generative AI research taxonomy. Section 8 identifies emerging research priorities in generative AI. Section 10 discusses the academic challenges of the rapid surge of preprints in AI. The paper concludes in Section 11, summarizing the overall effects of these developments in generative AI.

2. Background: Evolution of Generative AI

The ascent of generative AI has been marked by significant milestones, with each new model paving the way for the next evolutionary leap. From single-purpose algorithms to LLMs like OpenAI's ChatGPT and the latest multimodal systems, the AI landscape has been transformed, while countless other fields have been disrupted.

2.1. The Evolution of Language Models

Language models have undergone a transformative journey (Figure 3), evolving from rudimentary statistical methods to the complex neural network architectures that underpin today's LLMs [39,40]. This evolution has been driven by a relentless quest for models that more accurately reflect the nuances of human language, as well as the desire to push the boundaries of what machines can understand and generate [39–41]. However, this rapid advancement has not been without its challenges. As language models have grown in capability, so too have the ethical and safety concerns surrounding their use, prompting a reevaluation of how these models are developed and the purposes for which they are employed [39,42–45].

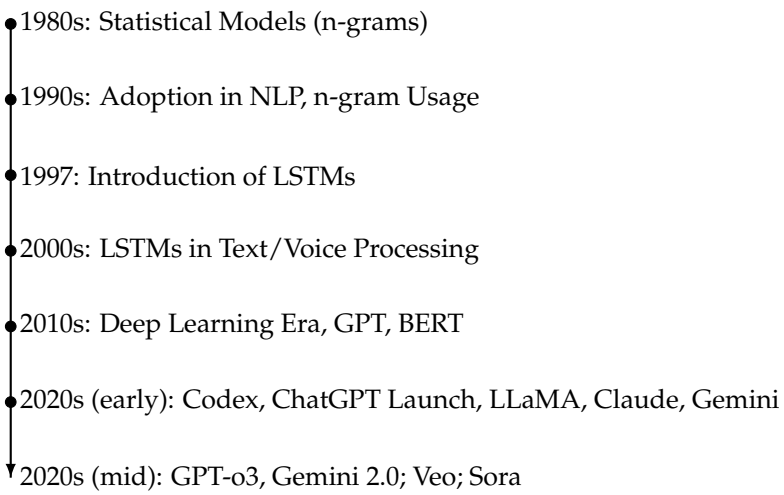

Figure 3. Timeline of key developments in language model evolution.

2.1.1. Language Models as Precursors

The inception of language modeling can be traced to the statistical approaches of the late 1980s, which was a period marked by a transition from rule-based to machine learning algorithms in natural language processing (NLP) [46–50]. Early models, primarily *n*-gram based, calculated the probability of word sequences in a corpus, thus providing a rudimentary understanding of language structure [46]. Those models, simplistic yet groundbreaking, laid the groundwork for future advances in language understanding. With the increase in computational power, the late 1980s witnessed a revolution in NLP, pivoting toward statistical models capable of 'soft' probabilistic decisions, as opposed to the rigid, 'handwritten' rule-based systems that dominated early NLP systems [48]. IBM's development of complicated statistical models throughout this period signified the growing importance and success of these approaches. In the subsequent decade, the popularity and applicability of statistical models surged, proving invaluable in managing the flourishing flow of digital text. The 1990s saw statistical methods firmly established in NLP research, with n-grams becoming instrumental in numerically capturing linguistic patterns. The introduction of Long Short-Term Memory (LSTM) networks in 1997 [51] and their application to voice and text processing a decade later [52–54] marked significant milestones, leading to the current era where neural network models represent the cutting edge of NLP research and development.

2.1.2. Large Language Models: Technical Advancement and Commercial Success

The advent of deep learning has revolutionized the field of NLP, leading to the development of early iterations of LLMs such as GPT and BERT, which laid the foundation for subsequent advancements. Codex, introduced by OpenAI [55], marked a significant step forward by fine-tuning language models for code generation, showcasing the versatility and domain-specific capabilities of LLMs. Notably, conversational agents like ChatGPT have further popularized the applications of these models. Recent innovations, including GPT-4o and the open-source LLaMA 3 family, have pushed the boundaries by integrating sophisticated techniques such as transformer architectures and advanced natural language understanding, illustrating the rapid evolution in this field [40]. These models represent a significant leap in NLP capabilities, leveraging vast computational resources and extensive datasets to achieve new heights in language understanding and generation [40,56]. ChatGPT has shown impressive conversational skills and contextual understanding with a broad spectrum of functional uses in many areas, as evidenced by its technical and commercial success, including rapid adoption by over 100 million users

shortly after launch, which emphasizes a robust market demand for natural language AI and has catalyzed interdisciplinary research into its applications in sectors like education, healthcare, and commerce [35,56–59]. In education, ChatGPT offers innovative approaches to personalized learning and interactive teaching [57,60–62], while in commerce, it has revolutionized customer service and content creation [63,64]. The widespread use of ChatGPT, Google Bard/Gemini, Anthropic Claude, and similar commercial LLMs has reignited important debates in the field of AI, particularly concerning AI consciousness and safety, as its human-like interaction capabilities raise significant ethical questions and highlight the need for robust governance and safety measures in AI development [12,65–67]. Such influence appears to extend beyond its technical achievements, shaping cultural and societal discussions about the role and future of AI in our world.

The advancements in LLMs, including the development of models like GPT and BERT, have paved the way for the conceptualization of Q*-capabilities centered around complex reasoning and early-stage planning capabilities. Specifically, the scalable architecture, advanced reinforcement learning regimens, and test-time compute flexibilities that characterize these models are foundational to the newly acquired capabilities of frontier models like the GPT o-series and Gemini 2.0 models. Similarly, the emergence of multimodal systems capable of integrating text, images, audio, and video reflects an evolutionary path that that is being opened toward omni-modal models, combining the versatility of LLMs for more holistic and potentially fully sensory and embodied AI solutions that are able for the first time to acquire real-world models that current models lack [68] (Table 1).

Table 1. Comparison of selected proprietary vs. open-source LLMs.

Model	Type	Developer	Key Features/Remarks
GPT-4o	Proprietary	OpenAI	State-of-the-art performance but limited transparency and restricted model access.
Gemini	Proprietary	Google	Multimodal focus, Agentic AI-capabilities; currently closed-source with advanced MMLU scores.
Claude	Proprietary	Anthropic	Emphasizes AI safety and interpretability; closed-source but widely used in industry.
GPT-NeoX [69]	Open-Source	EleutherAI	Community-driven LLM with openly released weights and configurations. Encourages reproducibility.
Llama 3 [70]	Open-Source (restricted)	Meta	Large-scale model with partial open license, fosters academic innovation with direct model access.
DeepSeek [21]	Open-Source	DeepSeek-AI	Large-scale MoE model, with multimodal versions DeepSeek-VL and DeepSeek-VL2, which integrate vision and language understanding.

2.1.3. Fine-Tuning, Hallucination Reduction, and Alignment in LLMs

The advancement of LLMs has underlined the significance of fine-tuning [71–74], hallucination reduction [75–78], and alignment [79–83]. These aspects are crucial in enhancing the functionality and reliability of LLMs. Fine-tuning, which involves adapting pretrained models to specific tasks, has seen significant progress: techniques like prompt-based and few-shot learning [84–87], alongside supervised fine-tuning on specialized datasets [71,88–90], have enhanced the adaptability of LLMs in various contexts, but challenges remain, particularly in bias mitigation and the generalization of models across diverse tasks [71,83,91]. Hallucination reduction is a persistent challenge in LLMs, which is characterized by the generation of confident but factually incorrect information [39]. Strategies such as confidence penalty regularization during fine-tuning have been implemented to mitigate overconfidence and improve accuracy [92–94]. Despite these efforts, the complexity of human language and the breadth of topics make completely eradicating hallucinations

a daunting task, especially in culturally sensitive contexts [10,39]. Alignment, ensuring LLM outputs are congruent with human values and ethics, is an area of ongoing research. Innovative approaches, from constrained optimization [95–99] to different types of reward modeling [100–103], aim to embed human preferences within AI systems. While advancements in fine-tuning, hallucination reduction, and alignment have propelled LLMs forward, these areas still present considerable challenges. The complexity of aligning AI with the diverse spectrum of human ethics and the persistence of hallucinations, particularly on culturally sensitive topics, highlight the need for continued interdisciplinary research in the development and application of LLMs [10,104].

2.1.4. Mixture of Experts: A Paradigm Shift

The adoption of the MoE architecture in LLMs marks a critical evolution in AI technology. This innovative approach, exemplified by advanced models like Google's Switch Transformer (https://huggingface.co/google/switch-c-2048; accessed on 10 January 2025) and MistralAI s Mixtral-8x7B (https://huggingface.co/mistralai/Mixtral-8x7B-v0.1; accessed on 10 January 2025), leverages multiple transformer-based expert modules for dynamic token routing, enhancing modeling efficiency and scalability. The primary advantage of the MoE lies in its ability to handle vast parameter scales, reducing memory footprint and computational costs significantly [105–113]. This is achieved through model parallelism across specialized experts, allowing for the training of models with trillions of parameters, and its specialization in handling diverse data distributions enhances its capability in few-shot learning and other complex tasks [106,107,114]. To illustrate the practicality of the MoE, consider its application in healthcare. For example, an MoE-based system could be used for personalized medicine, where different 'expert' modules specialize in various aspects of patient data analysis, including genomics, medical imaging, and electronic health records. This approach could significantly enhance diagnostic accuracy and treatment personalization. Similarly, in finance, MoE models can be deployed for risk assessment, where experts analyze distinct financial indicators, market trends, and regulatory compliance factors.

Despite its benefits, the MoE confronts challenges in dynamic routing complexity [115–118], expert imbalance [119–122], and probability dilution [123], and such technical hurdles demand sophisticated solutions to fully harness the MoE's potential. Moreover, while the MoE may offer performance gains, it does not inherently solve ethical alignment issues in AI [19,124,125]. The complexity and specialization of MoE models can obscure the decision-making processes, complicating efforts to ensure ethical compliance and alignment with human values [124,126]. Although the paradigm shift to MoE signifies a major leap in LLM development, offering significant scalability and specialization advantages while ensuring the safety, ethical alignment, and transparency of these models remains a paramount concern. The MoE architecture, while technologically advanced, entails continued interdisciplinary research and governance to align AI with broader societal values and ethical standards.

2.2. Multimodal AI and the Future of Interaction

The advent of multimodal AI marks a transformative era in AI development, revolutionizing how machines interpret and interact with a diverse array of human sensory inputs and contextual data.

2.3. Advancing Multimodal AI: A Comparative Perspective of Frontier LLMs

The emergence of advanced multimodal AI models represents a transformative shift in artificial intelligence, enabling systems to process and integrate information across diverse modalities such as text, images, audio, and video [127–129]. Among the most notable contri-

butions to this field are Google DeepMind's Gemini 2.0, OpenAI's GPT-4o, and Anthropic's Claude 3.5 Sonnet. While these models share the overarching ambition of advancing multimodal AI, they embody distinct architectural strategies, feature sets, and optimization goals, reflecting diverse priorities in their development. Google DeepMind's Gemini 2.0 is built on a sophisticated multimodal encoder–decoder framework [8,130,131], enabling seamless integration of text, images, audio, and video. This design enhances the model's ability to perform comprehensive, context-aware reasoning, making it highly effective for real-world, dynamic interactions. Similarly, OpenAI's GPT-4o employs an optimized multimodal architecture to prioritize real-time performance in response time for audio inputs, which facilitates latency-critical applications, and is mirrored by Anthropic's Claude 3.5 Sonnet.

Agentic capabilities further complement and differentiate these models by extending their potential beyond passive interaction into active task execution. Gemini 2.0 emphasises agentic AI [132], with capabilities to autonomously plan, execute, and adapt actions in complex, real-world scenarios. This feature aligns Gemini 2.0 with applications requiring autonomy and proactive decision making. Claude 3.5 Sonnet similarly incorporates agentic functionalities, enabling it to interact directly with computer systems to automate tasks such as opening applications, managing workflows, and executing commands. In contrast, while GPT-4o does not explicitly highlight agentic features, its multimodal reasoning and rapid processing make it a versatile solution for real-time adaptations, where user-driven applications featuring interactivity and efficiency are paramount.

The architectural priorities and optimization goals of these models highlight their contrasting strengths and use cases. Gemini 2.0 is designed as a comprehensive solution for multimodal integration, excelling in dynamic, context-aware scenarios that demand advanced reasoning and adaptability. GPT-4o, with its focus on cost efficiency and rapid response times, caters to industries requiring scalable, real-time multimodal solutions. Meanwhile, Claude 3.5 Sonnet, with its Vision API and agentic task automation capabilities, targets applications demanding high-speed multimodal processing and operational precision. The collective advancements represented by these models mark a significant milestone in the development of multimodal AI [8,133] and pave the way for future universal or omni-modal models that are able to scale to a full complement of sensory inputs.

2.3.1. Gemini 2.0: Advancing Multimodal AI Capabilities

Gemini 2.0 [132], an advanced multimodal AI model developed by Google DeepMind, represents a significant leap in AI technology by enhancing the integration and processing of diverse data types, including text, images, audio, and video. This progression builds upon its predecessor's foundation, introducing new features and improvements that set it apart in the realm of AI models.

The architecture of Gemini 2.0 incorporates a sophisticated multimodal encoder–decoder framework [8,130,131], enabling seamless understanding and generation across various modalities [8,127–129]. Key advancements include the following:

- Enhanced Multimodal Understanding: Gemini 2.0 exhibits improved capabilities in processing and reasoning across multiple modalities, allowing for more comprehensive and context-aware interactions.
- Native Image and Audio Generation: The model introduces native image generation and controllable text-to-speech functionalities, facilitating the creation of rich, multimodal content directly from textual inputs.
- Real-Time Multimodal Interactions: With the integration of the Multimodal Live API, Gemini 2.0 supports real-time audio and video streaming inputs, enabling dynamic and interactive AI experiences.

- Agentic Capabilities: Gemini 2.0 is designed for the agentic era, with the ability to plan, execute, and adapt actions autonomously, enhancing its utility in complex, real-world tasks.

These enhancements position Gemini 2.0 as a versatile and powerful AI model, capable of performing complex, multistep reasoning and providing more natural and efficient user interactions. Its deployment across various Google products and services highlights its significance in advancing AI applications and setting new benchmarks in the field [8,133].

2.3.2. Gemini: Redefining Benchmarks in Multimodality

Gemini, a pioneering multimodal conversational system, marks a significant shift in AI technology by surpassing traditional text-based LLMs like GPT-3 and even its multimodal counterpart, ChatGPT-4. Gemini's architecture has been designed to incorporate the processing of diverse data types such as text, images, audio, and video, a feat facilitated by its unique multimodal encoder, cross-modal attention network, and multimodal decoder [8,127–129]. The architectural core of Gemini is its dual-encoder structure, with separate encoders for visual and textual data, enabling sophisticated multimodal contextualization [8,130,131]. This architecture is believed to surpass the capabilities of single-encoder systems, allowing Gemini to associate textual concepts with image regions and achieve a compositional understanding of scenes [8]. Furthermore, Gemini integrates structured knowledge and employs specialized training paradigms for cross-modal intelligence, setting new benchmarks in AI [8,133]. In [8], Google has claimed and demonstrated that Gemini distinguishes itself from ChatGPT-4 through several key features:

- Breadth of Modalities: Unlike ChatGPT-4, which primarily focuses on text, documents, images, and code, Gemini handles a wider range of modalities, including audio and video. This extensive range allows Gemini to tackle complex tasks and understand real-world contexts more effectively.
- Performance: Gemini Ultra excels in key multimodality benchmarks, notably in massive multitask language understanding (MMLU) which encompasses a diverse array of domains like science, law, and medicine, outperforming ChatGPT-4.
- Scalability and Accessibility: Gemini is available in three tailored versions—Ultra, Pro, and Nano—catering to a range of applications from data centers to on-device tasks and featuring a level of flexibility not yet seen in ChatGPT-4.
- Code Generation: Gemini's proficiency in understanding and generating code across various programming languages is more advanced, offering practical applications beyond ChatGPT-4's capabilities.
- Transparency and Explainability: A focus on explainability sets Gemini apart, as it provides justifications for its outputs, enhancing user trust and understanding of the AI's reasoning process.

Despite these advancements, Gemini's real-world performance in complex reasoning tasks that require the integration of commonsense knowledge across modalities remains to be thoroughly evaluated.

2.3.3. Technical Challenges in Multimodal Systems

The development of multimodal AI systems faces several technical hurdles, including creating robust and diverse datasets, managing scalability, and enhancing user trust and system interpretability [134–136]. Challenges like data skew and bias are prevalent due to data acquisition and annotation issues, which require effective dataset management by employing strategies such as data augmentation, active learning, and transfer learning [91,134,136,137]. A significant challenge is the computational demands of processing various data streams simultaneously, requiring powerful hardware and optimized model

architectures for multiple encoders [138,139]. Advanced algorithms and multimodal attention mechanisms are needed to balance attention across different input media and resolve conflicts between modalities, especially when they provide contradictory information [139–141]. Scalability issues, due to the extensive computational resources needed, are exacerbated by limited high-performance hardware availability [142,143]. There is also a pressing need for calibrated multimodal encoders for compositional scene understanding and data integration [141]. Refining evaluation metrics for these systems is necessary to accurately assess performance in real-world tasks, calling for comprehensive datasets and unified benchmarks to enhance user trust and system interpretability through explainable AI in multimodal contexts. Addressing these challenges is vital for the advancement of multimodal AI systems, enabling seamless and intelligent interaction aligned with human expectations.

2.3.4. Multimodal AI: Beyond Text in Ethical and Social Contexts

The expansion of multimodal AI systems introduces both benefits and complex ethical and social challenges that extend beyond those faced by text-based AI. In commerce, multimodal AI can transform customer engagement by integrating visual, textual, and auditory data [144–146]. For autonomous vehicles, multimodality can enhance safety and navigation by synthesizing data from various sensors, including visual, radar, and Light Detection and Ranging (LIDAR) [146–148]. Still, DeepFake technology's ability to generate convincingly realistic videos, audio, and images is a critical concern in multimodality, as it poses risks of misinformation and manipulation that significantly impact public opinion, political landscapes, and personal reputations, thereby compromising the authenticity of digital media and raising issues in social engineering and digital forensics where distinguishing genuine from AI-generated content becomes increasingly challenging [149,150]. Privacy concerns are amplified in multimodal AI due to its ability to process and correlate diverse data sources, potentially leading to intrusive surveillance and profiling, which raises questions about the consent and rights of individuals, especially when personal media are used without permission for AI training or content creation [134,151,152]. Moreover, multimodal AI can propagate and amplify biases and stereotypes across different modalities, and if unchecked, this can perpetuate discrimination and social inequities, making it imperative to address algorithmic bias effectively [153–155]. The ethical development of multimodal AI systems requires robust governance frameworks focusing on transparency, consent, data handling protocols, and public awareness, where ethical guidelines must evolve to address the unique challenges posed by these technologies, including setting standards for data usage and safeguarding against the nonconsensual exploitation of personal information [156,157]. Additionally, the development of AI literacy programs will be crucial in helping society understand and responsibly interact with multimodal AI technologies [134,156]. As the field progresses, interdisciplinary collaboration will be key in ensuring these systems are developed and deployed in a manner that aligns with societal values and ethical principles [134].

2.3.5. Technical Challenges and Limitations of Current LLMs

While LLMs have demonstrated exceptional capabilities, their transformer-based architectural design and operational paradigms posses significant limitations when applied to tasks requiring planning [31,68], forming dynamic models of the world [158], and conducting intricate, single-pass reasoning during inference [32]. These challenges, as well as those surrounding their lack of long-term memory and inability to accurately process and retain long-term dependencies [159], expose a gap between their remarkable generative capabilities and the complex demands of real-world decision making necessary for AGI.

A fundamental limitation of LLMs lies in their inability to form persistent, dynamic models of the world. Unlike humans or symbolic AI systems that can represent, simulate, and adapt to changing environments, LLMs rely exclusively on static patterns encoded in their training data [160]. This reliance prevents them from modeling and truly understanding causality, predicting long-term outcomes, or adapting their reasoning to real-world complexities. Additionally, planning, which is a marker of sophisticated intelligence and requires recourse to real-world models, therefore poses a significant challenge for current LLMs [31,68]. Effective planning requires creating and evaluating sequences of actions to achieve specific goals, often under uncertainty and with constraints. LLMs, however, operate as single forward-pass systems, inherently limiting their ability to engage in deliberative, sequential reasoning and causal inferences. Techniques like Chain-of-Thought (CoT) prompting and Tree-of-Thought (ToT) reasoning have enabled LLMs to emulate aspects of planning by breaking down problems into smaller reasoning steps. Yet, these methods rely heavily on prompt engineering and external scaffolding, as well as new approaches to reinforcement learning for automating them, thus lacking intrinsic mechanisms for true deliberative planning. As a result, LLMs generally struggle with tasks that demand strategic foresight, iterative problem-solving, or dynamic scenario evaluation.

Recent advancements, such as test-time compute scaling, have sought to address these challenges by dynamically allocating computational resources during inference. Models like OpenAI's o1 and o3, as well as Gemini's 2.0, exemplify this approach, enhancing task performance by allowing models to enter a "think" mode [132]. This entails generating a large set of competing possible chains of reasoning steps at inference time, which constitute a solution space from which a most suitable solution needs to be searched and found. However, while such a test-time compute scaling technique represents a significant step forward, it is resource-intensive and constrained by the existing architectures of LLMs, which can at best emulate planning and deliberative capabilities via extensive searches over possible solutions.

2.4. Advances in Complex Reasoning and Emerging Trends

The evolution of artificial intelligence continues to push the boundaries of complex reasoning and quasi-planning capabilities, as exemplified by recent advancements in models like OpenAI's o1 and o3 and Google DeepMind's Gemini 2.0. These models represent significant strides in integrating reinforcement learning, Chain-of-Thought (CoT) reasoning, and search-based techniques, addressing critical limitations in traditional LLM architectures. Here, we continue to use the term "Q*" as a symbolic representation to encapsulate the unification of structured reasoning, generative capabilities, and adaptive planning that these advancements aim to achieve.

From AlphaGo's Legacy to the Symbolism of Q*

The transition from AlphaGo's mastery of Go to the advanced reasoning capabilities symbolized by Q*-like techniques reflects a fundamental shift in AI research. AlphaGo showcased the potential of reinforcement learning coupled with Monte Carlo tree search to achieve superhuman performance in rule-based environments [161,162]. In contrast, today's frontier models are extending these principles further to domains requiring complex, context-aware reasoning by leveraging reinforcement learning not just for optimization within bounded environments but also for broader, unstructured problem-solving tasks. In addition, memory augmentation techniques are being developed for LLMs, which are something they have lacked. Behrouz et al. [159] recently introduced a neural long-term memory module that augments transformer-based systems to enable them to integrate extensive historical context during inference without the computational overhead typi-

cally associated with large-context attention mechanisms. This innovation represents a significant step toward enhancing the reasoning and planning capabilities of AI systems by addressing limitations in handling long-range dependencies. The integration of these advancements, which includes bridging reinforcement learning and search-based algorithms during test-time compute, such as A*, can be encapsulated as realizations of Q* aspirations. Unlike AlphaGo's focus on predefined rules, Q*-like methodologies aim to operate in open-ended, multimodal contexts and go further by introducing structured reasoning and adaptive planning, enabling AI systems to autonomously optimize decision paths, refine strategies through interaction, and seamlessly blend logical and generative capabilities.

2.5. Speculative Advances and Chronological Trends

In the dynamic landscape of AI, the emerging capabilities of the Q* techniques, blending LLMs, Q learning, and A* (A-Star algorithm), embody a significant leap forward. This section explores the evolutionary trajectory from game-centric AI systems to the broad applications anticipated with the ongoing technical breakthroughs of Q* approaches to mitigate the limitations of existing LLM architectures.

2.5.1. From AlphaGo's Groundtruth to Q-Star's Exploration

The journey from AlphaGo, a game-centric AI, to the Q*-like capabilities of new frontier models, represents a significant paradigm shift in AI. AlphaGo's mastery in the game of Go highlighted the effectiveness of deep learning and tree search algorithms within well-defined rule-based environments, underscoring the potential of AI in complex strategy and decision making [161,162]. Q-Star, however, is speculated to move beyond these confines, aiming to amalgamate the strengths of reinforcement learning (as seen in AlphaGo), with the knowledge, NLG, creativity, and versatility of LLMs, as well as the strategic efficiency of pathfinding algorithms like A*. This blend, merging pathfinding algorithms and LLMs, could enable AI systems to transcend board game confines and, with Q-Star's natural language processing, interact with human language, enabling nuanced interactions and marking a leap toward AI adept in both structured tasks and complex human-like communication and reasoning. Moreover, the incorporation of Q learning and A* algorithms would enable Q-Star to optimize decision paths and learn from its interactions, making it more adaptable and intelligent over time. The combination of these technologies could lead to AI that is not only more efficient in problem solving but also creative and insightful in its approach. This speculative advancement from the game-focused power of AlphaGo to the comprehensive potential of Q-Star illustrates the dynamic and ever-evolving nature of AI research, and it opens up possibilities for AI applications that are more integrated with human life and capable of handling a broader range of tasks with greater autonomy and sophistication.

2.5.2. Bridging Structured Learning with Creativity

The anticipated Q* project, blending Q learning and A* algorithms with the creativity of LLMs, embodies a groundbreaking step in AI, potentially surpassing recent innovations like Gemini. The fusion suggested in Q* points to an integration of structured, goal-oriented learning with generative, creative capabilities, forming a combination that could transcend the existing achievements of Gemini. While Gemini represents a significant leap in multimodal AI, combining various forms of data inputs such as text, images, audio, and video, Q* is speculated to bring a more profound integration of creative reasoning and structured problem solving. This would be achieved by merging the precision and efficiency of algorithms like A* with the learning adaptability of Q learning, as well as the complex understanding of human language and context offered by LLMs. Such an integration could enable AI systems to not only process and analyze complex multimodal data but also

to autonomously navigate through structured tasks while engaging in creative problem solving and knowledge generation, mirroring the multifaceted nature of human cognition. The implications of this potential advancement are vast, suggesting applications that span beyond the capabilities of current multimodal systems like Gemini. By aligning the deterministic aspects of traditional AI algorithms with the creative and generative potential of LLMs, Q* could offer a more holistic approach to AI development. This could bridge the gap between the logical, rule-based processing of AI and the creative, abstract thinking characteristic of human intelligence. The anticipated unveiling of Q*, merging structured learning techniques and creative problem solving in a singular, advanced framework, holds the promise of not only extending but also significantly surpassing the multimodal capabilities of systems like Gemini, thus heralding another game-changing era in the domain of generative AI and showcasing its potential as a crucial development eagerly awaited in the ongoing evolution of AI.

3. The Current Generative AI Research Taxonomy

The field of Generative AI is evolving rapidly, which necessitates a comprehensive taxonomy that encompasses the breadth and depth of research within this domain. Detailed in Table 2, this taxonomy categorizes the key areas of inquiry and innovation in generative AI and serves as a foundational framework to understand the current state of the field, guiding through the complexities of evolving model architectures, advanced training methodologies, diverse application domains, ethical implications, and the frontiers of emerging technologies.

Table 2. Comprehensive taxonomy of current generative AI and LLM research.

Domain	Subdomain	Key Focus	Description
Model Architecture	Transformer Models	Efficiency, Scalability	Optimizing network structures for faster processing and larger datasets.
	Recurrent Neural Networks	Sequence Processing	Handling sequences of data, like text, for improved contextual understanding.
	MoE	Specialization, Efficiency	Leveraging multiple expert modules for enhanced efficiency and task-specific performance.
	Multimodal Models	Sensory Integration	Integrating text, vision, and audio inputs for comprehensive understanding.
Training Techniques	Supervised Learning	Data Labeling, Accuracy	Using labeled datasets to train models for precise predictions.
	Unsupervised Learning	Pattern Discovery	Finding patterns and structures from unlabeled data.
	Reinforcement Learning	Adaptability, Optimization	Training models through feedback mechanisms for optimal decision making.
	Transfer Learning	Versatility, Generalization	Applying knowledge gained in one task to different but related tasks.
Application Domains	Natural Language Understanding	Comprehension, Contextualization	Enhancing the ability to understand and interpret human language in context.
	Natural Language Generation	Creativity, Coherence	Generating coherent and contextually relevant text responses.
	Conversational AI	Interaction, Naturalness	Developing systems for natural and contextually relevant human–computer conversations.
	Creative AI	Innovation, Artistic Generation	Generating creative content, including text, art, and music.

Table 2. *Cont.*

Domain	Subdomain	Key Focus	Description
Compliance and Ethical Considerations	Bias Mitigation	Fairness, Representation	Addressing and reducing biases in AI outputs.
	Data Security	Data Protection, Confidentiality	Ensuring data confidentiality, integrity, and availability security in AI models and outputs.
	AI Ethics	Fairness, Accountability	Addressing ethical issues such as bias, fairness, and accountability in AI systems.
	Privacy Preservation	Privacy Compliance, Anonymization	Protecting data privacy in model training and outputs.
Advanced Learning	Self-supervised Learning	Autonomy, Efficiency	Utilizing unlabeled data for model training, enhancing learning efficiency.
	Meta-learning	Rapid Adaptation	Enabling AI models to quickly adapt to new tasks with minimal data.
	Fine Tuning	Domain-Specific Tuning, Personalization	Adapting models to specific domains or user preferences for enhanced relevance and accuracy.
	Human Value Alignment	Ethical Integration, Societal Alignment	Aligning AI outputs with human ethics and societal norms, ensuring decisions are ethically and socially responsible.
Emerging Trends	Multimodal Learning	Integration with Vision, Audio	Combining language models with other sensory data types for richer understanding, leading to fully omni-modal models.
	Interactive and Cooperative AI	Collaboration, Human–AI Interaction	Enhancing AI's ability to work alongside humans in collaborative tasks.
	AGI Development	Holistic Understanding	Pursuing the development of AI systems with comprehensive, human-like understanding.
	AGI Containment	Safety Protocols, Control Mechanisms	Developing methods to contain and control AGI systems to prevent unintended consequences.
	'Thinking Models'	Test-Time Compute, Inference Efficiency, Advanced Reinforcement Learning	Enabling AI systems to dynamically allocate computational resources during inference, enhancing flexibility and performance in complex reasoning tasks achieved by learned reasoning via reinforcement learning.
	Agentic AI	Autonomy, Task Execution	Empowering AI systems with the ability to autonomously plan, execute various tools, and adapt actions in real-world environments, aligning with human intentions and goals.
	Test Time Optimization	Adaptation, Memory	Adaptation at inference time via model updates and improved memory and ability to handle long dependencies.

3.1. Model Architectures

Generative AI model architectures have seen significant developments, with four key domains standing out:

- **Transformer Models:** Transformer models have significantly revolutionized the field of AI, especially in NLP, due to their higher efficiency and scalability [163–165]. They employ advanced attention mechanisms to achieve enhanced contextual processing, allowing for more subtle understanding and interaction [166–168]. These models have also made notable strides in computer vision, as evidenced by the development of vision transformers like EfficientViT [169,170] and YOLOv8 [171–173]. These inno-

vations symbolize the extended capabilities of transformer models in areas such as object detection, offering not only improved performance but also increased computational efficiency.

- **Recurrent Neural Networks (RNNs):** RNNs excel in the realm of sequence modeling, making them particularly effective for tasks involving language and temporal data, as their architecture is specifically designed to process sequences of data, such as text, enabling them to capture the context and order of the input effectively [174–178]. This proficiency in handling sequential information renders them indispensable in applications that require a deep understanding of the temporal dynamics within data, such as natural language tasks and time series analysis [179,180]. RNNs' ability to maintain a sense of continuity over sequences is a critical asset in the broader field of AI, especially in scenarios where context and historical data play crucial roles [181].
- **MoE:** MoE models can significantly enhance efficiency by deploying model parallelism across multiple specialized expert modules, which enables these models to leverage transformer-based modules for dynamic token routing and to scale to trillions of parameters, thereby reducing both memory footprint and computational costs [106,115,182–185]. MoE models stand out for their ability to divide computational loads among various experts, each specializing in different aspects of the data, which allows for handling vast scales of parameters more effectively, leading to a more efficient and specialized handling of complex tasks [106,186].
- **Multimodal Models:** Multimodal models, which integrate a variety of sensory inputs such as text, vision, and audio, are crucial in achieving a comprehensive understanding of complex datasets, making them particularly transformative in fields like medical imaging [8,134,136,187]. These models facilitate accurate and data-efficient analysis by employing multiview pipelines and cross-attention blocks [188,189]. This integration of diverse sensory inputs allows for a more nuanced and detailed interpretation of data, enhancing the model's ability to accurately analyze and understand various types of information [190]. The combination of different data types, processed concurrently, enables these models to provide a holistic view, making them especially effective in applications that require a deep and multifaceted understanding of complex scenarios [134,190–192].

3.2. Training Techniques

The training of generative AI models leverages four key techniques, each contributing uniquely to the field:

- **Supervised Learning:** Supervised learning, a foundational approach in AI, uses labeled datasets to guide models towards accurate predictions, and it has been integral to various applications, including image recognition and NLP [193–195]. Recent advancements have focused on developing sophisticated loss functions and regularization techniques, aimed at enhancing the performance and generalization capabilities of supervised learning models, ensuring they remain robust and effective across a wide range of tasks and data types [196–198].
- **Unsupervised Learning:** Unsupervised learning is essential in AI for uncovering patterns within unlabeled data, which is a process central to tasks like feature learning and clustering [199,200]. This method has seen significant advancements with the introduction of autoencoders [201,202] and Generative Adversarial Networks (GANs) [203–205], which have notably expanded unsupervised learning's applicability, enabling more sophisticated data generation and representation learning capabilities. Such innovations are crucial for understanding and leveraging the com-

plex structures often inherent in unstructured datasets, highlighting the growing versatility and depth of unsupervised learning techniques.

- **Reinforcement Learning:** Reinforcement learning, characterized by its adaptability and optimization capabilities, has become increasingly vital in decision making and autonomous systems [206,207]. This training technique has undergone significant advancements, particularly with the development of Deep Q-Networks (DQNs) [208–210] and Proximal Policy Optimization (PPO) algorithms [211–213]. These enhancements have been crucial in improving the efficacy and applicability of reinforcement learning, especially in complex and dynamic environments. By optimizing decisions and policies through interactive feedback loops, reinforcement learning has established itself as a crucial tool for training AI systems in scenarios that demand a high degree of adaptability and precision in decision making [214,215].
- **Transfer Learning:** Transfer learning emphasizes versatility and efficiency in AI training, allowing models to apply knowledge acquired from one task to different yet related tasks, which significantly reduces the need for large labeled datasets [216,217]. Transfer learning, through the use of pretrained networks, streamlines the training process by allowing models to be efficiently fine-tuned for specific applications, thereby enhancing adaptability and performance across diverse tasks and proving particularly beneficial in scenarios where acquiring extensive labeled data is impractical or unfeasible [218,219].

3.3. Application Domains

The application domains of Generative AI are remarkably diverse and evolving, encompassing both established and emerging areas of research and application. These domains have been significantly influenced by recent advancements in AI technology and the expanding scope of AI applications.

- **Natural Language Understanding (NLU)**: NLU is central to enhancing the comprehension and contextualization of human language in AI systems, and it involves key capabilities such as semantic analysis, named entity recognition, sentiment analysis, textual entailment, and machine reading comprehension [220–223]. Advances in NLU have been crucial in improving AI's proficiency in interpreting and analyzing language across a spectrum of contexts, ranging from straightforward conversational exchanges to intricate textual data [220,222,223]. NLU is fundamental in applications like sentiment analysis, language translation, information extraction, and more [224,225]. Recent advancements have prominently featured large transformer-based models like BERT and GPT-3, which have significantly advanced the field by enabling a deeper and more complex understanding of language subtleties [226,227].
- **Natural Language Generation (NLG)**: NLG emphasizes the training of models to generate coherent, contextually relevant, and creative text responses, which are critical components in chatbots, virtual assistants, and automated content creation tools [39,228–230]. NLG encompasses challenges such as topic modeling, discourse planning, concept-to-text generation, style transfer, and controllable text generation [39,231]. The recent surge in NLG capabilities, exemplified by advanced models like GPT-3, has significantly enhanced the sophistication and nuance of text generation, which enable AI systems to produce text that closely mirrors human writing styles, thereby broadening the scope and applicability of NLG in various interactive and creative contexts [57,61,232].

- **Conversational AI**: This subdomain is dedicated to developing AI systems capable of smooth, natural, and context-aware human–computer interactions by focusing on dialogue modeling, question answering, user intent recognition, and multiturn context tracking [233–236]. In finance and cybersecurity, AI's predictive analytics have transformed risk assessment and fraud detection, leading to more secure and efficient operations [34,234]. The advancements in this area, demonstrated by large pretrained models like Meena (https://neptune.ai/blog/transformer-nlp-models-meena-lamda-chatbots; accessed on 10 January 2025) and BlenderBot (https://blenderbot.ai; accessed on 10 January 2025), have significantly enhanced the empathetic and responsive capabilities of AI interactions. These systems not only improve user engagement and satisfaction, but also maintain the flow of conversation over multiple turns, providing coherent, contextually relevant, and engaging experiences [237,238].
- **Creative AI**: This emerging subdomain spans across text, art, music, and more, pushing the boundaries of AI's creative and innovative potential across various modalities including images, audio, and video by engaging in the generation of artistic content, encompassing applications in idea generation, storytelling, poetry, music composition, visual arts, and creative writing, which have resulted in commercial successes like MidJourney and DALL-E [239–241], as well as most recently with state-of-the-art video generation models, namely, OpenAI's Sora [242] and Google's Veo 2 [243]. The challenges in this field involve finding suitable data representations, algorithms, and evaluation metrics to effectively assess and foster creativity [241,244]. Creative AI serves not only as a tool for automating and enhancing artistic processes but also as a medium for exploring new forms of artistic expression, enabling the creation of novel and diverse creative outputs [241]. This domain represents a significant leap in AI's capability to engage in and contribute to creative endeavors, redefining the intersection of technology and art, with Sora and Veo 2 representing significant breakthroughs in the field of high-definition video generation that convincingly simulates real-world physics.

3.4. Compliance and Ethical Considerations

As AI technologies rapidly evolve and become more integrated into various sectors, ethical considerations and legal compliance have become increasingly crucial, which require a focus on developing 'Ethical AI Frameworks'—a new category in our taxonomy reflecting the trend toward responsible AI development in generative AI [28,245–248]. Such frameworks are crucial in ensuring AI systems are built with a core emphasis on ethical considerations, fairness, and transparency, as they address critical aspects such as bias mitigation for fairness, privacy and security concerns for data protection, and AI ethics for accountability; thus, responding to the evolving landscape where accountability in AI is of paramount importance [28,245]. The need for rigorous approaches to uphold ethical integrity and legal conformity has never been more pressing, reflecting the complexity and multifaceted challenges introduced by the adoption of these technologies [28].

- **Bias Mitigation:** Bias mitigation in AI systems is a critical endeavor to ensure fairness and representation, which involves not only balanced data collection to avoid skewed perspectives but also involves implementing algorithmic adjustments and regularization techniques to minimize biases [249,250]. Continuous monitoring and bias testing are essential to identify and address any biases that may emerge from AI's predictive patterns [250,251]. A significant challenge in this area is dealing with intersectional biases [252–254] and understanding the causal interactions that may contribute to these biases [255–258].

- **Data Security:** In AI data security, key requirements and challenges include ensuring data confidentiality, adhering to consent norms, and safeguarding against vulnerabilities like membership inference attacks [259,260]. Compliance with stringent legal standards within applicable jurisdictions, such as the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA), is essential, necessitating purpose limitation and data minimization [261–264]. Additionally, issues of data sovereignty and copyright emphasize the need for robust encryption, access control, and continuous security assessments [265,266]. These efforts are critical for maintaining the integrity of AI systems and protecting user privacy in an evolving digital landscape.
- **AI Ethics:** The field of AI ethics focuses on fairness, accountability, and societal impact; addresses the surge in ethical challenges posed by AI's increasing complexity and potential misalignment with human values; and requires ethical governance frameworks, multidisciplinary collaborations, and technological solutions [28,245,267,268]. Furthermore, AI ethics involves ensuring traceability, auditability, and transparency throughout the model development lifecycle, as well as employing practices such as algorithmic auditing, establishing ethics boards, and adhering to documentation standards and model cards [268,269]. However, the adoption of these initiatives remains uneven, highlighting the ongoing need for comprehensive and consistent ethical practices in AI development and deployment [245].
- **Privacy Preservation:** This domain focuses on maintaining data confidentiality and integrity, as well as employing strategies like anonymization and federated learning to minimize direct data exposure, especially when the rise of generative AI poses risks of user profiling [270,271]. Despite these efforts, challenges such as achieving true anonymity against correlation attacks highlight the complexities in effectively protecting against intrusive surveillance [272,273]. Ensuring compliance with privacy laws and implementing secure data handling practices are crucial in this context, demonstrating the continuous need for robust privacy preservation mechanisms.

3.5. Advanced Learning

Advanced learning techniques, including self-supervised learning, meta-learning, and fine-tuning, are at the forefront of AI research, enhancing the autonomy, efficiency, and versatility of AI models.

- **Self-supervised Learning:** This method emphasizes autonomous model training using unlabeled data, reducing manual labeling efforts and model biases [195,274,275]. It incorporates generative models like autoencoders and GANs for data distribution learning and original input reconstruction [276–278], as well as also includes contrastive methods such as SimCLR [279] and MoCo [280], which are designed to differentiate between positive and negative sample pairs. Further, it employs self-prediction strategies, inspired by NLP, using techniques like masking for input reconstruction, which are significantly enhanced by recent vision transformers developments [195,281,282]. This integration of varied methods highlights self-supervised learning's role in advancing AI's autonomous training capabilities.
- **Meta-learning:** Meta-learning, or 'learning to learn', centers on equipping AI models with the ability to rapidly adapt to new tasks and domains using limited data samples [283,284]. This technique involves mastering the optimization process and is critical in situations with limited data availability to ensure models can quickly adapt and perform across diverse tasks, which are essential capacities in the current data-driven landscape [285,286]. It focuses on few-shot generalization, enabling AI

to handle a wide range of tasks with minimal data, underlining its importance in developing versatile and adaptable AI systems [286–289].

- **Fine-tuning:** This involves customizing pretrained models to specific domains or user preferences, enhancing accuracy and relevance for niche applications [71,290,291]. Its two primary approaches are end-to-end fine-tuning, which adjusts all weights of the encoder and classifier [292,293], and feature extraction fine-tuning, where the encoder weights are frozen to extract features for a downstream classifier [294–296]. This technique ensures that generative models are more effectively adapted to specific user needs or domain requirements, making them more versatile and applicable across various contexts.
- **Human Value Alignment:** This emerging aspect concentrates on harmonizing AI models with human ethics and values to ensure that their decisions and actions mirror societal norms and ethical standards, involving the integration of ethical decision-making processes and the adaptation of AI outputs to conform with human moral values [100,297,298]. This is increasingly important in scenarios where AI interacts closely with humans, such as in healthcare, finance, and using personal assistants, to ensure that AI systems make decisions that are not only technically sound, but also ethically and socially responsible, which means human value alignment is becoming crucial in developing AI systems that are trusted and accepted by society [100,104,299].

3.6. Emerging Trends

Emerging trends in generative AI research are shaping the future of technology and human interaction, and they indicate a dynamic shift towards more integrated, interactive, and intelligent AI systems, driving forward the boundaries of what is possible in the realm of AI. Key developments in this area include the following:

- **Multimodal Learning:** Multimodal learning in AI, a rapidly evolving subdomain, focuses on combining language understanding with computer vision and audio processing to achieve a richer, multisensory context awareness [135,300]. Recent developments like the Gemini model have set new benchmarks by demonstrating state-of-the-art performance in various multimodal tasks, including natural image, audio, and video understanding, as well as mathematical reasoning [8]. Gemini's inherently multimodal design exemplifies the seamless integration and operation across different information types [8]. Despite the advancements, the field of multimodal learning still confronts ongoing challenges, such as refining the architectures to handle diverse data types more effectively [301,302], developing comprehensive datasets that accurately represent multifaceted information [301,303], and establishing benchmarks for evaluating the performance of these complex systems [13,304,305]. Omni-modal models represent a progression from multimodal systems, aiming to achieve universal sensory integration by enabling seamless processing across all conceivable modalities, including text, images, audio, video, and even tactile or environmental data. These models signify a shift toward unified architectures that transcend the limitations of modality-specific processing, enhancing generalizability and contextual reasoning by integrating diverse sensory inputs into a cohesive understanding. Recent research, such as OmniBench [306], has introduced benchmarks designed to evaluate models' ability to interpret and reason across visual, acoustic, and textual inputs simultaneously. Studies [307] in this domain highlight the need for robust multimodal integration techniques and training strategies to enhance performance across diverse modalities, which address challenges in establishing holistic understanding and reasoning in real-world scenarios.

- **Interactive and Cooperative AI:** This subdomain aims to enhance the capabilities of AI models to collaborate effectively with humans in complex tasks [38,308]. This trend focuses on developing AI systems that can work alongside humans, thereby improving user experience and efficiency across various applications, including productivity and healthcare [309–311]. Core aspects of this subdomain involve advancing AI in areas such as explainability [312], understanding human intentions and behavior (theory of mind) [313,314], and scalable coordination between AI systems and humans, which involve a collaborative approach crucial in creating more intuitive and interactive AI systems capable of assisting and augmenting human capabilities in diverse contexts [38,315].
- **AGI Development:** AGI, representing the visionary goal of crafting AI systems that emulate the comprehensive and multifaceted aspects of human cognition, is a subdomain focused on developing AI with the capability for holistic understanding and complex reasoning that closely aligns with the depth and breadth of human cognitive abilities [67,316,317]. AGI is not just about replicating human intelligence but also involves crafting systems that can autonomously perform a variety of tasks, demonstrating adaptability and learning capabilities akin to those of humans [316,317]. The pursuit of AGI is a long-term aspiration, continually pushing the boundaries of AI research and development.
- **AGI Containment:** AGI safety and containment acknowledges the potential risks associated with highly advanced AI systems, focused on ensuring that these advanced systems are not only technically proficient but also ethically aligned with human values and societal norms [12,28,67]. As we progress towards developing superintelligent systems, it becomes crucial to establish rigorous safety protocols and control mechanisms [12]. Key areas of concern include mitigating representational biases, addressing distribution shifts, and correcting spurious correlations within AI models [12,318]. The objective is to prevent unintended societal consequences by aligning AI development with responsible and ethical standards.
- **"Thinking" Models:** Thinking models represent a transformative trend in generative AI, focusing on integrating test-time compute scaling that enables computational flexibility during inference. These approaches allow AI systems to dynamically allocate additional computational resources based on task complexity, enhancing reasoning capabilities for problems requiring step-by-step deliberation. By allocating extra computational resources for reasoning-intensive tasks, these models achieve greater adaptability and efficiency, bridging the gap between traditional static inference and the demands of real-time, complex decision making. The development of thinking models highlights the potential to create systems capable of iterative, context-aware reasoning, addressing limitations in single-pass architectures while paving the way for more robust AI applications.
- **Agentic AI:** Agentic AI represents a critical advancement in empowering AI systems with autonomy, enabling them to plan, execute, and adapt actions in dynamic, real-world environments [319]. These systems leverage advanced tooling capabilities, allowing direct interaction with external software and hardware to achieve predefined objectives. Agentic AI also incorporates decision making and actionability, where the AI autonomously determines the optimal sequence of steps to achieve a goal while adhering to constraints. The development of agentic AI aligns with human-centric design principles, ensuring these systems are ethically aligned and augment human capabilities in diverse fields ranging from healthcare to logistics. However, the rise of agentic AI necessitates addressing challenges such as value alignment, safety protocols, and mitigating unintended consequences in autonomous decision making [319].

- **Test Time Optimization:** The ability for LLMs to dynamically adjust model parameters at inference time is an important research direction for LLMs. Test Time Training (TTT) [320,321] addresses the limitations of static, pretrained models by enabling LLMs to update parameters for smaller models during inference, especially for task-specific requirements. This facilitates real-time learning, allowing models to refine their understanding of a task as new inputs are encountered. This shift toward more adaptable systems allows LLMs to learn from each unique context and improve responses dynamically while also addressing computational limitations associated with long sequences. Approaches like TTT also mitigate the problem of memorization by learning how to memorize, rather than simply memorizing training data, enhancing generalization. This adaptability is can be seen as being complemented by enhanced long-term neural memory [159], which stores and recalls past information, moving beyond the limitations of short context windows in standard attention. These solutions are composed of short-term memory (core attention), long-term memory that learns to memorize at test time, and persistent memory, which stores task-specific knowledge. These advancements are pivotalm as they enable more flexible, context-aware AI systems capable of improved accuracy, better generalization, and more dynamic reasoning in complex, real-world applications.

4. Innovative Horizon of MoE

Recent advances in LLMs, multimodal frameworks, and agentic AI have renewed interest inMo) architectures as a crucial approach for scaling model capacity. This section provides an updated examination of MoE's core concepts (Figure 4), recent progress, and prospective research directions in light of the latest literature, including improved parallelization and integration with advanced AI paradigms such as thinking models (test-time compute) and multistep planning.

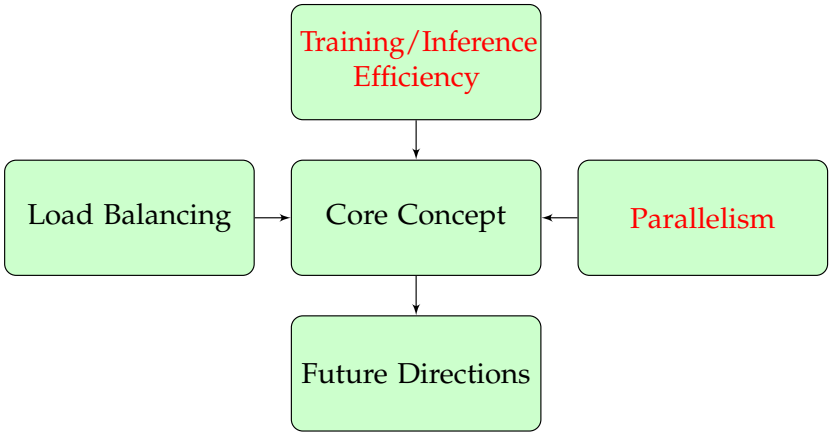

Figure 4. Conceptual diagram of MoE innovation.

4.1. Core Concept and Structure

The MoE model architecture continues to emerge as a *scalable and specialized* framework in transformer-era LLMs, enabling networks to partition complex tasks into specialized "expert" pathways [125,322,323]. Instead of relying on fully dense layers, the MoE replaces certain blocks with multiple feedforward "experts", and a learnable *router* mechanism assigns each input token to only a subset of these experts, thus enforcing computational sparsity [106,324].

- **Sparse Routing:** Because only a few experts are activated per token, the MoE scales model capacity significantly without linearly increasing compute. This "conditional

computation" is particularly valuable for tasks where a large portion of the model parameters may be unnecessary for every input.

- **Transformer Integration:** Recent LLM research incorporates MoE layers interleaved with attention blocks to maintain the contextual strengths of transformers while boosting efficiency for large-scale training [186].

As a result, MoE-based systems exhibit enhanced specialization across diverse tasks and can scale to trillions of parameters. However, key challenges such as dynamic load balancing, training stability, and memory overhead remain focal points for the community [106,325]. Furthermore, synergy with agentic AI demands robust gating logic that can flexibly adapt to environment-driven queries.

4.2. Training and Inference Efficiency

Recent LLMs confirm that MoE architectures significantly reduce training overhead while boosting performance across multilingual, coding, and domain-specific tasks [322]. By activating only a fraction of experts per token, these sparse layers lower the overall floating-point operations (FLOPs) while preserving or even improving accuracy relative to equally sized dense models.

Training Cost and Scalability: Studies have shown that advanced MoE setups can achieve up to 3–5× speedups in pretraining relative to dense transformer baselines [186,326]. Tools like Lina [327] tackle all-to-all communications to mitigate routing overhead, enhancing large-scale training parallelization. Even so, fine-tuning continues to impose heavy memory requirements in many MoE designs, as all experts may reside in GPU VRAM, though new work on hierarchical or load-on-demand experts is attempting to reduce this cost [125,328].

Inference Efficiency: Efficient serving of MoE-based LLMs is facilitated by advanced parallelism modes (e.g., expert parallelism or model compression) [322]. Compression strategies and micro-batching significantly decrease latency, with some efforts achieving up to 7–8× speedups at inference time [106]. Dynamic gating also opens the possibility of test-time compute scaling: the network can allocate additional experts for complex inputs, aligning with the concept of "thinking models" that allocate more compute to reason-intensive queries.

4.3. Load Balancing and Router Optimization

The gating network, or router, is integral to MoE's effectiveness and stability [325,329]. Recent works introduced load-balancing losses (such as "Z-loss") that prevent over-reliance on a small subset of experts, reducing training instabilities [330]. Techniques like capacity-limited experts also ensure that no single expert becomes saturated, maintaining throughput efficiency.

Robust Routing in Adversarial Settings: As LLMs become more agentic and are deployed in diverse real-world scenarios, adversarial examples or unusual input distributions may trigger degenerate routing. Ongoing research investigates robust gating to preserve balanced workloads even under adversarial attacks [326]. The synergy between gating and Chain-of-Thought or multimodal inputs remains a frontier area, where dynamic allocation of cross-modal "expert" modules can adapt in real time.

4.4. Parallelism and Serving Techniques

Realizing efficient MoE deployment at scale involves bridging HPC techniques with sophisticated software frameworks:

- *DeepSpeed-MoE* [322] introduces three main parallelism strategies—data parallelism, expert parallelism, and tensor slicing—to manage immense models (e.g., 1–2 trillion parameters) without exploding memory or latency.
- **Expert Ensemble Approaches:** In certain large-scale LLM tasks, MoE experts can function similarly to an ensemble model, boosting performance across specialized tasks like code generation, multilingual translation [331–333], or domain adaptation.
- **MoE Model Compression:** Hybrid or partial expert load can further reduce runtime, enabling more cost-effective serving. This direction aligns with industry demands for on-device or edge-based LLM inference where VRAM is at a premium [325].

Consequently, state-of-the-art MoE solutions can yield sublinear scaling of compute versus total model parameters [125,186,322,325], effectively democratizing trillion-parameter LLM capabilities. However, continued refinements—particularly in scheduling, caching, and dynamic routing—are needed to deliver truly universal and resource-flexible solutions for real-time multimodal and agentic applications.

4.5. Future Directions and Applications

Building on these innovations, the MoE paradigm is rapidly expanding in scope. We outline the following key open questions:

- **Sparse Fine-Tuning and Instruction Tuning:** Future research may explore gating-based fine-tuning that updates only a small subset of experts, enhancing computational and memory efficiency while retaining model quality. Instruction tuning can also be integrated, with specialized experts employed for each "instruction domain".
- **Multimodal Integration:** The MoE can unify specialized experts for text, image, audio, and even 3D data, offering a flexible "mixture-of-modalities" approach. This is especially relevant for agentic AI systems requiring simultaneous processing of multiple data types.
- **Adaptive Test-Time Compute:** Research on on-demand expert activation—akin to recent "Thinking Model" frameworks—could allow LLMs to allocate more experts for complex tasks and fewer for simpler queries, optimizing both cost and performance.
- **Calibration and Safety:** Because experts can specialize in sensitive domains (e.g., healthcare, finance, etc.), robust mechanisms for bias mitigation and alignment with human values remain vital [80].

These directions underscore the MoE's potential to serve as the backbone for next-generation AI ecosystems, spanning from massive cloud-based inference to resource-constrained edge scenarios. Aligning with emerging agentic AI and multimodal "LLM 2.0" paradigms, the MoE's sparse design promises to deliver both scale and specialization, without incurring the prohibitive costs typical of traditional dense architectures. Through continued research in gating optimization, parallelization, compression, and dynamic inference, the MoE stands poised to redefine the frontier of large-scale neural network design and deployment.

5. Capabilities of Q*-Enabled Models

In light of recent advances in LLMs discussed in the previous section, the following formulations illustrate plausible ways that modules for reasoning, memory, and external retrieval could be interconnected in an advanced AI system referred to as "Q^*" (Figure 5). These equations are intended as conceptual frameworks rather than rigorous proofs, highlighting how modern LLM-based systems combine external memory, step-by-step reasoning, and reinforcement signals to address complex tasks.

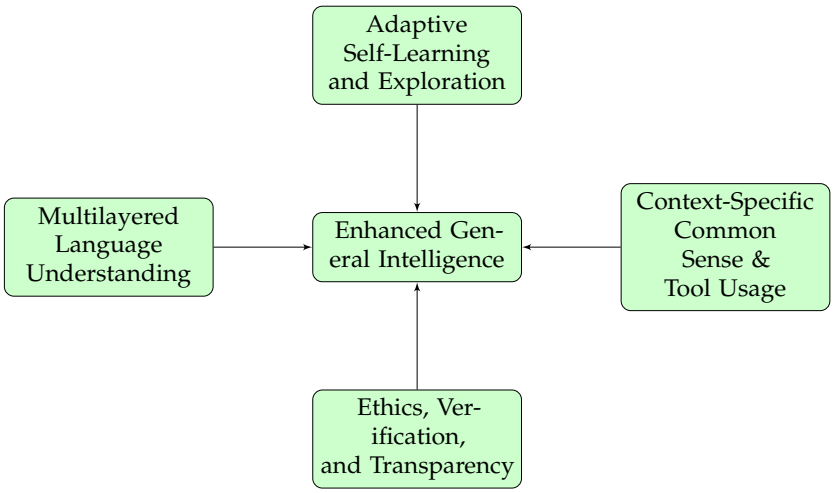

Figure 5. Conceptual diagram of the revised Q* capabilities.

5.1. Enhanced General Intelligence

Recent architectures frequently merge large-scale parameterized models with external retrieval and Chain-of-Thought techniques [15,16]. We capture this synergy as

$$\mathrm{EGI}(Q*) \;=\; \bigoplus_{i=1}^{n} \Big[\, \mathrm{LLM}_i(\cdot) \;\otimes\; \mathrm{CoT}_i(\cdot) \;\oplus\; \mathrm{Retr}_i \Big], \tag{1}$$

Therein, we have the following:

- $\mathrm{EGI}(\cdot)$ denotes a generalized intelligence operator integrating multiple modules;
- LLM_i is a large-scale language model trained on broad corpora;
- $\mathrm{CoT}_i(\cdot)$ stands for Chain-of-Thought prompting and reasoning [15];
- Retr_i indicates retrieval modules for external knowledge (e.g., a vector database);
- $\otimes$ and $\oplus$ indicate functional integrations at different processing stages (e.g., fusing retrieval results with step-by-step reasoning).

This formulation highlights the current trend of combining parameter-rich models with lightweight retrieval and reasoning components to tackle broad, open-ended tasks efficiently.

5.2. Adaptive Self-Learning and Exploration

Building upon progress in self-play, hierarchical reinforcement learning, and reflection loops [334], we consider the following:

$$\mathrm{ASLE}(Q*) \;=\; \mathrm{HRL}\big(\mathrm{PNN}, \mathrm{Mem}\big) \;\times\; \mathrm{Reflex}, \tag{2}$$

Therein, we have the following:

- $\mathrm{ASLE}(\cdot)$ is an *adaptive self-learning and exploration* operator;
- $\mathrm{HRL}(\cdot)$ represents hierarchical reinforcement learning, splitting learning into high-level and low-level policies;
- PNN is a policy neural network interfacing with the hierarchical controllers;
- Mem denotes a knowledge buffer or episodic memory for iterative policy updates [335];
- Reflex implements reflection-based self-critique steps [334].

This highlights how next-generation RL systems can iteratively refine both high-level goals and low-level execution, leveraging memory and self-reflection mechanisms for continual improvement.

5.3. Multilayered Language Understanding and Reasoning

Rather than emulating strict human-level metrics, contemporary work emphasizes layered reasoning involving factual, logical, and commonsense modalities [80]. A possible abstraction is the following:

$$\mathrm{MLLU}(Q*) = \sum_{m \in \mathcal{M}_{\mathrm{Reason}}} \Big[\mathrm{LLM}(\mathrm{Knowledge}_m) \oplus \mathrm{Align}_m \Big], \tag{3}$$

Therein, we have the following:

- $\mathrm{MLLU}(\cdot)$ denotes a *multilayered language understanding* operator;
- $\mathcal{M}_{\mathrm{Reason}}$ enumerates various modes of reasoning (expert logic, common sense, etc.);
- $\mathrm{LLM}(\mathrm{Knowledge}_m)$ processes specialized or curated knowledge resources;
- Align_m is an alignment or value filter [80] ensuring compliance with domain or ethical constraints;
- $\oplus$ merges each reasoning layer's output.

This structured approach reflects the current focus on multistage, alignment-aware pipelines in advanced language reasoning systems.

5.4. Context-Specific Common Sense and Tool Usage

The research on tool usage and external plugins extends "common sense" beyond internal reasoning [336]. We propose the following:

$$\mathrm{CSCU}(Q*) = (\mathrm{CSEngine} \otimes \mathrm{ToolAPI}) \oplus \mathrm{WorldK}, \tag{4}$$

Therein, we have the following:

- $\mathrm{CSCU}(\cdot)$ captures *context-specific common sense and tool usage*;
- CSEngine is a core symbolic or Chain-of-Thought logic unit for everyday reasoning;
- ToolAPI encapsulates callable external utilities (e.g., search engines, calculators, code interpreters, etc.);
- WorldK is a dynamic knowledge base or curated fact store;
- $\otimes$ and $\oplus$ denote flexible chaining of these resources on demand.

This formulation aligns with frameworks like ToolFormer and plugin-based systems, enabling flexible, context-driven resource usage.

5.5. Ethics, Verification, and Transparency

Finally, modern AI systems integrate interpretability, safety, and ethics checking directly into their generative workflows [337]. We define the following:

$$\mathrm{EVT}(Q*) = \mathrm{Verifier}\big(\mathrm{LLM}\big) \otimes \mathrm{Explainer} \otimes \mathrm{EthicsCheck}, \tag{5}$$

Therein, we have the following:

- $\mathrm{EVT}(\cdot)$ stands for *ethics, verification, and transparency*;
- Verifier includes factual consistency or theorem-proving tools;
- Explainer produces Chain-of-Thought or post hoc rationales for interpretability;
- EthicsCheck embeds societal or policy-based constraints.

This ensures that advanced generative models maintain accountability and transparency in parallel with their raw computational capabilities.

6. Projected Capabilities of AGI

Artificial General Intelligence (AGI) is increasingly discussed as a prospective leap in AI research, aiming to replicate (or at least approximate) the breadth of human cognitive abilities within computational systems [67,316,317]. Recent progress in large-scale language models, multimodal integration, and agentic AI suggest possible pathways toward more autonomous systems, although significant conceptual and technical barriers remain. Figure 6 offers a schematic view of how AGI might encompass autonomous learning, broad cognitive faculties, common sense reasoning, and holistic knowledge integration.

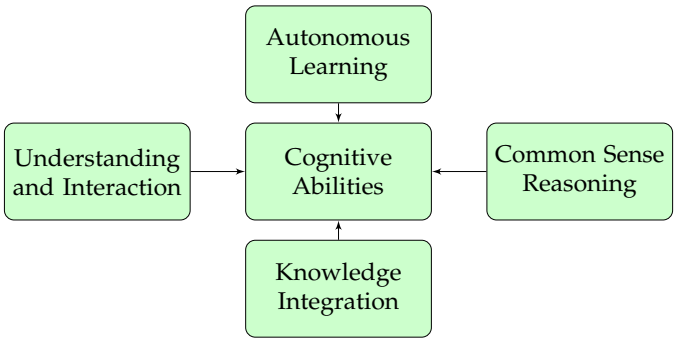

Figure 6. Conceptual diagram of projected AGI capabilities.

6.1. Revolution in Autonomous Learning

AGI research aspires to push beyond static, dataset-centric training regimes by introducing systems that can continuously learn from interactions and environmental feedback [67,316,317,338]. Methods such as Proximal Policy Optimization (PPO) or other reinforcement-based strategies allow these models to adapt their parameters dynamically, indicating a paradigm shift from "train-then-infer" cycles to ongoing self-improvement [211,339]. This approach aims to reduce reliance on frequent human-initiated retraining, making the system more robust and responsive to nonstationary or novel scenarios.

6.2. Broadening of Cognitive Abilities

By integrating multiple architectures—ranging from large transformer-based models (e.g., GPT, BERT, etc.) to multimodal pipelines—AGI could exhibit a more holistic intelligence akin to human cognition [316,340]. Recent progress in multistep reasoning, "Chain-of-Thought" prompting, and agentic AI frameworks suggests that future systems might autonomously coordinate specialized modules or "universal adapters" to assimilate diverse information types in real time [316,341]. This adaptability is already beginning to show promise in fields like healthcare, where advanced diagnostic models indicate how context-rich learning could transform medical analysis and treatment planning.

6.3. Elevating Understanding and Interaction

AGI is projected to reach deeper linguistic and socio-emotional comprehension, leveraging large-scale models with sophisticated inference and possibly multisensory integration [316,338,342]. Beyond textual or visual modalities, such systems may interpret pragmatic cues—tone, affect, or user intent—enabling complex, empathetic, and context-sensitive engagements. This expansion from purely symbolic or pattern-based processing toward contextually grounded interactions marks a critical step in building AI systems that can meaningfully collaborate with humans in scientific research, creative endeavors, and decision making.

6.4. Advanced Common Sense Reasoning

While current LLMs often exhibit "hallucinations" or lack robust world modeling, a future AGI might incorporate symbolic reasoning, probabilistic inference, and real-time environment data to achieve common sense understanding [316,343,344]. Such systems could navigate real-world tasks more effectively, bridging the gap between narrowly trained AI and human-like flexibility. Symbolic submodules or knowledge graphs may augment generative models, offering explicit reasoning chains that align more closely with how humans conceptualize cause–effect relationships and domain knowledge.

6.5. Holistic Integration of Knowledge

As AGI efforts progress, new methods for fusing diverse knowledge representations—textual, visual, and multimodal—are emerging [316,342]. Such comprehensive synthesis could allow AGI systems to handle complex, cross-disciplinary challenges: from climate change modeling to economic policy simulations. At the same time, formal verification techniques and advanced alignment protocols aim to ensure that AGI's outputs are not only accurate but also ethically or legally compliant [28,67]. If realized, these capabilities could yield impactful solutions in fields ranging from environmental sustainability to large-scale social planning [317,340].

6.6. Challenges and Opportunities in AGI Development

Despite the optimism surrounding AGI, significant hurdles persist. Models still grapple with issues like representational bias, interpretability, and the high computational overhead of large-scale architectures [28,67]. Balancing these challenges requires robust data governance, more efficient resource usage, and governance frameworks that address societal concerns ranging from privacy to misuse [340]. Furthermore, it remains unclear whether scaling existing architectures or forging entirely new paradigms (e.g., brain-inspired spiking networks) will prove most effective in achieving AGI [67,345].

Experts caution that we should not conflate current LLM-based "sparks of intelligence" with the full spectrum of human-level cognition [316]. Indeed, the gap between practical AI systems and truly general intelligence remains considerable. Nonetheless, the trajectory of agentic AI, multimodal reasoning, and dynamic training points to a future where AI could expand its autonomy across various domains [28,316]. Realizing AGI's transformative potential thus requires diligent research, transparent collaboration among stakeholders, and the development of ethical guardrails—ensuring responsible innovation as these capabilities evolve.

7. Impact Analysis on Generative AI Research Taxonomy

This section presents an updated analysis of how recent developments and insights in generative AI—including thinking models, agentic AI, and test time optimization—influence existing research domains across model architectures, training techniques, application areas, compliance/ethics, advanced learning paradigms, and emerging trends. We first outline the criteria used to categorize the shifting research landscape and then provide a revised overview of how each subdomain in our taxonomy is impacted.

7.1. Criteria for Impact Analysis

The rapidly evolving landscape of generative AI catalyzes transformative shifts across diverse research domains. To systematically evaluate the influence of these advancements, we have established a robust set of criteria detailed in Table 3. These criteria serve as analytical lenses to quantify and categorize the impact of generative AI innovations that

are grounded in the dynamic interplay between technological progress and the evolving paradigms of research focus areas.

Our analytical framework employs a gradient scale ranging from emergent to obsolete, reflecting the extent to which areas of generative AI research are being reshaped. The categorization into five distinct classes—*Emerging Direction*, *Requiring Redirection*, *Still Relevant*, *Likely to Become Redundant*, and *Inherently Unresolvable*—allows for a nuanced assessment, acknowledging that not all research areas are uniformly affected. This multitiered approach is informed by historical patterns of technological disruption and the adaptability of scientific inquiry [346,347].

Table 3. Criteria for analyzing impact on generative AI research.

Symbol	Criteria	Score	Definition	Justification
↗	Emerging Direction	5	New research areas expected to arise as a direct consequence of AI advancements.	Emphasizes novel research domains emerging from AI breakthroughs [346,348].
↪	Requiring Redirection	4	Areas that need to shift focus or methodology to stay relevant with new AI developments.	Technological shifts necessitate re-evaluation and redirection in AI research [346,347].
↔	Still Relevant	3	Areas where the advancements have minimal or no impact, maintaining their current status and methodologies.	Observes the persistence of certain AI research areas despite technological advancements [347].
↘	Likely to Become Redundant	2	Areas that may lose relevance or become obsolete with the advent of new AI technologies.	Discusses rapid obsolescence in AI methodologies due to new technologies [349].
△	Inherently Unresolvable	1	Challenges that may remain unresolved due to complexities like subjective human perspectives and diverse cultural values.	Inherent difficulties in issues such as aligning AI with diverse human values and ethics [350,351].

7.1.1. Emerging Direction (↗)

At the apex of our evaluative hierarchy, *Emerging Direction* encapsulates the advent of uncharted research vistas propelled by ongoing AI breakthroughs. These are new or rapidly growing research areas that are highly influenced by current innovations, such as advanced reasoning capabilities and agentic AI systems [346,348]. This category highlights novel subfields driven by cutting-edge technologies, signaling significant shifts in research focus and the potential for groundbreaking discoveries.

7.1.2. Requiring Redirection (↪)

Requiring Redirection pertains to established research areas that must adapt or pivot to remain relevant in the face of new AI developments. These areas are at an inflection point, necessitating a strategic overhaul of traditional methodologies to integrate emergent paradigms like thinking models, multimodality, or advanced LLMs [346,347]. For instance, the transition from rule-based expert systems to adaptive machine learning frameworks exemplifies the need for redirection in existing research domains.

7.1.3. Still Relevant (↔)

The *Still Relevant* classification affirms the resilience of select research domains that retain their core significance despite new technological advancements. These areas maintain their relevance by addressing persistent scientific inquiries or through their inherent flexibility to incorporate new techniques and best practices [347]. Longstanding topics, such as fundamental algorithmic improvements or theoretical foundations, fall under this category, demonstrating the enduring nature of certain research pursuits.

7.1.4. Likely to Become Redundant (↘)

Areas categorized as *Likely to Become Redundant* face the risk of obsolescence due to the emergence of more efficient or generalizable methods. These research domains may be overshadowed by powerful new LLM architectures or emergent frameworks that

render previous approaches less effective or relevant [349]. This classification serves as a warning for researchers to anticipate potential reductions in relevance and consider strategic foresight and resource reallocation to prevent scientific stagnation.

7.1.5. Inherently Unresolvable (△)

Finally, *Inherently Unresolvable* challenges represent topics that remain incomplete or irreconcilable, often due to complex human or cultural factors. These enduring dilemmas within AI research, such as fully universal ethics, are deeply rooted in the diverse tapestry of human values and societal imperatives [350,351]. This category highlights the perpetual nature of certain challenges, highlighting the necessity for ongoing dialogue and interdisciplinary approaches to navigate these complex issues.

The comprehensive categorization provided by these criteria enables a structured and detailed assessment of how Generative AI innovations influence various research domains. By systematically applying these labels, researchers can identify trends, anticipate shifts, and strategically align their work with the evolving landscape of AI advancements.

7.2. Overview of Impact Analysis

This subsection provides an updated and detailed overview of the impact analysis carried out on the research taxonomy within the realm of generative AI. We expand on the existing paradigms of *MoEs*, *multimodality*, and *AGI* by incorporating the newly emerging frameworks of *thinking models* (involving test-time compute) and *agentic AI*. Our goal is to illustrate how these collective advancements influence multiple facets of generative AI research, spanning from model architectures and sophisticated training methodologies to ethical, compliance, and governance issues. By synthesizing both quantitative and qualitative assessments across different domains and subdomains of LLM research, we highlight how each area is compelled to adapt, remain steadfast, or potentially yield to more advanced approaches.

In carrying out this analysis, several factors were considered. First, we investigated whether newly introduced techniques and system architectures (e.g., sparse routing, test time optimization, Chain-of-Thought, and real-time memory augmentation) have triggered *emerging directions* in established subfields. Second, we asked if existing research directions must *redirect* their approaches or frameworks to align with the capabilities and demands of next-generation LLMs. Third, we observed the continued relevance of foundational methodologies that may remain robust despite the introduction of more sophisticated generative systems. Finally, the analysis reveals whether certain lines of research have become less vital in light of these advancements or if core ethical dilemmas remain *inherently unresolvable*, owing to the complex nature of societal, psychological, or philosophical challenges. Table 4 encapsulates these findings, including revised scores for each subdomain, thereby presenting a holistic map of how generative AI has evolved in recent months.

Table 4. Impact of MoE, multimodality, and AGI on generative AI research.

Domain	Subdomain	MoE	Multimodality	AGI	Overall Score
Model Architecture	Transformer Models	↪ (4)	↔ (3)	↪ (4)	11
	Recurrent Neural Networks	↘ (2)	↔ (3)	↘ (2)	7
	MoE	↔ (3)	↗ (5)	↪ (4)	12
	Multimodal Models	↗ (5)	↔ (3)	↗ (5)	13

Table 4. *Cont.*

Domain	Subdomain	MoE	Multimodality	AGI	Overall Score
Training Techniques	Supervised Learning	↪ (4)	↔ (3)	↘ (2)	9
	Unsupervised Learning	↪ (4)	↔ (3)	↪ (4)	11
	Reinforcement Learning	↔ (3)	↪ (4)	↗ (5)	12
	Transfer Learning	↔ (3)	↗ (5)	↪ (4)	12
Application Domains	Natural Language Understanding	↔ (3)	↔ (3)	↗ (5)	11
	Natural Language Generation	↔ (3)	↪ (4)	↗ (5)	12
	Conversational AI	↪ (4)	↗ (5)	↗ (5)	14
	Creative AI	↪ (4)	↗ (5)	↗ (5)	14
Compliance and Ethical Considerations	Bias Mitigation	↪ (4)	↪ (4)	↗ (5)	13
	Data Security	↔ (3)	↔ (3)	↔ (3)	9
	AI Ethics	↪ (4)	↪ (4)	△ (1)	9
	Privacy Preservation	↪ (4)	↪ (4)	↪ (4)	12
Advanced Learning	Self-supervised Learning	↪ (4)	↗ (5)	↔ (3)	12
	Meta-learning	↔ (3)	↔ (3)	↗ (5)	11
	Fine-tuning	↔ (3)	↔ (3)	↘ (2)	8
	Human Value Alignment	△ (1)	△ (1)	△ (1)	3
Emerging Trends	Multimodal Learning	↗ (5)	↔ (3)	↗ (5)	13
	Interactive and Cooperative AI	↪ (4)	↔ (3)	↗ (5)	12
	AGI Development	↪ (4)	↪ (4)	↔ (3)	11
	AGI Containment	△ (1)	△ (1)	↗ (5)	7
	Thinking Models	↪ (4)	↗ (5)	↗ (5)	14
	Agentic AI	↪ (4)	↪ (4)	↗ (5)	13
	Test Time Optimization	↗ (5)	↪ (4)	↗ (5)	14

7.2.1. Impact on Model Architecture

Transformer Models: Transformer models have been scored with a *redirection requirement* (↪) of 4 in both MoE and AGI contexts and a *relevance* (↔) of 3 in multimodality, leading to an overall score of 11. These models, forming the backbone of many AI architectures, continue to be effective for handling complex input sequences. Nevertheless, to harness the emergent multistep reasoning (as in thinking models) and modular expansions (e.g., MoE gating layers), transformers require strategic adaptation in design and training methodologies.

Recurrent Neural Networks (RNNs): RNNs face a potential decmidrule in relevance, as indicated by their scores: *likely to become redundant* (↘) 2 in both MoE and AGI contexts while *still relevant* (↔) 3 in multimodality, resulting in a total of 7. While RNNs excel at certain types of sequential or stream-based data, their limitations in capturing long-range dependencies and the rise of more efficient transformer-based approaches continue to overshadow RNNs in large-scale tasks. They may retain a niche for highly specialized or constrained scenarios.

MoE: MoE models have scored a consistent *relevance* (↔) of 3 for their own development and a score of 5 (*emerging direction*, ↗) in multimodality. In the context of AGI, they require *redirection* (↪) of 4, summing to 12 overall. Their sparse gating approach is particularly well suited to specialized subtasks and high-parameter efficiency, making them key candidates for advanced LLMs and multimodal deployments. At the same

time, integrating MoE models into agentic frameworks demands dynamic load balancing, flexible routing, and possibly test-time compute expansions to adapt to more generalized cognitive tasks.

Multimodal Models: These models received high *emerging direction* (↗) scores of 5 in both MoE and AGI domains, together with a *relevance* (↔) of 3 in existing multimodality frameworks, leading to an overall score of 13. The integration of MoE and the push toward AGI create new pathways for multimodal processing, including omni-modal intelligence that spans text, image, audio, and possibly video streams. Recent frontier models like Gemini and GPT-4o highlight how advanced architectures can concurrently handle multiple sensory inputs for more robust reasoning and agentic interactions.

7.2.2. Impact on Training Techniques

Supervised Learning: Supervised learning scored a *redirection* (↪) of 4, *relevance* (↔) of 3 in multimodality, and *potential redundancy* (↘) of 2 in AGI, summing to 9. Although supervised datasets remain essential in many applications, LLMs and agentic systems rely increasingly on self-supervised or reinforcement-driven approaches. Consequently, supervised learning is gradually being supplanted by these more autonomous paradigms in certain advanced or generalizable tasks.

Unsupervised Learning: Unsupervised learning obtains a *redirection requirement* (↪) of 4 in MoE and AGI and maintains *relevance* (↔) of 3 in multimodality, totaling 11. New forms of generative pretraining, such as masked-token modeling and contrastive approaches, remain central to large-scale LLMs. The inherently flexible nature of unsupervised learning is especially critical in multimodal expansions, where labeled data are often sparse.

Reinforcement Learning: Reinforcement learning is *still relevant* (↔) with 3 in MoE, *requiring redirection* (↪) 4 in multimodality, and *emerging* (↗) 5 in AGI, yielding 12. Of particular note is the agentic usage of RL for fine-tuning LLMs with environmental feedback (e.g., RLHF or tool usage). In multimodal environments, RL can coordinate multiple streams of perception and action. As AGI-level capabilities mature, RL frameworks are likely to feature even more prominently in how these models learn and act in unstructured tasks.

Transfer Learning: Transfer learning shows *relevance* (↔) of 3 in MoE, *emerging* (↗) 5 in multimodality, and *redirection* (↪) 4 in AGI, totaling 12. Given the explosive scale of LLM pretraining, transfer learning proves indispensable for tailoring these giant networks to downstream tasks or domains. In a future AGI paradigm, it may expand to encompass not merely static fine-tuning but also real-time domain assimilation, bridging specialized knowledge across multiple tasks.

7.2.3. Impact on Application Domains

Natural Language Understanding (NLU): NLU remains *relevant* (↔) with 3 in MoE and multimodality and *emerging* (↗) 5 in AGI, resulting in a total of 11. The MoE's capacity to handle massive amounts of language data enhances NLU's precision; multimodal setups enable more holistic language understanding, including cross-linguistic or context-imbued tasks. Future AGI progress is poised to radically expand NLU into areas requiring advanced semantics, pragmatics, and reasoning.

Natural Language Generation (NLG): NLG remains *relevant* (↔) of 3 in MoE, *requiring redirection* (↪) 4 in multimodality, and *emerging* (↗) 5 in AGI, summing to 12. NLG stands at the heart of new product categories in content creation and interactive media. Achieving fluid, contextually coherent, and ethically aligned generation demands advanced gating, Chain-of-Thought reasoning, and methodical integration of multimodal signals.

Conversational AI: Conversational AI is assigned *redirection* (↪) of 4 in MoE and *emerging* (↗) 5 in both multimodality and AGI, leading to 14. Frontier LLMs such as ChatGPT, Bard, Claude, and Gemini have already showcased how combining a LLMl with tool usage, advanced search, or multimodal inputs can yield more human-like dialogue, including deep contextual understanding and real-time adaptation.

Creative AI: Creative AI scores 4 (*redirection*) for MoE and 5 (*emerging*) in both multimodality and AGI (totaling 14). With the integration of multimodal data, generative video, audio, and 3D content expand beyond mere text and image creation. Developments like Google's Veo or OpenAI's Sora highlight the growing interest in bridging large language modeling with creative endeavors, potentially leading to novel forms of multimodal artistic generation.

7.2.4. Impact on Compliance and Ethical Considerations

Bias Mitigation: Bias mitigation shows a *redirection* (↪) of 4 in MoE and multimodality and *emerging* (↗) 5 in AGI, totaling 13. As systems become more modular (via MoE) and integrate multimodal data, new forms of bias can emerge or amplify. Researchers must embed advanced bias detection and auditing mechanisms early in the development cycle, especially for Q*-like or AGI-like pipelines.

Data Security: Data security remains *relevant* (↔) 3 across MoE, multimodality, and AGI. Even though distributed or multiexpert architectures amplify potential vulnerabilities, foundational practices like encryption, access control, and robust data governance persist as mainstays, with possible incremental updates needed for emerging large-scale or agentic systems.

AI Ethics: AI ethics is marked *redirection* (↪) 4 in MoE and multimodality but *inherently unresolvable* (△) 1 in AGI (score of 9 overall). As soon as LLMs gain agentic capabilities, existential and normative questions escalate—ranging from accountability to moral agency. The complexities of AGI-level systems suggest that fully "resolving" ethics may remain elusive, likely requiring ongoing cross-disciplinary discourse.

Privacy Preservation: Privacy preservation sees a *redirection* (↪) of 4 across MoE, multimodality, and AGI, reaching 12. The MoE's distributed nature or agentic AI's autonomous data usage can complicate the enforcement of data minimization and confidentiality principles. More sophisticated solutions, possibly hardware-accelerated and cryptographically enriched, are needed to maintain user trust and legal compliance.

7.2.5. Impact on Advanced Learning

Self-supervised Learning: For the MoE, self-supervision demands *redirection* (↪) 4; in multimodality, it is *emerging* (↗) 5; and it is *still relevant* (↔) 3 in AGI contexts (totaling 12). While these massive generative systems have scaled primarily through self-supervised paradigms, future developments in multimodal and agentic contexts may require novel objectives that handle dynamic, heterogeneous data and real-time tasks.

Meta-learning: Meta-learning remains *relevant* (↔) 3 for MoE and multimodality but is *emerging* (↗) 5 in AGI (score of 11). Adaptive "learning to learn" becomes more critical as systems approach general intelligence. Techniques that allow for the quick assimilation of new tasks, especially in real-world open-ended domains, become a priority research area.

Fine-tuning: Fine-tuning remains *still relevant* (↔) 3 in MoE and multimodality but is *likely to become redundant* (↘) 2 in AGI, resulting in a total of 8. While domain-specific fine-tuning (RLHF, specialized data, etc.) is widely used today, advanced LLMs with agentic or test time adaptive abilities may gradually diminish its necessity, particularly for broad tasks requiring real-time self-updating strategies.

Human Value Alignment: Human value alignment remains an *inherently unresolvable* (△) challenge across MoE, multimodality, and AGI, each scoring 1 (totaling 3). As LLMs expand in scope and autonomy, reconciling cultural, moral, and individual values continues to defy definitive technical solutions. Instead, ongoing participatory governance and multistakeholder collaborations are imperative.

7.2.6. Impact on Emerging Trends

Multimodal Learning: Multimodal learning earns *emerging* (↗) 5 in MoE and AGI and retains *relevance* (↔) 3 in existing multimodal frameworks, totaling 13. The surge of large-scale multimodal LLMs reflects the growing realization that intelligence emerges more richly when multiple sensory channels are integrated. Innovations like visual–textual Chain-of-RTought or audio–text coherence are key research thrusts.

Interactive and Cooperative AI: This subdomain is *requiring redirection* (↪) 4 in MoE and *still relevant* (↔) 3 in multimodality but *emerging* (↗) 5 in AGI, reaching 12 overall. Multiagent scenarios, co-creative systems, and user-centric interactions demand new approaches to model interpretability, memory integration, and scenario planning—particularly in bridging agentic LLMs with human collaborators.

AGI Development and Containment: AGI development requires *redirection* (↪) 4 in MoE and multimodality yet remains *at its own frontier* (↔) 3 in AGI, scoring 11. Expanding beyond domain-specific tasks, AGI aims for universal reasoning and planning. Meanwhile, AGI containment is *not required to be solved* (△) in MoE and multimodality contexts but *emerging* (↗) 5 in AGI (totaling 7). As AI moves closer to agentic capabilities, safe deployment and fail-safe control measures (both technical and policy-based) become significantly more urgent.

Thinking Models and Agentic AI: *Thinking models* (test-time compute, multistep reasoning, etc.) and *agentic AI* (tool usage, autonomous planning, etc.) represent newly emphasized trends that heavily reshape the research landscape.

- **Thinking Models**: Demand re-engineering of existing architectures, such as the MoE, to enable iterative solution-finding during inference. They also push multimodal and AGI work toward expanded context windows, hierarchical reasoning, and memory-augmented transformations.
- **Agentic AI**: Extends LLM capabilities by enabling them to plan actions, execute external commands, and adapt to real-time feedback. This raises new research questions about controllability, misalignment, and safety, linking to the emergent concept of test time optimization and dynamic resource allocation.

Combined, these two strands accelerate the evolution of generative AI into more interactive, autonomous, and context-resilient systems, further highlighting the necessity for robust ethical and governance frameworks.

Another emerging trend, *thinking models*, is scored as requiring redirection (↪) in MoE (4) because iterative reasoning steps must integrate effectively with sparse expert layers, while in multimodality and AGI, it is an emerging research direction (↗), with each having a score of 5. This reflects the heightened need for multistep and context-dependent reasoning across diverse sensory data, as well as the pursuit of self-reflective intelligence.

Similarly, *agentic AI* is identified as requiring redirection (↪, 4) in both MoE and multimodality, highlighting that autonomous goal-driven behavior imposes new design and training considerations for these architectures. For AGI (↗, 5), agentic capabilities represent a key priority, driving research toward adaptive planning and self-directed action.

Lastly, *test time optimization* is seen as an emerging direction (↗, 5) in both MoE and AGI, indicating that adaptive inference methods—where models dynamically adjust compute based on task complexity—can synergize well with sparse expert systems and

more advanced reasoning. In multimodality, it requires redirection (↪, 4), underscoring the necessity to balance on-demand resource allocation across multiple data types.

Overall, the impact analysis underscores a rapidly evolving landscape where older paradigms (e.g., RNN-based sequence modeling, purely supervised training, etc.) are partially supplanted by more flexible, large-scale, and hierarchical approaches. Meanwhile, new frontiers in multistep reasoning, agentic behaviors, and multimodal integration open both opportunities for real-world innovation and urgent ethical/safety considerations. As such, generative AI research appears poised to undergo further seismic shifts, especially as developers and researchers race to adapt existing methods to the complexities of multidomain, multimodal, and multistep problem solving that define the next generation of advanced AI systems.

8. Emergent Research Priorities in Generative AI

Recent breakthroughs in LLMs, multimodal learning, and agentic AI have reshaped the generative AI research landscape, prompting new directions and emphasizing the scalability and specialization of MoE approaches [106,186,324]. While the previous sections highlight the importance of the MoE, reinforcement learning, and multimodality, this section integrates these topics within broader trends in generative AI and AGI.

8.1. Emergent Research Priorities in MoE

(1) Advanced MoE Architectures for Multimodal Models in Model Architecture:

MoE designs, which distribute computational load across specialized experts, increasingly combine textual, visual, and other data modalities to address tasks with complex data streams [322]. Such integration aims to strengthen the model's capacity for domain-specific and cross-domain reasoning. Efforts to incorporate agentic frameworks and test-time compute have demonstrated the potential of MoE systems to dynamically route inputs across diverse experts or modalities, which is essential for both specialized and generalized AI scenarios.

(2) Synergy Between MoE and Multimodal Learning:

Building on progress in large-scale LLMs, MoE in multimodal learning leverages gating mechanisms to selectively activate relevant experts when processing text, image, audio, or video streams [125,323]. This capacity aligns well with the growth of generative models that can synthesize richer content, indicating new directions in combining MoE with interactive and real-time inference for advanced applications (e.g., robotics or immersive AI).

(3) Investment Trends in MoE:

Funding patterns reflect a shift toward specialized architectures capable of multisensory integration and real-time adaptation. This trend is evident as major tech companies and research organizations invest in large-scale MoE-based systems [325], driving further innovation in training optimization (e.g., load balancing, memory-efficient gating, etc.) and performance enhancements for extremely large models.

8.2. Emergent Research Priorities in Multimodality

Multimodal generative AI entails developing models that effectively fuse text, images, audio, and more. Building on the synergy with MoE, we identify the following key developments:

- **MoE-Driven Multimodal Pipelines:** MoE models integrated within transformers can help manage complex data distributions by assigning modalities or subtasks to

specialized experts [106]. This approach mitigates the computational overhead of large multimodal embeddings while preserving flexibility in learning.

- **Transfer Learning for Cross-Modality:** Transfer learning remains vital, enabling models to reuse representations from one modality to augment learning in another [87]. Emerging research focuses on bridging these modalities via universal encoders or decoders, which is in line with large LLM expansions into audio, vision, and beyond.
- **Conversational and Creative AI Integration:** With the advent of advanced generative models (e.g., ChatGPT, Sora, Gemini, etc.), multimodal AI increasingly supports natural interactions involving both text and visuals, further stimulating research on collaborative content generation and deeper user engagement.
- **Self-Supervised Paradigms:** Self-supervised learning plays a pivotal role in scaling to massive unlabeled multimodal datasets. Future work explores how gating or MoE logic can tailor representations to modality-specific nuances while still maintaining a shared latent space.

These advancements also drive transformations in educational programs, encouraging AI curricula to emphasize cross-disciplinary expertise, particularly around multimodal fusion and large-scale data management.

8.3. Emergent Research Priorities in AGI

The quest for AGI has sparked a surge in interdisciplinary research, focusing on human-like understanding, autonomy, and creative problem solving [67,316]. Below are key areas experiencing renewed attention:

- **Reinforcement Learning for Agentic Reasoning:** AGI aspirations hinge on adaptable models that learn from and respond to unstructured environments [352]. Combining MoE-based LLMs with reinforcement signals is emerging as a way to achieve both specialized skill (via experts) and generalizable decision making.
- **Complex Application Domains:** AGI extends beyond standard NLP tasks, targeting advanced medical diagnosis, scientific discovery, and highly creative endeavors. Natural Language Understanding (NLU), generative text, and multisensor integration underscore the potential for bridging multiple fields—while simultaneously raising new ethical and safety concerns [28].
- **Bias Mitigation and Ethical Frameworks:** As systems scale, so do risks of amplified biases, security loopholes, and accountability gaps. Aligning AI with fairness and transparency remains a priority, requiring cross-disciplinary collaboration [340].
- **Meta-Learning and Emergent Trends:** AGI solutions increasingly explore meta-learning, aiming to equip models with the capacity to rapidly adapt to new tasks or contexts [338]. Equally crucial is the integration of "AGI containment" or safety strategies, reflecting the need to manage advanced autonomous models responsibly.

Along with these trends, funding trajectories reveal a growing interest in AGI's foundational technologies, such as MoE-based inference pipelines, large-scale reinforcement learning environments, and advanced domain adaptation. By directing investment toward robust, transparent, and ethically sound research, the field moves closer to harnessing generative AI's broader societal impact.

9. Practical Implications and Limitations of Generative AI Technologies

Recent advancements in generative AI—such as ultra-large Mixture-of-Experts (MoE) models, multimodal architectures, and increasingly agentic AI frameworks—underscore both the vast potential and the operational constraints of these technologies. This section expands on the computational challenges inherent in training and deploying advanced AI models while detailing real-world applications, market readiness, and key limitations.

9.1. Computational Complexity and Real-World Applications of Generative AI Technologies

9.1.1. Computational Complexity

Generative AI systems, encompassing MoE models, multimodality, and prospective AGI paradigms, present unique computational challenges in terms of scaling, memory requirements, and inference throughput. Pertaining challenges are listed below:

- **Processing Power Requirements**: Advanced generative AI models, including MoE architectures and large-scale LLMs with multistep reasoning, typically demand substantial compute resources [353]. The need for efficient GPU/TPU provisioning grows more acute in scenarios involving multimodal data or test time scaling, potentially leading to significant infrastructure costs.
- **Memory Usage in AI Modeling**: A critical challenge in training and deploying large-scale AI models, particularly those with sparsely activated experts or multisensory inputs, lies in substantial GPU and VRAM requirements. Unlike main system memory, VRAM remains relatively constrained, posing a bottleneck for massive multimodal models or large MoE solutions. Novel methods that offload expert parameters or use dynamic expert loading can mitigate these challenges.
- **Scalability and Efficiency**: Addressing scalability in generative AI, especially in MoE-based and AGI contexts, involves optimizing load management and parallel processing [322]. Research on specialized routing, advanced pipeline parallelism, and memory-optimizing strategies is critical for real-world deployments in healthcare, finance, and education.

9.1.2. Real-World Application Examples of Generative AI Technologies

Despite the computational cost, generative AI's potential continues to reshape diverse sectors, often outpacing conventional analytics or rule-based systems. The following lists some applications:

- **Healthcare**: Generative AI powers diagnostic imaging, personalized treatment recommendations, and disease forecasting [354]. However, reliance on large training corpora raises privacy questions and can demand specialized hardware for medical-grade model inference.
- **Finance**: AI-driven fraud detection and algorithmic trading illustrate high-impact, data-intensive applications [355]. At the same time, decision-making transparency and robust data governance become essential to counter possible biases or hidden vulnerabilities in generative pipelines.
- **Education**: Generative AI fosters adaptive learning, automated content creation, and personalized feedback loops. While these systems promise scalable tutoring and resource-saving benefits, ethical debates around academic integrity persist. Concerns arise regarding the potential of AI-generated content (AIGC) to displace educators or compromise academic authenticity.

9.2. Commercial Viability and Industry Solutions

9.2.1. Market Readiness

Market adoption of generative AI depends on factors, including cost, domain alignment, and workforce readiness, to integrate advanced ML pipelines into existing workflows.

- **Cost Analysis**: Running large generative models—especially MoE or multimodal frameworks—can strain budgets due to hardware needs and specialized engineering talent.
- **Accessibility and Deployment**: Enterprise readiness varies. Some industries (e.g., tech, finance, etc.) rapidly embrace AI, while others face skill gaps and infrastructural barriers.

- **User Adoption Trends**: Public-facing generative AI tools, such as ChatGPT, show swift user adoption. However, domain-specific solutions require specialized training data, expansions in interpretability, and robust MLOps pipelines to gain user trust.

9.2.2. Existing Industry Solutions

Generative AI reshapes sectors by offering unprecedented automation and content generation:

- **Sector-Wise Deployment**: From content creation (marketing, design, etc.) to code generation and robotic process automation, generative AI demands re-evaluation of creative ownership and IP rights.
- **Impact on Market Dynamics**: Traditional players face competition from AI-native startups, while new business models, such as AI-as-a-Service (AIaaS), form around generative capabilities.
- **Challenges and Constraints**: Fundamental issues like scalability, data management complexity, and privacy concerns persist, underscoring the importance of robust governance, standard-setting bodies, and multistakeholder dialogues.

9.3. Limitations and Future Directions in Generative AI

9.3.1. Technical Limitations

As generative AI expands into real-time, multimodal, and agentic domains, several technical gaps become increasingly salient:

- **Contextual Understanding**: Despite LLM breakthroughs, common sense reasoning and long-horizon context modeling remain weak points, spurring research into memory-augmented architectures.
- **Handling Ambiguous Data**: Real-world data often contain inconsistencies or missing attributes. Reliability in these conditions may demand robust inference techniques, e.g., Bayesian or reinforcement-based training, to handle uncertainty.
- **Navigating Human Judgment**: Even though generative AI can interpret policies and procedures, it falls short in reproducing nuanced human values or in assessing complex legal/political implications. Biased or manipulative usage of AI-generated content (AIGC) can lead to skewed decision making.

9.3.2. Future Research Directions

Addressing these shortcomings—while harnessing generative AI's strengths—depends on continued multidisciplinary innovation:

- **Improved Contextual Understanding**: Research on Chain-of-Thought prompting, hierarchical gating, and multihop inference can deepen AI's ability to interpret extended or complex queries.
- **Robust Handling of Ambiguity**: Model architectures that incorporate reliability estimates or adversarial training may yield more stable performance under ambiguous data conditions.
- **Ethical Integration of AIGC in Legal and Political Arenas**: As generative models permeate high-stakes environments, developing frameworks that bolster transparency, mitigate bias, and validate correctness is imperative [356]. Researchers must also consider socio-technical complexities—such as corruption or manipulative usage—where advanced generative models could inadvertently amplify systemic inequities.

In conclusion, generative AI's commercial viability and real-world impact hinge on strategic investment in computational efficiency, regulatory guidance, and ethical oversight. Although future models may significantly augment or automate tasks in healthcare, finance, and education, they also challenge existing norms around accountability, interpretability,

and trust. Addressing these technological and societal questions will define the next phase of AI's adoption and ensure that generative systems serve as a beneficial rather than disruptive force.

10. Impact of Generative AI on Preprints Across Disciplines

The widespread success and rapid commercial availability of generative AI models—exemplified by ChatGPT—have accelerated the production and dissemination of academic manuscripts to unprecedented levels. This phenomenon is particularly evident in the surge of preprints on platforms like arXiv (Figure 7), where submissions in the *cs.AI* category demonstrate a steep climb [357,358]. As generative AI becomes integral to both research and writing processes, the traditional mechanisms of peer review struggle to keep pace, fueling concerns about overall scientific rigor and the potential propagation of unverified or substandard results.

A notable consequence is the bottleneck now observed in scholarly communication: the peer-review workflow, historically a gatekeeper for quality assurance, lacks the bandwidth to manage the deluge of rapidly generated manuscripts. Parallelly, generative tools (e.g., ChatGPT, Claude, etc.) expedite the manuscript drafting process, sometimes obfuscating author contributions or overshadowing methodological weaknesses [357,358]. This proliferation of AI-facilitated or AI-authored manuscripts—spanning not only computer science but also disciplines like biology, social sciences, and engineering—poses a critical test for the resilience and adaptability of academia.

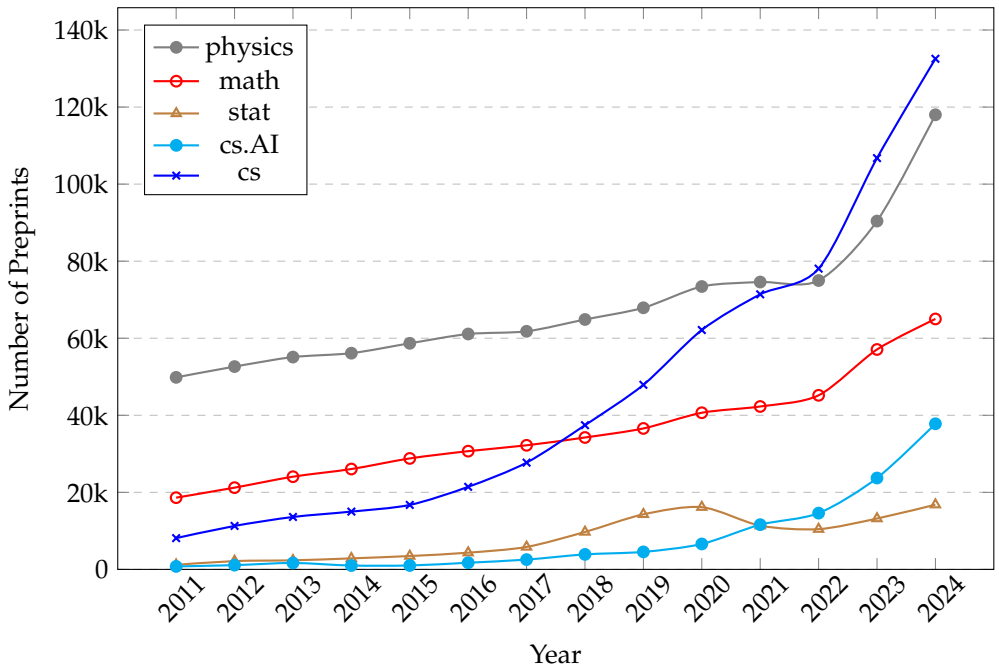

Figure 7. Annual preprint submissions to different categories on arXiv.org.

Furthermore, the swift release of preprints often outpaces the capacity for thorough validation or replication, complicating evidence synthesis and systematic reviews [359,360]. High-volume preprint posting can saturate scholarly communities, particularly in emerging fields such as multimodal AI, Mixture-of-Experts (MoE), and even the nascent exploration of AGI. Consequently, this exponential growth can diffuse focus, rendering it challenging for experts to track crucial developments or filter out lower-quality submissions. More problematic is that many preprints lack a robust editorial or retraction infrastructure, potentially perpetuating flawed methods and biases.

Peer Review at a Crossroads: The accelerated publishing cycle has magnified the limitations of traditional peer review, which is a bottleneck that is underscored by limited reviewer availability, extended turnaround times, and the specialized nature of modern AI research [361]. In areas such as MoE-based language modeling or agentic AI, only a handful of qualified reviewers may be available, resulting in possible conflicts of interest or undue delays. Consequently, the academic community grapples with the risk of unverified findings circulating widely and influentially.

Hybrid Models of Validation: Amid these challenges, novel approaches to vetting emerging AI research are under active exploration. One proposal envisions hybrid models of review that combine preliminary, community-driven assessments—akin to product review or code repository feedback—with more formal peer review [362]. In this structure (Figure 8), preprints receive rapid user commentary, highlighting potential strengths or flaws early in the dissemination process. Editors and formal reviewers subsequently refine these critiques, integrating peer evaluations focused on methodological rigor, ethical considerations, and replicability. Advanced AI tools might also be enlisted to flag textual inconsistencies or duplication, which is a measure that could streamline triage and mitigate unethical practices like ghostwriting or auto-generation of entire sections.

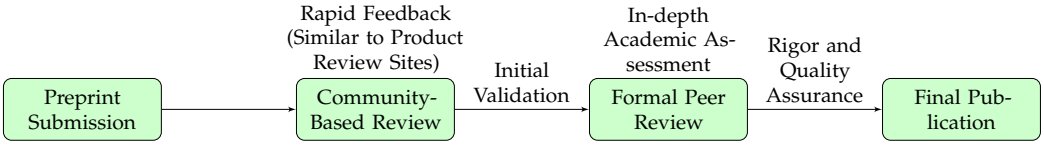

Figure 8. Possible convergence between traditional peer review and the preprint ecosystem.

While no single strategy can fully address the scale and complexity of today's AI publishing surge, these emerging paradigms offer a blueprint for adapting academic norms. The shift to more open, community-oriented, and continuous review processes could provide a partial remedy to the "publish-first, validate-later" culture enabled by generative AI. Equally important is the establishment of best practices, such as labeling AI-generated text, clarifying author roles, and adopting consistent ethics statements. Taken together, these innovations may help sustain academic rigor amid the accelerating influx of AI-driven research.

11. Conclusions

This roadmap survey has embarked on an exploration of the transformative trends in generative AI research, particularly inspired by recent advancements in producing Large Language Models that can solve increasingly complex reasoning tasks via a scaling mechanism that can flexibly apportion more compute resources at inference time. Our analysis highlights that such progress has led to a paradigm shift, which is complemented by additional technological innovations such as Mixture of Experts, agentic AI, multimodal and adaptable AI systems, and the pursuit of AGI. These advancements signal a future where AI systems could significantly extend their capabilities in reasoning, contextual understanding, and creative problem solving. This study reflects on AI's dual potential to either contribute to or impede global equity and justice. The equitable distribution of AI benefits and its role in decision-making processes raise crucial questions about fairness and inclusivity. It is imperative to thoughtfully integrate AI into societal structures to enhance justice and reduce disparities. Despite these advancements, several open questions and research gaps remain. These include ensuring the ethical alignment of advanced AI systems with human values and societal norms, which is a challenge compounded by their increasing autonomy. The safety and robustness of AGI systems in diverse environments

also remain as significant research gaps to explore. Addressing these challenges requires a multidisciplinary approach, incorporating ethical, social, and philosophical perspectives.

Our survey has highlighted key areas for future interdisciplinary research in AI, emphasizing the integration of ethical, sociological, and technical perspectives. This approach will foster collaborative research, bridging the gap between technological advancement and societal needs and ensuring that AI development is aligned with human values and global welfare. The roles of recent technological breakthroughs in generative AI have been identified as significant, as their advancements can enhance model performance and versatility, as well as pave the way for future research in areas like ethical AI alignment and AGI. As we forge ahead, the balance between AI advancements and human creativity is not just a goal but a necessity, ensuring AI's role as a complementary force that amplifies our capacity to innovate and solve complex challenges. Our responsibility is to guide these advancements toward enriching the human experience, aligning technological progress with ethical standards and societal well-being.

Author Contributions: Conceptualization, T.R.M., T.S. and M.N.H.; methodology, T.R.M., T.S. and M.N.H.; software, T.S. and D.L.; validation, D.L., T.L. and P.W.; formal analysis, T.R.M. and P.W.; resources, T.L.; data curation, T.S. and D.X.; writing—original draft preparation, T.R.M., T.L. and M.N.H.; writing—review and editing, T.S., P.W. and D.X.; visualization, T.S.; supervision, T.R.M., T.L. and M.N.H.; project administration, T.R.M. and P.W.; funding acquisition, T.L. and M.N.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data are available upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AGI	Artificial General Intelligence
AI	Artificial Intelligence
AIGC	AI-generated content
BERT	Bidirectional Encoder Representations from Transformers
CCPA	California Consumer Privacy Act
DQN	Deep Q-Networks
EU	European Union
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
GPT	Generative Pretrained Transformers
GPU	Graphics Processing Unit
LIDAR	Light Detection and Ranging
LLM	Large Language Model
LSTM	Long Short-Term Memory
MCTS	Monte Carlo Tree Search
ML	Machine Learning
MoE	Mixture of Experts
NLG	Natural Language Generation
NLP	Natural Language Processing
NLU	Natural Language Understanding
NN	Neural Network
PPO	Proximal Policy Optimization
RNNs	Recurrent Neural Networks

VNN	Value Neural Network
VRAM	Video Random Access Memory

References

1. Turing, A. Computing machinery and intelligence. *Mind* **1950**, *59*, 433. [CrossRef]
2. McDermott, D. Artificial intelligence meets natural stupidity. *Acm Sigart Bull.* **1976**, *57*, 4–9. [CrossRef]
3. Minsky, M. Steps toward artificial intelligence. *Proc. IRE* **1961**, *49*, 8–30. [CrossRef]
4. Yann, L.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
5. Minsky, M.; Papert, S. An introduction to computational geometry. *Camb. Tiass. HIT* **1969**, *479*, 104.
6. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
7. OpenAI. OpenAI o3 Model Announcement, 2024. Available online: https://chatgpt.com (accessed on 13 January 2025).
8. Gemini Team, Google. Gemini: A Family of Highly Capable Multimodal Models, 2023. Available online: https://gemini.google.com (accessed on 17 December 2023).
9. Anthropic. Model Card and Evaluations for Claude Models, 2023. Available online: https://claude.ai (accessed on 13 January 2025).
10. McIntosh, T.R.; Liu, T.; Susnjak, T.; Watters, P.; Ng, A.; Halgamuge, M.N. A Culturally Sensitive Test to Evaluate Nuanced GPT Hallucination. *IEEE Trans. Artif. Intell.* **2023**, *5*, 2739 –2751. [CrossRef]
11. Morris, M.R.; Sohl-dickstein, J.; Fiedel, N.; Warkentin, T.; Dafoe, A.; Faust, A.; Farabet, C.; Legg, S. Levels of AGI: Operationalizing Progress on the Path to AGI. *arXiv* **2023**, arXiv:2311.02462.
12. Schuett, J.; Dreksler, N.; Anderljung, M.; McCaffary, D.; Heim, L.; Bluemke, E.; Garfinkel, B. Towards best practices in AGI safety and governance: A survey of expert opinion. *arXiv* **2023**, arXiv:2305.07153.
13. McIntosh, T.R.; Susnjak, T.; Liu, T.; Watters, P.; Halgamuge, M.N. Inadequacies of large language model benchmarks in the era of generative artificial intelligence. *arXiv* **2024**, arXiv:2402.09880.
14. Singh, A.; Ehtesham, A.; Kumar, S.; Khoei, T.T. Enhancing AI Systems with Agentic Workflows Patterns in Large Language Model. In Proceedings of the 2024 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 29–31 May 2024; pp. 527–532. [CrossRef]
15. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.; Le, Q.; Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *arXiv* **2022**, arXiv:2201.11903. http://arxiv.org/abs/2201.11903.
16. Yao, S.; Yu, D.; Zhao, J.; Shafran, I.; Griffiths, T.L.; Cao, Y.; Narasimhan, K. Tree of Thoughts: Deliberate Problem Solving with Large Language Models. *arXiv* **2023**, arXiv:2305.10601. http://arxiv.org/abs/2305.10601.
17. Bi, Z.; Han, K.; Liu, C.; Tang, Y.; Wang, Y. Forest-of-Thought: Scaling Test-Time Compute for Enhancing LLM Reasoning. *arXiv* **2024**, arXiv:2412.09078.
18. Li, X.; Dong, G.; Jin, J.; Zhang, Y.; Zhou, Y.; Zhu, Y.; Zhang, P.; Dou, Z. Search-o1: Agentic Search-Enhanced Large Reasoning Models. *arXiv* **2025**, arXiv:2501.05366. http://arxiv.org/abs/2501.05366.
19. Chen, Z.; Deng, Y.; Wu, Y.; Gu, Q.; Li, Y. Towards Understanding the Mixture-of-Experts Layer in Deep Learning. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 23049–23062.
20. OpenAI. Learning to Reason with LLMs, 2024. Available online: https://openai.com/index/learning-to-reason-with-llms/ (accessed on 13 January 2025).
21. DeepSeek-AI.; Liu, A.; Feng, B.; Xue, B.; Wang, B.; Wu, B.; Lu, C.; Zhao, C.; Deng, C.; Zhang, C.; et al. DeepSeek-V3 Technical Report. *arXiv* **2024**, arXiv:2412.19437. http://arxiv.org/abs/2412.19437.
22. Zeng, Z.; Cheng, Q.; Yin, Z.; Wang, B.; Li, S.; Zhou, Y.; Guo, Q.; Huang, X.; Qiu, X. Scaling of Search and Learning: A Roadmap to Reproduce o1 from Reinforcement Learning Perspective. *arXiv* **2024**, arXiv:2412.14135. http://arxiv.org/abs/2412.14135.
23. Watch, R. Sleuths Spur Cleanup at Journal with Nearly 140 Retractions and Counting, 2024. Available online: https://retractionwatch.com/2024/08/22/sleuths-spur-cleanup-at-journal-with-nearly-140-retractions-and-counting/ (accessed on 13 January 2025).
24. Watch, R. Springer Nature Journal Has Retracted over 200 Papers Since September, 2024. Available online: https://retractionwatch.com/2024/10/15/springer-nature-journal-has-retracted-over-200-papers-since-september/ (accessed on 13 January 2025).
25. Watch, R. Signs of Undeclared ChatGPT Use in Papers Mounting, 2023. Available online: https://retractionwatch.com/2023/10/06/signs-of-undeclared-chatgpt-use-in-papers-mounting/ (accessed on 13 January 2025).
26. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–13.
27. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. *Improving Language Understanding by Generative Pre-Training*; OpenAI: San Francisco, CA, USA, 2018.

28. Huang, C.; Zhang, Z.; Mao, B.; Yao, X. An overview of artificial intelligence ethics. *IEEE Trans. Artif. Intell.* **2022**, *4*, 799–819. [CrossRef]
29. Besançon, L.; Peiffer-Smadja, N.; Segalas, C.; Jiang, H.; Masuzzo, P.; Smout, C.; Billy, E.; Deforet, M.; Leyrat, C. Open science saves lives: Lessons from the COVID-19 pandemic. *BMC Med. Res. Methodol.* **2021**, *21*, 177. [CrossRef] [PubMed]
30. Triggle, C.R.; MacDonald, R.; Triggle, D.J.; Grierson, D. Requiem for impact factors and high publication charges. *Account. Res.* **2022**, *29*, 133–164. [CrossRef] [PubMed]
31. LeCun, Y. A Path Towards Autonomous Machine Intelligence Version; 2022. Available online: https://openreview.net/pdf?id=BZ5a1r-kVsf (accessed on 13 January 2025).
32. Kambhampati, S. Can large language models reason and plan? *Ann. N. Y. Acad. Sci.* **2024**, *1534*, 15–18. [CrossRef]
33. McIntosh, T.; Kayes, A.; Chen, Y.P.P.; Ng, A.; Watters, P. Ransomware mitigation in the modern era: A comprehensive review, research challenges, and future directions. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 197. [CrossRef]
34. McIntosh, T.; Liu, T.; Susnjak, T.; Alavizadeh, H.; Ng, A.; Nowrozy, R.; Watters, P. Harnessing GPT-4 for generation of cybersecurity GRC policies: A focus on ransomware attack mitigation. *Comput. Secur.* **2023**, *134*, 103424. [CrossRef]
35. Maddigan, P.; Susnjak, T. Chat2vis: Generating data visualisations via natural language using chatgpt, codex and gpt-3 large language models. *IEEE Access* **2023**, *11*, 45181–45193. [CrossRef]
36. Dwivedi, Y.K.; Hughes, L.; Ismagilova, E.; Aarts, G.; Coombs, C.; Crick, T.; Duan, Y.; Dwivedi, R.; Edwards, J.; Eirug, A.; et al. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int. J. Inf. Manag.* **2021**, *57*, 101994. [CrossRef]
37. Gabriel, I. Artificial intelligence, values, and alignment. *Minds Mach.* **2020**, *30*, 411–437. [CrossRef]
38. Shaban-Nejad, A.; Michalowski, M.; Bianco, S.; Brownstein, J.S.; Buckeridge, D.L.; Davis, R.L. Applied artificial intelligence in healthcare: Listening to the winds of change in a post-COVID-19 world. *Exp. Biol. Med.* **2022**, *247*, 1969–1971. [CrossRef]
39. Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y.J.; Madotto, A.; Fung, P. Survey of hallucination in natural language generation. *ACM Comput. Surv.* **2023**, *55*, 1–38. [CrossRef]
40. Min, B.; Ross, H.; Sulem, E.; Veyseh, A.P.B.; Nguyen, T.H.; Sainz, O.; Agirre, E.; Heintz, I.; Roth, D. Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Comput. Surv.* **2023**, *56*, 1–40. [CrossRef]
41. Li, J.; Cheng, X.; Zhao, W.X.; Nie, J.Y.; Wen, J.R. Halueval: A large-scale hallucination evaluation benchmark for large language models. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Singapore, 6–10 December 2023; pp. 6449–6464.
42. Weidinger, L.; Mellor, J.; Rauh, M.; Griffin, C.; Uesato, J.; Huang, P.S.; Cheng, M.; Glaese, M.; Balle, B.; Kasirzadeh, A.; et al. Ethical and social risks of harm from language models. *arXiv* **2021**, arXiv:2112.04359.
43. Xi, Z.; Rui, Z.; Tao, G. Safety and ethical concerns of large language models. In Proceedings of the 22nd Chinese National Conference on Computational Linguistics (Volume 4: Tutorial Abstracts), Harbin, China, 3–5 August 2023; pp. 9–16.
44. Mahmud, D.; Hajmohamed, H.; Almentheri, S.; Alqaydi, S.; Aldhaheri, L.; Khalil, R.A.; Saeed, N. Integrating LLMs with ITS: Recent Advances, Potentials, Challenges, and Future Directions. *arXiv* **2025**, arXiv:2501.04437. [CrossRef]
45. Sujan, M.; Slater, D.; Crumpton, E. How can large language models assist with a FRAM analysis? *Saf. Sci.* **2025**, *181*, 106695. [CrossRef]
46. Brown, P.F.; Della Pietra, V.J.; Desouza, P.V.; Lai, J.C.; Mercer, R.L. Class-based n-gram models of natural language. *Comput. Linguist.* **1992**, *18*, 467–480.
47. Katz, S. Estimation of probabilities from sparse data for the language model component of a speech recognizer. *IEEE Trans. Acoust. Speech Signal Process.* **1987**, *35*, 400–401. [CrossRef]
48. Kneser, R.; Ney, H. Improved backing-off for m-gram language modeling. In Proceedings of the 1995 International Conference on Acoustics, Speech, And Signal Processing, Detroit, MI, USA, 9–12 May 1995; Volume 1, pp. 181–184.
49. Kuhn, R.; De Mori, R. A cache-based natural language model for speech recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 570–583. [CrossRef]
50. Ney, H.; Essen, U.; Kneser, R. On structuring probabilistic dependences in stochastic language modelling. *Comput. Speech Lang.* **1994**, *8*, 1–38. [CrossRef]
51. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
52. Nammous, M.K.; Saeed, K. Natural language processing: Speaker, language, and gender identification with LSTM. In *Advanced Computing and Systems for Security*; Springer: Singapore, 2019; Volume 883, pp. 143–156.
53. Wei, D.; Wang, B.; Lin, G.; Liu, D.; Dong, Z.; Liu, H.; Liu, Y. Research on unstructured text data mining and fault classification based on RNN-LSTM with malfunction inspection report. *Energies* **2017**, *10*, 406. [CrossRef]
54. Yao, L.; Guan, Y. An improved LSTM structure for natural language processing. In Proceedings of the 2018 IEEE International Conference of Safety Produce Informatization (IICSPI), Chongqing, China, 10–12 December 2018; pp. 565–569.
55. Chen, M.; Tworek, J.; Jun, H.; Yuan, Q.; Pinto, H.P.d.O.; Kaplan, J.; Edwards, H.; Burda, Y.; Joseph, N.; Brockman, G.; et al. Evaluating Large Language Models Trained on Code. *arXiv* **2021**, arXiv:2107.03374.

56. Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. Training language models to follow instructions with human feedback. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 27730–27744.
57. Susnjak, T. Beyond Predictive Learning Analytics Modelling and onto Explainable Artificial Intelligence with Prescriptive Analytics and ChatGPT. *Int. J. Artif. Intell. Educ.* **2023**, *34*, 452–482. [CrossRef]
58. Susnjak, T.; Griffin, E.; McCutcheon, M.; Potter, K. Towards Clinical Prediction with Transparency: An Explainable AI Approach to Survival Modelling in Residential Aged Care. *arXiv* **2023**, arXiv:2312.00271.
59. Yang, R.; Tan, T.F.; Lu, W.; Thirunavukarasu, A.J.; Ting, D.S.W.; Liu, N. Large language models in health care: Development, applications, and challenges. *Health Care Sci.* **2023**, *2*, 255–263. [CrossRef]
60. Baidoo-Anu, D.; Ansah, L.O. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *J. AI* **2023**, *7*, 52–62. [CrossRef]
61. Susnjak, T. ChatGPT: The end of online exam integrity? *arXiv* **2022**, arXiv:2212.09292.
62. Tlili, A.; Shehata, B.; Adarkwah, M.A.; Bozkurt, A.; Hickey, D.T.; Huang, R.; Agyemang, B. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learn. Environ.* **2023**, *10*, 15. [CrossRef]
63. AlAfnan, M.A.; Dishari, S.; Jovic, M.; Lomidze, K. Chatgpt as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *J. Artif. Intell. Technol.* **2023**, *3*, 60–68. [CrossRef]
64. George, A.S.; George, A.H. A review of ChatGPT AI's impact on several business sectors. *Partners Univers. Int. Innov. J.* **2023**, *1*, 9–23.
65. Hadfield, G.K.; Clark, J. Regulatory Markets: The Future of AI Governance. *arXiv* **2023**, arXiv:2304.04914.
66. LaGrandeur, K. How safe is our reliance on AI, and should we regulate it? *AI Ethics* **2021**, *1*, 93–99. [CrossRef]
67. McLean, S.; Read, G.J.; Thompson, J.; Baber, C.; Stanton, N.A.; Salmon, P.M. The risks associated with Artificial General Intelligence: A systematic review. *J. Exp. Theor. Artif. Intell.* **2023**, *35*, 649–663. [CrossRef]
68. Mitchell, M. AI's challenge of understanding the world. *Science* **2023**, *382*, eadm8175. [CrossRef] [PubMed]
69. Black, S.; Gao, L.; Wang, P.; Leahy, C.; Biderman, S. GPT-NeoX-20B: An Open-Source Autoregressive Language Model, 2022. Available online: https://github.com/EleutherAI/gpt-neox (accessed on 9 January 2025).
70. Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. Llama 2: Open Foundation and Fine-Tuned Chat Models. *arXiv* **2023**, arXiv:2307.09288.
71. Bakker, M.; Chadwick, M.; Sheahan, H.; Tessler, M.; Campbell-Gillingham, L.; Balaguer, J.; McAleese, N.; Glaese, A.; Aslanides, J.; Botvinick, M.; et al. Fine-tuning language models to find agreement among humans with diverse preferences. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 38176–38189.
72. Hu, Z.; Lan, Y.; Wang, L.; Xu, W.; Lim, E.P.; Lee, R.K.W.; Bing, L.; Poria, S. LLM-Adapters: An Adapter Family for Parameter-Efficient Fine-Tuning of Large Language Models. *arXiv* **2023**, arXiv:2304.01933.
73. Liu, H.; Tam, D.; Muqeeth, M.; Mohta, J.; Huang, T.; Bansal, M.; Raffel, C.A. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 1950–1965.
74. Zheng, H.; Shen, L.; Tang, A.; Luo, Y.; Hu, H.; Du, B.; Tao, D. Learn From Model Beyond Fine-Tuning: A Survey. *arXiv* **2023**, arXiv:2310.08184.
75. Manakul, P.; Liusie, A.; Gales, M.J. Selfcheckgpt: Zero-resource black-box hallucination detection for generative large language models. *arXiv* **2023**, arXiv:2303.08896.
76. Martino, A.; Iannelli, M.; Truong, C. Knowledge injection to counter large language model (llm) hallucination. In Proceedings of the European Semantic Web Conference, Hersonissos, Crete, Greece, 31 May 2023; pp. 182–185.
77. Yao, J.Y.; Ning, K.P.; Liu, Z.H.; Ning, M.N.; Yuan, L. Llm lies: Hallucinations are not bugs, but features as adversarial examples. *arXiv* **2023**, arXiv:2310.01469.
78. Zhang, Y.; Li, Y.; Cui, L.; Cai, D.; Liu, L.; Fu, T.; Huang, X.; Zhao, E.; Zhang, Y.; Chen, Y.; et al. Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. *arXiv* **2023**, arXiv:2309.01219.
79. Ji, J.; Liu, M.; Dai, J.; Pan, X.; Zhang, C.; Bian, C.; Sun, R.; Wang, Y.; Yang, Y. Beavertails: Towards improved safety alignment of llm via a human-preference dataset. *arXiv* **2023**, arXiv:2307.04657.
80. Liu, Y.; Yao, Y.; Ton, J.F.; Zhang, X.; Cheng, R.G.; Klochkov, Y.; Taufiq, M.F.; Li, H. Trustworthy LLMs: A Survey and Guideline for Evaluating Large Language Models' Alignment. *arXiv* **2023**, arXiv:2308.05374.
81. Wang, Y.; Zhong, W.; Li, L.; Mi, F.; Zeng, X.; Huang, W.; Shang, L.; Jiang, X.; Liu, Q. Aligning large language models with human: A survey. *arXiv* **2023**, arXiv:2307.12966.
82. Sun, Z.; Shen, Y.; Zhou, Q.; Zhang, H.; Chen, Z.; Cox, D.; Yang, Y.; Gan, C. Principle-driven self-alignment of language models from scratch with minimal human supervision. *arXiv* **2023**, arXiv:2305.03047.
83. Wolf, Y.; Wies, N.; Levine, Y.; Shashua, A. Fundamental limitations of alignment in large language models. *arXiv* **2023**, arXiv:2304.11082.
84. Dang, H.; Mecke, L.; Lehmann, F.; Goller, S.; Buschek, D. How to prompt? Opportunities and challenges of zero-and few-shot learning for human-AI interaction in creative applications of generative models. *arXiv* **2022**, arXiv:2209.01390.

85. Ma, R.; Zhou, X.; Gui, T.; Tan, Y.; Li, L.; Zhang, Q.; Huang, X. Template-free prompt tuning for few-shot NER. *arXiv* **2021**, arXiv:2109.13532.
86. Qin, C.; Joty, S. LFPT5: A unified framework for lifelong few-shot language learning based on prompt tuning of t5. *arXiv* **2021**, arXiv:2110.07298.
87. Wang, S.; Tang, L.; Majety, A.; Rousseau, J.F.; Shih, G.; Ding, Y.; Peng, Y. Trustworthy assertion classification through prompting. *J. Biomed. Inform.* **2022**, *132*, 104139. [CrossRef] [PubMed]
88. Fan, Y.; Jiang, F.; Li, P.; Li, H. GrammarGPT: Exploring Open-Source LLMs for Native Chinese Grammatical Error Correction with Supervised Fine-Tuning. In Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Foshan, China, 12–15 October 2023; pp. 69–80.
89. Liga, D.; Robaldo, L. Fine-tuning GPT-3 for legal rule classification. *Comput. Law Secur. Rev.* **2023**, *51*, 105864. [CrossRef]
90. Liu, Y.; Singh, A.; Freeman, C.D.; Co-Reyes, J.D.; Liu, P.J. Improving Large Language Model Fine-tuning for Solving Math Problems. *arXiv* **2023**, arXiv:2310.10047.
91. Talat, Z.; Névéol, A.; Biderman, S.; Clinciu, M.; Dey, M.; Longpre, S.; Luccioni, S.; Masoud, M.; Mitchell, M.; Radev, D.; et al. You reap what you sow: On the challenges of bias evaluation under multilingual settings. In Proceedings of the BigScience Episode# 5–Workshop on Challenges & Perspectives in Creating Large Language Models, Dublin, Ireland, 27 May 2022; pp. 26–41.
92. Liu, Y.; Yu, S.; Lin, T. Hessian regularization of deep neural networks: A novel approach based on stochastic estimators of Hessian trace. *Neurocomputing* **2023**, *536*, 13–20. [CrossRef]
93. Lu, Y.; Bo, Y.; He, W. Confidence adaptive regularization for deep learning with noisy labels. *arXiv* **2021**, arXiv:2108.08212.
94. Pereyra, G.; Tucker, G.; Chorowski, J.; Kaiser, Ł.; Hinton, G. Regularizing neural networks by penalizing confident output distributions. *arXiv* **2017**, arXiv:1701.06548.
95. Chen, E.; Hong, Z.W.; Pajarinen, J.; Agrawal, P. Redeeming intrinsic rewards via constrained optimization. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 4996–5008.
96. Jiang, Y.; Li, Z.; Tan, M.; Wei, S.; Zhang, G.; Guan, Z.; Han, B. A stable block adjustment method without ground control points using bound constrained optimization. *Int. J. Remote Sens.* **2022**, *43*, 4708–4722. [CrossRef]
97. Kachuee, M.; Lee, S. Constrained policy optimization for controlled self-learning in conversational AI systems. *arXiv* **2022**, arXiv:2209.08429.
98. Song, Z.; Wang, H.; Jin, Y. A Surrogate-Assisted Evolutionary Framework With Regions of Interests-Based Data Selection for Expensive Constrained Optimization. *IEEE Trans. Syst. Man Cybern. Syst.* **2023**, *53*, 6268–6280. [CrossRef]
99. Yu, J.; Xu, T.; Rong, Y.; Huang, J.; He, R. Structure-aware conditional variational auto-encoder for constrained molecule optimization. *Pattern Recognit.* **2022**, *126*, 108581. [CrossRef]
100. Butlin, P. AI alignment and human reward. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, Virtual, 19–21 May 2021; pp. 437–445.
101. Faal, F.; Schmitt, K.; Yu, J.Y. Reward modeling for mitigating toxicity in transformer-based language models. *Appl. Intell.* **2023**, *53*, 8421–8435. [CrossRef]
102. Leike, J.; Krueger, D.; Everitt, T.; Martic, M.; Maini, V.; Legg, S. Scalable agent alignment via reward modeling: A research direction. *arXiv* **2018**, arXiv:1811.07871.
103. Li, L.; Chai, Y.; Wang, S.; Sun, Y.; Tian, H.; Zhang, N.; Wu, H. Tool-Augmented Reward Modeling. *arXiv* **2023**, arXiv:2310.01045.
104. McIntosh, T.R.; Susnjak, T.; Liu, T.; Watters, P.; Halgamuge, M.N. The Inadequacy of Reinforcement Learning from Human Feedback - Radicalizing Large Language Models via Semantic Vulnerabilities. *IEEE Trans. Cogn. Dev. Syst.* **2024**, *16*, 1561–1574. [CrossRef]
105. Barreto, F.; Moharkar, L.; Shirodkar, M.; Sarode, V.; Gonsalves, S.; Johns, A. Generative Artificial Intelligence: Opportunities and Challenges of Large Language Models. In Proceedings of the International Conference on Intelligent Computing and Networking, Mumbai, India, 24–25 February 2023; pp. 545–553.
106. Chen, Z.; Wang, Z.; Wang, Z.; Liu, H.; Yin, Z.; Liu, S.; Sheng, L.; Ouyang, W.; Qiao, Y.; Shao, J. Octavius: Mitigating Task Interference in MLLMs via MoE. *arXiv* **2023**, arXiv:2311.02684.
107. Dun, C.; Garcia, M.D.C.H.; Zheng, G.; Awadallah, A.H.; Kyrillidis, A.; Sim, R. Sweeping Heterogeneity with Smart MoPs: Mixture of Prompts for LLM Task Adaptation. *arXiv* **2023**, arXiv:2310.02842.
108. Naveed, H.; Khan, A.U.; Qiu, S.; Saqib, M.; Anwar, S.; Usman, M.; Barnes, N.; Mian, A. A comprehensive overview of large language models. *arXiv* **2023**, arXiv:2307.06435.
109. Xue, F.; Fu, Y.; Zhou, W.; Zheng, Z.; You, Y. To Repeat or Not To Repeat: Insights from Scaling LLM under Token-Crisis. *arXiv* **2023**, arXiv:2305.13230.
110. Vats, A.; Raja, R.; Jain, V.; Chadha, A. The Evolution of Mixture of Experts: A Survey from Basics to Breakthroughs. *Preprints* **2024**, 2024080583.
111. Lin, B.; Tang, Z.; Ye, Y.; Cui, J.; Zhu, B.; Jin, P.; Zhang, J.; Ning, M.; Yuan, L. Moe-llava: Mixture of experts for large vision-language models. *arXiv* **2024**, arXiv:2401.15947.

112. Tian, Y.; Xia, F.; Song, Y. Dialogue summarization with mixture of experts based on large language models. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, Bangkok, Thailand, 11–16 August 2024; pp. 7143–7155.
113. Li, Y.; Jiang, S.; Hu, B.; Wang, L.; Zhong, W.; Luo, W.; Ma, L.; Zhang, M. Uni-MoE: Scaling Unified Multimodal LLMs with Mixture of Experts. *arXiv* **2024**, arXiv:2405.11273.
114. Fedus, W.; Dean, J.; Zoph, B. A review of sparse expert models in deep learning. *arXiv* **2022**, arXiv:2209.01667.
115. Nowaz Rabbani Chowdhury, M.; Zhang, S.; Wang, M.; Liu, S.; Chen, P.Y. Patch-level Routing in Mixture-of-Experts is Provably Sample-efficient for Convolutional Neural Networks. *arXiv* **2023**, arXiv:2306.04073.
116. Santos, C.N.d.; Lee-Thorp, J.; Noble, I.; Chang, C.C.; Uthus, D. Memory Augmented Language Models through Mixture of Word Experts. *arXiv* **2023**, arXiv:2311.10768.
117. Wang, W.; Ma, G.; Li, Y.; Du, B. Language-Routing Mixture of Experts for Multilingual and Code-Switching Speech Recognition. *arXiv* **2023**, arXiv:2307.05956.
118. Zhao, X.; Chen, X.; Cheng, Y.; Chen, T. Sparse MoE with Language Guided Routing for Multilingual Machine Translation. In Proceedings of the Twelfth International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024.
119. Huang, W.; Zhang, H.; Peng, P.; Wang, H. Multi-gate Mixture-of-Expert Combined with Synthetic Minority Over-sampling Technique for Multimode Imbalanced Fault Diagnosis. In Proceedings of the 2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Rio de Janeiro, Brazil, 24–26 May 2023; pp. 456–461.
120. Liu, B.; Ding, L.; Shen, L.; Peng, K.; Cao, Y.; Cheng, D.; Tao, D. Diversifying the Mixture-of-Experts Representation for Language Models with Orthogonal Optimizer. *arXiv* **2023**, arXiv:2310.09762.
121. Wang, W.; Lai, Z.; Li, S.; Liu, W.; Ge, K.; Liu, Y.; Shen, A.; Li, D. Prophet: Fine-grained Load Balancing for Parallel Training of Large-scale MoE Models. In Proceedings of the 2023 IEEE International Conference on Cluster Computing (CLUSTER), Santa Fe, NM, USA, 31 October–3 November 2023; pp. 82–94.
122. Yao, X.; Liang, S.; Han, S.; Huang, H. Enhancing Molecular Property Prediction via Mixture of Collaborative Experts. *arXiv* **2023**, arXiv:2312.03292.
123. Xiao, Z.; Jiang, Y.; Tang, G.; Liu, L.; Xu, S.; Xiao, Y.; Yan, W. Adversarial mixture of experts with category hierarchy soft constraint. In Proceedings of the 2021 IEEE 37th International Conference on Data Engineering (ICDE), Chania, Greece, 19–22 April 2021; pp. 2453–2463.
124. Agbese, M.; Mohanani, R.; Khan, A.; Abrahamsson, P. Implementing AI Ethics: Making Sense of the Ethical Requirements. In Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering, Oulu, Finland, 14–16 June 2023; pp. 62–71.
125. Zhou, Y.; Lei, T.; Liu, H.; Du, N.; Huang, Y.; Zhao, V.; Dai, A.M.; Le, Q.V.; Laudon, J. Mixture-of-experts with expert choice routing. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 7103–7114.
126. Guha, N.; Lawrence, C.; Gailmard, L.A.; Rodolfa, K.; Surani, F.; Bommasani, R.; Raji, I.; Cuéllar, M.F.; Honigsberg, C.; Liang, P.; et al. AI Regulation Has Its Own Alignment Problem: The Technical and Institutional Feasibility of Disclosure, Registration, Licensing, and Auditing. *Georg. Wash. Law Rev.* **2024**, *92*, 1473–1557.
127. Team, G.; Georgiev, P.; Lei, V.I.; Burnell, R.; Bai, L.; Gulati, A.; Tanzer, G.; Vincent, D.; Pan, Z.; Wang, S. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv* **2024**, arXiv:2403.05530.
128. Team, G.; Anil, R.; Borgeaud, S.; Alayrac, J.B.; Yu, J.; Soricut, R.; Schalkwyk, J.; Dai, A.M.; Hauth, A.; Millican, K. Gemini: A family of highly capable multimodal models. *arXiv* **2023**, arXiv:2312.11805.
129. Alsajri, A.; Salman, H.A.; Steiti, A. Generative Models in Natural Language Processing: A Comparative Study of ChatGPT and Gemini. *Babylon. J. Artif. Intell.* **2024**, *2024*, 134–145. [CrossRef]
130. Yang, L.; Xu, S.; Sellergren, A.; Kohlberger, T.; Zhou, Y.; Ktena, I.; Kiraly, A.; Ahmed, F.; Hormozdiari, F.; Jaroensri, T. Advancing multimodal medical capabilities of Gemini. *arXiv* **2024**, arXiv:2405.03162.
131. Lewandowska-Tomaszczyk, B.; Liebeskind, C. Opinion events and stance types: Advances in LLM performance with ChatGPT and Gemini. *Lodz Pap. Pragmat.* **2024**, *20*, 413–432. [CrossRef]
132. Google DeepMind. *Introducing Gemini 2.0: Our New AI Model for the Agentic Era*; Google: Mountain View, CA, USA, 2024.
133. Akhtar, Z.B. From bard to Gemini: An investigative exploration journey through Google's evolution in conversational AI and generative AI. *Comput. Artif. Intell.* **2024**, *2*, 1378. [CrossRef]
134. Acosta, J.N.; Falcone, G.J.; Rajpurkar, P.; Topol, E.J. Multimodal biomedical AI. *Nat. Med.* **2022**, *28*, 1773–1784. [CrossRef]
135. Qi, S.; Cao, Z.; Rao, J.; Wang, L.; Xiao, J.; Wang, X. What is the limitation of multimodal LLMs? A deeper look into multimodal LLMs through prompt probing. *Inf. Process. Manag.* **2023**, *60*, 103510. [CrossRef]
136. Xu, B.; Kocyigit, D.; Grimm, R.; Griffin, B.P.; Cheng, F. Applications of artificial intelligence in multimodality cardiovascular imaging: A state-of-the-art review. *Prog. Cardiovasc. Dis.* **2020**, *63*, 367–376. [CrossRef] [PubMed]
137. Birhane, A.; Prabhu, V.U.; Kahembwe, E. Multimodal datasets: Misogyny, pornography, and malignant stereotypes. *arXiv* **2021**, arXiv:2110.01963.

138. Li, Y.; Li, W.; Li, N.; Qiu, X.; Manokaran, K.B. Multimodal information interaction and fusion for the parallel computing system using AI techniques. *Int. J. High Perform. Syst. Archit.* **2021**, *10*, 185–196. [CrossRef]
139. Zhang, C.; Yang, Z.; He, X.; Deng, L. Multimodal intelligence: Representation learning, information fusion, and applications. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 478–493. [CrossRef]
140. Qiao, H.; Liu, V.; Chilton, L. Initial images: Using image prompts to improve subject representation in multimodal ai generated art. In Proceedings of the 14th Conference on Creativity and Cognition, Venice, Italy, 20–23 June 2022; pp. 15–28.
141. Stewart, A.E.; Keirn, Z.; D'Mello, S.K. Multimodal modeling of collaborative problem-solving facets in triads. *User Model. User-Adapt. Interact.* **2021**, *34*, 713–751. [CrossRef]
142. Xue, L.; Yu, N.; Zhang, S.; Li, J.; Martín-Martín, R.; Wu, J.; Xiong, C.; Xu, R.; Niebles, J.C.; Savarese, S. ULIP-2: Towards Scalable Multimodal Pre-training For 3D Understanding. *arXiv* **2023**, arXiv:2305.08275.
143. Yan, L.; Zhao, L.; Gasevic, D.; Martinez-Maldonado, R. Scalability, sustainability, and ethicality of multimodal learning analytics. In Proceedings of the LAK22: 12th International Learning Analytics and Knowledge Conference, Online, 21–25 March 2022; pp. 13–23.
144. Liu-Thompkins, Y.; Okazaki, S.; Li, H. Artificial empathy in marketing interactions: Bridging the human-AI gap in affective and social customer experience. *J. Acad. Mark. Sci.* **2022**, *50*, 1198–1218. [CrossRef]
145. Rahman, M.S.; Bag, S.; Hossain, M.A.; Fattah, F.A.M.A.; Gani, M.O.; Rana, N.P. The new wave of AI-powered luxury brands online shopping experience: The role of digital multisensory cues and customers' engagement. *J. Retail. Consum. Serv.* **2023**, *72*, 103273. [CrossRef]
146. Sachdeva, E.; Agarwal, N.; Chundi, S.; Roelofs, S.; Li, J.; Dariush, B.; Choi, C.; Kochenderfer, M. Rank2tell: A multimodal driving dataset for joint importance ranking and reasoning. *arXiv* **2023**, arXiv:2309.06597.
147. Cui, C.; Ma, Y.; Cao, X.; Ye, W.; Zhou, Y.; Liang, K.; Chen, J.; Lu, J.; Yang, Z.; Liao, K.D.; et al. A survey on multimodal large language models for autonomous driving. *arXiv* **2023**, arXiv:2311.12320.
148. Temsamani, A.B.; Chavali, A.K.; Vervoort, W.; Tuytelaars, T.; Radevski, G.; Van Hamme, H.; Mets, K.; Hutsebaut-Buysse, M.; De Schepper, T.; Latré, S. A multimodal AI approach for intuitively instructable autonomous systems: A case study of an autonomous off-highway vehicle. In Proceedings of the Eighteenth International Conference on Autonomic and Autonomous Systems, ICAS 2022, Venice, Italy, 22–26 May 2022; pp. 31–39.
149. Lee, J.; Shin, S.Y. Something that they never said: Multimodal disinformation and source vividness in understanding the power of AI-enabled deepfake news. *Media Psychol.* **2022**, *25*, 531–546. [CrossRef]
150. Muppalla, S.; Jia, S.; Lyu, S. Integrating Audio-Visual Features for Multimodal Deepfake Detection. *arXiv* **2023**, arXiv:2310.03827.
151. Kumar, S.; Chaube, M.K.; Nenavath, S.N.; Gupta, S.K.; Tetarave, S.K. Privacy preservation and security challenges: A new frontier multimodal machine learning research. *Int. J. Sens. Netw.* **2022**, *39*, 227–245. [CrossRef]
152. Marchang, J.; Di Nuovo, A. Assistive multimodal robotic system (AMRSys): Security and privacy issues, challenges, and possible solutions. *Appl. Sci.* **2022**, *12*, 2174. [CrossRef]
153. Peña, A.; Serna, I.; Morales, A.; Fierrez, J.; Ortega, A.; Herrarte, A.; Alcantara, M.; Ortega-Garcia, J. Human-centric multimodal machine learning: Recent advances and testbed on AI-based recruitment. *SN Comput. Sci.* **2023**, *4*, 434. [CrossRef]
154. Wolfe, R.; Caliskan, A. American==White in multimodal language-and-image ai. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, Oxford, UK, 1–3 August 2022; pp. 800–812.
155. Wolfe, R.; Yang, Y.; Howe, B.; Caliskan, A. Contrastive language-vision ai models pretrained on web-scraped multimodal data exhibit sexual objectification bias. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, 12–15 June 2023; pp. 1174–1185.
156. Afshar, M.; Sharma, B.; Dligach, D.; Oguss, M.; Brown, R.; Chhabra, N.; Thompson, H.M.; Markossian, T.; Joyce, C.; Churpek, M.M.; et al. Development and multimodal validation of a substance misuse algorithm for referral to treatment using artificial intelligence (SMART-AI): A retrospective deep learning study. *Lancet Digit. Health* **2022**, *4*, e426–e435. [CrossRef] [PubMed]
157. Alwahaby, H.; Cukurova, M.; Papamitsiou, Z.; Giannakos, M. The evidence of impact and ethical considerations of Multimodal Learning Analytics: A Systematic Literature Review. In *The Multimodal Learning Analytics Handbook*; Springer: Cham, Switzerland, 2022; pp. 289–325.
158. Wu, W.; Mao, S.; Zhang, Y.; Xia, Y.; Dong, L.; Cui, L.; Wei, F. Visualization-of-Thought Elicits Spatial Reasoning in Large Language Models. *arXiv* **2024**, arXiv:2404.03622.
159. Behrouz, A.; Zhong, P.; Mirrokni, V. Titans: Learning to Memorize at Test Time. *arXiv* **2024**, arXiv:2501.00663.
160. Mitchell, M. How do we know how smart AI systems are? *Science* **2023**, *381*, eadj5957. [CrossRef]
161. Miao, Q.; Zheng, W.; Lv, Y.; Huang, M.; Ding, W.; Wang, F.Y. DAO to HANOI via DeSci: AI paradigm shifts from AlphaGo to ChatGPT. *IEEE/CAA J. Autom. Sin.* **2023**, *10*, 877–897. [CrossRef]
162. Rong, Y. Roadmap of AlphaGo to AlphaStar: Problems and challenges. In Proceedings of the 2nd International Conference on Artificial Intelligence, Automation, and High-Performance Computing (AIAHPC 2022), Zhuhai, China, 25–27 February 2022; Volume 12348, pp. 904–914.

163. Gao, Y.; Zhou, M.; Liu, D.; Yan, Z.; Zhang, S.; Metaxas, D.N. A data-scalable transformer for medical image segmentation: Architecture, model efficiency, and benchmark. *arXiv* **2022**, arXiv:2203.00131.
164. Peebles, W.; Xie, S. Scalable diffusion models with transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 4195–4205.
165. Pope, R.; Douglas, S.; Chowdhery, A.; Devlin, J.; Bradbury, J.; Heek, J.; Xiao, K.; Agrawal, S.; Dean, J. Efficiently scaling transformer inference. In Proceedings of the 6th Machine Learning and Systems, Miami, FL, USA, 4–8 June 2023.
166. Ding, Y.; Jia, M. Convolutional transformer: An enhanced attention mechanism architecture for remaining useful life estimation of bearings. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3515010. [CrossRef]
167. Ding, Y.; Jia, M.; Miao, Q.; Cao, Y. A novel time–frequency Transformer based on self–attention mechanism and its application in fault diagnosis of rolling bearings. *Mech. Syst. Signal Process.* **2022**, *168*, 108616. [CrossRef]
168. Wang, G.; Zhao, Y.; Tang, C.; Luo, C.; Zeng, W. When shift operation meets vision transformer: An extremely simple alternative to attention mechanism. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2022; Volume 36, pp. 2423–2430.
169. Cai, H.; Li, J.; Hu, M.; Gan, C.; Han, S. EfficientViT: Lightweight Multi-Scale Attention for High-Resolution Dense Prediction. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 17302–17313.
170. Liu, X.; Peng, H.; Zheng, N.; Yang, Y.; Hu, H.; Yuan, Y. EfficientViT: Memory Efficient Vision Transformer with Cascaded Group Attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 14420–14430.
171. Li, Y.; Fan, Q.; Huang, H.; Han, Z.; Gu, Q. A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition. *Drones* **2023**, *7*, 304. [CrossRef]
172. Talaat, F.M.; ZainEldin, H. An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Comput. Appl.* **2023**, *35*, 20939–20954. [CrossRef]
173. Tamang, S.; Sen, B.; Pradhan, A.; Sharma, K.; Singh, V.K. Enhancing covid-19 safety: Exploring yolov8 object detection for accurate face mask classification. *Int. J. Intell. Syst. Appl. Eng.* **2023**, *11*, 892–897.
174. Lu, J.; Xiong, R.; Tian, J.; Wang, C.; Hsu, C.W.; Tsou, N.T.; Sun, F.; Li, J. Battery degradation prediction against uncertain future conditions with recurrent neural network enabled deep learning. *Energy Storage Mater.* **2022**, *50*, 139–151. [CrossRef]
175. Onan, A. Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification. *J. King Saud-Univ.-Comput. Inf. Sci.* **2022**, *34*, 2098–2117. [CrossRef]
176. Shan, F.; He, X.; Armaghani, D.J.; Zhang, P.; Sheng, D. Success and challenges in predicting TBM penetration rate using recurrent neural networks. *Tunn. Undergr. Space Technol.* **2022**, *130*, 104728. [CrossRef]
177. Sridhar, C.; Pareek, P.K.; Kalidoss, R.; Jamal, S.S.; Shukla, P.K.; Nuagah, S.J. Optimal medical image size reduction model creation using recurrent neural network and GenPSOWVQ. *J. Healthc. Eng.* **2022**, *2022*, 2354866. [CrossRef]
178. Zhu, J.; Jiang, Q.; Shen, Y.; Qian, C.; Xu, F.; Zhu, Q. Application of recurrent neural network to mechanical fault diagnosis: A review. *J. Mech. Sci. Technol.* **2022**, *36*, 527–542. [CrossRef]
179. Lin, S.; Lin, W.; Wu, W.; Zhao, F.; Mo, R.; Zhang, H. Segrnn: Segment recurrent neural network for long-term time series forecasting. *arXiv* **2023**, arXiv:2308.11200.
180. Wei, Z.; Zhang, X.; Sun, M. Extracting weighted finite automata from recurrent neural networks for natural languages. In Proceedings of the International Conference on Formal Engineering Methods, ICFEM 2022, Madrid, Spain, 24–27 October 2022; pp. 370–385.
181. Bonassi, F.; Farina, M.; Xie, J.; Scattolini, R. On recurrent neural networks for learning-based control: Recent results and ideas for future developments. *J. Process. Control* **2022**, *114*, 92–104. [CrossRef]
182. Büchel, J.; Vasilopoulos, A.; Simon, W.A.; Boybat, I.; Tsai, H.; Burr, G.W.; Castro, H.; Filipiak, B.; Le Gallo, M.; Rahimi, A. Efficient scaling of large language models with mixture of experts and 3D analog in-memory computing. *Nat. Comput. Sci.* **2025**, *5*, 13–26. [CrossRef]
183. Liao, X.; Sun, Y.; Tian, H.; Wan, X.; Jin, Y.; Wang, Z.; Ren, Z.; Huang, X.; Li, W.; Tse, K.F. mFabric: An Efficient and Scalable Fabric for Mixture-of-Experts Training. *arXiv* **2025**, arXiv:2501.03905.
184. Park, B.; Go, H.; Kim, J.Y.; Woo, S.; Ham, S.; Kim, C. *Switch Diffusion Transformer: Synergizing Denoising Tasks with Sparse Mixture-of-Experts*; Springer: Cham, Switzerland, 2025; pp. 461–477.
185. Di Sario, F.; Renzulli, R.; Tartaglione, E.; Grangetto, M. *Boost Your NeRF: A Model-Agnostic Mixture of Experts Framework for High Quality and Efficient Rendering*; Springer: Cham, Switzerland, 2025; pp. 176–192.
186. Du, N.; Huang, Y.; Dai, A.M.; Tong, S.; Lepikhin, D.; Xu, Y.; Krikun, M.; Zhou, Y.; Yu, A.W.; Firat, O.; et al. Glam: Efficient scaling of language models with mixture-of-experts. In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 5547–5569.

187. Dhivya, K.; Kumar, S.N.; Victoria, D.R.S.; Sherly, S.I.; Durgadevi, G. Advanced Neural Networks for Multimodal Data Fusion in Interdisciplinary Research. In *Advanced Interdisciplinary Applications of Deep Learning for Data Science*; IGI Global Scientific Publishing: Hershey, PA, USA, 2025; pp. 201–232.
188. Guo, Z.; Tang, Y.; Zhang, R.; Wang, D.; Wang, Z.; Zhao, B.; Li, X. Viewrefer: Grasp the multi-view knowledge for 3d visual grounding. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 15372–15383.
189. Pan, C.; He, Y.; Peng, J.; Zhang, Q.; Sui, W.; Zhang, Z. BAEFormer: Bi-Directional and Early Interaction Transformers for Bird's Eye View Semantic Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 9590–9599.
190. Xu, P.; Zhu, X.; Clifton, D.A. Multimodal learning with transformers: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 12113–12132. [CrossRef] [PubMed]
191. Molenaar, I.; de Mooij, S.; Azevedo, R.; Bannert, M.; Järvelä, S.; Gašević, D. Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. *Comput. Hum. Behav.* **2023**, *139*, 107540. [CrossRef]
192. Steyaert, S.; Pizurica, M.; Nagaraj, D.; Khandelwal, P.; Hernandez-Boussard, T.; Gentles, A.J.; Gevaert, O. Multimodal data fusion for cancer biomarker discovery with deep learning. *Nat. Mach. Intell.* **2023**, *5*, 351–362. [CrossRef]
193. Rani, V.; Nabi, S.T.; Kumar, M.; Mittal, A.; Kumar, K. Self-supervised learning: A succinct review. *Arch. Comput. Methods Eng.* **2023**, *30*, 2761–2775. [CrossRef] [PubMed]
194. Schiappa, M.C.; Rawat, Y.S.; Shah, M. Self-supervised learning for videos: A survey. *ACM Comput. Surv.* **2023**, *55*, 228. [CrossRef]
195. Yu, J.; Yin, H.; Xia, X.; Chen, T.; Li, J.; Huang, Z. Self-supervised learning for recommender systems: A survey. *IEEE Trans. Knowl. Data Eng.* **2023**, *36*, 335–355. [CrossRef]
196. Bharti, V.; Kumar, A.; Purohit, V.; Singh, R.; Singh, A.K.; Singh, S.K. A Label Efficient Semi Self-Supervised Learning Framework for IoT Devices in Industrial Process. *IEEE Trans. Ind. Inform.* **2023**, *20*, 2253–2262. [CrossRef]
197. Sam, D.; Kolter, J.Z. Losses over labels: Weakly supervised learning via direct loss construction. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 9695–9703.
198. Wang, M.; Xie, P.; Du, Y.; Hu, X. T5-Based Model for Abstractive Summarization: A Semi-Supervised Learning Approach with Consistency Loss Functions. *Appl. Sci.* **2023**, *13*, 7111. [CrossRef]
199. Li, Q.; Peng, X.; Qiao, Y.; Hao, Q. Unsupervised person re-identification with multi-label learning guided self-paced clustering. *Pattern Recognit.* **2022**, *125*, 108521. [CrossRef]
200. Nancy, P.; Pallathadka, H.; Naved, M.; Kaliyaperumal, K.; Arumugam, K.; Garchar, V. Deep Learning and Machine Learning Based Efficient Framework for Image Based Plant Disease Classification and Detection. In Proceedings of the 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), Coimbatore, India, 4–5 March 2022; pp. 1–6.
201. An, P.; Wang, Z.; Zhang, C. Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection. *Inf. Process. Manag.* **2022**, *59*, 102844. [CrossRef]
202. Yan, S.; Shao, H.; Xiao, Y.; Liu, B.; Wan, J. Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises. *Robot. Comput.-Integr. Manuf.* **2023**, *79*, 102441. [CrossRef]
203. Ayanoglu, E.; Davaslioglu, K.; Sagduyu, Y.E. Machine learning in NextG networks via generative adversarial networks. *IEEE Trans. Cogn. Commun. Netw.* **2022**, *8*, 480–501. [CrossRef]
204. Yan, K.; Chen, X.; Zhou, X.; Yan, Z.; Ma, J. Physical model informed fault detection and diagnosis of air handling units based on transformer generative adversarial network. *IEEE Trans. Ind. Inform.* **2022**, *19*, 2192–2199. [CrossRef]
205. Zhou, N.R.; Zhang, T.F.; Xie, X.W.; Wu, J.Y. Hybrid quantum–classical generative adversarial networks for image generation via learning discrete distribution. *Signal Process. Image Commun.* **2023**, *110*, 116891. [CrossRef]
206. Ladosz, P.; Weng, L.; Kim, M.; Oh, H. Exploration in deep reinforcement learning: A survey. *Inf. Fusion* **2022**, *85*, 1–22. [CrossRef]
207. Matsuo, Y.; LeCun, Y.; Sahani, M.; Precup, D.; Silver, D.; Sugiyama, M.; Uchibe, E.; Morimoto, J. Deep learning, reinforcement learning, and world models. *Neural Netw.* **2022**, *152*, 267–275. [CrossRef] [PubMed]
208. Bertoin, D.; Zouitine, A.; Zouitine, M.; Rachelson, E. Look where you look! Saliency-guided Q-networks for generalization in visual Reinforcement Learning. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 30693–30706.
209. Hafiz, A. A Survey of Deep Q-Networks used for Reinforcement Learning: State of the Art. In *Intelligent Communication Technologies and Virtual Mobile Networks, Proceedings of ICICV 2022, Tamil Nadu, India, 10–11 February 2022*; Springer: Singapore, 2022; pp. 393–402.
210. Hafiz, A.; Hassaballah, M.; Alqahtani, A.; Alsubai, S.; Hameed, M.A. Reinforcement Learning with an Ensemble of Binary Action Deep Q-Networks. *Comput. Syst. Sci. Eng.* **2023**, *46*, 2651–2666. [CrossRef]
211. Alagha, A.; Singh, S.; Mizouni, R.; Bentahar, J.; Otrok, H. Target localization using multi-agent deep reinforcement learning with proximal policy optimization. *Future Gener. Comput. Syst.* **2022**, *136*, 342–357. [CrossRef]

212. Hassan, S.S.; Park, Y.M.; Tun, Y.K.; Saad, W.; Han, Z.; Hong, C.S. 3TO: THz-enabled throughput and trajectory optimization of UAVs in 6G networks by proximal policy optimization deep reinforcement learning. In Proceedings of the ICC 2022-IEEE International Conference on Communications, Seoul, Republic of Korea, 16–20 May 2022; pp. 5712–5718.
213. Jayant, A.K.; Bhatnagar, S. Model-based safe deep reinforcement learning via a constrained proximal policy optimization algorithm. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 24432–24445.
214. Lin, B. Reinforcement learning and bandits for speech and language processing: Tutorial, review and outlook. *Expert Syst. Appl.* **2023**, *238*, 122254. [CrossRef]
215. Luo, B.; Wu, Z.; Zhou, F.; Wang, B.C. Human-in-the-loop reinforcement learning in continuous-action space. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 15735–15744. [CrossRef]
216. Raza, A.; Tran, K.P.; Koehl, L.; Li, S. Designing ecg monitoring healthcare system with federated transfer learning and explainable ai. *Knowl.-Based Syst.* **2022**, *236*, 107763. [CrossRef]
217. Siahpour, S.; Li, X.; Lee, J. A novel transfer learning approach in remaining useful life prediction for incomplete dataset. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3509411. [CrossRef]
218. Guo, Z.; Lin, K.; Chen, X.; Chit, C.Y. Transfer learning for angle of arrivals estimation in massive mimo system. In Proceedings of the 2022 IEEE/CIC International Conference on Communications in China (ICCC), Sanshui, Foshan, China, 11–13 August 2022; pp. 506–511.
219. Liu, S.; Lu, Y.; Zheng, P.; Shen, H.; Bao, J. Adaptive reconstruction of digital twins for machining systems: A transfer learning approach. *Robot. Comput.-Integr. Manuf.* **2022**, *78*, 102390. [CrossRef]
220. Liu, H.; Liu, J.; Cui, L.; Teng, Z.; Duan, N.; Zhou, M.; Zhang, Y. Logiqa 2.0—An improved dataset for logical reasoning in natural language understanding. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2023**, *31*, 2947–2962. [CrossRef]
221. Meng, Y.; Huang, J.; Zhang, Y.; Han, J. Generating training data with language models: Towards zero-shot language understanding. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 462–477.
222. Samant, R.M.; Bachute, M.R.; Gite, S.; Kotecha, K. Framework for deep learning-based language models using multi-task learning in natural language understanding: A systematic literature review and future directions. *IEEE Access* **2022**, *10*, 17078–17097. [CrossRef]
223. Weld, H.; Huang, X.; Long, S.; Poon, J.; Han, S.C. A survey of joint intent detection and slot filling models in natural language understanding. *ACM Comput. Surv.* **2022**, *55*, 156. [CrossRef]
224. Ajmal, S.; Ahmed, A.A.I.; Jalota, C. Natural Language Processing in Improving Information Retrieval and Knowledge Discovery in Healthcare Conversational Agents. *J. Artif. Intell. Mach. Learn. Manag.* **2023**, *7*, 34–47.
225. Montejo-Ráez, A.; Jiménez-Zafra, S.M. Current Approaches and Applications in Natural Language Processing. *Appl. Sci.* **2022**, *12*, 4859. [CrossRef]
226. Manning, C.D. Human language understanding & reasoning. *Daedalus* **2022**, *151*, 127–138.
227. Peng, W.; Xu, D.; Xu, T.; Zhang, J.; Chen, E. Are GPT Embeddings Useful for Ads and Recommendation? In Proceedings of the International Conference on Knowledge Science, Engineering and Management, KSEM 2023, Guangzhou, China, 16–18 August 2023; pp. 151–162.
228. Erdem, E.; Kuyu, M.; Yagcioglu, S.; Frank, A.; Parcalabescu, L.; Plank, B.; Babii, A.; Turuta, O.; Erdem, A.; Calixto, I.; et al. Neural natural language generation: A survey on multilinguality, multimodality, controllability and learning. *J. Artif. Intell. Res.* **2022**, *73*, 1131–1207. [CrossRef]
229. Qian, J.; Dong, L.; Shen, Y.; Wei, F.; Chen, W. Controllable natural language generation with contrastive prefixes. *arXiv* **2022**, arXiv:2202.13257.
230. Rashkin, H.; Nikolaev, V.; Lamm, M.; Aroyo, L.; Collins, M.; Das, D.; Petrov, S.; Tomar, G.S.; Turc, I.; Reitter, D. Measuring attribution in natural language generation models. *Comput. Linguist.* **2023**, *49*, 777–840. [CrossRef]
231. Pandey, A.K.; Roy, S.S. Natural Language Generation Using Sequential Models: A Survey. *Neural Process. Lett.* **2023**, *55*, 7709–7742. [CrossRef]
232. Khan, J.Y.; Uddin, G. Automatic code documentation generation using gpt-3. In Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering, Rochester, MI, USA, 10–14 October 2022; pp. 1–6.
233. Dwivedi, Y.K.; Kshetri, N.; Hughes, L.; Slade, E.L.; Jeyaraj, A.; Kar, A.K.; Baabdullah, A.M.; Koohang, A.; Raghavan, V.; Ahuja, M.; et al. "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Int. J. Inf. Manag.* **2023**, *71*, 102642. [CrossRef]
234. Fu, T.; Gao, S.; Zhao, X.; Wen, J.r.; Yan, R. Learning towards conversational AI: A survey. *AI Open* **2022**, *3*, 14–28. [CrossRef]
235. Ji, H.; Han, I.; Ko, Y. A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *J. Res. Technol. Educ.* **2023**, *55*, 48–63. [CrossRef]
236. Wan, Y.; Wang, W.; He, P.; Gu, J.; Bai, H.; Lyu, M.R. Biasasker: Measuring the bias in conversational ai system. In Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, San Francisco, CA, USA, 5–7 December 2023; pp. 515–527.

237. Kusal, S.; Patil, S.; Choudrie, J.; Kotecha, K.; Mishra, S.; Abraham, A. AI-based conversational agents: A scoping review from technologies to future directions. *IEEE Access* **2022**, *10*, 92337–92356. [CrossRef]
238. Xiao, Z. Seeing Us Through Machines: Designing and Building Conversational AI to Understand Humans. Ph.D. Thesis, University of Illinois at Urbana-Champaign, Champaign, IL, USA, 2023.
239. Ko, H.K.; Park, G.; Jeon, H.; Jo, J.; Kim, J.; Seo, J. Large-scale text-to-image generation models for visual artists' creative works. In Proceedings of the 28th International Conference on Intelligent User Interfaces, Sydney, NSW, Australia, 27–31 March 2023; pp. 919–933.
240. Pearson, A. The rise of CreAltives: Using AI to enable and speed up the creative process. *J. AI Robot. Workplace Autom.* **2023**, *2*, 101–114. [CrossRef]
241. Rezwana, J.; Maher, M.L. Designing creative AI partners with COFI: A framework for modeling interaction in human-AI co-creative systems. *ACM Trans. Comput.-Hum. Interact.* **2023**, *30*, 67. [CrossRef]
242. Brooks, T.; Peebles, B.; Holmes, C.; DePue, W.; Guo, Y.; Jing, L.; Schnurr, D.; Taylor, J.; Luhman, T.; Luhman, E.; et al. *Video Generation Models as World Simulators*; OpenAI: San Francisco, CA, USA, 2024.
243. Google DeepMind. *Veo 2: Google's Advanced Video Generation Model*; Google: Mountain View, CA, USA, 2024.
244. Sharma, S.; Bvuma, S. Generative Adversarial Networks (GANs) for Creative Applications: Exploring Art and Music Generation. *Int. J. Multidiscip. Innov. Res. Methodol.* **2023**, *2*, 29–33.
245. Attard-Frost, B.; De los Ríos, A.; Walters, D.R. The ethics of AI business practices: A review of 47 AI ethics guidelines. *AI Ethics* **2023**, *3*, 389–406. [CrossRef]
246. Gardner, A.; Smith, A.L.; Steventon, A.; Coughlan, E.; Oldfield, M. Ethical funding for trustworthy AI: Proposals to address the responsibilities of funders to ensure that projects adhere to trustworthy AI practice. *AI Ethics* **2022**, *2*, 277–291. [CrossRef]
247. Schuett, J. Three lines of defense against risks from AI. *AI Soc.* **2023**. [CrossRef]
248. Sloane, M.; Zakrzewski, J. German AI Start-Ups and "AI Ethics": Using A Social Practice Lens for Assessing and Implementing Socio-Technical Innovation. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, 21–24 June 2022; pp. 935–947.
249. Vasconcelos, M.; Cardonha, C.; Gonçalves, B. Modeling epistemological principles for bias mitigation in AI systems: An illustration in hiring decisions. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, New Orleans, LA, USA, 2–3 February 2018; pp. 323–329.
250. Yang, Y.; Gupta, A.; Feng, J.; Singhal, P.; Yadav, V.; Wu, Y.; Natarajan, P.; Hedau, V.; Joo, J. Enhancing fairness in face detection in computer vision systems by demographic bias mitigation. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, Oxford, UK, 1–3 August 2022; pp. 813–822.
251. Schwartz, R.; Vassilev, A.; Greene, K.; Perine, L.; Burt, A.; Hall, P. *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*; NIST Special Publication; NIST: Gaithersburg, MD, USA, 2022; Volume 1270.
252. Guo, W.; Caliskan, A. Detecting emergent intersectional biases: Contextualized word embeddings contain a distribution of human-like biases. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, Virtual, 19–21 May 2021; pp. 122–133.
253. Kong, Y. Are "intersectionally fair" ai algorithms really fair to women of color? A philosophical analysis. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, 21–24 June 2022; pp. 485–494.
254. Tan, Y.C.; Celis, L.E. Assessing social and intersectional biases in contextualized word representations. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 1–12.
255. Cheng, L.; Mosallanezhad, A.; Sheth, P.; Liu, H. Causal learning for socially responsible AI. *arXiv* **2021**, arXiv:2104.12278.
256. Correa, J.D.; Tian, J.; Bareinboim, E. Identification of causal effects in the presence of selection bias. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 2744–2751.
257. Ghai, B.; Mueller, K. D-BIAS: A causality-based human-in-the-loop system for tackling algorithmic bias. *IEEE Trans. Vis. Comput. Graph.* **2022**, *29*, 473–482. [CrossRef] [PubMed]
258. Yan, J.N.; Gu, Z.; Lin, H.; Rzeszotarski, J.M. Silva: Interactively assessing machine learning fairness using causality. In Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–13.
259. Bertino, E.; Kantarcioglu, M.; Akcora, C.G.; Samtani, S.; Mittal, S.; Gupta, M. AI for Security and Security for AI. In Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy, Virtual, 26–28 April 2021; pp. 333–334.
260. Susanto, H.; Yie, L.F.; Rosiyadi, D.; Basuki, A.I.; Setiana, D. Data security for connected governments and organisations: Managing automation and artificial intelligence. In *Web 2.0 and Cloud Technologies for Implementing Connected Government*; IGI Global: Hershey, PA, USA, 2021; pp. 229–251.
261. Dilmaghani, S.; Brust, M.R.; Danoy, G.; Cassagnes, N.; Pecero, J.; Bouvry, P. Privacy and security of big data in AI systems: A research and standards perspective. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 5737–5743.

262. McIntosh, T. Intercepting Ransomware Attacks with Staged Event-Driven Access Control. Ph.D. Thesis, La Trobe University, Melbourne, VIC, Australia, 2022.
263. McIntosh, T.; Kayes, A.; Chen, Y.P.P.; Ng, A.; Watters, P. Applying staged event-driven access control to combat ransomware. *Comput. Secur.* **2023**, *128*, 103160. [CrossRef]
264. McIntosh, T.R.; Susnjak, T.; Liu, T.; Watters, P.; Nowrozy, R.; Halgamuge, M.N. From COBIT to ISO 42001: Evaluating Cybersecurity Frameworks for Opportunities, Risks, and Regulatory Compliance in Commercializing Large Language Models. *arXiv* **2024**, arXiv:2402.15770. [CrossRef]
265. Hummel, P.; Braun, M.; Tretter, M.; Dabrock, P. Data sovereignty: A review. *Big Data Soc.* **2021**, *8*, 2053951720982012. [CrossRef]
266. Lukings, M.; Habibi Lashkari, A. Data Sovereignty. In *Understanding Cybersecurity Law in Data Sovereignty and Digital Governance: An Overview from a Legal Perspective*; Springer: Cham, Switzerland, 2022; pp. 1–38.
267. Hickok, M. Lessons learned from AI ethics principles for future actions. *AI Ethics* **2021**, *1*, 41–47. [CrossRef]
268. Zhou, J.; Chen, F. AI ethics: From principles to practice. *AI Soc.* **2022**. [CrossRef]
269. Kroll, J.A. Outlining traceability: A principle for operationalizing accountability in computing systems. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual, 3–10 March 2021; pp. 758–771.
270. Oseni, A.; Moustafa, N.; Janicke, H.; Liu, P.; Tari, Z.; Vasilakos, A. Security and privacy for artificial intelligence: Opportunities and challenges. *arXiv* **2021**, arXiv:2102.04661.
271. Stahl, B.C.; Wright, D. Ethics and privacy in AI and big data: Implementing responsible research and innovation. *IEEE Secur. Priv.* **2018**, *16*, 26–33. [CrossRef]
272. Ma, C.; Li, J.; Wei, K.; Liu, B.; Ding, M.; Yuan, L.; Han, Z.; Poor, H.V. Trusted ai in multiagent systems: An overview of privacy and security for distributed learning. *Proc. IEEE* **2023**, *111*, 1097–1132. [CrossRef]
273. Song, M.; Wang, Z.; Zhang, Z.; Song, Y.; Wang, Q.; Ren, J.; Qi, H. Analyzing user-level privacy attack against federated learning. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 2430–2444. [CrossRef]
274. Misra, I.; Maaten, L.v.d. Self-supervised learning of pretext-invariant representations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 14–19 June 2020; pp. 6707–6717.
275. Zhai, X.; Oliver, A.; Kolesnikov, A.; Beyer, L. S4l: Self-supervised semi-supervised learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1476–1485.
276. Chen, T.; Zhai, X.; Ritter, M.; Lucic, M.; Houlsby, N. Self-supervised gans via auxiliary rotation loss. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 12154–12163.
277. Jenni, S.; Favaro, P. Self-supervised feature learning by learning to spot artifacts. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2733–2742.
278. Patel, P.; Kumari, N.; Singh, M.; Krishnamurthy, B. Lt-gan: Self-supervised gan with latent transformation detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 5–9 January 2021; pp. 3189–3198.
279. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the International Conference on Machine Learning, Virtual, 12–18 July 2020; pp. 1597–1607.
280. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 14–19 June 2020; pp. 9729–9738.
281. Liu, A.T.; Li, S.W.; Lee, H.y. Tera: Self-supervised learning of transformer encoder representation for speech. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 2351–2366. [CrossRef]
282. Pang, Y.; Wang, W.; Tay, F.E.; Liu, W.; Tian, Y.; Yuan, L. Masked autoencoders for point cloud self-supervised learning. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 604–621.
283. Hospedales, T.; Antoniou, A.; Micaelli, P.; Storkey, A. Meta-learning in neural networks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 5149–5169. [CrossRef]
284. Vilalta, R.; Drissi, Y. A perspective view and survey of meta-learning. *Artif. Intell. Rev.* **2002**, *18*, 77–95. [CrossRef]
285. Al-Shedivat, M.; Li, L.; Xing, E.; Talwalkar, A. On data efficiency of meta-learning. In Proceedings of the International Conference on Artificial Intelligence and Statistics, Virtual, 13–15 April 2021; pp. 1369–1377.
286. Hu, Y.; Liu, R.; Li, X.; Chen, D.; Hu, Q. Task-sequencing meta learning for intelligent few-shot fault diagnosis with limited data. *IEEE Trans. Ind. Inform.* **2021**, *18*, 3894–3904. [CrossRef]
287. Baik, S.; Choi, J.; Kim, H.; Cho, D.; Min, J.; Lee, K.M. Meta-learning with task-adaptive loss function for few-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montréal, BC, Canada, 11–17 October 2021; pp. 9465–9474.
288. Chen, Y.; Liu, Z.; Xu, H.; Darrell, T.; Wang, X. Meta-baseline: Exploring simple meta-learning for few-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montréal, BC, Canada, 11–17 October 2021; pp. 9062–9071.
289. Jamal, M.A.; Qi, G.J. Task agnostic meta-learning for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 11719–11727.

290. Behnia, R.; Ebrahimi, M.R.; Pacheco, J.; Padmanabhan, B. EW-Tune: A Framework for Privately Fine-Tuning Large Language Models with Differential Privacy. In Proceedings of the 2022 IEEE International Conference on Data Mining Workshops (ICDMW), Orlando, FL, USA, 28 November–1 December 2022; pp. 560–566.
291. Wei, J.; Bosma, M.; Zhao, V.Y.; Guu, K.; Yu, A.W.; Lester, B.; Du, N.; Dai, A.M.; Le, Q.V. Finetuned language models are zero-shot learners. *arXiv* **2021**, arXiv:2109.01652.
292. Kuang, W.; Qian, B.; Li, Z.; Chen, D.; Gao, D.; Pan, X.; Xie, Y.; Li, Y.; Ding, B.; Zhou, J. Federatedscope-llm: A comprehensive package for fine-tuning large language models in federated learning. *arXiv* **2023**, arXiv:2309.00363.
293. Nguyen, M.; Kishan, K.; Nguyen, T.; Chadha, A.; Vu, T. Efficient Fine-Tuning Large Language Models for Knowledge-Aware Response Planning. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Turin, Italy, 18–22 September 2023; pp. 593–611.
294. Engelbach, M.; Klau, D.; Scheerer, F.; Drawehn, J.; Kintz, M. Fine-tuning and aligning question answering models for complex information extraction tasks. *arXiv* **2023**, arXiv:2309.14805.
295. Nguyen, T.T.; Wilson, C.; Dalins, J. Fine-tuning llama 2 large language models for detecting online sexual predatory chats and abusive texts. *arXiv* **2023**, arXiv:2308.14683.
296. Zhou, Q.; Yu, C.; Zhang, S.; Wu, S.; Wang, Z.; Wang, F. RegionBLIP: A Unified Multi-modal Pre-training Framework for Holistic and Regional Comprehension. *arXiv* **2023**, arXiv:2308.02299.
297. Arnold, T.; Kasenberg, D. Value Alignment or Misalignment—What Will Keep Systems Accountable? In Proceedings of the AAAI Workshop on AI, Ethics, and Society, San Francisco, CA, USA, 4 February 2017.
298. Gabriel, I.; Ghazavi, V. The challenge of value alignment: From fairer algorithms to AI safety. *arXiv* **2021**, arXiv:2101.06060.
299. Nyholm, S. Responsibility gaps, value alignment, and meaningful human control over artificial intelligence. In *Risk and Responsibility in Context*; Routledge: Oxfordshire, UK, 2023; pp. 191–213.
300. Wu, S.; Fei, H.; Qu, L.; Ji, W.; Chua, T.S. Next-gpt: Any-to-any multimodal llm. *arXiv* **2023**, arXiv:2309.05519.
301. Bayoudh, K.; Knani, R.; Hamdaoui, F.; Mtibaa, A. A survey on deep multimodal learning for computer vision: Advances, trends, applications, and datasets. *Vis. Comput.* **2021**, *38*, 2939–2970. [CrossRef] [PubMed]
302. Hu, P.; Zhen, L.; Peng, D.; Liu, P. Scalable deep multimodal learning for cross-modal retrieval. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Paris, France, 21–25 July 2019; pp. 635–644.
303. Rahate, A.; Walambe, R.; Ramanna, S.; Kotecha, K. Multimodal co-learning: Challenges, applications with datasets, recent advances and future directions. *Inf. Fusion* **2022**, *81*, 203–239. [CrossRef]
304. Che, L.; Wang, J.; Zhou, Y.; Ma, F. Multimodal federated learning: A survey. *Sensors* **2023**, *23*, 6986. [CrossRef]
305. Liang, P.P.; Lyu, Y.; Fan, X.; Wu, Z.; Cheng, Y.; Wu, J.; Chen, L.; Wu, P.; Lee, M.A.; Zhu, Y.; et al. Multibench: Multiscale benchmarks for multimodal representation learning. *arXiv* **2021**, arXiv:2107.07502.
306. Li, Y.; Zhang, G.; Ma, Y.; Yuan, R.; Zhu, K.; Guo, H.; Liang, Y.; Liu, J.; Yang, J.; Wu, S.; et al. OmniBench: Towards The Future of Universal Omni-Language Models. *arXiv* **2024**, arXiv:2409.15272.
307. Chen, L.; Hu, H.; Zhang, M.; Chen, Y.; Wang, Z.; Li, Y.; Shyam, P.; Zhou, T.; Huang, H.; Yang, M.H.; et al. Omni×R: Evaluating Omni-modality Language Models on Reasoning across Modalities. *arXiv* **2024**, arXiv:2410.12219.
308. Ashktorab, Z.; Liao, Q.V.; Dugan, C.; Johnson, J.; Pan, Q.; Zhang, W.; Kumaravel, S.; Campbell, M. Human-ai collaboration in a cooperative game setting: Measuring social perception and outcomes. *Proc. ACM Hum.-Comput. Interact.* **2020**, *4*, 1–20. [CrossRef]
309. Esmaeilzadeh, P.; Mirzaei, T.; Dharanikota, S. Patients' perceptions toward human–artificial intelligence interaction in health care: Experimental study. *J. Med. Internet Res.* **2021**, *23*, e25856. [CrossRef]
310. Nazar, M.; Alam, M.M.; Yafi, E.; Su'ud, M.M. A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques. *IEEE Access* **2021**, *9*, 153316–153348. [CrossRef]
311. Rajawat, A.S.; Rawat, R.; Barhanpurkar, K.; Shaw, R.N.; Ghosh, A. Robotic process automation with increasing productivity and improving product quality using artificial intelligence and machine learning. In *Artificial Intelligence for Future Generation Robotics*; Elsevier: Berlin/Heidelberg, Germany, 2021; pp. 1–13.
312. Mohseni, S.; Zarei, N.; Ragan, E.D. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Trans. Interact. Intell. Syst. (TiiS)* **2021**, *11*, 1–45. [CrossRef]
313. Buehler, M.C.; Weisswange, T.H. Theory of mind based communication for human agent cooperation. In Proceedings of the 2020 IEEE International Conference on Human-Machine Systems (ICHMS), Rome, Italy, 7–9 September 2020; pp. 1–6.
314. Çelikok, M.M.; Peltola, T.; Daee, P.; Kaski, S. Interactive AI with a Theory of Mind. *arXiv* **2019**, arXiv:1912.05284.
315. Dafoe, A.; Hughes, E.; Bachrach, Y.; Collins, T.; McKee, K.R.; Leibo, J.Z.; Larson, K.; Graepel, T. Open problems in cooperative AI. *arXiv* **2020**, arXiv:2012.08630.
316. Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Lee, P.; Lee, Y.T.; Li, Y.; Lundberg, S.; et al. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv* **2023**, arXiv:2303.12712.

317. Fei, N.; Lu, Z.; Gao, Y.; Yang, G.; Huo, Y.; Wen, J.; Lu, H.; Song, R.; Gao, X.; Xiang, T.; et al. Towards artificial general intelligence via a multimodal foundation model. *Nat. Commun.* **2022**, *13*, 3094. [CrossRef] [PubMed]
318. Williams, R.; Yampolskiy, R. Understanding and avoiding ai failures: A practical guide. *Philosophies* **2021**, *6*, 53. [CrossRef]
319. Shavit, Y.; Agarwal, S.; Brundage, M.; Adler, S.; O'Keefe, C.; Campbell, R.; Lee, T.; Mishkin, P.; Eloundou, T.; Hickey, A.; et al. *Practices for Governing Agentic AI Systems*; OpenAI White Paper; OpenAI: San Francisco, CA, USA, 2023.
320. Sun, Y.; Wang, X.; Liu, Z.; Miller, J.; Efros, A.; Hardt, M. Test-time training with self-supervision for generalization under distribution shifts. In Proceedings of the International Conference on Machine Learning, Virtual, 13–18 July 2020; pp. 9229–9248.
321. Akyürek, E.; Damani, M.; Qiu, L.; Guo, H.; Kim, Y.; Andreas, J. The Surprising Effectiveness of Test-Time Training for Abstract Reasoning. *arXiv* **2024**, arXiv:2411.07279.
322. Rajbhandari, S.; Li, C.; Yao, Z.; Zhang, M.; Aminabadi, R.Y.; Awan, A.A.; Rasley, J.; He, Y. Deepspeed-moe: Advancing mixture-of-experts inference and training to power next-generation ai scale. In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 18332–18346.
323. Shen, L.; Wu, Z.; Gong, W.; Hao, H.; Bai, Y.; Wu, H.; Wu, X.; Bian, J.; Xiong, H.; Yu, D.; et al. Se-moe: A scalable and efficient mixture-of-experts distributed training and inference system. *arXiv* **2022**, arXiv:2205.10034.
324. Fedus, W.; Zoph, B.; Shazeer, N. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *J. Mach. Learn. Res.* **2022**, *23*, 5232–5270.
325. Hwang, C.; Cui, W.; Xiong, Y.; Yang, Z.; Liu, Z.; Hu, H.; Wang, Z.; Salas, R.; Jose, J.; Ram, P.; et al. Tutel: Adaptive mixture-of-experts at scale. In Proceedings of the 6th Machine Learning and Systems, Miami, FL, USA, 4–8 June 2023.
326. Puigcerver, J.; Jenatton, R.; Riquelme, C.; Awasthi, P.; Bhojanapalli, S. On the adversarial robustness of mixture of experts. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 9660–9671.
327. Li, J.; Jiang, Y.; Zhu, Y.; Wang, C.; Xu, H. Accelerating Distributed {MoE} Training and Inference with Lina. In Proceedings of the 2023 USENIX Annual Technical Conference (USENIX ATC 23), Boston, MA, USA, 10–12 July 2023; pp. 945–959.
328. Wang, Y.; Mukherjee, S.; Liu, X.; Gao, J.; Awadallah, A.H.; Gao, J. Adamix: Mixture-of-adapter for parameter-efficient tuning of large language models. *arXiv* **2022**, arXiv:2205.12410.
329. Chi, Z.; Dong, L.; Huang, S.; Dai, D.; Ma, S.; Patra, B.; Singhal, S.; Bajaj, P.; Song, X.; Mao, X.L.; et al. On the representation collapse of sparse mixture of experts. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 34600–34613.
330. Zoph, B.; Bello, I.; Kumar, S.; Du, N.; Huang, Y.; Dean, J.; Shazeer, N.; Fedus, W. Designing effective sparse expert models. *arXiv* **2022**, *2*, arXiv:2202.08906.
331. Fan, Z.; Sarkar, R.; Jiang, Z.; Chen, T.; Zou, K.; Cheng, Y.; Hao, C.; Wang, Z. M^3vit: Mixture-of-experts vision transformer for efficient multi-task learning with model-accelerator co-design. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 28441–28457.
332. Zadouri, T.; Üstün, A.; Ahmadian, A.; Ermiş, B.; Locatelli, A.; Hooker, S. Pushing mixture of experts to the limit: Extremely parameter efficient moe for instruction tuning. *arXiv* **2023**, arXiv:2309.05444.
333. Zhu, J.; Zhu, X.; Wang, W.; Wang, X.; Li, H.; Wang, X.; Dai, J. Uni-perceiver-moe: Learning sparse generalist models with conditional moes. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 2664–2678.
334. Shinn, M.; Gu, S.S.; Bengio, Y. Reflexion: Language Agents with Verbal Reinforcement. *arXiv* **2023**, arXiv:2303.11366. http://arxiv.org/abs/2303.11366.
335. Mialon, T.; Denoyer, L.; Hoffmann, J.; Casanova, H.; Chalumeau, N.; Razdaibiedina, A.; Piantanida, P.; Caron, M.; Lhoest, Q.; de Vries, H.; et al. Augmented Language Models: A Survey. *arXiv* **2023**, arXiv:2302.07842.
336. Parisi, G.P.; Canton-Ferrer, C.; Palenicek, D.; Goodwin, R.; Gordon, J. TALM: Tool Augmented Language Models. In Proceedings of the 2022 EMNLP Workshop on Structured and Visually Grounded Supervision for Statistical Parsing of Scenes and Language (SV-PARSE), Abu Dhabi, United Arab Emirates, 7–8 December 2022.
337. Weidinger, L.; Mellor, J.; Bentham, M.R.; Redkin, V.; Zain, G.; Taylor, R.; Chadwick, M.; Everitt, T.; Guardino, C.; Grünbaum, B.B.; et al. Ethical and Social Risks of Harm from Language Models. *arXiv* **2022**, arXiv:2112.04359.
338. Dou, F.; Ye, J.; Yuan, G.; Lu, Q.; Niu, W.; Sun, H.; Guan, L.; Lu, G.; Mai, G.; Liu, N.; et al. Towards artificial general intelligence (agi) in the internet of things (iot): Opportunities and challenges. *arXiv* **2023**, arXiv:2309.07438.
339. Jia, Z.; Li, X.; Ling, Z.; Liu, S.; Wu, Y.; Su, H. Improving policy optimization with generalist-specialist learning. In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 10104–10119.
340. Simeone, M. Unknown Future, Repeated Present: A Narrative-Centered Analysis of Long-Term AI Discourse. *Hum. Stud. Digit. Age* **2022**, *7*, 1–13. [CrossRef]
341. Nair, A.; Banaei-Kashani, F. Bridging the Gap between Artificial Intelligence and Artificial General Intelligence: A Ten Commandment Framework for Human-Like Intelligence. *arXiv* **2022**, arXiv:2210.09366.
342. Jarrahi, M.H.; Askay, D.; Eshraghi, A.; Smith, P. Artificial intelligence and knowledge management: A partnership between human and AI. *Bus. Horizons* **2023**, *66*, 87–99. [CrossRef]

343. Edwards, D.J.; McEnteggart, C.; Barnes-Holmes, Y. A functional contextual account of background knowledge in categorization: Implications for artificial general intelligence and cognitive accounts of general knowledge. *Front. Psychol.* **2022**, *13*, 745306. [CrossRef] [PubMed]
344. McCarthy, J. Artificial Intelligence, Logic, and Formalising Common Sense. In *Philosophical Logic and Artificial Intelligence*; Springer: Dordrecht, The Netherlands, 2022; pp. 69–90.
345. Friederich, S. Symbiosis, not alignment, as the goal for liberal democracies in the transition to artificial general intelligence. *AI Ethics* **2023**, *4*, 315–324. [CrossRef]
346. Makridakis, S. The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures* **2017**, *90*, 46–60. [CrossRef]
347. Verma, S.; Sharma, R.; Deb, S.; Maitra, D. Artificial intelligence in marketing: Systematic review and future research direction. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100002. [CrossRef]
348. Pal, S.; Kumari, K.; Kadam, S.; Saha, A. *The AI Revolution*; IARA Publication: Tiruchirappalli, India, 2023.
349. Budhwar, P.; Chowdhury, S.; Wood, G.; Aguinis, H.; Bamber, G.J.; Beltran, J.R.; Boselie, P.; Lee Cooke, F.; Decker, S.; DeNisi, A.; et al. Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Hum. Resour. Manag. J.* **2023**, *33*, 606–659. [CrossRef]
350. Telkamp, J.B.; Anderson, M.H. The implications of diverse human moral foundations for assessing the ethicality of Artificial Intelligence. *J. Bus. Ethics* **2022**, *178*, 961–976. [CrossRef]
351. Zhou, X.; Liu, C.; Zhai, L.; Jia, Z.; Guan, C.; Liu, Y. Interpretable and robust ai in eeg systems: A survey. *arXiv* **2023**, arXiv:2304.10755.
352. McIntosh, T.R.; Susnjak, T.; Liu, T.; Watters, P.; Ng, A.; Halgamuge, M.N. A game-theoretic approach to containing artificial general intelligence: Insights from highly autonomous aggressive malware. *IEEE Trans. Artif. Intell.* **2024**, *5*, 6290–6303. [CrossRef]
353. Zhang, C.; Zhang, C.; Li, C.; Qiao, Y.; Zheng, S.; Dam, S.K.; Zhang, M.; Kim, J.U.; Kim, S.T.; Choi, J.; et al. One small step for generative ai, one giant leap for agi: A complete survey on chatgpt in aigc era. *arXiv* **2023**, arXiv:2304.06488.
354. Singhal, K.; Tu, T.; Gottweis, J.; Sayres, R.; Wulczyn, E.; Hou, L.; Clark, K.; Pfohl, S.; Cole-Lewis, H.; Neal, D.; et al. Towards expert-level medical question answering with large language models. *arXiv* **2023**, arXiv:2305.09617. [CrossRef] [PubMed]
355. Wu, S.; Irsoy, O.; Lu, S.; Dabravolski, V.; Dredze, M.; Gehrmann, S.; Kambadur, P.; Rosenberg, D.; Mann, G. Bloomberggpt: A large language model for finance. *arXiv* **2023**, arXiv:2303.17564.
356. Henderson, P.; Sinha, K.; Angelard-Gontier, N.; Ke, N.R.; Fried, G.; Lowe, R.; Pineau, J. Ethical challenges in data-driven dialogue systems. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, New Orleans, LA, USA, 2–3 February 2018; pp. 123–129.
357. Bin-Nashwan, S.A.; Sadallah, M.; Bouteraa, M. Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technol. Soc.* **2023**, *75*, 102370. [CrossRef]
358. Liu, N.; Brown, A. AI Increases the Pressure to Overhaul the Scientific Peer Review Process. Comment on "Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened". *J. Med. Internet Res.* **2023**, *25*, e50591. [CrossRef] [PubMed]
359. Siddaway, A.P.; Wood, A.M.; Hedges, L.V. How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annu. Rev. Psychol.* **2019**, *70*, 747–770. [CrossRef]
360. Landhuis, E. Scientific literature: Information overload. *Nature* **2016**, *535*, 457–458. [CrossRef] [PubMed]
361. Chloros, G.D.; Giannoudis, V.P.; Giannoudis, P.V. Peer-reviewing in surgical journals: Revolutionize or perish? *Ann. Surg.* **2022**, *275*, e82–e90. [CrossRef] [PubMed]
362. Allen, K.A.; Reardon, J.; Lu, Y.; Smith, D.V.; Rainsford, E.; Walsh, L. Towards improving peer review: Crowd-sourced insights from Twitter. *J. Univ. Teach. Learn. Pract.* **2022**, *19*, 2.