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

Timothy R. McIntosh, Teo Susnjak, Tong Liu, Paul Watters, *Senior Member, IEEE*, and Malka N. Halgamuge, *Senior Member, IEEE*

***Abstract*—This comprehensive survey explored the evolving landscape of generative Artificial Intelligence (AI), with a specific focus on the transformative impacts of Mixture of Experts (MoE), multimodal learning, and the speculated advancements towards Artificial General Intelligence (AGI). It critically examined the current state and future trajectory of generative Artificial Intelligence (AI), exploring how innovations like Google's Gemini and the anticipated OpenAI Q* project are reshaping research priorities and applications across various domains, including an impact analysis on the generative AI research taxonomy. It 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. It 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 MoE, multimodality, and AGI in generative AI.**

***Index Terms*—AI Ethics, Artificial General Intelligence (AGI), Artificial Intelligence (AI), Gemini, Generative AI, Mixture of Experts (MoE), Multimodality, Q* (Q-star), Research Impact Analysis.**

I. Introduction

THE historical context of AI, tracing back to Alan Turing's "Imitation Game" [1], early computational theories [2], [3], and the development of the first neural networks and machine learning [4], [5], [6], has set the foundation for today's advanced models. This evolution, accentuated by crucial moments such as the rise of deep learning and reinforcement learning, has been vital in shaping the contemporary trends in AI, including the sophisticated Mixture of Experts (MoE) models and multimodal AI systems, illustrating the field's dynamic and continuously evolving character. These advancements are a testament to the dynamic and ever-evolving nature of AI technology. The evolution of Artificial Intelligence (AI) has witnessed a crucial turn with the advent of Large Language Models (LLMs), notably ChatGPT, developed by OpenAI, and the recent unveiling of Google's Gemini [7], [8]. This technology has not only revolutionized the industry and academia, but has also reignited critical discussions concerning AI consciousness and its potential threats to humanity [9], [10], [11]. The development of such advanced AI systems, including notable competitors like Anthropic's Claude, and now Gemini, which demonstrates several advances over previous models like GPT-3 and Google's own LaMDA, has reshaped the research landscape. Gemini's ability to learn from two-way conversations and its "spike-and-slab" attention method, which allows it to focus on relevant parts of the context during multi-turn conversations, represents a significant leap in developing models that are better equipped for multidomain conversational applications[1]. These innovations in LLMs, including the mixture-of-experts methods employed by Gemini, signal a move towards models that can handle a diversity of inputs and foster multimodal approaches. Amidst this backdrop, speculations of an OpenAI project known as Q* (Q-Star) have surfaced, allegedly combining the power of LLMs with sophisticated algorithms such as Q-learning and A* (A-Star algorithm), further contributing to the dynamic research environment[2].

A. Changing AI Research Popularity

As the field of LLMs continues to evolve, exemplified by innovations such as Gemini and Q*, 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 towards multimodal capabilities and conversation-driven learning, as demonstrated by Gemini. The propagation of preprints has expedited knowledge sharing, but also brings the risk of reduced academic scrutiny. Issues like inherent biases, noted by Retraction Watch, along with concerns about plagiarism and forgery, present substantial hurdles [12]. The academic world, therefore, stands at an intersection, necessitating a unified drive

Manuscript received December 19, 2023. *(Corresponding author: Timothy R. McIntosh.)*

Timothy McIntosh is with Academies Australasia Polytechnic, Melbourne, VIC 3000, Australia (e-mail: t.mcintosh@aapoly.edu.au).

Teo Susnjak and Tong Liu are with Massey University, Auckland 0632, New Zealand (e-mail: t.liu@massey.ac.nz; t.susnjak@massey.ac.nz).

Paul Watters is with Cyberstronomy Pty Ltd, Ballarat, VIC 3350, Australia (e-mail: ceo@cyberstronomy.com).

Malka N. Halgamuge is with RMIT University, Melbourne, VIC 3000, Australia (e-mail: malka.halgamuge@rmit.edu.au).

[1]https://deepmind.google/technologies/gemini/

[2]https://www.forbes.com/sites/lanceeliot/2023/11/26/about-that-mysterious-ai-breakthrough-known-as-q-by-openai-that-allegedly-attains-true-ai-or-is-on-the-path-toward-artificial-general-intelligence-agi

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 generative models like GPT and the widespread commercial success of ChatGPT have 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 [13], the GPT model in 2018 [14], 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 models like GPT and LLaMA redefine what's 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 [15].

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.

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 [16], [17]. 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.

B. Objectives

The impetus for this investigation is the official unveiling of Gemini and the speculative discourse surrounding Q* project, which prompts a timely examination of the prevailing currents in generative AI research. This paper specifically contributes to the understanding of how MoE, multimodality, and Artificial General Intelligence (AGI) are impacting generative AI models, offering detailed analysis and future directions for each of these three key areas. This study does not aim to perpetuate conjecture about the unrevealed Q-Star initiative, but rather to critically appraise the potential for obsolescence or insignificance in extant research themes, whilst 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 towards data theft strategies utilizing varied attack vectors, relegating contemporary studies on crypto-ransomware to the status of latecomers [18], [19]. Advances in AI are anticipated to not only enhance capabilities in language analysis and knowledge synthesis but also to pioneer in areas like Mixture of Experts (MoE) [20], [21], [22], [23], [24], [25], multimodality [26], [27], [28], [29], [30], and Artificial General Intelligence (AGI) [31], [32], [10], [11], and has already heralded the obsolescence of conventional, statistics-driven natural language processing techniques in many domains [8]. Nonetheless, the perennial imperative for AI to align with human ethics and values persists as a fundamental tenet [33], [34], [35], and the conjectural Q-Star initiative offers an unprecedented opportunity to instigate discourse on how such advancements might reconfigure the LLM research topography. Within this milieu, insights from Dr. Jim Fan (senior research scientist & lead of AI agents at NVIDIA) on Q*, particularly concerning the amalgamation of learning and search algorithms, furnish an invaluable perspective on the prospective technical construct and proficiencies of such an undertaking[4]. Our research methodology involved a structured literature search using key terms like 'Large Language Models' and 'Generative AI'. We utilized filters across several academic databases such as IEEE Xplore, Scopus, ACM Digital Library, ScienceDirect, Web of Science, and ProQuest Central, tailored to identify relevant articles published in the timeframe from 2017 (the release of the "Transformer" model) to 2023 (the writing time of this manuscript). This paper aspires to dissect the technical ramifications of Gemini and Q*, probing how they (and similar technologies whose emergence is now inevitable) may transfigure research trajectories and disclose new vistas in the domain of AI. In doing so, we have pinpointed three nascent research domains—MoE, multimodality, and AGI—that stand to reshape the generative AI research landscape profoundly. This investigation adopts a survey-style approach, systematically mapping out a research roadmap that synthesizes and analyzes the current and emergent trends in generative AI.

The major contributions of this study is as follows:

1) Detailed examination of the evolving landscape in generative AI, emphasizing the advancements and innovations in technologies like Gemini and Q*, and their wide-ranging implications within the AI domain.

[3]The legend entries correspond to the keywords used in the search query, which is constructed as: *"(AI OR artificial OR (machine learning) OR (neural network) OR computer OR software) AND ([specific keyword])"*.

[4]https://twitter.com/DrJimFan/status/1728100123862004105

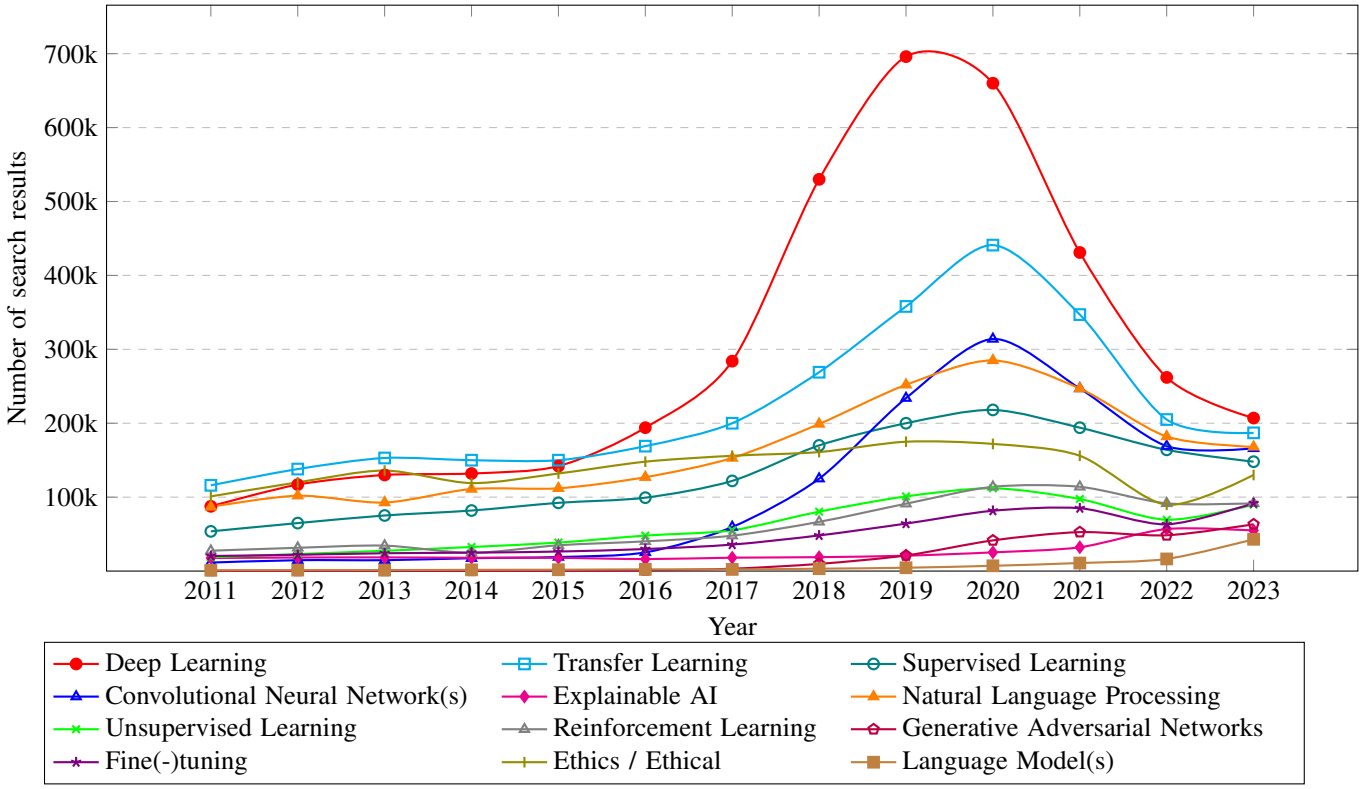

Figure 1: Number of search results on Google Scholar with different keywords by year [3]

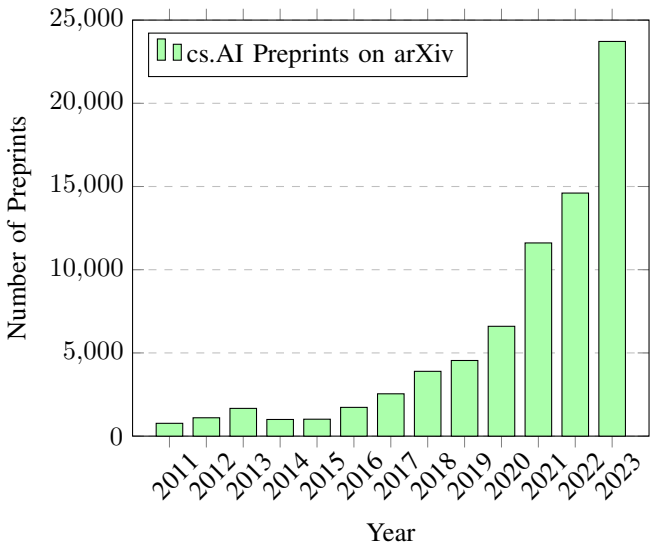

Figure 2: Annual number of preprints posted under the cs.AI category on arXiv.org

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 II explores the historical development of Generative AI. Section III presents a taxonomy of current Generative AI research. Section IV explores the Mixture of Experts (MoE) model architecture, its innovative features, and its impact on transformer-based language models. Section V discusses the speculated capabilities of the Q* project. Section VI discusses the projected capabilities of AGI. Section VII examines the impact of recent advancements on the Generative AI research taxonomy. Section VIII identifies emerging research priorities in Generative AI. Section X discusses the academic challenges of the rapid surge of preprints in AI. The paper concludes in Section XI, summarizing the overall effects of these developments in generative AI.

II. 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.

A. The Evolution of Language Models

Language models have undergone a transformative journey (Fig. 3), evolving from rudimentary statistical methods to the

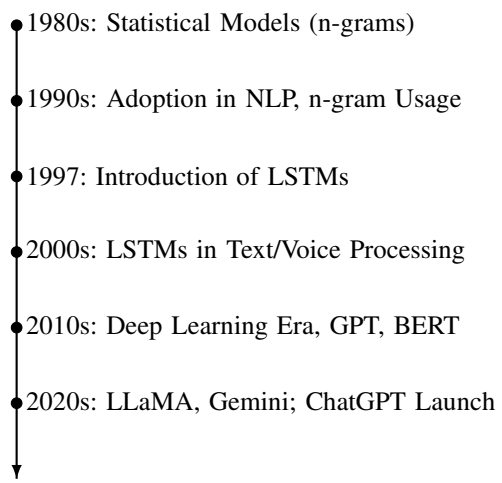

Figure 3: Timeline of Key Developments in Language Model Evolution

complex neural network architectures that underpin today's LLMs [36], [37]. 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 [36], [38], [37]. 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 [36], [39], [40].

1) Language Models as Precursors: The inception of language modeling can be traced to the statistical approaches of the late 1980s, a period marked by a transition from rule-based to machine learning algorithms in Natural Language Processing (NLP) [41], [42], [43], [44], [45]. Early models, primarily n-gram based, calculated the probability of word sequences in a corpus, thus providing a rudimentary understanding of language structure [41]. Those models, simplistic yet groundbreaking, laid the groundwork for future advances in language understanding. With the increase of computational power, the late 1980s witnessed a revolution in NLP, pivoting towards statistical models capable of 'soft' probabilistic decisions, as opposed to the rigid, 'handwritten' rule-based systems that dominated early NLP systems [43]. 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 [46], and their application to voice and text processing a decade later [47], [48], [49], marked a significant milestone, leading to the current era where neural network models represent the cutting edge of NLP research and development.

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 LLMs like GPT, BERT, and notably, OpenAI's ChatGPT. Recent models such as GPT-4 and LLaMA have pushed the boundaries by integrating sophisticated techniques like transformer architectures and advanced natural language understanding, illustrating the rapid evolution in this field [37]. 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 [37], [50]. 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 underscores a robust market demand for natural language AI and has catalyzed interdisciplinary research into its applications in sectors like education, healthcare, and commerce [8], [50], [51], [52], [53]. In education, ChatGPT offers innovative approaches to personalized learning and interactive teaching [54], [51], [55], [56], while in commerce, it revolutionizes customer service and content creation [57], [58]. The widespread use of ChatGPT, Google Bard, 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 [59], [31], [32], [11]. 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*. Specifically, the scalable architecture and extensive training data that characterize these models are foundational to the proposed capabilities of Q*. The success of ChatGPT in contextual understanding and conversational AI, for example, informs the design principles of Q*, suggesting a trajectory towards more sophisticated, context-aware, and adaptive language processing capabilities. Similarly, the emergence of multimodal systems like Gemini, capable of integrating text, images, audio, and video, reflects an evolutionary path that Q* could extend, combining the versatility of LLMs with advanced learning and pathfinding algorithms for a more holistic AI solution.

3) Fine-tuning, Hallucination Reduction, and Alignment in LLMs: The advancement of LLMs has underlined the significance of fine-tuning [60], [61], [62], [63], hallucination reduction [64], [65], [66], [67], and alignment [68], [69], [70], [71], [72]. These aspects are crucial in enhancing the functionality and reliability of LLMs. Fine-tuning, which involves adapting pre-trained models to specific tasks, has seen significant progress: techniques like prompt-based and few-shot learning [73], [74], [75], [76], alongside supervised fine-tuning on specialized datasets [60], [77], [78], [79], have enhanced the adaptability of LLMs in various contexts, but challenges remain, particularly in bias mitigation and the

generalization of models across diverse tasks [60], [80], [72]. Hallucination reduction is a persistent challenge in LLMs, characterized by the generation of confident but factually incorrect information [36]. Strategies such as confidence penalty regularization during fine-tuning have been implemented to mitigate overconfidence and improve accuracy [81], [82], [83]. 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 [36], [9]. Alignment, ensuring LLM outputs are congruent with human values and ethics, is an area of ongoing research. Innovative approaches, from constrained optimization [84], [85], [86], [87], [88], to different types of reward modeling [89], [90], [91], [92], 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 [9].

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[5] and MistralAI s Mixtral-8x7B[6], leverages multiple transformer-based expert modules for dynamic token routing, enhancing modeling efficiency and scalability. The primary advantage of MoE lies in its ability to handle vast parameter scales, reducing memory footprint and computational costs significantly [93], [94], [95], [96], [97]. This is achieved through model parallelism across specialized experts, allowing 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 [94], [95]. To illustrate the practicality of 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, MoE confronts challenges in dynamic routing complexity [98], [99], [100], [101], [102], expert imbalance [103], [104], [105], [106], and probability dilution [107], and such technical hurdles demand sophisticated solutions to fully harness MoE's potential. Moreover, while MoE may offer performance gains, it does not inherently solve ethical alignment issues in AI [108], [109], [110]. The complexity and specialization of MoE models can obscure the decision-making processes, complicating efforts to ensure ethical compliance and alignment with human values [108], [111]. Although the paradigm shift to MoE signifies a major leap in LLM development, offering significant scalability and specialization advantages, 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.

[5]https://huggingface.co/google/switch-c-2048
[6]https://huggingface.co/mistralai/Mixtral-8x7B-v0.1

B. 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.

1) 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 [112]. The architectural core of Gemini is its dual-encoder structure, with separate encoders for visual and textual data, enabling sophisticated multimodal contextualization [112]. 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 [112]. Furthermore, Gemini integrates structured knowledge and employs specialized training paradigms for cross-modal intelligence, setting new benchmarks in AI [112]. In [112], 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, 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 integration

of commonsense knowledge across modalities remains to be thoroughly evaluated.

2) 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 [113], [114], [115]. Challenges like data skew and bias are prevalent due to data acquisition and annotation issues, which requires effective dataset management by employing strategies such as data augmentation, active learning, and transfer learning [113], [116], [80], [115]. A significant challenge is the computational demands of processing various data streams simultaneously, requiring powerful hardware and optimized model architectures for multiple encoders [117], [118]. 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 [119], [120], [118]. Scalability issues, due to the extensive computational resources needed, are exacerbated by limited high-performance hardware availability [121], [122]. There is also a pressing need for calibrated multimodal encoders for compositional scene understanding and data integration [120]. Refining evaluation metrics for these systems is necessary to accurately assess performance in real-world tasks, calling for comprehensive datasets and unified benchmarks, and for enhancing 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.

3) 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 [123], [124], [125]. For autonomous vehicles, multimodality can enhance safety and navigation by synthesizing data from various sensors, including visual, radar, and Light Detection and Ranging (LIDAR) [126], [125], [127]. 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 [128], [129]. 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 is used without permission for AI training or content creation [113], [130], [131]. 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 [132], [133], [134]. The ethical development of multimodal AI systems requires robust governance frameworks focusing on transparency, consent, data handling protocols, and public awareness, when 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 [135], [136]. Additionally, the development of AI literacy programs will be crucial in helping society understand and responsibly interact with multimodal AI technologies [113], [135]. 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 [113].

C. Speculative Advances and Chronological Trends

In the dynamic landscape of AI, the speculative capabilities of the Q* project, blending LLMs, Q-learning, and A* (A-Star algorithm), embodies a significant leap forward. This section explores the evolutionary trajectory from game-centric AI systems to the broad applications anticipated with Q*.

1) From AlphaGo's Groundtruth to Q-Star's Exploration: The journey from AlphaGo, a game-centric AI, to the conceptual Q-Star project 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 [137], [138]. 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, and 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 towards 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 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) 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, 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, and 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, showcasing its potential as a crucial development eagerly awaited in the ongoing evolution of AI.

III. 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 I, 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.

A. 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 [139], [140], [141]. They employ advanced attention mechanisms to achieve enhanced contextual processing, allowing for more subtle understanding and interaction [142], [143], [144]. These models have also made notable strides in computer vision, as evidenced by the development of vision transformers like EfficientViT [145], [146] and YOLOv8 [147], [148], [149]. These innovations 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 [150], [151], [152], [153], [154]. 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 [155], [156]. 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 [157].
- **Mixture of Experts (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 [94], [98]. 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 [94], [21].
- **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 data sets, particularly transformative in fields like medical imaging [113], [112], [115]. These models facilitate accurate and data-efficient analysis by employing multi-view pipelines and cross-attention blocks [158], [159]. 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 [160]. 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 [113], [161], [162], [160].

B. 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 [163], [164], [165]. 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 [166], [167], [168].

Table I: 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.
	Mixture of Experts	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.
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.
	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.

- **Unsupervised Learning:** Unsupervised learning is essential in AI for uncovering patterns within unlabeled data, a process central to tasks like feature learning and clustering [169], [170]. This method has seen significant advancements with the introduction of autoencoders [171], [172] and Generative Adversarial Networks (GANs) [173], [174], [175], 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 complex 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 [176], [177]. This training technique has undergone significant advancements, particularly with the development of Deep Q-Networks (DQN) [178], [179], [180] and Proximal Policy Optimization (PPO) algorithms [181], [182], [183]. 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 [184], [185].
- **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 [186], [187]. Transfer learning, through the use of pre-trained 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 [188], [189].

C. 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 involves key capabilities such as semantic analysis, named entity recognition, sentiment analysis, textual entailment, and machine reading comprehension [190], [191], [192], [193]. 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 [190], [192], [193]. NLU is fundamental in applications like sentiment analysis, language translation, information extraction, and more [194], [195], [196]. 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 [197], [198].
- **Natural Language Generation (NLG)**: NLG emphasizes the training of models to generate coherent, contextually-relevant, and creative text responses, a critical component in chatbots, virtual assistants, and automated content creation tools [199], [36], [200], [201]. NLG encompasses challenges such as topic modeling, discourse planning, concept-to-text generation, style transfer, and controllable text generation [36], [202]. 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 [203], [55], [51].
- **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 multi-turn context tracking [204], [205], [206], [207]. In finance and cybersecurity, AI's predictive analytics have transformed risk assessment and fraud detection, leading to more secure and efficient operations [205], [19]. The advancements in this area, demonstrated by large pre-trained models like Meena[7] and BlenderBot[8], 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 [208], [209].
- **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, and has resulted in commercial success like MidJourney and DALL-E [210], [211], [212]. The challenges in this field involve finding suitable data representations, algorithms, and evaluation metrics to effectively assess and foster creativity [212], [213]. 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 [212]. 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.

D. 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 requires a focus on developing 'Ethical AI Frameworks', a new category in our taxonomy reflecting the trend towards responsible AI development in generative AI [214], [215], [15], [216], [217]. 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 [214], [15]. 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 [15].

- **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 [218], [219]. Continuous monitoring and bias testing are essential to identify and address any biases that may emerge from AI's predictive patterns [220], [219]. A significant challenge in this area is dealing with intersectional biases [221], [222], [223] and

[7]https://neptune.ai/blog/transformer-nlp-models-meena-lamda-chatbots
[8]https://blenderbot.ai

understanding the causal interactions that may contribute to these biases [224], [225], [226], [227].

- **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 [228], [229]. 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 [230], [231], [232]. Additionally, issues of data sovereignty and copyright emphasize the need for robust encryption, access control, and continuous security assessments [233], [234]. 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 [214], [235], [15], [236]. Furthermore, AI Ethics involves ensuring traceability, auditability, and transparency throughout the model development lifecycle, employing practices such as algorithmic auditing, establishing ethics boards, and adhering to documentation standards and model cards [237], [236]. However, the adoption of these initiatives remains uneven, highlighting the ongoing need for comprehensive and consistent ethical practices in AI development and deployment [214].
- **Privacy Preservation:** This domain focuses on maintaining data confidentiality and integrity, employing strategies like anonymization and federated learning to minimize direct data exposure, especially when the rise of generative AI poses risks of user profiling [238], [239]. Despite these efforts, challenges such as achieving true anonymity against correlation attacks highlight the complexities in effectively protecting against intrusive surveillance [240], [241]. 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.

E. 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 [242], [165], [243]. It incorporates generative models like autoencoders and GANs for data distribution learning and original input reconstruction [244], [245], [246], and also includes contrastive methods such as SimCLR [247] and MoCo [248], 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, significantly enhanced by recent Vision Transformers developments [249], [250], [165]. 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 [251], [252]. 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, essential in the current data-driven landscape [253], [254]. 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 [255], [256], [254], [257].
- **Fine Tuning:** Involves customizing pre-trained models to specific domains or user preferences, enhancing accuracy and relevance for niche applications [60], [258], [259]. Its two primary approaches are end-to-end fine-tuning, which adjusts all weights of the encoder and classifier [260], [261], and feature-extraction fine-tuning, where the encoder weights are frozen to extract features for a downstream classifier [262], [263], [264]. 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 [265], [89], [266]. This is increasingly important in scenarios where AI interacts closely with humans, such as in healthcare, finance, and 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 [89], [267].

F. 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:

- **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, multi-sensory context

awareness [114], [268]. Recent developments like Gemini model have set new benchmarks by demonstrating state-of-the-art performance in various multimodal tasks, including natural image, audio, and video understanding, and mathematical reasoning [112]. Gemini's inherently multimodal design exemplifies the seamless integration and operation across different information types [112]. Despite the advancements, the field of multimodal learning still confronts ongoing challenges, such as refining the architectures to handle diverse data types more effectively [269], [270], developing comprehensive datasets that accurately represent multifaceted information [269], [271], and establishing benchmarks for evaluating the performance of these complex systems [272], [273].

- **Interactive and Cooperative AI:** This subdomain aims to enhance the capabilities of AI models to collaborate effectively with humans in complex tasks [274], [35]. 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 [275], [276], [277]. Core aspects of this subdomain involve advancing AI in areas such as explainability [278], understanding human intentions and behavior (theory of mind) [279], [280], and scalable coordination between AI systems and humans, a collaborative approach crucial in creating more intuitive and interactive AI systems, capable of assisting and augmenting human capabilities in diverse contexts [281], [35].
- **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 [282], [283], [32]. 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 [282], [283]. 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 [15], [32], [11]. As we progress towards developing superintelligent systems, it becomes crucial to establish rigorous safety protocols and control mechanisms [11]. Key areas of concern include mitigating representational biases, addressing distribution shifts, and correcting spurious correlations within AI models [11], [284]. The objective is to prevent unintended societal consequences by aligning AI development with responsible and ethical standards.

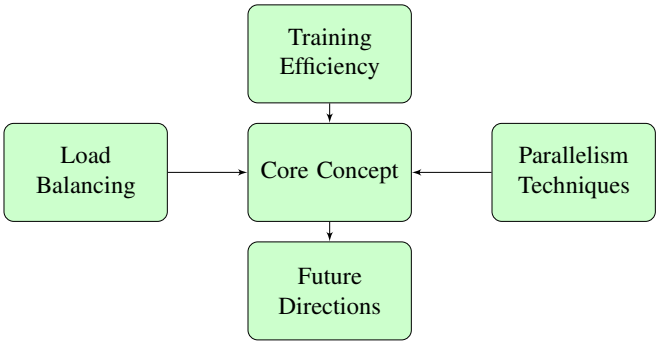

Figure 4: Conceptual Diagram of MoE's Innovation

IV. Innovative Horizon of MoE

The MoE model architecture represents a pioneering advancement in transformer-based language models, offering unparalleled scalability and efficiency (Fig. 4). As evidenced by recent models like the 1.6 trillion parameter Switch Transformer [285] and the 8x7B parameter Mixtra [286], MoE-based designs are rapidly redefining the frontiers of model scale and performance across diverse language tasks.

A. Core Concept and Structure

MoE models represent a significant innovation in neural network design, offering enhanced scalability and efficiency in training and inference [287], [288], [110]. At their core, MoE models utilize a sparsity-driven architecture by replacing dense layers with sparse MoE layers comprising multiple expert networks, where each expert is dedicated to a specific subset of the training data or task, and a trainable gating mechanism dynamically allocates input tokens to these experts, thereby optimizing computational resources and effectively adapting to the task's complexity [94], [21], [110]. MoE models demonstrate a substantial advantage in terms of pretraining speed, outperforming dense models [94], [287]. However, they face challenges in fine-tuning and require substantial memory for inference due to the necessity of loading all experts into Video Random Access Memory (VRAM) [289], [290], [110]. The structure of MoE involves alternating transformer layers with router layers containing gating networks for expert routing, leading to an architecture that allows significant parameter scaling and advanced specialization in problem-solving [291], [21].

A distinguishing characteristic of MoE models is their flexibility in managing large datasets, capable of amplifying model capacity by over a thousand times while only experiencing minor reductions in computational efficiency [289], [292]. The Sparsely-Gated Mixture-of-Experts Layer, a key component of these models, comprises numerous simple feed-forward expert networks and a trainable gating network responsible for expert selection, which can facilitate the dynamic and sparse activation of experts for each input instance, maintaining high computational efficiency [293], [294], [110].

Recent advancements in MoE models, such as those in the Switch Transformer, have highlighted the significant benefits of intelligent routing, when the router's ability to intelligently

route tokens to appropriate experts confers considerable advantages to MoE models, allowing them to scale up model sizes while keeping compute time constant [295], [296], [297]. Experimental evidence suggests that routers learn to route inputs according to data clusters, demonstrating their potential in real-world applications [295], [289]. The core concept and structure of MoE models lie in their dynamic routing and specialization capabilities, offering promising avenues for scaling up neural networks and enhancing their efficiency and adaptability in various tasks, but the robustness of the router must be protected against adversarial attacks [289], [298].

B. Training and Inference Efficiency

MoE models, notably Mixtral 8x7B, are renowned for their superior pretraining speed compared to dense models, yet they face hurdles in fine-tuning and demand considerable VRAM for inference, owing to the requirement of loading all experts [289], [290], [110]. Recent advancements in MoE architecture have resulted in notable training cost efficiencies, especially in encoder-decoder models, with evidence showing cost savings of up to fivefold in certain contexts when compared to dense models [21], [289], [298], [287]. Innovations like DeepSpeed-MoE [287] offered new architectural designs and model compression, decreasing the MoE model size by approximately 3.7x and optimizing inference to achieve up to 7.3x better latency and cost efficiency. The progression in distributed MoE training and inference, notably with innovations like Lina [299], has effectively tackled the all-to-all communication bottleneck by enhancing tensor partitioning, which not only improves all-to-all communication and training step time, but also optimizes resource scheduling during inference, leading to a substantial reduction in training step time by up to 1.73 times and lowering the 95th percentile inference time by an average of 1.63 times compared to existing systems. These developments have marked a crucial shift in the large model landscape, from dense to sparse MoE models, expanding the potential applications of AI by training higher-quality models with fewer resources.

C. Load Balancing and Router Optimization

Effective load balancing is essential in MoE models to guarantee a uniform distribution of computational load among experts, with the router network in MoE layers, responsible for selecting the appropriate experts for processing specific tokens, playing a pivotal role in achieving this balance, which is fundamental to the stability and overall performance of MoE models [293], [289], [288], [300], [110]. Developments in router Z-loss regularization techniques plays a crucial role in addressing expert imbalance in MoE models by fine-tuning the gating mechanism, ensuring a more equitable workload distribution across experts and fostering a stable training environment, thereby enhancing model performance and reducing training time and computational overhead [301], [302]. Concurrently, the integration of expert capacity management strategies, emerges as a crucial approach in MoE models to regulate the processing abilities of individual experts by setting thresholds on the number of tokens each can handle, effectively averting bottlenecks and ensuring a more efficient and streamlined model operation, leading to improved training processes and heightened performance during complex computational tasks [293], [303], [289].

D. Parallelism and Serving Techniques

Recent developments in MoE models highlighted their efficiency in parallelism and serving techniques, significantly influencing large-scale neural networks. DeepSpeed-MoE, for instance, introduces advanced parallelism modes like data parallelism, tensor-slicing for non-expert parameters, and expert parallelism for expert parameters, enhancing model efficiency, as their approach optimizes both latency and throughput in MoE model inference, offering scalable solutions in production environments using multiple Graphics Processing Unit (GPU) devices [287]. MoE models, versatile in applications like multilingual tasks and coding, demonstrated impressive capabilities in handling complex tasks due to their ensemble-like structure within a single framework [304], [305], [306]. Notably, models like Mixtral and Switch Transformer, with over 1.6 trillion parameters, achieved computational efficiency equivalent to a 10 billion-parameter dense model, because they benefited from the sublinear scaling of MoE compute versus model size, leading to substantial accuracy gains within fixed compute budgets [21], [289], [287], [110]. Moreover, DeepSpeed-MoE included model compression techniques, reducing model size by up to 3.7x while maintaining accuracy, and an end-to-end MoE training and inference solution, part of the DeepSpeed library, which was instrumental in serving large-scale MoE models with enhanced speed and cost-efficiency [287]. These innovations open new directions in AI, shifting from dense to sparse MoE models, where training and deploying higher-quality models with fewer resources become more widely achievable.

E. Future Directions and Applications

Emerging research on MoE architectures could focus on advancing sparse fine-tuning techniques, exploring instruction tuning methods, and improving routing algorithms to fully utilize performance and efficiency gains. As models scale over one billion parameters, MoE represents a paradigm shift for vastly expanding capabilities across scientific, medical, creative, and real-world applications. Frontier work could also aim to refine auto-tuning of hyperparameters during fine-tuning to optimize accuracy, calibration, and safety. MoE research continues to push model scale limits while maintaining specialization for transfer learning. Adaptive sparse access allows coordinating thousands of experts to cooperate on tasks ranging from reasoning to open domain dialogue. Continued analysis of routing mechanisms seeks to balance load across experts and minimize redundant computation. As the AI community further investigates MoE methods at scale, these models hold promise for new breakthroughs in language, code generation, reasoning, and multimodal applications. There is great interest in evaluating implications across education, healthcare, financial analysis, and other fields. Outcomes may yield insights not only into model optimization but also for understanding principles behind combinatorial generalization.

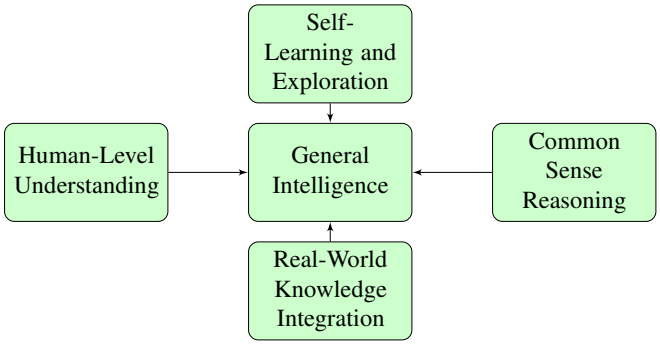

Figure 5: Conceptual Diagram of Speculated Q* Capabilities

V. Speculated Capabilities of Q*

In the burgeoning realm of AI, the anticipated Q* project stands as a beacon of potential breakthroughs, heralding advancements that could redefine the landscape of AI capabilities (Fig. 5).

A. Enhanced General Intelligence

Q*'s development in the arena of general intelligence represents a paradigm shift from specialized to holistic AI, indicating a broadening of the model's cognitive abilities akin to human intelligence. This advanced form of general intelligence involves integrating diverse neural network architectures and machine learning techniques, enabling the AI to process and synthesize multifaceted information seamlessly. The universal adapter approach, mirroring models like T0, could endow Q* with the capability to rapidly assimilate knowledge from various domains. This method allows Q* to learn adaptable module plugins, enhancing its ability to tackle new data types while preserving existing skills, leading to an AI model that combines narrow specializations into a comprehensive, adaptive, and versatile reasoning system. The corresponding quasi-mathematical formulation can be expressed as:

$$EGI(Q*) = \bigoplus_{i=1}^{n} (NN_i \odot MLT_i) \tag{1}$$

Where:

- EGI: "Enhanced General Intelligence"
- NN_i: a diverse set of neural network architectures.
- MLT_i: various machine learning techniques.
- $\bigoplus$: the integration of these components.
- $\odot$: a functional interaction between neural networks and machine learning techniques.

Such advancements in AI suggest the emergence of an intelligence that not only parallels but potentially exceeds human cognitive flexibility, with far-reaching implications in facilitating cross-disciplinary innovations and complex problem-solving. The speculated capabilities of Q* bring forth complex ethical implications and governance challenges. As AI systems approach higher levels of autonomy and decision-making, it is crucial to establish robust ethical frameworks and governance structures to ensure responsible and transparent AI development. This involves mitigating potential risks associated with advanced AI capabilities, emphasizing the need for comprehensive and dynamic ethical guidelines that evolve in tandem with AI advancements.

B. Advanced Self-Learning and Exploration

In the realm of advanced AI development, Q* is anticipated to represent a significant evolution in self-learning and exploration capabilities. It is speculated to utilize sophisticated Policy Neural Networks (NNs), similar to those in AlphaGo, but with substantial enhancements to handle the complexities of language and reasoning tasks. These networks are expected to employ advanced reinforcement learning techniques like Proximal Policy Optimization (PPO), which stabilizes policy updates and improves sample efficiency, a crucial factor in autonomous learning. The integration of these NNs with cutting-edge search algorithms, potentially including novel iterations of Tree or Graph of Thought, is predicted to enable Q* to autonomously navigate and assimilate complex information. This approach might be augmented with graph neural networks to bolster meta-learning capacities, allowing Q* to rapidly adapt to new tasks and environments while retaining previously acquired knowledge. The corresponding quasi-mathematical formulation can be represented as:

$$ASLE(Q*) = RL(PNN, SA) \times GNN \tag{2}$$

Where:

- $ASLE$: "Advanced Self-Learning and Exploration"
- RL: to reinforcement learning algorithms, particularly Proximal Policy Optimization (PPO).
- PNN: Policy Neural Networks, adapted for language and reasoning tasks.
- SA: sophisticated search algorithms, like Tree or Graph of Thought.
- GNN: the incorporation of Graph Neural Networks for meta-learning.
- $\times$: the cross-functional enhancement of RL with GNN.

Such capabilities indicate a model not limited to understanding existing data but equipped to actively seek and synthesize new knowledge, effectively adapting to evolving scenarios without the need for frequent retraining. This signifies a leap beyond current AI models, embedding a level of autonomy and efficiency previously unattained.

C. Superior Human-Level Understanding

Q*'s aspiration to achieve superior human-level understanding is speculated to hinge on an advanced integration of multiple neural networks, including a Value Neural Network (VNN), paralleling the evaluative components found in systems like AlphaGo. This network would extend beyond assessing accuracy and relevance in language and reasoning processes, delving into the subtleties of human communication. The model's deep comprehension capabilities may be enhanced by advanced natural language processing algorithms and techniques, such as those found in transformer architectures like DeBERTa. These algorithms would empower Q* to interpret not just the text but also the nuanced socio-emotional aspects such as intent, emotion, and underlying meanings.

Incorporating sentiment analysis and natural language inference, Q* could navigate layers of socio-emotional insights, including empathy, sarcasm, and attitude. The corresponding quasi-mathematical formulation can be expressed as:

$$SHLU(Q*) = \sum_{alg \in NLP} (VNN \oplus alg) \tag{3}$$

Where:

- $SHLU$: "Superior Human-Level Understanding".
- VNN: the Value Neural Network, similar to evaluative components in systems like AlphaGo.
- NLP: a set of advanced NLP algorithms.
- $\oplus$: the combination of VNN evaluation with NLP algorithms.
- alg: individual algorithms within the NLP set.

This level of understanding, surpassing current language models, would position Q* to excel in empathetic, context-aware interactions, thus enabling a new echelon of personalization and user engagement in AI applications.

D. Advanced Common Sense Reasoning

Q*'s anticipated development in advanced common sense reasoning is predicted to integrate sophisticated logic and decision-making algorithms, potentially combining elements of symbolic AI and probabilistic reasoning. This integration aims to endow Q* with an intuitive grasp of everyday logic and an understanding akin to human common sense, thus bridging a significant gap between artificial and natural intelligence. Enhancements in Q*'s reasoning abilities might involve graph-structured world knowledge, incorporating physics and social engines similar to those in models like CogSKR. This approach, grounded in physical reality, is expected to capture and interpret the everyday logic often absent in contemporary AI systems. By leveraging large-scale knowledge bases and semantic networks, Q* could effectively navigate and respond to complex social and practical scenarios, aligning its inferences and decisions more closely with human experiences and expectations. The corresponding quasi-mathematical formulation can be represented as:

$$ACSR(Q*) = LogicAI \odot ProbAI \odot WorldK \tag{4}$$

Where:

- $ACSR$: "Advanced Common Sense Reasoning".
- $LogicAI$ and $ProbAI$: symbolic AI and probabilistic reasoning components, respectively.
- $WorldK$: the integration of graph-structured world knowledge.
- $\odot$: the integrated operation of these elements for common sense reasoning.

E. Extensive Real-World Knowledge Integration

Q*'s approach to integrating extensive real-world knowledge is speculated to involve the use of advanced formal verification systems, which would provide a robust basis for validating its logical and factual reasoning. This method, when coupled with sophisticated neural network architectures and dynamic learning algorithms, would enable Q* to engage deeply with the complexities of the real world, transcending conventional AI limitations. Additionally, Q* might employ mathematical theorem proving techniques for validation, ensuring that its reasoning and outputs are not only accurate but also ethically grounded. The incorporation of Ethics classifiers in this process further strengthens its capacity to deliver reliable and responsible understanding and interaction with real-world scenarios. The corresponding quasi-mathematical formulation can be represented as:

$$ERWKI(Q*) = FVS \otimes NN \otimes LTP \otimes EC \tag{5}$$

Where:

- $ERWKI$: "Extensive Real-World Knowledge Integration".
- FVS: Formal Verification Systems.
- NN: neural network architectures.
- LTP: mathematical theorem proving for logical and factual validation.
- EC: the incorporation of Ethics classifiers.
- $\otimes$: the comprehensive integration for knowledge synthesis and ethical alignment.

Furthermore, the speculated capabilities of Q* have the potential to significantly reshape the job market and labor dynamics. With its advanced functionalities, Q* could automate complex tasks, leading to a shift in job requirements and the emergence of new skill demands. This necessitates a re-evaluation of workforce strategies and educational paradigms, aligning them with the evolving technological landscape and ensuring that the workforce is equipped to interact with and complement these advanced AI systems.

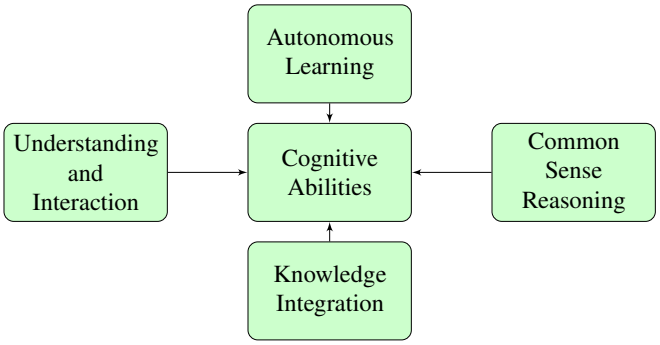

Figure 6: Conceptual Diagram of Projected AGI Capabilities

VI. Projected Capabilities of AGI

AGI stands as a transformative leap in AI, endeavoring to mirror human cognitive abilities in a software paradigm (Fig. 6). AGI's evolution is marked by advanced self-learning capabilities, utilizing policy neural networks and sophisticated reinforcement learning techniques for autonomous adaptation. The integration of algorithms like Tree/Graph of Thought with these networks suggests a future where AGI can independently acquire and apply knowledge across diverse domains.

A. Revolution in Autonomous Learning

AGI is anticipated to revolutionize self-learning and exploration [282], [307], [283], [32]. By incorporating methods like PPO, AGI models are positioned to achieve a level of autonomous learning and problem-solving that exceeds the current AI models' dependence on training data, indicating a potential paradigm shift towards reducing the need for frequent retraining and facilitating dynamic adaptation in response to evolving scenarios [181], [308].

B. Broadening of Cognitive Abilities

Envisaged to integrate various architectures, AGI could promise a level of general intelligence that replicates the multifaceted nature of human cognition [282], [309]. The universal adapter approach, mirroring models like GPT and BERT, could facilitate rapid assimilation of diverse information, positioning AGI as a system capable of performing tasks across multiple domains with an adaptability akin to human intellect [282], [310]. While AGI's full capabilities remain speculative, current trends suggest its potential application in advanced healthcare diagnostics, which is evidenced by recent breakthroughs in AI-driven predictive medicine models, indicating AGI's potential to revolutionize medical diagnosis and treatment.

C. Elevating Understanding and Interaction

AGI is projected to achieve an unparalleled understanding of human language and socio-emotional subtleties, leveraging algorithms like those in transformer architectures, which would enable AGI to engage in complex, empathetic, and contextually aware interactions, suggesting potential applications that revolutionize how AI systems communicate and interact [282], [307], [311].

D. Advanced Common Sense Reasoning

Symbolic AI and probabilistic reasoning, integrated into AGI, could imbue these systems with an innate grasp of common sense, to bridge the gap between artificial and natural intelligence, enabling AGI to navigate and respond effectively to real-world scenarios with reasoning aligned closely with human thought processes [282], [312], [313].

E. Holistic Integration of Knowledge

AGI's potential in integrating extensive real-world knowledge, guided by formal verification systems, hints at future capabilities where AGI's outputs are not only accurate but ethically grounded, suggesting AGI's ability for responsible interaction with real-world complexities [282], [311]. The projected capabilities of AGI extend to addressing significant global challenges, such as climate change, in which AGI's advanced data analysis and predictive modeling can play a better and more crucial role in environmental monitoring, forecasting climate patterns, and devising sustainable solutions, contributing significantly to global ecological efforts [282], [283], [32].

F. Challenges and Opportunities in AGI Development

The development of AGI encompasses both challenges and opportunities. While AGI promises productivity boosts in creative fields and innovations in cross-modal generation techniques, substantial challenges like data bias, computational efficiency, and ethical implications persist [15], [32]. These challenges necessitate a balanced approach in AGI development, focusing on data curation, efficient systems, and societal impacts [309].

In the context of AGI development, experts from various domains caution against overestimating current AI capabilities and highlight the gap between the theoretical framework of AGI and the practical realities of today's AI [314], [32]. The envisioned autonomy and cognitive abilities of AGI separate it from current AI models, suggesting a future where AI systems could perform tasks across various domains without human intervention [282]. This development trajectory underscores the importance of ethical considerations and technological breakthroughs in AGI's journey towards becoming a transformative force in society [15], [32]. While projecting the timeline for achieving true AGI remains speculative, recognizing potential roadblocks is crucial, such as the current limitations in computational power, and the complexity of replicating human-like cognitive abilities. These emphasize the need for sustained research and ethical considerations in the pursuit of AGI, ensuring responsible and conscientious development.

VII. Impact Analysis on Generative AI Research Taxonomy

With the advent of advanced AI developments such as MoE, multimodality, and AGI, the landscape of Generative AI research is undergoing a significant transformation. This section analyzes how these developments are reshaping the research taxonomy in generative AI.

A. Criteria for Impact Analysis

The continuously evolving landscape of Generative AI, which instigates transformative changes across various research domains, necessitates a systematic evaluation of these advancements' influence, for which we have established a set of criteria detailed in Table II, serving as analytical lenses to quantify and categorize the impact, deeply rooted in the dynamic interplay between technological progress and the evolving paradigms of research focus areas. Our analysis framework has been constructed on 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 allows for a complex assessment, acknowledging that not all areas will be uniformly affected. This multi-tiered approach is informed by historical patterns of technological disruption and the adaptability of scientific inquiry.

At the apex of our evaluative hierarchy, 'Emerging Direction' encapsulates the advent of uncharted research vistas, propelled by ongoing AI breakthroughs, which is predicated not on conjecture, but on a historical continuum of AI evolution, where each surge in technological power unfurls new

Table II: Criteria for Analyzing Impact on Generative AI Research

Symbol	Criteria	Score	Definition	Justification
$\nearrow$	Emerging Direction	5	New research areas expected to arise as a direct consequence of AI advancements.	Emphasizes novel research domains emerging from AI breakthroughs [315], [316].
$\hookrightarrow$	Requiring Redirection	4	Areas that need to shift focus or methodology to stay relevant with new AI developments.	Technological shifts necessitate reevaluation and redirection in AI research [315], [317].
$\leftrightarrow$	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 [317].
$\searrow$	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 [318].
$\triangle$	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 [319], [320].

scientific enigmas and avenues [315], [316]. 'Areas Requiring Redirection' denote research spheres that, though established, find themselves at an inflection point, necessitating a strategic pivot to assimilate emergent AI paradigms and an overhaul of traditional methodologies, akin to the transition from rule-based expert systems to adaptive machine learning frameworks [315], [317]. The 'Still Relevant' classification affirms the tenacity of select research domains that, by addressing persistent scientific inquiries or through their inherent malleability, remain impervious to the tides of AI innovation [317]. In contrast, domains categorized as 'Likely to Become Redundant' confront potential obsolescence, inviting strategic foresight and resource reallocation to forestall scientific stagnation [318]. Lastly, 'Inherently Unresolvable' challenges serve as a sobering reminder of the perpetual dilemmas within AI research that defy resolution, rooted in the complex web of human ethics and cultural diversity, thus anchoring the pursuit of AI within the intractable tapestry of human values and societal imperatives [319], [320].

B. Overview of Impact Analysis

This subsection offers a detailed overview of the impact analysis carried out on the research taxonomy within the realm of generative AI, with a specific focus on recent progress in MoE, multimodality, and AGI, aiming to evaluate the impact of these innovative developments on various facets of generative AI research, ranging from model architecture to sophisticated learning methodologies, and includes both quantitative and qualitative assessments across a multitude of domains and subdomains in LLM research, shedding light on the extent to which each area is influenced by these technological advancements. This evaluation considered factors such as the emergence of new research directions, the necessity for redirection in existing research areas, the continued relevance of certain methodologies, and the potential redundancy of others, and has encapsulated in Table III.

1) Impact On Model Architecture: Transformer Models have been scored with a redirection requirement ($\hookrightarrow$) of 4 in both MoE and AGI, and a relevance ($\leftrightarrow$) of 3 in multimodality, leading to an overall score of 11. These models, forming the backbone of many current AI architectures, continue to be relevant for handling complex input sequences. However, the emergence of MoE and AGI indicates a shift towards more dynamic and specialized architectures. While transformers remain essential, there is a need for them to evolve and integrate with these advanced systems for enhanced performance and adaptability.

Recurrent Neural Networks (RNNs) are facing a potential decline in relevance, as indicated by their scores: likely to become redundant ($\searrow$) 2 in both MoE and AGI contexts and still relevant ($\leftrightarrow$) 3 in multimodality, totaling a score of 7. Although effective for sequence processing, RNNs are challenged by their limitations in handling long-range dependencies and lower efficiency compared to newer models like transformers. They may retain some relevance in multimodal tasks involving sequential data but are generally overshadowed by more advanced architectures.

The MoE models have scored a consistent relevance ($\leftrightarrow$) of 3 in their own development and a score of 5 ($\nearrow$) in multimodality, combined with a redirection score ($\hookrightarrow$) of 4 in the context of AGI, amounting to an overall score of 12. MoE models are at the forefront of emerging research in multimodality due to their ability to handle diverse data types. For AGI, these models will require adjustments to effectively integrate into systems exhibiting general intelligence, especially in areas beyond their initial specialization.

Multimodal Models have received high scores for emerging research directions ($\nearrow$) of 5 in both MoE and AGI contexts, alongside a score of 3 ($\leftrightarrow$) for current relevance in multimodality, culminating in an overall score of 13. The integration of MoE and the pursuit of AGI are opening new pathways for research in multimodal models. These developments are crucial for enhancing the ability to process and synthesize information from multiple modalities, a key aspect for both specialized and generalized AI systems.

2) Impact On Training Techniques: Supervised Learning has been assigned a redirection score ($\hookrightarrow$) of 4, a relevance score ($\leftrightarrow$) of 3 in multimodality, and a score indicating potential redundancy ($\searrow$) of 2 in the context of AGI, culminating in an overall score of 9. While supervised learning requires adaptation to fit the MoE framework, it remains relevant for multimodal AI models that depend on labeled data. However, with the shift towards more autonomous learning methods in AGI, the dependence on extensive labeled datasets typically associated with supervised learning may diminish, leading to its potential decrease in significance.

Table III: 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
	Mixture of Experts	↔ (3)	↗ (5)	↪ (4)	12
	Multimodal Models	↗ (5)	↔ (3)	↗ (5)	13
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

Unsupervised Learning scores a redirection requirement (↪) of 4 in both MoE and AGI contexts and maintains its relevance (↔) with a score of 3 in multimodality, resulting in a total score of 11. In the MoE architecture, unsupervised learning methods may need adjustments, particularly in managing dynamic task allocation. It remains crucial for understanding unlabeled data across various modalities. In AGI, unsupervised learning is expected to evolve beyond traditional techniques, focusing on more advanced self-discovery and intrinsic learning mechanisms.

Reinforcement Learning is rated as still relevant (↔) with a score of 3 in MoE, requiring redirection (↪) with a score of 4 in multimodality, and identified as an emerging research area (↗) with a score of 5 in AGI, giving it a total score of 12. This technique continues to play a significant role in optimizing MoE model structures. In the realm of multimodality, it necessitates a strategic shift to effectively manage complex interactions between different modalities. As for AGI, reinforcement learning is emerging as a crucial area, particularly in the development of autonomous systems that learn from their environment.

Transfer Learning receives a consistent relevance score (↔) of 3 in MoE, a high score for emerging research directions (↗) of 5 in multimodality, and a redirection requirement (↪) of 4 in AGI, accumulating to an overall score of 12. It remains important in the MoE framework for leveraging knowledge across different experts. In multimodal contexts, transfer learning is becoming increasingly crucial as it facilitates the transfer of learning between different modalities. With the evolution of AGI, this technique is expected to undergo significant changes to cater to broader and more generalized knowledge applications.

3) Impact On Application Domains: Natural Language Understanding holds steady relevance (↔) with a score of 3 in both MoE and multimodality, and an emerging direction (↗) score of 5 in AGI, totaling an overall score of 11. MoE models support the relevance of NLU by enhancing its precision and depth through their ability to handle large, diverse datasets. In multimodal AI, NLU remains a critical component for comprehending language in diverse data formats. With AGI's progress, NLU is expected to undergo significant expansion, moving towards more advanced, human-like comprehension and interpretation capabilities.

Natural Language Generation maintains relevance (↔) with a score of 3 in MoE, requires redirection (↪) with a score of 4 in multimodality, and is identified as an emerging research area (↗) with a score of 5 in AGI, resulting in a total score of 12. MoE's scalability is crucial for enhancing NLG, while in multimodal contexts, NLG may need strategic adjustments to align effectively with other modalities. As AGI evolves, NLG is anticipated to venture into new research domains, especially in creating content that reflects human-like creativity and adaptability.

Conversational AI is marked for redirection (↪) with a score of 4 in MoE, emerging research directions (↗) with a score of 5 in both multimodality and AGI, accumulating an overall score of 14. While MoE enhances conversational AI, it may require strategic changes to fully utilize MoE's distributed expertise. The integration of multiple modalities opens new avenues for conversational AI, expanding its scope to include various sensory data. The development of AGI is set to bring revolutionary advancements in this domain, paving the way for more autonomous, context-aware, and human-like interactions.

Creative AI scores a redirection requirement (↪) of 4 in

MoE, and high scores for emerging research directions (↗) of 5 in both multimodality and AGI, leading to a total score of 14. In the context of MoE, Creative AI may need to be realigned to capitalize on MoE's capacity for generating novel content. The combination of different modalities in creative AI presents exciting new research opportunities, enabling the creation of more intricate and diverse outputs. As AGI progresses, it is expected to significantly broaden the capabilities of creative AI, potentially surpassing existing boundaries and exploring new realms of creativity.

4) Impact On Compliance and Ethical Considerations: Bias Mitigation in the context of MoE, multimodality, and AGI scores a redirection requirement (↪) of 4 in both MoE and multimodality, and an emerging research direction (↗) with a score of 5 in AGI, resulting in an overall score of 13. MoE architectures demand a new approach in bias mitigation due to the diversity of expert networks, which could otherwise amplify biases. In multimodal systems, bias mitigation requires novel strategies to address biases in various data types, including non-textual forms like images and audio. With AGI's broad cognitive capabilities, a comprehensive approach towards understanding and addressing biases across diverse domains is emerging as a critical research area.

Data Security maintains a consistent relevance (↔) with a score of 3 across MoE, multimodality, and AGI, leading to a total score of 9. The fundamental principles of data security remain crucial despite the advancements in MoE, which may necessitate tailored strategies for its distributed nature. In multimodal AI, the secure handling of diverse data types continues to be of paramount importance. The core tenets of data security are sustained even with the advancement of AGI, though the complexity and scope of security measures are likely to increase.

AI Ethics is marked for redirection (↪) with a score of 4 in both MoE and multimodality, and faces inherently unresolvable challenges (△) with a score of 1 in AGI, accumulating a total score of 9. The decision-making processes and transparency of MoE models necessitate a reevaluation of ethical considerations. In multimodal AI, ethical concerns, particularly in the interpretation and use of multimodal data, require new approaches. The ethical challenges in AGI are expected to be complex and involve deep philosophical and societal implications that might be difficult to fully resolve.

Privacy Preservation scores a redirection need (↪) of 4 across MoE, multimodality, and AGI, leading to an overall score of 12. The distributed nature of MoE systems requires a reassessment of privacy preservation techniques to handle data processed by multiple experts. Multimodal AI systems, especially those handling sensitive data such as images and sounds, necessitate tailored privacy strategies. With the extensive data processing capabilities of AGI, advanced and potentially new approaches to privacy preservation are called for.

5) Impact On Advanced Learning: In the context of MoE, self-supervised learning requires redirection (↪) with a score of 4, signaling the need to adapt to the evolving architecture. Emerging research directions (↗) with a score of 5 are identified in multimodality, suggesting the integration of various autonomous data types like text, image, and audio. For AGI, self-supervised learning remains relevant (↔) with a score of 3, contributing to the system's autonomy and adaptability, though likely to be integrated with more complex strategies. The overall impact score is 12.

Meta-learning maintains consistent relevance (↔) with a score of 3 across MoE and multimodality, aligning well with the dynamic nature of MoE and aiding quick adaptation to varying data types and tasks in multimodal contexts. In AGI, it is marked as an emerging research direction (↗) with a score of 5, suggesting novel research in achieving human-like adaptability and learning efficiency. The total score for meta-learning is 11.

Fine tuning continues to be relevant (↔) with a score of 3 in both MoE and multimodality, being essential for adapting pre-trained models to specific tasks and tailoring multimodal models. However, in AGI, it is likely to become redundant (↘) with a score of 2, as AGI aims to develop systems that autonomously understand and learn across a broad range of domains, reducing the need for traditional fine-tuning processes. The overall impact score for fine tuning is 8.

Aligning AI with human values poses inherently unresolvable challenges (△) in all contexts—MoE, multimodality, and AGI—with a score of 1. This reflects the complexity and diversity of tasks MoE models handle, the integration of various data types in multimodal AI, and the broad range of cognitive abilities encompassed by AGI. These factors contribute to the significant ongoing challenges in aligning AI with human values, resulting in a total score of 3.

6) Impact On Emerging Trends: Multimodal learning is marked as an emerging research direction (↗) with a score of 5 in both MoE and AGI contexts, reflecting its capacity to integrate various data types such as text, images, and audio. This integration is crucial for specialized tasks in MoE and processing diverse forms of data in AGI. In the realm of multimodality, it remains a core aspect (↔) with a score of 3, being essential for ongoing multimodal AI development. The overall impact score is 13.

Interactive and Cooperative AI requires redirection (↪) in MoE with a score of 4, as MoE models adapt to include more interactive elements for broader applications. In multimodality, interaction and cooperation continue to be central (↔) with a score of 3, especially in fields like robotics and virtual assistants. AGI's evolution includes significant advancements in interactive AI, marking it as an emerging research area (↗) with a score of 5. The total score for this trend is 12.

The development of AGI necessitates redirection (↪) in both MoE and multimodality, each with a score of 4, indicating the need for more integrated and complex systems. AGI remains at the forefront of its own field (↔) with a score of 3, with each breakthrough directly influencing its progress. The overall impact score for AGI development is 11.

AGI containment is identified as a challenge not required to be solved (△) in both MoE and multimodality, with a score of 1, as these areas are not expected to reach the levels of autonomy and complexity associated with AGI. However, as AGI progresses, the emerging need for effective containment strategies is marked (↗) with a score of 5, highlighting the

importance of ensuring safe and controlled AI deployment. The total impact score is 7.

VIII. Emergent Research Priorities in Generative AI

As we are likely to approach the precipice of a new era marked by the advent of Q*, nudging us closer to the realization of usable AGI, the research landscape in generative AI is undergoing a crucial transformation.

A. Emergent Research Priorities in MoE

The MoE domain is increasingly focusing on two critical areas:

- **Multimodal Models in Model Architecture:** The integration of MoE and AGI is opening new pathways for research in multimodal models. These developments are enhancing the capability to process and synthesize information from multiple modalities, which is crucial for both specialized and generalized AI systems.
- **Multimodal Learning in Emerging Trends:** MoE is at the forefront of multimodal learning, integrating diverse data types like text, images, and audio for specialized tasks. This trend is directly impacting the enhancement of the field.

Furthermore, an analysis of funding trends and investment patterns in AI research could indicate a substantial shift towards areas like multimodal models in MoE. This trend, characterized by increased capital flow into fields involving complex data processing and autonomous systems, is shaping the direction of future research priorities. It underscores the growing interest and investment in the potential of generative AI, influencing both academic and industry-led initiatives.

B. Emergent Research Priorities in Multimodality

In the realm of multimodality, several areas are identified as emerging research priorities:

- **MoE in Model Architecture:** MoE models are becoming increasingly relevant for handling diverse data types in multimodal contexts.
- **Transfer Learning in Training Techniques:** Transfer learning is emerging as a key research direction, especially for learning between different modalities.
- **Conversational AI and Creative AI in Application Domains:** Both conversational AI and creative AI are expanding in multimodal contexts, encompassing visual, auditory, and other sensory data integration.
- **Self-Supervised Learning in Advanced Learning:** New research directions in self-supervised learning are emerging, focusing on the integration of various data types autonomously.

Additionally, the rise of generative AI, particularly in multimodal contexts, can significantly impact educational curricula and skill development. There is a growing need to update academic programs to include comprehensive AI literacy, with a focus on multimodal AI technologies. This evolution in education is aimed at preparing future professionals to effectively engage with and leverage the advancements in AI, equipping them with the necessary skills to navigate its complexities and innovations.

C. Emergent Research Priorities in AGI

The AGI domain is witnessing a surge in research priorities across multiple areas:

- **Multimodal Models in Model Architecture:** Similar to MoE, multimodal models are crucial in AGI, enabling deeper and more nuanced understanding.
- **Reinforcement Learning in Training Techniques:** Emerging as a key area in AGI, reinforcement learning focuses on developing autonomous systems learning from their environment.
- **Application Domains:** AGI is extending the boundaries of natural language understanding and generation, conversational AI, and creative AI, with a focus on human-like comprehension and creativity.
- **Bias Mitigation in Compliance and Ethical Considerations:** New directions in bias mitigation are focusing on a comprehensive approach to addressing biases across diverse domains in AGI.
- **Meta-Learning in Advanced Learning:** AGI's pursuit of human-like adaptability is leading to novel research in meta-learning.
- **Emerging Trends:** Multimodal learning, interactive and cooperative AI, and AGI containment strategies are becoming crucial research areas as AGI progresses.

In line with these developments in AGI, a noticeable trend in AI research funding and investment patterns is evident. There is a significant inclination towards supporting projects and studies in AGI, particularly in areas such as natural language understanding and generation, and autonomous systems. This funding trend not only mirrors the escalating interest in the capabilities of AGI but also directs the trajectory of future research, shaping both academic exploration and industry-driven projects.

IX. Practical Implications and Limitations of Generative AI Technologies

Generative AI technologies, encompassing MoE, multimodality, and AGI, present unique computational challenges. This section explores the processing power requirements, memory usage, and scalability concerns inherent in these advanced AI models.

A. Computational Complexity and Real-world Applications of Generative AI Technologies

1) Computational Complexity: Generative AI technologies, encompassing MoE, multimodality, and AGI, present unique computational challenges. This section explores the processing power requirements, memory usage, and scalability concerns inherent in these advanced AI models.

- **Processing Power Requirements**: Advanced generative AI models, including MoE architectures and AGI systems, require significant processing power [321]. The

demand for GPUs and TPUs is accentuated, particularly when handling complex computations and large datasets typical in multimodal AI applications.
- **Memory Usage in AI Modeling**: A critical challenge in training and deploying large-scale AI models, particularly in multimodal and AGI systems executed on GPUs, lies in the substantial GPU and VRAM requirements. Unlike computer RAM, VRAM often cannot be expanded easily on many platforms, posing significant constraints. Developing strategies for GPU and VRAM optimization and efficient model scaling is thus crucial for the practical deployment of these AI technologies.
- **Scalability and Efficiency in AI Deployment**: Addressing scalability challenges in generative AI, especially in MoE and AGI contexts, involves optimizing load management and parallel processing techniques. This is vital for their practical application in fields like healthcare, finance, and education.

2) Real-world Application Examples of Generative AI Technologies: The application of generative AI models in real-world scenarios demonstrates their transformative potential and challenges in various sectors.
- **Healthcare**: In healthcare, generative AI facilitates advancements in diagnostic imaging and personalized medicine, but also raises significant concerns regarding data privacy and the potential for misuse of sensitive health information [322].
- **Finance**: The use of AI for fraud detection and algorithmic trading in finance underlines its efficiency and accuracy, while at the same time, it raises ethical concerns, particularly in automated decision-making processes, which may lack transparency and accountability [323].
- **Education**: Generative AI's role in creating personalized learning experiences offers immense benefits in terms of educational accessibility and tailored instruction. However, it poses challenges in equitable access to technology, potential biases in AI-Generated Content (AIGC), and could reduce demand for human educators. Additionally, there's a growing concern about educators who are against the use of AIGC, fearing it may undermine traditional teaching methodologies and the role of educators.

B. Commercial Viability and Industry Solutions in Generative AI Technologies

1) Market Readiness: Assessing the market readiness of generative AI technologies involves analyzing cost, accessibility, deployment challenges, and user adoption trends.
- **Cost Analysis**: The financial aspects of deploying generative AI, including MoE, multimodality, and AGI, are crucial for market adoption.
- **Accessibility and Deployment**: Integration of these technologies into existing systems and the technical expertise required are key factors influencing their adoption.
- **User Adoption Trends**: Understanding current adoption patterns provides insights into market acceptance and the role of user trust and perceived benefits.

2) Existing Industry Solutions: Generative AI is reshaping various industries by offering innovative solutions and altering market dynamics.
- **Sector-Wise Deployment**: The diverse applications of generative AI, from digital content creation to process streamlining, also raise questions about originality and intellectual property rights.
- **Impact on Market Dynamics**: The effect of AI solutions on traditional industry structures and the introduction of novel business models are significant considerations.
- **Challenges and Constraints**: Addressing limitations such as scalability, data management complexity, privacy concerns, and ethical implications is essential for robust governance frameworks.

C. Limitations and Future Directions in Generative AI Technologies

1) Technical Limitations: Identifying and addressing technical limitations in generative AI models is crucial for their advancement and reliability.
- **Contextual Understanding**: Enhancing AI's ability to understand and interpret context, especially in natural language processing and image recognition, is a key area for improvement.
- **Handling Ambiguous Data**: Developing better algorithms for processing ambiguous or incomplete data sets is essential for decision-making accuracy and reliability.
- **Navigating Human Judgment**: Despite generative AI's accuracy in interpreting policies and procedures, its impact is limited in replacing human judgment. This is especially true in legal and political contexts where decision-makers might selectively use AIGC, leading to biased outcomes. Thus, the effectiveness of generative AI in such scenarios should be realistically assessed.

2) Future Research Directions to Enhance the Practicality of Generative AI: Future research in generative AI should focus on addressing current limitations and expanding its practical applications.
- **Improved Contextual Understanding**: Research should aim at developing models with better contextual awareness, particularly in complex natural language and image processing tasks.
- **Robust Handling of Ambiguous Data**: Investigating techniques for effective processing of ambiguous data is vital for advancing the decision-making capabilities of AI models.
- **Ethical Integration of AIGC in Legal and Political Arenas**: Future research should focus on the ethical integration of AI-generated content into legal and political decision-making processes, which involves developing frameworks that utilize AIGC in a supportive role, ensuring it enhances human judgment and contributes to transparency and fairness [324]. Importantly, researchers should consider the biases and limitations inherent in AI [324], alongside the potential for human fallibility, ethical complexities, and possible corruption in these domains.

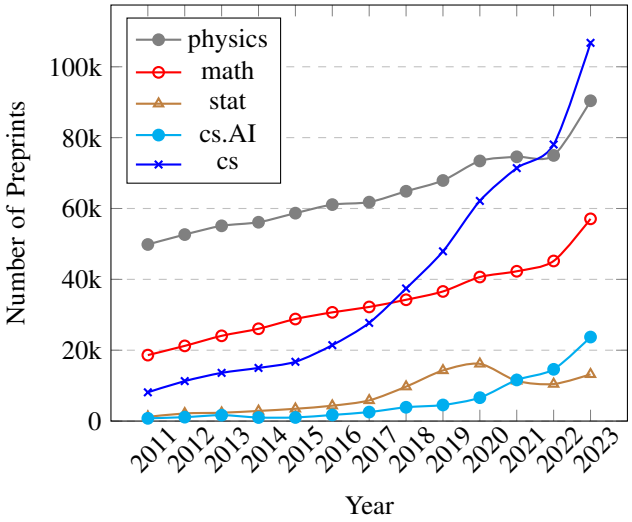

Figure 7: Annual preprint submissions to different categories on arXiv.org

X. Impact of Generative AI on Preprints Across Disciplines

The challenges detailed in this section are not directly related to the knowledge domains within generative AI, but are fueled by the success of Generative AI, particularly the commercialization of ChatGPT. The proliferation of preprints in the field of AI (Fig. 7), especially in the cs.AI category on platforms like arXiv, has introduced a set of academic challenges that merit careful consideration and strategic response. The rapid commercialization and adoption of tools such as ChatGPT, as evidenced by over 55,700 entries on Google Scholar mentioning "ChatGPT" within just one year of its commercialization, exemplify the accelerated pace at which the field is advancing. This rapid development is not mirrored in the traditional peer-review process, which is considerably slower. The peer-review process now appears to be overwhelmed with manuscripts that are either generated with ChatGPT (or other LLMs), or whose writing processes have been significantly accelerated by such LLMs, contributing to a bottleneck in scholarly communication [325], [326]. This situation is further compounded by the fact that many journals in disciplines outside of computer science are also experiencing longer review times and higher rates of desk rejections. Additionally, the flourishing trend of manuscripts and preprints, either generated by or significantly expedited using tools like ChatGPT, extends beyond computer science into diverse academic disciplines. This trend presents a looming challenge, potentially overwhelming both the traditional peer-review process and the flourishing preprint ecosystem with a volume of work that may not always adhere to established academic standards.

The sheer volume of preprints has made the task of selecting and scrutinizing research exceedingly demanding. In the current research era, the exploration of scientific literature has become increasingly complex, as knowledge has continued to expand and disseminate exponentially, while concurrently, integrative research efforts attempting to distill these vast literature, attempt to identify and understand a smaller sets of core contributions [327]. Thus, the rapid expansion of academic literature across various fields presents a significant challenge for researchers seeking to perform evidence syntheses over the increasingly vast body of available knowledge [328]. Furthermore, this explosion in publication volume poses a distinct challenge for literature reviews and surveys, where the human capacity for manually selecting, understanding, and critically evaluating articles is increasingly strained, potentially leading to gaps in synthesizing comprehensive knowledge landscapes. Although reproduction of results is a theoretical possibility, practical constraints such as the lack of technical expertise, computational resources, or access to proprietary datasets hinder rigorous evaluation. This is concerning, as the inability to thoroughly assess preprint research undermines the foundation of scientific reliability and validity. Furthermore, the peer-review system, a cornerstone of academic rigour, is under the threat of being further overwhelmed [325], [329]. The potential consequences are significant, with unvetted preprints possibly perpetuating biases or errors within the scientific community and beyond. The absence of established retraction mechanisms for preprints, akin to those for published articles, exacerbates the risk of persistent dissemination of flawed research.

The academic community is at a crossroads, necessitating an urgent and thoughtful discourse on navigating this emerging "mess" — a situation that risks spiraling out of control if left unaddressed. In this context, the role of peer review becomes increasingly crucial, as it serves as a critical checkpoint for quality and validity, ensuring that the rapid production of AI research is rigorously studied for scientific accuracy and relevance. However, the current *modus operandi* of traditional peer review does not appear to be sustainable, primarily due to its inability to keep pace with the exponential growth in AI-themed research and Generative-AI-accelerated research submissions, and the increasingly specialized nature of emerging AI topics [325], [326]. This situation is compounded by a finite pool of qualified reviewers, leading to delays, potential biases, and a burden on the scholarly community. This reality demands an exploration of new paradigms for peer review and dissemination of research that can keep pace with swift advancements in AI. Innovative models for community-driven vetting processes, enhanced reproducibility checks, and dynamic frameworks for post-publication scrutiny and correction may be necessary. Efforts to incorporate automated tools and AI-assisted review processes could also be explored to alleviate the strain on human reviewers.

In this rapidly evolving landscape, envision a convergence between the traditional peer review system and the flourishing preprint ecosystem, which could involve creating hybrid models (Fig. 8), where preprints undergo a preliminary community-based review, harnessing the collective expertise and rapid feedback of the academic community, similar to product review websites and Twitter [330]. This approach could provide an initial layer of validation, offering additional insights on issues that may be overlooked by a limited number of peer reviewers. The Editors-in-Chief (EICs) could consider the major criticisms and suggestions of an article from the community-based review, ensuring a more

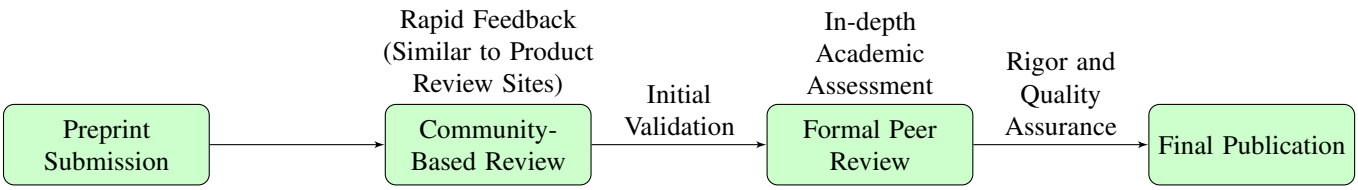

Figure 8: Possible Convergence Between Traditional Peer Review and the Preprint Ecosystem

thorough and diverse evaluation. Subsequent, more formal peer review processes could then refine and endorse these preprints for academic rigor and quality assurance. This hybrid model would require robust technological support, possibly leveraging AI and machine learning tools to assist in initial screening and identification of suitable reviewers. The aim would be to establish a seamless continuum from rapid dissemination to validated publication, ensuring both the speed of preprints and the credibility of peer-reviewed research. A balanced approach must be struck to harness the benefits of preprints—such as rapid dissemination of findings and open access—while mitigating their drawbacks. The development of new infrastructure and norms could be instrumental in steering the academic community towards a sustainable model that upholds the integrity and trustworthiness of scientific research in the age of Generative AI.

XI. Conclusions

This roadmap survey has embarked on an exploration of the transformative trends in generative AI research, particularly focusing on speculated advancements like Q* and the progressive strides towards AGI. Our analysis highlights a crucial paradigm shift, driven by innovations such as MoE, multimodal learning, 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, a challenge compounded by their increasing autonomy. The safety and robustness of AGI systems in diverse environments also remain a significant research gap. 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, ensuring that AI development is aligned with human values and global welfare. The roles of MoE, multimodal, and AGI in reshaping generative AI have been identified as significant, as their advancements can enhance model performance and versatility, and 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 towards enriching the human experience, aligning technological progress with ethical standards and societal well-being.

Disclaimer

The authors hereby declare no conflict of interest.

Abbreviations

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 Pre-trained 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] A. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, no. 236, p. 433, 1950.
[2] D. McDermott, "Artificial intelligence meets natural stupidity," *Acm Sigart Bulletin*, no. 57, pp. 4–9, 1976.
[3] M. Minsky, "Steps toward artificial intelligence," *Proceedings of the IRE*, vol. 49, no. 1, pp. 8–30, 1961.
[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[5] M. Minsky and S. Papert, "An introduction to computational geometry," *Cambridge tiass., HIT*, vol. 479, no. 480, p. 104, 1969.
[6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[7] G.-G. Lee, L. Shi, E. Latif, Y. Gao, A. Bewersdorf, M. Nyaaba, S. Guo, Z. Wu, Z. Liu, H. Wang *et al.*, "Multimodality of ai for education: Towards artificial general intelligence," *arXiv preprint arXiv:2312.06037*, 2023.

[8] P. Maddigan and T. Susnjak, "Chat2vis: Generating data visualisations via natural language using chatgpt, codex and gpt-3 large language models," *IEEE Access*, 2023.

[9] T. R. McIntosh, T. Liu, T. Susnjak, P. Watters, A. Ng, and M. N. Halgamuge, "A culturally sensitive test to evaluate nuanced gpt hallucination," *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 01, pp. 1–13, 2023.

[10] M. R. Morris, J. Sohl-dickstein, N. Fiedel, T. Warkentin, A. Dafoe, A. Faust, C. Farabet, and S. Legg, "Levels of agi: Operationalizing progress on the path to agi," *arXiv preprint arXiv:2311.02462*, 2023.

[11] J. Schuett, N. Dreksler, M. Anderljung, D. McCaffary, L. Heim, E. Bluemke, and B. Garfinkel, "Towards best practices in agi safety and governance: A survey of expert opinion," *arXiv preprint arXiv:2305.07153*, 2023.

[12] X. Shuai, J. Rollins, I. Moulinier, T. Custis, M. Edmunds, and F. Schilder, "A multidimensional investigation of the effects of publication retraction on scholarly impact," *Journal of the Association for Information Science and Technology*, vol. 68, no. 9, pp. 2225–2236, 2017.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[14] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever *et al.*, "Improving language understanding by generative pre-training," 2018.

[15] C. Huang, Z. Zhang, B. Mao, and X. Yao, "An overview of artificial intelligence ethics," *IEEE Transactions on Artificial Intelligence*, 2022.

[16] L. Besançon, N. Peiffer-Smadja, C. Segalas, H. Jiang, P. Masuzzo, C. Smout, E. Billy, M. Deforet, and C. Leyrat, "Open science saves lives: lessons from the covid-19 pandemic," *BMC Medical Research Methodology*, vol. 21, no. 1, pp. 1–18, 2021.

[17] C. R. Triggle, R. MacDonald, D. J. Triggle, and D. Grierson, "Requiem for impact factors and high publication charges," *Accountability in Research*, vol. 29, no. 3, pp. 133–164, 2022.

[18] T. McIntosh, A. Kayes, Y.-P. P. Chen, A. Ng, and P. Watters, "Ransomware mitigation in the modern era: A comprehensive review, research challenges, and future directions," *ACM Computing Surveys (CSUR)*, vol. 54, no. 9, pp. 1–36, 2021.

[19] T. McIntosh, T. Liu, T. Susnjak, H. Alavizadeh, A. Ng, R. Nowrozy, and P. Watters, "Harnessing gpt-4 for generation of cybersecurity grc policies: A focus on ransomware attack mitigation," *Computers & Security*, vol. 134, p. 103424, 2023.

[20] H. Bao, W. Wang, L. Dong, Q. Liu, O. K. Mohammed, K. Aggarwal, S. Som, S. Piao, and F. Wei, "Vlmo: Unified vision-language pre-training with mixture-of-modality-experts," *Advances in Neural Information Processing Systems*, vol. 35, pp. 32 897–32 912, 2022.

[21] N. Du, Y. Huang, A. M. Dai, S. Tong, D. Lepikhin, Y. Xu, M. Krikun, Y. Zhou, A. W. Yu, O. Firat *et al.*, "Glam: Efficient scaling of language models with mixture-of-experts," in *International Conference on Machine Learning*. PMLR, 2022, pp. 5547–5569.

[22] S. Masoudnia and R. Ebrahimpour, "Mixture of experts: a literature survey," *Artificial Intelligence Review*, vol. 42, pp. 275–293, 2014.

[23] C. Riquelme, J. Puigcerver, B. Mustafa, M. Neumann, R. Jenatton, A. Susano Pinto, D. Keysers, and N. Houlsby, "Scaling vision with sparse mixture of experts," *Advances in Neural Information Processing Systems*, vol. 34, pp. 8583–8595, 2021.

[24] S. E. Yuksel, J. N. Wilson, and P. D. Gader, "Twenty years of mixture of experts," *IEEE transactions on neural networks and learning systems*, vol. 23, no. 8, pp. 1177–1193, 2012.

[25] L. Zhang, S. Huang, W. Liu, and D. Tao, "Learning a mixture of granularity-specific experts for fine-grained categorization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 8331–8340.

[26] D. Martin, S. Malpica, D. Gutierrez, B. Masia, and A. Serrano, "Multimodality in vr: A survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 10s, pp. 1–36, 2022.

[27] Q. Sun, Q. Yu, Y. Cui, F. Zhang, X. Zhang, Y. Wang, H. Gao, J. Liu, T. Huang, and X. Wang, "Generative pretraining in multimodality," *arXiv preprint arXiv:2307.05222*, 2023.

[28] L. Wei, L. Xie, W. Zhou, H. Li, and Q. Tian, "Mvp: Multimodality-guided visual pre-training," in *European Conference on Computer Vision*. Springer, 2022, pp. 337–353.

[29] J. Wu, W. Zhou, X. Qian, J. Lei, L. Yu, and T. Luo, "Menet: Lightweight multimodality enhancement network for detecting salient objects in rgb-thermal images," *Neurocomputing*, vol. 527, pp. 119–129, 2023.

[30] Q. Ye, H. Xu, G. Xu, J. Ye, M. Yan, Y. Zhou, J. Wang, A. Hu, P. Shi, Y. Shi *et al.*, "mplug-owl: Modularization empowers large language models with multimodality," *arXiv preprint arXiv:2304.14178*, 2023.

[31] K. LaGrandeur, "How safe is our reliance on ai, and should we regulate it?" *AI and Ethics*, vol. 1, pp. 93–99, 2021.

[32] S. McLean, G. J. Read, J. Thompson, C. Baber, N. A. Stanton, and P. M. Salmon, "The risks associated with artificial general intelligence: A systematic review," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 35, no. 5, pp. 649–663, 2023.

[33] Y. K. Dwivedi, L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, Y. Duan, R. Dwivedi, J. Edwards, A. Eirug, V. Galanos, P. V. Ilavarasan, M. Janssen, P. Jones, A. K. Kar, H. Kizgin, B. Kronemann, B. Lal, B. Lucini, R. Medaglia, K. Le Meunier-FitzHugh, L. C. Le Meunier-FitzHugh, S. Misra, E. Mogaji, S. K. Sharma, J. B. Singh, V. Raghavan, R. Raman, N. P. Rana, S. Samothrakis, J. Spencer, K. Tamilmani, A. Tubadji, P. Walton, and M. D. Williams, "Artificial intelligence (ai): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *International Journal of Information Management*, vol. 57, p. 101994, 2021.

[34] I. Gabriel, "Artificial intelligence, values, and alignment," *Minds and Machines*, vol. 30, pp. 411–437, 2020.

[35] A. Shaban-Nejad, M. Michalowski, S. Bianco, J. S. Brownstein, D. L. Buckeridge, and R. L. Davis, "Applied artificial intelligence in healthcare: Listening to the winds of change in a post-covid-19 world," pp. 1969–1971, 2022.

[36] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung, "Survey of hallucination in natural language generation," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–38, 2023.

[37] B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz, E. Agirre, I. Heintz, and D. Roth, "Recent advances in natural language processing via large pre-trained language models: A survey," *ACM Computing Surveys*, vol. 56, no. 2, pp. 1–40, 2023.

[38] J. Li, X. Cheng, W. X. Zhao, J.-Y. Nie, and J.-R. Wen, "Halueval: A large-scale hallucination evaluation benchmark for large language models," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023, pp. 6449–6464.

[39] L. Weidinger, J. Mellor, M. Rauh, C. Griffin, J. Uesato, P.-S. Huang, M. Cheng, M. Glaese, B. Balle, A. Kasirzadeh *et al.*, "Ethical and social risks of harm from language models," *arXiv preprint arXiv:2112.04359*, 2021.

[40] X. Zhiheng, Z. Rui, and G. Tao, "Safety and ethical concerns of large language models," in *Proceedings of the 22nd Chinese National Conference on Computational Linguistics (Volume 4: Tutorial Abstracts)*, 2023, pp. 9–16.

[41] P. F. Brown, V. J. Della Pietra, P. V. Desouza, J. C. Lai, and R. L. Mercer, "Class-based n-gram models of natural language," *Computational linguistics*, vol. 18, no. 4, pp. 467–480, 1992.

[42] S. Katz, "Estimation of probabilities from sparse data for the language model component of a speech recognizer," *IEEE transactions on acoustics, speech, and signal processing*, vol. 35, no. 3, pp. 400–401, 1987.

[43] R. Kneser and H. Ney, "Improved backing-off for m-gram language modeling," in *1995 international conference on acoustics, speech, and signal processing*, vol. 1. IEEE, 1995, pp. 181–184.

[44] R. Kuhn and R. De Mori, "A cache-based natural language model for speech recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 12, no. 6, pp. 570–583, 1990.

[45] H. Ney, U. Essen, and R. Kneser, "On structuring probabilistic dependences in stochastic language modelling," *Computer Speech & Language*, vol. 8, no. 1, pp. 1–38, 1994.

[46] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[47] M. K. Nammous and K. Saeed, "Natural language processing: speaker, language, and gender identification with lstm," *Advanced Computing and Systems for Security: Volume Eight*, pp. 143–156, 2019.

[48] D. Wei, B. Wang, G. Lin, D. Liu, Z. Dong, H. Liu, and Y. Liu, "Research on unstructured text data mining and fault classification based on rnn-lstm with malfunction inspection report," *Energies*, vol. 10, no. 3, p. 406, 2017.

[49] L. Yao and Y. Guan, "An improved lstm structure for natural language processing," in *2018 IEEE International Conference of Safety Produce Informatization (IICSPI)*. IEEE, 2018, pp. 565–569.

[50] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, "Training language

models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27 730–27 744, 2022.
[51] T. Susnjak, "Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and chatgpt," *International Journal of Artificial Intelligence in Education*, pp. 1–31, 2023.
[52] T. Susnjak, E. Griffin, M. McCutcheon, and K. Potter, "Towards clinical prediction with transparency: An explainable ai approach to survival modelling in residential aged care," *arXiv preprint arXiv:2312.00271*, 2023.
[53] R. Yang, T. F. Tan, W. Lu, A. J. Thirunavukarasu, D. S. W. Ting, and N. Liu, "Large language models in health care: Development, applications, and challenges," *Health Care Science*, vol. 2, no. 4, pp. 255–263, 2023.
[54] D. Baidoo-Anu and L. O. Ansah, "Education in the era of generative artificial intelligence (ai): Understanding the potential benefits of chatgpt in promoting teaching and learning," *Journal of AI*, vol. 7, no. 1, pp. 52–62, 2023.
[55] T. Susnjak, "Chatgpt: The end of online exam integrity?" *arXiv preprint arXiv:2212.09292*, 2022.
[56] A. Tlili, B. Shehata, M. A. Adarkwah, A. Bozkurt, D. T. Hickey, R. Huang, and B. Agyemang, "What if the devil is my guardian angel: Chatgpt as a case study of using chatbots in education," *Smart Learning Environments*, vol. 10, no. 1, p. 15, 2023.
[57] M. A. AlAfnan, S. Dishari, M. Jovic, and K. Lomidze, "Chatgpt as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses," *Journal of Artificial Intelligence and Technology*, vol. 3, no. 2, pp. 60–68, 2023.
[58] A. S. George and A. H. George, "A review of chatgpt ai's impact on several business sectors," *Partners Universal International Innovation Journal*, vol. 1, no. 1, pp. 9–23, 2023.
[59] G. K. Hadfield and J. Clark, "Regulatory markets: The future of ai governance," *arXiv preprint arXiv:2304.04914*, 2023.
[60] M. Bakker, M. Chadwick, H. Sheahan, M. Tessler, L. Campbell-Gillingham, J. Balaguer, N. McAleese, A. Glaese, J. Aslanides, M. Botvinick *et al.*, "Fine-tuning language models to find agreement among humans with diverse preferences," *Advances in Neural Information Processing Systems*, vol. 35, pp. 38 176–38 189, 2022.
[61] Z. Hu, Y. Lan, L. Wang, W. Xu, E.-P. Lim, R. K.-W. Lee, L. Bing, and S. Poria, "Llm-adapters: An adapter family for parameter-efficient fine-tuning of large language models," *arXiv preprint arXiv:2304.01933*, 2023.
[62] H. Liu, D. Tam, M. Muqeeth, J. Mohta, T. Huang, M. Bansal, and C. A. Raffel, "Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 1950–1965, 2022.
[63] H. Zheng, L. Shen, A. Tang, Y. Luo, H. Hu, B. Du, and D. Tao, "Learn from model beyond fine-tuning: A survey," *arXiv preprint arXiv:2310.08184*, 2023.
[64] P. Manakul, A. Liusie, and M. J. Gales, "Selfcheckgpt: Zero-resource black-box hallucination detection for generative large language models," *arXiv preprint arXiv:2303.08896*, 2023.
[65] A. Martino, M. Iannelli, and C. Truong, "Knowledge injection to counter large language model (llm) hallucination," in *European Semantic Web Conference*. Springer, 2023, pp. 182–185.
[66] J.-Y. Yao, K.-P. Ning, Z.-H. Liu, M.-N. Ning, and L. Yuan, "Llm lies: Hallucinations are not bugs, but features as adversarial examples," *arXiv preprint arXiv:2310.01469*, 2023.
[67] Y. Zhang, Y. Li, L. Cui, D. Cai, L. Liu, T. Fu, X. Huang, E. Zhao, Y. Zhang, Y. Chen *et al.*, "Siren's song in the ai ocean: A survey on hallucination in large language models," *arXiv preprint arXiv:2309.01219*, 2023.
[68] J. Ji, M. Liu, J. Dai, X. Pan, C. Zhang, C. Bian, R. Sun, Y. Wang, and Y. Yang, "Beavertails: Towards improved safety alignment of llm via a human-preference dataset," *arXiv preprint arXiv:2307.04657*, 2023.
[69] Y. Liu, Y. Yao, J.-F. Ton, X. Zhang, R. G. H. Cheng, Y. Klochkov, M. F. Taufiq, and H. Li, "Trustworthy llms: a survey and guideline for evaluating large language models' alignment," *arXiv preprint arXiv:2308.05374*, 2023.
[70] Y. Wang, W. Zhong, L. Li, F. Mi, X. Zeng, W. Huang, L. Shang, X. Jiang, and Q. Liu, "Aligning large language models with human: A survey," *arXiv preprint arXiv:2307.12966*, 2023.
[71] Z. Sun, Y. Shen, Q. Zhou, H. Zhang, Z. Chen, D. Cox, Y. Yang, and C. Gan, "Principle-driven self-alignment of language models from scratch with minimal human supervision," *arXiv preprint arXiv:2305.03047*, 2023.
[72] Y. Wolf, N. Wies, Y. Levine, and A. Shashua, "Fundamental limitations of alignment in large language models," *arXiv preprint arXiv:2304.11082*, 2023.
[73] H. Dang, L. Mecke, F. Lehmann, S. Goller, and D. Buschek, "How to prompt? opportunities and challenges of zero-and few-shot learning for human-ai interaction in creative applications of generative models," *arXiv preprint arXiv:2209.01390*, 2022.
[74] R. Ma, X. Zhou, T. Gui, Y. Tan, L. Li, Q. Zhang, and X. Huang, "Template-free prompt tuning for few-shot ner," *arXiv preprint arXiv:2109.13532*, 2021.
[75] C. Qin and S. Joty, "Lfpt5: A unified framework for lifelong few-shot language learning based on prompt tuning of t5," *arXiv preprint arXiv:2110.07298*, 2021.
[76] S. Wang, L. Tang, A. Majety, J. F. Rousseau, G. Shih, Y. Ding, and Y. Peng, "Trustworthy assertion classification through prompting," *Journal of biomedical informatics*, vol. 132, p. 104139, 2022.
[77] Y. Fan, F. Jiang, P. Li, and H. Li, "Grammargpt: Exploring open-source llms for native chinese grammatical error correction with supervised fine-tuning," in *CCF International Conference on Natural Language Processing and Chinese Computing*. Springer, 2023, pp. 69–80.
[78] D. Liga and L. Robaldo, "Fine-tuning gpt-3 for legal rule classification," *Computer Law & Security Review*, vol. 51, p. 105864, 2023.
[79] Y. Liu, A. Singh, C. D. Freeman, J. D. Co-Reyes, and P. J. Liu, "Improving large language model fine-tuning for solving math problems," *arXiv preprint arXiv:2310.10047*, 2023.
[80] Z. Talat, A. Névéol, S. Biderman, M. Clinciu, M. Dey, S. Longpre, S. Luccioni, M. Masoud, M. Mitchell, D. Radev *et al.*, "You reap what you sow: On the challenges of bias evaluation under multilingual settings," in *Proceedings of BigScience Episode# 5–Workshop on Challenges & Perspectives in Creating Large Language Models*, 2022, pp. 26–41.
[81] Y. Liu, S. Yu, and T. Lin, "Hessian regularization of deep neural networks: A novel approach based on stochastic estimators of hessian trace," *Neurocomputing*, vol. 536, pp. 13–20, 2023.
[82] Y. Lu, Y. Bo, and W. He, "Confidence adaptive regularization for deep learning with noisy labels," *arXiv preprint arXiv:2108.08212*, 2021.
[83] G. Pereyra, G. Tucker, J. Chorowski, Ł. Kaiser, and G. Hinton, "Regularizing neural networks by penalizing confident output distributions," *arXiv preprint arXiv:1701.06548*, 2017.
[84] E. Chen, Z.-W. Hong, J. Pajarinen, and P. Agrawal, "Redeeming intrinsic rewards via constrained optimization," *Advances in Neural Information Processing Systems*, vol. 35, pp. 4996–5008, 2022.
[85] Y. Jiang, Z. Li, M. Tan, S. Wei, G. Zhang, Z. Guan, and B. Han, "A stable block adjustment method without ground control points using bound constrained optimization," *International Journal of Remote Sensing*, vol. 43, no. 12, pp. 4708–4722, 2022.
[86] M. Kachuee and S. Lee, "Constrained policy optimization for controlled self-learning in conversational ai systems," *arXiv preprint arXiv:2209.08429*, 2022.
[87] Z. Song, H. Wang, and Y. Jin, "A surrogate-assisted evolutionary framework with regions of interests-based data selection for expensive constrained optimization," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023.
[88] J. Yu, T. Xu, Y. Rong, J. Huang, and R. He, "Structure-aware conditional variational auto-encoder for constrained molecule optimization," *Pattern Recognition*, vol. 126, p. 108581, 2022.
[89] P. Butlin, "Ai alignment and human reward," in *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 2021, pp. 437–445.
[90] F. Faal, K. Schmitt, and J. Y. Yu, "Reward modeling for mitigating toxicity in transformer-based language models," *Applied Intelligence*, vol. 53, no. 7, pp. 8421–8435, 2023.
[91] J. Leike, D. Krueger, T. Everitt, M. Martic, V. Maini, and S. Legg, "Scalable agent alignment via reward modeling: a research direction," *arXiv preprint arXiv:1811.07871*, 2018.
[92] L. Li, Y. Chai, S. Wang, Y. Sun, H. Tian, N. Zhang, and H. Wu, "Tool-augmented reward modeling," *arXiv preprint arXiv:2310.01045*, 2023.
[93] F. Barreto, L. Moharkar, M. Shirodkar, V. Sarode, S. Gonsalves, and A. Johns, "Generative artificial intelligence: Opportunities and challenges of large language models," in *International Conference on Intelligent Computing and Networking*. Springer, 2023, pp. 545–553.
[94] Z. Chen, Z. Wang, Z. Wang, H. Liu, Z. Yin, S. Liu, L. Sheng, W. Ouyang, Y. Qiao, and J. Shao, "Octavius: Mitigating task interference in mllms via moe," *arXiv preprint arXiv:2311.02684*, 2023.
[95] C. Dun, M. D. C. H. Garcia, G. Zheng, A. H. Awadallah, A. Kyrillidis, and R. Sim, "Sweeping heterogeneity with smart mops: Mixture of prompts for llm task adaptation," *arXiv preprint arXiv:2310.02842*, 2023.

[96] H. Naveed, A. U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, N. Barnes, and A. Mian, "A comprehensive overview of large language models," *arXiv preprint arXiv:2307.06435*, 2023.

[97] F. Xue, Y. Fu, W. Zhou, Z. Zheng, and Y. You, "To repeat or not to repeat: Insights from scaling llm under token-crisis," *arXiv preprint arXiv:2305.13230*, 2023.

[98] M. Nowaz Rabbani Chowdhury, S. Zhang, M. Wang, S. Liu, and P.-Y. Chen, "Patch-level routing in mixture-of-experts is provably sample-efficient for convolutional neural networks," *arXiv e-prints*, pp. arXiv–2306, 2023.

[99] J. Peng, K. Zhou, R. Zhou, T. Hartvigsen, Y. Zhang, Z. Wang, and T. Chen, "Sparse moe as a new treatment: Addressing forgetting, fitting, learning issues in multi-modal multi-task learning," in *Conference on Parsimony and Learning (Recent Spotlight Track)*, 2023.

[100] C. N. d. Santos, J. Lee-Thorp, I. Noble, C.-C. Chang, and D. Uthus, "Memory augmented language models through mixture of word experts," *arXiv preprint arXiv:2311.10768*, 2023.

[101] W. Wang, G. Ma, Y. Li, and B. Du, "Language-routing mixture of experts for multilingual and code-switching speech recognition," *arXiv preprint arXiv:2307.05956*, 2023.

[102] X. Zhao, X. Chen, Y. Cheng, and T. Chen, "Sparse moe with language guided routing for multilingual machine translation," in *Conference on Parsimony and Learning (Recent Spotlight Track)*, 2023.

[103] W. Huang, H. Zhang, P. Peng, and H. Wang, "Multi-gate mixture-of-expert combined with synthetic minority over-sampling technique for multimode imbalanced fault diagnosis," in *2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. IEEE, 2023, pp. 456–461.

[104] B. Liu, L. Ding, L. Shen, K. Peng, Y. Cao, D. Cheng, and D. Tao, "Diversifying the mixture-of-experts representation for language models with orthogonal optimizer," *arXiv preprint arXiv:2310.09762*, 2023.

[105] W. Wang, Z. Lai, S. Li, W. Liu, K. Ge, Y. Liu, A. Shen, and D. Li, "Prophet: Fine-grained load balancing for parallel training of large-scale moe models," in *2023 IEEE International Conference on Cluster Computing (CLUSTER)*. IEEE, 2023, pp. 82–94.

[106] X. Yao, S. Liang, S. Han, and H. Huang, "Enhancing molecular property prediction via mixture of collaborative experts," *arXiv preprint arXiv:2312.03292*, 2023.

[107] Z. Xiao, Y. Jiang, G. Tang, L. Liu, S. Xu, Y. Xiao, and W. Yan, "Adversarial mixture of experts with category hierarchy soft constraint," in *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. IEEE, 2021, pp. 2453–2463.

[108] M. Agbese, R. Mohanani, A. Khan, and P. Abrahamsson, "Implementing ai ethics: Making sense of the ethical requirements," in *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering*, 2023, pp. 62–71.

[109] Z. Chen, Y. Deng, Y. Wu, Q. Gu, and Y. Li, "Towards understanding the mixture-of-experts layer in deep learning," *Advances in neural information processing systems*, vol. 35, pp. 23 049–23 062, 2022.

[110] Y. Zhou, T. Lei, H. Liu, N. Du, Y. Huang, V. Zhao, A. M. Dai, Q. V. Le, J. Laudon *et al.*, "Mixture-of-experts with expert choice routing," *Advances in Neural Information Processing Systems*, vol. 35, pp. 7103–7114, 2022.

[111] N. Guha, C. Lawrence, L. A. Gailmard, K. Rodolfa, F. Surani, R. Bommasani, I. Raji, M.-F. Cuéllar, C. Honigsberg, P. Liang *et al.*, "Ai regulation has its own alignment problem: The technical and institutional feasibility of disclosure, registration, licensing, and auditing," *George Washington Law Review, Forthcoming*, 2023.

[112] Gemini Team, Google, "Gemini: A family of highly capable multimodal models," 2023, accessed: 17 December 2023. [Online]. Available: https://storage.googleapis.com/deepmind-media/gemini/gemini_1_report.pdf

[113] J. N. Acosta, G. J. Falcone, P. Rajpurkar, and E. J. Topol, "Multimodal biomedical ai," *Nature Medicine*, vol. 28, no. 9, pp. 1773–1784, 2022.

[114] S. Qi, Z. Cao, J. Rao, L. Wang, J. Xiao, and X. Wang, "What is the limitation of multimodal llms? a deeper look into multimodal llms through prompt probing," *Information Processing & Management*, vol. 60, no. 6, p. 103510, 2023.

[115] B. Xu, D. Kocyigit, R. Grimm, B. P. Griffin, and F. Cheng, "Applications of artificial intelligence in multimodality cardiovascular imaging: a state-of-the-art review," *Progress in cardiovascular diseases*, vol. 63, no. 3, pp. 367–376, 2020.

[116] A. Birhane, V. U. Prabhu, and E. Kahembwe, "Multimodal datasets: misogyny, pornography, and malignant stereotypes," *arXiv preprint arXiv:2110.01963*, 2021.

[117] Y. Li, W. Li, N. Li, X. Qiu, and K. B. Manokaran, "Multimodal information interaction and fusion for the parallel computing system using ai techniques," *International Journal of High Performance Systems Architecture*, vol. 10, no. 3-4, pp. 185–196, 2021.

[118] C. Zhang, Z. Yang, X. He, and L. Deng, "Multimodal intelligence: Representation learning, information fusion, and applications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 3, pp. 478–493, 2020.

[119] H. Qiao, V. Liu, and L. Chilton, "Initial images: using image prompts to improve subject representation in multimodal ai generated art," in *Proceedings of the 14th Conference on Creativity and Cognition*, 2022, pp. 15–28.

[120] A. E. Stewart, Z. Keirn, and S. K. D'Mello, "Multimodal modeling of collaborative problem-solving facets in triads," *User Modeling and User-Adapted Interaction*, pp. 1–39, 2021.

[121] L. Xue, N. Yu, S. Zhang, J. Li, R. Martín-Martín, J. Wu, C. Xiong, R. Xu, J. C. Niebles, and S. Savarese, "Ulip-2: Towards scalable multimodal pre-training for 3d understanding," *arXiv preprint arXiv:2305.08275*, 2023.

[122] L. Yan, L. Zhao, D. Gasevic, and R. Martinez-Maldonado, "Scalability, sustainability, and ethicality of multimodal learning analytics," in *LAK22: 12th international learning analytics and knowledge conference*, 2022, pp. 13–23.

[123] Y. Liu-Thompkins, S. Okazaki, and H. Li, "Artificial empathy in marketing interactions: Bridging the human-ai gap in affective and social customer experience," *Journal of the Academy of Marketing Science*, vol. 50, no. 6, pp. 1198–1218, 2022.

[124] M. S. Rahman, S. Bag, M. A. Hossain, F. A. M. A. Fattah, M. O. Gani, and N. P. Rana, "The new wave of ai-powered luxury brands online shopping experience: The role of digital multisensory cues and customers' engagement," *Journal of Retailing and Consumer Services*, vol. 72, p. 103273, 2023.

[125] E. Sachdeva, N. Agarwal, S. Chundi, S. Roelofs, J. Li, B. Dariush, C. Choi, and M. Kochenderfer, "Rank2tell: A multimodal driving dataset for joint importance ranking and reasoning," *arXiv preprint arXiv:2309.06597*, 2023.

[126] C. Cui, Y. Ma, X. Cao, W. Ye, Y. Zhou, K. Liang, J. Chen, J. Lu, Z. Yang, K.-D. Liao *et al.*, "A survey on multimodal large language models for autonomous driving," *arXiv preprint arXiv:2311.12320*, 2023.

[127] A. B. Temsamani, A. K. Chavali, W. Vervoort, T. Tuytelaars, G. Radevski, H. Van Hamme, K. Mets, M. Hutsebaut-Buysse, T. De Schepper, and S. Latré, "A multimodal ai approach for intuitively instructable autonomous systems: a case study of an autonomous off-highway vehicle," in *The Eighteenth International Conference on Autonomic and Autonomous Systems, ICAS 2022, May 22-26, 2022, Venice, Italy*, 2022, pp. 31–39.

[128] J. Lee and S. Y. Shin, "Something that they never said: Multimodal disinformation and source vividness in understanding the power of ai-enabled deepfake news," *Media Psychology*, vol. 25, no. 4, pp. 531–546, 2022.

[129] S. Muppalla, S. Jia, and S. Lyu, "Integrating audio-visual features for multimodal deepfake detection," *arXiv preprint arXiv:2310.03827*, 2023.

[130] S. Kumar, M. K. Chaube, S. N. Nenavath, S. K. Gupta, and S. K. Tetarave, "Privacy preservation and security challenges: a new frontier multimodal machine learning research," *International Journal of Sensor Networks*, vol. 39, no. 4, pp. 227–245, 2022.

[131] J. Marchang and A. Di Nuovo, "Assistive multimodal robotic system (amrsys): security and privacy issues, challenges, and possible solutions," *Applied Sciences*, vol. 12, no. 4, p. 2174, 2022.

[132] A. Peña, I. Serna, A. Morales, J. Fierrez, A. Ortega, A. Herrarte, M. Alcantara, and J. Ortega-Garcia, "Human-centric multimodal machine learning: Recent advances and testbed on ai-based recruitment," *SN Computer Science*, vol. 4, no. 5, p. 434, 2023.

[133] R. Wolfe and A. Caliskan, "American== white in multimodal language-and-image ai," in *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 2022, pp. 800–812.

[134] R. Wolfe, Y. Yang, B. Howe, and A. Caliskan, "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*, 2023, pp. 1174–1185.

[135] M. Afshar, B. Sharma, D. Dligach, M. Oguss, R. Brown, N. Chhabra, H. M. Thompson, T. Markossian, C. Joyce, M. M. Churpek *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," *The Lancet Digital Health*, vol. 4, no. 6, pp. e426–e435, 2022.

[136] H. Alwahaby, M. Cukurova, Z. Papamitsiou, and M. Giannakos, "The evidence of impact and ethical considerations of multimodal learning analytics: A systematic literature review," *The Multimodal Learning Analytics Handbook*, pp. 289–325, 2022.

[137] Q. Miao, W. Zheng, Y. Lv, M. Huang, W. Ding, and F.-Y. Wang, "Dao to hanoi via desci: Ai paradigm shifts from alphago to chatgpt," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 4, pp. 877–897, 2023.

[138] Y. Rong, "Roadmap of alphago to alphastar: Problems and challenges," in *2nd International Conference on Artificial Intelligence, Automation, and High-Performance Computing (AIAHPC 2022)*, vol. 12348. SPIE, 2022, pp. 904–914.

[139] Y. Gao, M. Zhou, D. Liu, Z. Yan, S. Zhang, and D. N. Metaxas, "A data-scalable transformer for medical image segmentation: architecture, model efficiency, and benchmark," *arXiv preprint arXiv:2203.00131*, 2022.

[140] W. Peebles and S. Xie, "Scalable diffusion models with transformers," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4195–4205.

[141] R. Pope, S. Douglas, A. Chowdhery, J. Devlin, J. Bradbury, J. Heek, K. Xiao, S. Agrawal, and J. Dean, "Efficiently scaling transformer inference," *Proceedings of Machine Learning and Systems*, vol. 5, 2023.

[142] Y. Ding and M. Jia, "Convolutional transformer: An enhanced attention mechanism architecture for remaining useful life estimation of bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.

[143] Y. Ding, M. Jia, Q. Miao, and Y. Cao, "A novel time–frequency transformer based on self–attention mechanism and its application in fault diagnosis of rolling bearings," *Mechanical Systems and Signal Processing*, vol. 168, p. 108616, 2022.

[144] G. Wang, Y. Zhao, C. Tang, C. Luo, and W. Zeng, "When shift operation meets vision transformer: An extremely simple alternative to attention mechanism," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, 2022, pp. 2423–2430.

[145] H. Cai, J. Li, M. Hu, C. Gan, and S. Han, "Efficientvit: Lightweight multi-scale attention for high-resolution dense prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 17 302–17 313.

[146] X. Liu, H. Peng, N. Zheng, Y. Yang, H. Hu, and Y. Yuan, "Efficientvit: Memory efficient vision transformer with cascaded group attention," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14 420–14 430.

[147] Y. Li, Q. Fan, H. Huang, Z. Han, and Q. Gu, "A modified yolov8 detection network for uav aerial image recognition," *Drones*, vol. 7, no. 5, p. 304, 2023.

[148] F. M. Talaat and H. ZainEldin, "An improved fire detection approach based on yolo-v8 for smart cities," *Neural Computing and Applications*, vol. 35, no. 28, pp. 20 939–20 954, 2023.

[149] S. Tamang, B. Sen, A. Pradhan, K. Sharma, and V. K. Singh, "Enhancing covid-19 safety: Exploring yolov8 object detection for accurate face mask classification," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 2, pp. 892–897, 2023.

[150] J. Lu, R. Xiong, J. Tian, C. Wang, C.-W. Hsu, N.-T. Tsou, F. Sun, and J. Li, "Battery degradation prediction against uncertain future conditions with recurrent neural network enabled deep learning," *Energy Storage Materials*, vol. 50, pp. 139–151, 2022.

[151] A. Onan, "Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 5, pp. 2098–2117, 2022.

[152] F. Shan, X. He, D. J. Armaghani, P. Zhang, and D. Sheng, "Success and challenges in predicting tbm penetration rate using recurrent neural networks," *Tunnelling and Underground Space Technology*, vol. 130, p. 104728, 2022.

[153] C. Sridhar, P. K. Pareek, R. Kalidoss, S. S. Jamal, P. K. Shukla, S. J. Nuagah *et al.*, "Optimal medical image size reduction model creation using recurrent neural network and genpsowvq," *Journal of Healthcare Engineering*, vol. 2022, 2022.

[154] J. Zhu, Q. Jiang, Y. Shen, C. Qian, F. Xu, and Q. Zhu, "Application of recurrent neural network to mechanical fault diagnosis: A review," *Journal of Mechanical Science and Technology*, vol. 36, no. 2, pp. 527–542, 2022.

[155] S. Lin, W. Lin, W. Wu, F. Zhao, R. Mo, and H. Zhang, "Segrnn: Segment recurrent neural network for long-term time series forecasting," *arXiv preprint arXiv:2308.11200*, 2023.

[156] Z. Wei, X. Zhang, and M. Sun, "Extracting weighted finite automata from recurrent neural networks for natural languages," in *International Conference on Formal Engineering Methods*. Springer, 2022, pp. 370–385.

[157] F. Bonassi, M. Farina, J. Xie, and R. Scattolini, "On recurrent neural networks for learning-based control: recent results and ideas for future developments," *Journal of Process Control*, vol. 114, pp. 92–104, 2022.

[158] Z. Guo, Y. Tang, R. Zhang, D. Wang, Z. Wang, B. Zhao, and X. Li, "Viewrefer: Grasp the multi-view knowledge for 3d visual grounding," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 15 372–15 383.

[159] C. Pan, Y. He, J. Peng, Q. Zhang, W. Sui, and Z. Zhang, "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*, 2023, pp. 9590–9599.

[160] P. Xu, X. Zhu, and D. A. Clifton, "Multimodal learning with transformers: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[161] I. Molenaar, S. de Mooij, R. Azevedo, M. Bannert, S. Järvelä, and D. Gašević, "Measuring self-regulated learning and the role of ai: Five years of research using multimodal multichannel data," *Computers in Human Behavior*, vol. 139, p. 107540, 2023.

[162] S. Steyaert, M. Pizurica, D. Nagaraj, P. Khandelwal, T. Hernandez-Boussard, A. J. Gentles, and O. Gevaert, "Multimodal data fusion for cancer biomarker discovery with deep learning," *Nature Machine Intelligence*, vol. 5, no. 4, pp. 351–362, 2023.

[163] V. Rani, S. T. Nabi, M. Kumar, A. Mittal, and K. Kumar, "Self-supervised learning: A succinct review," *Archives of Computational Methods in Engineering*, vol. 30, no. 4, pp. 2761–2775, 2023.

[164] M. C. Schiappa, Y. S. Rawat, and M. Shah, "Self-supervised learning for videos: A survey," *ACM Computing Surveys*, vol. 55, no. 13s, pp. 1–37, 2023.

[165] J. Yu, H. Yin, X. Xia, T. Chen, J. Li, and Z. Huang, "Self-supervised learning for recommender systems: A survey," *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[166] V. Bharti, A. Kumar, V. Purohit, R. Singh, A. K. Singh, and S. K. Singh, "A label efficient semi self-supervised learning framework for iot devices in industrial process," *IEEE Transactions on Industrial Informatics*, 2023.

[167] D. Sam and J. Z. Kolter, "Losses over labels: Weakly supervised learning via direct loss construction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 8, 2023, pp. 9695–9703.

[168] M. Wang, P. Xie, Y. Du, and X. Hu, "T5-based model for abstractive summarization: A semi-supervised learning approach with consistency loss functions," *Applied Sciences*, vol. 13, no. 12, p. 7111, 2023.

[169] Q. Li, X. Peng, Y. Qiao, and Q. Hao, "Unsupervised person re-identification with multi-label learning guided self-paced clustering," *Pattern Recognition*, vol. 125, p. 108521, 2022.

[170] P. Nancy, H. Pallathadka, M. Naved, K. Kaliyaperumal, K. Arumugam, and V. Garchar, "Deep learning and machine learning based efficient framework for image based plant disease classification and detection," in *2022 International Conference on Advanced Computing Technologies and Applications (ICACTA)*. IEEE, 2022, pp. 1–6.

[171] P. An, Z. Wang, and C. Zhang, "Ensemble unsupervised autoencoders and gaussian mixture model for cyberattack detection," *Information Processing & Management*, vol. 59, no. 2, p. 102844, 2022.

[172] S. Yan, H. Shao, Y. Xiao, B. Liu, and J. Wan, "Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises," *Robotics and Computer-Integrated Manufacturing*, vol. 79, p. 102441, 2023.

[173] E. Ayanoglu, K. Davaslioglu, and Y. E. Sagduyu, "Machine learning in nextg networks via generative adversarial networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 480–501, 2022.

[174] K. Yan, X. Chen, X. Zhou, Z. Yan, and J. Ma, "Physical model informed fault detection and diagnosis of air handling units based on transformer generative adversarial network," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 2192–2199, 2022.

[175] N.-R. Zhou, T.-F. Zhang, X.-W. Xie, and J.-Y. Wu, "Hybrid quantum–classical generative adversarial networks for image generation via learning discrete distribution," *Signal Processing: Image Communication*, vol. 110, p. 116891, 2023.

[176] P. Ladosz, L. Weng, M. Kim, and H. Oh, "Exploration in deep reinforcement learning: A survey," *Information Fusion*, vol. 85, pp. 1–22, 2022.

[177] Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, E. Uchibe, and J. Morimoto, "Deep learning, reinforcement learning, and world models," *Neural Networks*, vol. 152, pp. 267–275, 2022.

[178] D. Bertoin, A. Zouitine, M. Zouitine, and E. Rachelson, "Look where you look! saliency-guided q-networks for generalization in visual reinforcement learning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 30 693–30 706, 2022.

[179] A. Hafiz, "A survey of deep q-networks used for reinforcement learning: State of the art," *Intelligent Communication Technologies and Virtual Mobile Networks: Proceedings of ICICV 2022*, pp. 393–402, 2022.

[180] A. Hafiz, M. Hassaballah, A. Alqahtani, S. Alsubai, and M. A. Hameed, "Reinforcement learning with an ensemble of binary action deep q-networks." *Computer Systems Science & Engineering*, vol. 46, no. 3, 2023.

[181] A. Alagha, S. Singh, R. Mizouni, J. Bentahar, and H. Otrok, "Target localization using multi-agent deep reinforcement learning with proximal policy optimization," *Future Generation Computer Systems*, vol. 136, pp. 342–357, 2022.

[182] S. S. Hassan, Y. M. Park, Y. K. Tun, W. Saad, Z. Han, and C. S. Hong, "3to: Thz-enabled throughput and trajectory optimization of uavs in 6g networks by proximal policy optimization deep reinforcement learning," in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, pp. 5712–5718.

[183] A. K. Jayant and S. Bhatnagar, "Model-based safe deep reinforcement learning via a constrained proximal policy optimization algorithm," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 432–24 445, 2022.

[184] B. Lin, "Reinforcement learning and bandits for speech and language processing: Tutorial, review and outlook," *Expert Systems with Applications*, p. 122254, 2023.

[185] B. Luo, Z. Wu, F. Zhou, and B.-C. Wang, "Human-in-the-loop reinforcement learning in continuous-action space," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[186] A. Raza, K. P. Tran, L. Koehl, and S. Li, "Designing ecg monitoring healthcare system with federated transfer learning and explainable ai," *Knowledge-Based Systems*, vol. 236, p. 107763, 2022.

[187] S. Siahpour, X. Li, and J. Lee, "A novel transfer learning approach in remaining useful life prediction for incomplete dataset," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.

[188] Z. Guo, K. Lin, X. Chen, and C.-Y. Chit, "Transfer learning for angle of arrivals estimation in massive mimo system," in *2022 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE, 2022, pp. 506–511.

[189] S. Liu, Y. Lu, P. Zheng, H. Shen, and J. Bao, "Adaptive reconstruction of digital twins for machining systems: A transfer learning approach," *Robotics and Computer-Integrated Manufacturing*, vol. 78, p. 102390, 2022.

[190] H. Liu, J. Liu, L. Cui, Z. Teng, N. Duan, M. Zhou, and Y. Zhang, "Logiqa 2.0—an improved dataset for logical reasoning in natural language understanding," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023.

[191] Y. Meng, J. Huang, Y. Zhang, and J. Han, "Generating training data with language models: Towards zero-shot language understanding," *Advances in Neural Information Processing Systems*, vol. 35, pp. 462–477, 2022.

[192] R. M. Samant, M. R. Bachute, S. Gite, and K. Kotecha, "Framework for deep learning-based language models using multi-task learning in natural language understanding: A systematic literature review and future directions," *IEEE Access*, vol. 10, pp. 17 078–17 097, 2022.

[193] H. Weld, X. Huang, S. Long, J. Poon, and S. C. Han, "A survey of joint intent detection and slot filling models in natural language understanding," *ACM Computing Surveys*, vol. 55, no. 8, pp. 1–38, 2022.

[194] S. Ajmal, A. A. I. Ahmed, and C. Jalota, "Natural language processing in improving information retrieval and knowledge discovery in healthcare conversational agents," *Journal of Artificial Intelligence and Machine Learning in Management*, vol. 7, no. 1, pp. 34–47, 2023.

[195] A. Montejo-Ráez and S. M. Jiménez-Zafra, "Current approaches and applications in natural language processing," *Applied Sciences*, vol. 12, no. 10, p. 4859, 2022.

[196] K. Vijayan, O. Anand, and A. Sahaj, "Language-agnostic text processing for information extraction," in *CS & IT Conference Proceedings*, vol. 12, no. 23. CS & IT Conference Proceedings, 2022.

[197] C. D. Manning, "Human language understanding & reasoning," *Daedalus*, vol. 151, no. 2, pp. 127–138, 2022.

[198] W. Peng, D. Xu, T. Xu, J. Zhang, and E. Chen, "Are gpt embeddings useful for ads and recommendation?" in *International Conference on Knowledge Science, Engineering and Management*. Springer, 2023, pp. 151–162.

[199] E. Erdem, M. Kuyu, S. Yagcioglu, A. Frank, L. Parcalabescu, B. Plank, A. Babii, O. Turuta, A. Erdem, I. Calixto *et al.*, "Neural natural language generation: A survey on multilinguality, multimodality, controllability and learning," *Journal of Artificial Intelligence Research*, vol. 73, pp. 1131–1207, 2022.

[200] J. Qian, L. Dong, Y. Shen, F. Wei, and W. Chen, "Controllable natural language generation with contrastive prefixes," *arXiv preprint arXiv:2202.13257*, 2022.

[201] H. Rashkin, V. Nikolaev, M. Lamm, L. Aroyo, M. Collins, D. Das, S. Petrov, G. S. Tomar, I. Turc, and D. Reitter, "Measuring attribution in natural language generation models," *Computational Linguistics*, pp. 1–64, 2023.

[202] A. K. Pandey and S. S. Roy, "Natural language generation using sequential models: A survey," *Neural Processing Letters*, pp. 1–34, 2023.

[203] J. Y. Khan and G. Uddin, "Automatic code documentation generation using gpt-3," in *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, 2022, pp. 1–6.

[204] Y. K. Dwivedi, N. Kshetri, L. Hughes, E. L. Slade, A. Jeyaraj, A. K. Kar, A. M. Baabdullah, A. Koohang, V. Raghavan, M. Ahuja *et al.*, ""so what if chatgpt wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy," *International Journal of Information Management*, vol. 71, p. 102642, 2023.

[205] T. Fu, S. Gao, X. Zhao, J.-r. Wen, and R. Yan, "Learning towards conversational ai: A survey," *AI Open*, vol. 3, pp. 14–28, 2022.

[206] H. Ji, I. Han, and Y. Ko, "A systematic review of conversational ai in language education: Focusing on the collaboration with human teachers," *Journal of Research on Technology in Education*, vol. 55, no. 1, pp. 48–63, 2023.

[207] Y. Wan, W. Wang, P. He, J. Gu, H. Bai, and M. R. Lyu, "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*, 2023, pp. 515–527.

[208] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, S. Mishra, and A. Abraham, "Ai-based conversational agents: A scoping review from technologies to future directions," *IEEE Access*, 2022.

[209] Z. Xiao, "Seeing us through machines: designing and building conversational ai to understand humans," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2023.

[210] H.-K. Ko, G. Park, H. Jeon, J. Jo, J. Kim, and J. Seo, "Large-scale text-to-image generation models for visual artists' creative works," in *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 2023, pp. 919–933.

[211] A. Pearson, "The rise of crealtives: Using ai to enable and speed up the creative process," *Journal of AI, Robotics & Workplace Automation*, vol. 2, no. 2, pp. 101–114, 2023.

[212] J. Rezwana and M. L. Maher, "Designing creative ai partners with cofi: A framework for modeling interaction in human-ai co-creative systems," *ACM Transactions on Computer-Human Interaction*, vol. 30, no. 5, pp. 1–28, 2023.

[213] S. Sharma and S. Bvuma, "Generative adversarial networks (gans) for creative applications: Exploring art and music generation," *International Journal of Multidisciplinary Innovation and Research Methodology, ISSN: 2960-2068*, vol. 2, no. 4, pp. 29–33, 2023.

[214] B. Attard-Frost, A. De los Ríos, and D. R. Walters, "The ethics of ai business practices: a review of 47 ai ethics guidelines," *AI and Ethics*, vol. 3, no. 2, pp. 389–406, 2023.

[215] A. Gardner, A. L. Smith, A. Steventon, E. Coughlan, and M. Oldfield, "Ethical funding for trustworthy ai: proposals to address the responsibilities of funders to ensure that projects adhere to trustworthy ai practice," *AI and Ethics*, pp. 1–15, 2022.

[216] J. Schuett, "Three lines of defense against risks from ai," *AI & SOCIETY*, pp. 1–15, 2023.

[217] M. Sloane and J. Zakrzewski, "German ai start-ups and "ai ethics": Using a social practice lens for assessing and implementing sociotechnical innovation," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022, pp. 935–947.

[218] M. Vasconcelos, C. Cardonha, and B. Gonçalves, "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*, 2018, pp. 323–329.

[219] Y. Yang, A. Gupta, J. Feng, P. Singhal, V. Yadav, Y. Wu, P. Natarajan, V. Hedau, and J. Joo, "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*, 2022, pp. 813–822.

[220] R. Schwartz, A. Vassilev, K. Greene, L. Perine, A. Burt, P. Hall *et al.*, "Towards a standard for identifying and managing bias in artificial intelligence," *NIST special publication*, vol. 1270, no. 10.6028, 2022.

[221] W. Guo and A. Caliskan, "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*, 2021, pp. 122–133.

[222] Y. Kong, "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*, 2022, pp. 485–494.

[223] Y. C. Tan and L. E. Celis, "Assessing social and intersectional biases in contextualized word representations," *Advances in neural information processing systems*, vol. 32, 2019.

[224] L. Cheng, A. Mosallanezhad, P. Sheth, and H. Liu, "Causal learning for socially responsible ai," *arXiv preprint arXiv:2104.12278*, 2021.

[225] J. D. Correa, J. Tian, and E. Bareinboim, "Identification of causal effects in the presence of selection bias," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 2744–2751.

[226] B. Ghai and K. Mueller, "D-bias: a causality-based human-in-the-loop system for tackling algorithmic bias," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 1, pp. 473–482, 2022.

[227] J. N. Yan, Z. Gu, H. Lin, and J. M. Rzeszotarski, "Silva: Interactively assessing machine learning fairness using causality," in *Proceedings of the 2020 chi conference on human factors in computing systems*, 2020, pp. 1–13.

[228] E. Bertino, M. Kantarcioglu, C. G. Akcora, S. Samtani, S. Mittal, and M. Gupta, "Ai for security and security for ai," in *Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy*, 2021, pp. 333–334.

[229] H. Susanto, L. F. Yie, D. Rosiyadi, A. I. Basuki, and D. Setiana, "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, 2021, pp. 229–251.

[230] S. Dilmaghani, M. R. Brust, G. Danoy, N. Cassagnes, J. Pecero, and P. Bouvry, "Privacy and security of big data in ai systems: A research and standards perspective," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 5737–5743.

[231] T. McIntosh, "Intercepting ransomware attacks with staged event-driven access control," Ph.D. dissertation, La Trobe, 2022.

[232] T. McIntosh, A. Kayes, Y.-P. P. Chen, A. Ng, and P. Watters, "Applying staged event-driven access control to combat ransomware," *Computers & Security*, vol. 128, p. 103160, 2023.

[233] P. Hummel, M. Braun, M. Tretter, and P. Dabrock, "Data sovereignty: A review," *Big Data & Society*, vol. 8, no. 1, p. 2053951720982012, 2021.

[234] M. Lukings and A. Habibi Lashkari, "Data sovereignty," in *Understanding Cybersecurity Law in Data Sovereignty and Digital Governance: An Overview from a Legal Perspective*. Springer, 2022, pp. 1–38.

[235] M. Hickok, "Lessons learned from ai ethics principles for future actions," *AI and Ethics*, vol. 1, no. 1, pp. 41–47, 2021.

[236] J. Zhou and F. Chen, "Ai ethics: From principles to practice," *AI & SOCIETY*, pp. 1–11, 2022.

[237] J. A. Kroll, "Outlining traceability: A principle for operationalizing accountability in computing systems," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, pp. 758–771.

[238] A. Oseni, N. Moustafa, H. Janicke, P. Liu, Z. Tari, and A. Vasilakos, "Security and privacy for artificial intelligence: Opportunities and challenges," *arXiv preprint arXiv:2102.04661*, 2021.

[239] B. C. Stahl and D. Wright, "Ethics and privacy in ai and big data: Implementing responsible research and innovation," *IEEE Security & Privacy*, vol. 16, no. 3, pp. 26–33, 2018.

[240] C. Ma, J. Li, K. Wei, B. Liu, M. Ding, L. Yuan, Z. Han, and H. V. Poor, "Trusted ai in multiagent systems: An overview of privacy and security for distributed learning," *Proceedings of the IEEE*, vol. 111, no. 9, pp. 1097–1132, 2023.

[241] M. Song, Z. Wang, Z. Zhang, Y. Song, Q. Wang, J. Ren, and H. Qi, "Analyzing user-level privacy attack against federated learning," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 10, pp. 2430–2444, 2020.

[242] I. Misra and L. v. d. Maaten, "Self-supervised learning of pretext-invariant representations," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 6707–6717.

[243] X. Zhai, A. Oliver, A. Kolesnikov, and L. Beyer, "S4l: Self-supervised semi-supervised learning," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1476–1485.

[244] T. Chen, X. Zhai, M. Ritter, M. Lucic, and N. Houlsby, "Self-supervised gans via auxiliary rotation loss," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 12 154–12 163.

[245] S. Jenni and P. Favaro, "Self-supervised feature learning by learning to spot artifacts," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2733–2742.

[246] P. Patel, N. Kumari, M. Singh, and B. Krishnamurthy, "Lt-gan: Self-supervised gan with latent transformation detection," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2021, pp. 3189–3198.

[247] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[248] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9729–9738.

[249] A. T. Liu, S.-W. Li, and H.-y. Lee, "Tera: Self-supervised learning of transformer encoder representation for speech," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 2351–2366, 2021.

[250] Y. Pang, W. Wang, F. E. Tay, W. Liu, Y. Tian, and L. Yuan, "Masked autoencoders for point cloud self-supervised learning," in *European conference on computer vision*. Springer, 2022, pp. 604–621.

[251] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-learning in neural networks: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 9, pp. 5149–5169, 2021.

[252] R. Vilalta and Y. Drissi, "A perspective view and survey of meta-learning," *Artificial intelligence review*, vol. 18, pp. 77–95, 2002.

[253] M. Al-Shedivat, L. Li, E. Xing, and A. Talwalkar, "On data efficiency of meta-learning," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 1369–1377.

[254] Y. Hu, R. Liu, X. Li, D. Chen, and Q. Hu, "Task-sequencing meta learning for intelligent few-shot fault diagnosis with limited data," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 3894–3904, 2021.

[255] S. Baik, J. Choi, H. Kim, D. Cho, J. Min, and K. M. Lee, "Meta-learning with task-adaptive loss function for few-shot learning," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 9465–9474.

[256] Y. Chen, Z. Liu, H. Xu, T. Darrell, and X. Wang, "Meta-baseline: Exploring simple meta-learning for few-shot learning," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 9062–9071.

[257] M. A. Jamal and G.-J. Qi, "Task agnostic meta-learning for few-shot learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 719–11 727.

[258] R. Behnia, M. R. Ebrahimi, J. Pacheco, and B. Padmanabhan, "Ew-tune: A framework for privately fine-tuning large language models with differential privacy," in *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2022, pp. 560–566.

[259] J. Wei, M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le, "Finetuned language models are zero-shot learners," *arXiv preprint arXiv:2109.01652*, 2021.

[260] W. Kuang, B. Qian, Z. Li, D. Chen, D. Gao, X. Pan, Y. Xie, Y. Li, B. Ding, and J. Zhou, "Federatedscope-llm: A comprehensive package for fine-tuning large language models in federated learning," *arXiv preprint arXiv:2309.00363*, 2023.

[261] M. Nguyen, K. Kishan, T. Nguyen, A. Chadha, and T. Vu, "Efficient fine-tuning large language models for knowledge-aware response planning," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2023, pp. 593–611.

[262] M. Engelbach, D. Klau, F. Scheerer, J. Drawehn, and M. Kintz, "Fine-tuning and aligning question answering models for complex information extraction tasks," *arXiv preprint arXiv:2309.14805*, 2023.

[263] T. T. Nguyen, C. Wilson, and J. Dalins, "Fine-tuning llama 2 large language models for detecting online sexual predatory chats and abusive texts," *arXiv preprint arXiv:2308.14683*, 2023.

[264] Q. Zhou, C. Yu, S. Zhang, S. Wu, Z. Wang, and F. Wang, "Regionblip: A unified multi-modal pre-training framework for holistic and regional comprehension," *arXiv preprint arXiv:2308.02299*, 2023.

[265] T. Arnold and D. Kasenberg, "Value alignment or misalignment - what will keep systems accountable?" in *AAAI Workshop on AI, Ethics, and Society*, 2017.

[266] I. Gabriel and V. Ghazavi, "The challenge of value alignment: From fairer algorithms to ai safety," *arXiv preprint arXiv:2101.06060*, 2021.

[267] S. Nyholm, "Responsibility gaps, value alignment, and meaningful human control over artificial intelligence," in *Risk and responsibility in context*. Routledge, 2023, pp. 191–213.

[268] S. Wu, H. Fei, L. Qu, W. Ji, and T.-S. Chua, "Next-gpt: Any-to-any multimodal llm," *arXiv preprint arXiv:2309.05519*, 2023.

[269] K. Bayoudh, R. Knani, F. Hamdaoui, and A. Mtibaa, "A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets," *The Visual Computer*, pp. 1–32, 2021.

[270] P. Hu, L. Zhen, D. Peng, and P. Liu, "Scalable deep multimodal learning for cross-modal retrieval," in *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, 2019, pp. 635–644.

[271] A. Rahate, R. Walambe, S. Ramanna, and K. Kotecha, "Multimodal co-learning: Challenges, applications with datasets, recent advances and future directions," *Information Fusion*, vol. 81, pp. 203–239, 2022.

[272] L. Che, J. Wang, Y. Zhou, and F. Ma, "Multimodal federated learning: A survey," *Sensors*, vol. 23, no. 15, p. 6986, 2023.

[273] P. P. Liang, Y. Lyu, X. Fan, Z. Wu, Y. Cheng, J. Wu, L. Chen, P. Wu, M. A. Lee, Y. Zhu *et al.*, "Multibench: Multiscale benchmarks for multimodal representation learning," *arXiv preprint arXiv:2107.07502*, 2021.

[274] Z. Ashktorab, Q. V. Liao, C. Dugan, J. Johnson, Q. Pan, W. Zhang, S. Kumaravel, and M. Campbell, "Human-ai collaboration in a cooperative game setting: Measuring social perception and outcomes," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–20, 2020.

[275] P. Esmaeilzadeh, T. Mirzaei, and S. Dharanikota, "Patients' perceptions toward human–artificial intelligence interaction in health care: experimental study," *Journal of medical Internet research*, vol. 23, no. 11, p. e25856, 2021.

[276] M. Nazar, M. M. Alam, E. Yafi, and M. M. Su'ud, "A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques," *IEEE Access*, vol. 9, pp. 153 316–153 348, 2021.

[277] A. S. Rajawat, R. Rawat, K. Barhanpurkar, R. N. Shaw, and A. Ghosh, "Robotic process automation with increasing productivity and improving product quality using artificial intelligence and machine learning," in *Artificial Intelligence for Future Generation Robotics*. Elsevier, 2021, pp. 1–13.

[278] S. Mohseni, N. Zarei, and E. D. Ragan, "A multidisciplinary survey and framework for design and evaluation of explainable ai systems," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 11, no. 3-4, pp. 1–45, 2021.

[279] M. C. Buehler and T. H. Weisswange, "Theory of mind based communication for human agent cooperation," in *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*. IEEE, 2020, pp. 1–6.

[280] M. M. Çelikok, T. Peltola, P. Daee, and S. Kaski, "Interactive ai with a theory of mind," *arXiv preprint arXiv:1912.05284*, 2019.

[281] A. Dafoe, E. Hughes, Y. Bachrach, T. Collins, K. R. McKee, J. Z. Leibo, K. Larson, and T. Graepel, "Open problems in cooperative ai," *arXiv preprint arXiv:2012.08630*, 2020.

[282] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg *et al.*, "Sparks of artificial general intelligence: Early experiments with gpt-4," *arXiv preprint arXiv:2303.12712*, 2023.

[283] N. Fei, Z. Lu, Y. Gao, G. Yang, Y. Huo, J. Wen, H. Lu, R. Song, X. Gao, T. Xiang *et al.*, "Towards artificial general intelligence via a multimodal foundation model," *Nature Communications*, vol. 13, no. 1, p. 3094, 2022.

[284] R. Williams and R. Yampolskiy, "Understanding and avoiding ai failures: A practical guide," *Philosophies*, vol. 6, no. 3, p. 53, 2021.

[285] W. Fedus, B. Zoph, and N. Shazeer, "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity," *The Journal of Machine Learning Research*, vol. 23, no. 1, pp. 5232–5270, 2022.

[286] S. Shen, L. Hou, Y. Zhou, N. Du, S. Longpre, J. Wei, H. W. Chung, B. Zoph, W. Fedus, X. Chen *et al.*, "Mixture-of-experts meets instruction tuning: A winning combination for large language models," *arXiv preprint arXiv:2305.14705*, 2023.

[287] S. Rajbhandari, C. Li, Z. Yao, M. Zhang, R. Y. Aminabadi, A. A. Awan, J. Rasley, and Y. He, "Deepspeed-moe: Advancing mixture-of-experts inference and training to power next-generation ai scale," in *International Conference on Machine Learning*. PMLR, 2022, pp. 18 332–18 346.

[288] L. Shen, Z. Wu, W. Gong, H. Hao, Y. Bai, H. Wu, X. Wu, J. Bian, H. Xiong, D. Yu *et al.*, "Se-moe: A scalable and efficient mixture-of-experts distributed training and inference system," *arXiv preprint arXiv:2205.10034*, 2022.

[289] C. Hwang, W. Cui, Y. Xiong, Z. Yang, Z. Liu, H. Hu, Z. Wang, R. Salas, J. Jose, P. Ram *et al.*, "Tutel: Adaptive mixture-of-experts at scale," *Proceedings of Machine Learning and Systems*, vol. 5, 2023.

[290] Y. Wang, S. Mukherjee, X. Liu, J. Gao, A. H. Awadallah, and J. Gao, "Adamix: Mixture-of-adapter for parameter-efficient tuning of large language models," *arXiv preprint arXiv:2205.12410*, vol. 1, no. 2, p. 4, 2022.

[291] T. Chen, Z. Zhang, A. Jaiswal, S. Liu, and Z. Wang, "Sparse moe as the new dropout: Scaling dense and self-slimmable transformers," *arXiv preprint arXiv:2303.01610*, 2023.

[292] H. Zhu, B. He, and X. Zhang, "Multi-gate mixture-of-experts stacked autoencoders for quality prediction in blast furnace ironmaking," *ACS omega*, vol. 7, no. 45, pp. 41 296–41 303, 2022.

[293] Z. Chi, L. Dong, S. Huang, D. Dai, S. Ma, B. Patra, S. Singhal, P. Bajaj, X. Song, X.-L. Mao *et al.*, "On the representation collapse of sparse mixture of experts," *Advances in Neural Information Processing Systems*, vol. 35, pp. 34 600–34 613, 2022.

[294] S. Gupta, S. Mukherjee, K. Subudhi, E. Gonzalez, D. Jose, A. H. Awadallah, and J. Gao, "Sparsely activated mixture-of-experts are robust multi-task learners," *arXiv preprint arXiv:2204.07689*, 2022.

[295] N. Dikkala, N. Ghosh, R. Meka, R. Panigrahy, N. Vyas, and X. Wang, "On the benefits of learning to route in mixture-of-experts models," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023, pp. 9376–9396.

[296] N. Dryden and T. Hoefler, "Spatial mixture-of-experts," *Advances in Neural Information Processing Systems*, vol. 35, pp. 11 697–11 713, 2022.

[297] Z. You, S. Feng, D. Su, and D. Yu, "Speechmoe2: Mixture-of-experts model with improved routing," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 7217–7221.

[298] J. Puigcerver, R. Jenatton, C. Riquelme, P. Awasthi, and S. Bhojanapalli, "On the adversarial robustness of mixture of experts," *Advances in Neural Information Processing Systems*, vol. 35, pp. 9660–9671, 2022.

[299] J. Li, Y. Jiang, Y. Zhu, C. Wang, and H. Xu, "Accelerating distributed {MoE} training and inference with lina," in *2023 USENIX Annual Technical Conference (USENIX ATC 23)*, 2023, pp. 945–959.

[300] L. Wu, M. Liu, Y. Chen, D. Chen, X. Dai, and L. Yuan, "Residual mixture of experts," *arXiv preprint arXiv:2204.09636*, 2022.

[301] B. Zoph, I. Bello, S. Kumar, N. Du, Y. Huang, J. Dean, N. Shazeer, and W. Fedus, "Designing effective sparse expert models," *arXiv preprint arXiv:2202.08906*, vol. 2, 2022.

[302] ——, "St-moe: Designing stable and transferable sparse expert models," *arXiv preprint arXiv:2202.08906*, 2022.

[303] Y. Chow, A. Tulepbergenov, O. Nachum, M. Ryu, M. Ghavamzadeh, and C. Boutilier, "A mixture-of-expert approach to rl-based dialogue management," *arXiv preprint arXiv:2206.00059*, 2022.

[304] Z. Fan, R. Sarkar, Z. Jiang, T. Chen, K. Zou, Y. Cheng, C. Hao, Z. Wang *et al.*, "M^3vit: Mixture-of-experts vision transformer for efficient multi-task learning with model-accelerator co-design," *Advances in Neural Information Processing Systems*, vol. 35, pp. 28 441–28 457, 2022.

[305] T. Zadouri, A. Üstün, A. Ahmadian, B. Ermiş, A. Locatelli, and S. Hooker, "Pushing mixture of experts to the limit: Extremely parameter efficient moe for instruction tuning," *arXiv preprint arXiv:2309.05444*, 2023.

[306] J. Zhu, X. Zhu, W. Wang, X. Wang, H. Li, X. Wang, and J. Dai, "Uni-perceiver-moe: Learning sparse generalist models with conditional moes," *Advances in Neural Information Processing Systems*, vol. 35, pp. 2664–2678, 2022.

[307] F. Dou, J. Ye, G. Yuan, Q. Lu, W. Niu, H. Sun, L. Guan, G. Lu, G. Mai, N. Liu *et al.*, "Towards artificial general intelligence (agi) in the internet of things (iot): Opportunities and challenges," *arXiv preprint arXiv:2309.07438*, 2023.

[308] Z. Jia, X. Li, Z. Ling, S. Liu, Y. Wu, and H. Su, "Improving policy optimization with generalist-specialist learning," in *International Conference on Machine Learning*. PMLR, 2022, pp. 10 104–10 119.
[309] M. Simeone, "Unknown future, repeated present: A narrative-centered analysis of long-term ai discourse," *Humanist Studies & the Digital Age*, vol. 7, no. 1, 2022.
[310] A. Nair and F. Banaei-Kashani, "Bridging the gap between artificial intelligence and artificial general intelligence: A ten commandment framework for human-like intelligence," *arXiv preprint arXiv:2210.09366*, 2022.
[311] M. H. Jarrahi, D. Askay, A. Eshraghi, and P. Smith, "Artificial intelligence and knowledge management: A partnership between human and ai," *Business Horizons*, vol. 66, no. 1, pp. 87–99, 2023.
[312] D. J. Edwards, C. McEnteggart, and Y. Barnes-Holmes, "A functional contextual account of background knowledge in categorization: Implications for artificial general intelligence and cognitive accounts of general knowledge," *Frontiers in Psychology*, vol. 13, p. 745306, 2022.
[313] J. McCarthy, "Artificial intelligence, logic, and formalising common sense," *Machine Learning and the City: Applications in Architecture and Urban Design*, pp. 69–90, 2022.
[314] S. Friederich, "Symbiosis, not alignment, as the goal for liberal democracies in the transition to artificial general intelligence," *AI and Ethics*, pp. 1–10, 2023.
[315] S. Makridakis, "The forthcoming artificial intelligence (ai) revolution: Its impact on society and firms," *Futures*, vol. 90, pp. 46–60, 2017.
[316] S. Pal, K. Kumari, S. Kadam, and A. Saha, "The ai revolution," *IARA Publication*, 2023.
[317] S. Verma, R. Sharma, S. Deb, and D. Maitra, "Artificial intelligence in marketing: Systematic review and future research direction," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100002, 2021.
[318] P. Budhwar, S. Chowdhury, G. Wood, H. Aguinis, G. J. Bamber, J. R. Beltran, P. Boselie, F. Lee Cooke, S. Decker, A. DeNisi *et al.*, "Human resource management in the age of generative artificial intelligence: Perspectives and research directions on chatgpt," *Human Resource Management Journal*, vol. 33, no. 3, pp. 606–659, 2023.
[319] J. B. Telkamp and M. H. Anderson, "The implications of diverse human moral foundations for assessing the ethicality of artificial intelligence," *Journal of Business Ethics*, vol. 178, no. 4, pp. 961–976, 2022.
[320] X. Zhou, C. Liu, L. Zhai, Z. Jia, C. Guan, and Y. Liu, "Interpretable and robust ai in eeg systems: A survey," *arXiv preprint arXiv:2304.10755*, 2023.
[321] C. Zhang, C. Zhang, C. Li, Y. Qiao, S. Zheng, S. K. Dam, M. Zhang, J. U. Kim, S. T. Kim, J. Choi *et al.*, "One small step for generative ai, one giant leap for agi: A complete survey on chatgpt in aigc era," *arXiv preprint arXiv:2304.06488*, 2023.
[322] K. Singhal, T. Tu, J. Gottweis, R. Sayres, E. Wulczyn, L. Hou, K. Clark, S. Pfohl, H. Cole-Lewis, D. Neal *et al.*, "Towards expert-level medical question answering with large language models," *arXiv preprint arXiv:2305.09617*, 2023.
[323] S. Wu, O. Irsoy, S. Lu, V. Dabravolski, M. Dredze, S. Gehrmann, P. Kambadur, D. Rosenberg, and G. Mann, "Bloomberggpt: A large language model for finance," *arXiv preprint arXiv:2303.17564*, 2023.
[324] P. Henderson, K. Sinha, N. Angelard-Gontier, N. R. Ke, G. Fried, R. Lowe, and J. Pineau, "Ethical challenges in data-driven dialogue systems," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 2018, pp. 123–129.
[325] S. A. Bin-Nashwan, M. Sadallah, and M. Bouteraa, "Use of chatgpt in academia: Academic integrity hangs in the balance," *Technology in Society*, vol. 75, p. 102370, 2023.
[326] N. Liu, A. Brown *et al.*, "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*, vol. 25, p. e50591, 2023.
[327] A. P. Siddaway, A. M. Wood, and L. V. Hedges, "How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses," *Annual review of psychology*, vol. 70, pp. 747–770, 2019.
[328] E. Landhuis, "Scientific literature: Information overload," *Nature*, vol. 535, no. 7612, pp. 457–458, 2016.
[329] G. D. Chloros, V. P. Giannoudis, and P. V. Giannoudis, "Peer-reviewing in surgical journals: revolutionize or perish?" *Annals of surgery*, vol. 275, no. 1, pp. e82–e90, 2022.
[330] K.-A. Allen, J. Reardon, Y. Lu, D. V. Smith, E. Rainsford, and L. Walsh, "Towards improving peer review: Crowd-sourced insights from twitter," *Journal of university teaching & learning practice*, vol. 19, no. 3, p. 02, 2022.