A Survey of Advanced AI Models: Large Language Models, Multimodal Models, and Cognitive Computing

www.surveyx.cn

Abstract

The survey examines the transformative potential of advanced artificial intelligence (AI) models, focusing on Large Language Models (LLMs), multimodal models, and cognitive computing. It highlights the necessity of dynamic auditing frameworks and tailored ethical guidelines to address challenges posed by LLMs, ensuring responsible deployment and mitigating societal inequities. The advancements in LLMs, exemplified by models like ChatGPT, suggest possibilities for achieving general intelligence and innovative applications. The integration of personalized frameworks like PlanFitting and the importance of interoperability in AI platforms are emphasized for maximizing potential through collaborative efforts. Future research should prioritize ability-oriented assessments reflecting AI's cognitive capabilities and explore AI's moral discourse in various cultural contexts. The survey underscores the effectiveness of models like RAM in dynamic knowledge acquisition, highlighting the potential for AI systems to learn and adapt continuously. Interdisciplinary collaboration is vital for advancing AI technologies, fostering innovation, and addressing multifaceted challenges. The conclusion emphasizes MarineGPT's improvement in generating accurate marine-related responses and suggests future research directions for enhancing model capabilities. Key findings indicate a 50

1 Introduction

1.1 Overview of Advanced AI Fields

The field of artificial intelligence (AI) is undergoing a significant transformation driven by advancements in Large Language Models (LLMs), multimodal models, and cognitive computing. LLMs, such as ChatGPT, enhance human-AI interactions by integrating physiological data, thereby improving empathy and interpretation of mental and emotional states [1]. This integration is vital for creating intuitive AI systems, particularly as studies reveal generative AI's impact on younger demographics [2].

Simultaneously, multimodal data integration has become crucial, allowing AI systems to process and synthesize diverse datasets. The Valley model exemplifies this capability by combining video, image, and language understanding [3]. This is further illustrated by innovative techniques for decoding imagined speech using high-density functional near-infrared spectroscopy (fNIRS) [4]. Domain-specific models like MarineGPT also highlight the effectiveness of specialized vision-language models in addressing unique challenges in fields like marine research [5].

Cognitive computing aims to replicate human cognitive processes to enhance decision-making and problem-solving. The integration of Generative AI (GenAI) with Evolutionary Algorithms (EAs) showcases this potential, especially in large-scale multi-objective optimization tasks [6]. Additionally,

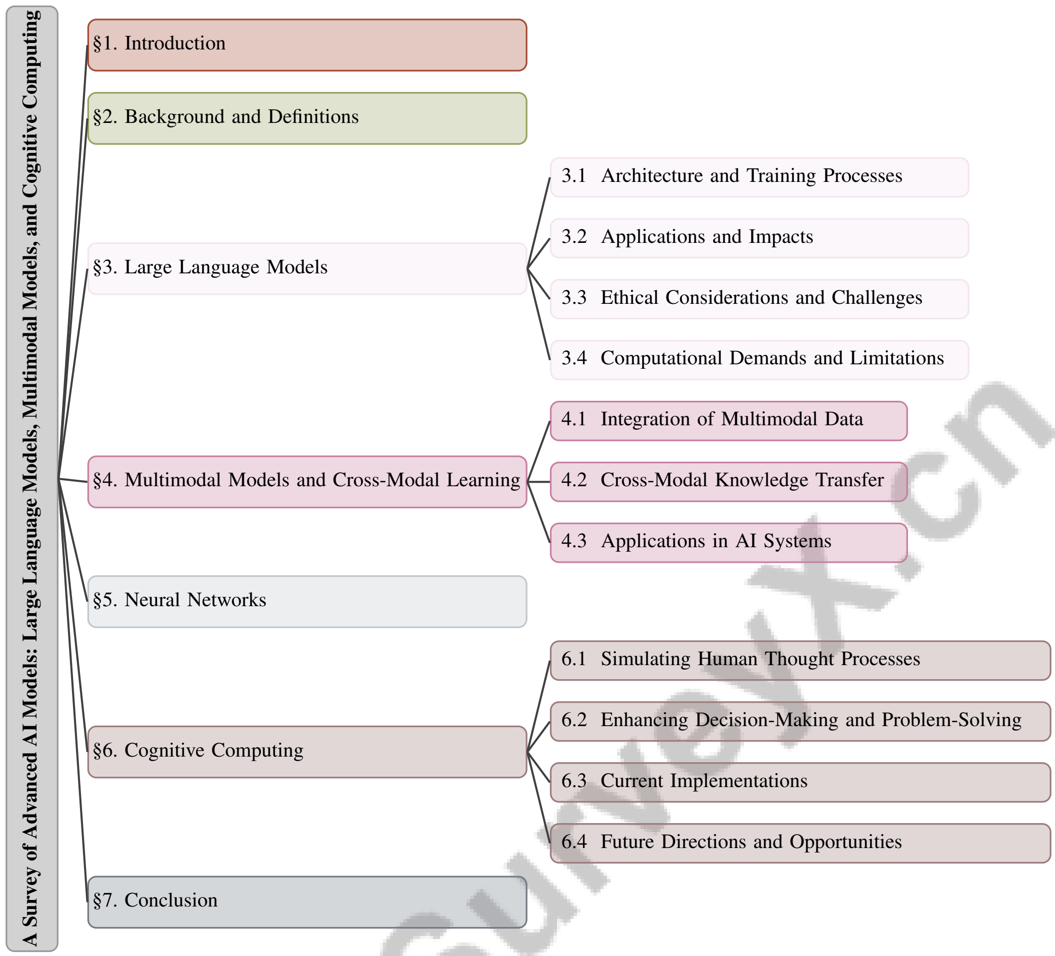


Figure 1: chapter structure

research into 'neural erosion' for simulating neurodegeneration offers insights into the limitations and possibilities of current AI technologies [7].

Generative AI tools are transforming fields like education and mental health, exemplified by LLMs providing personalized therapy, which underscores both opportunities and limitations in mental health treatment [8]. In educational contexts, AI tutors designed with effective personas demonstrate enhanced learning outcomes and user trust.

The demand for explainable AI is increasingly recognized across advanced fields, as stakeholders—including citizens, regulators, and domain experts—seek transparency to foster trust and address concerns about AI interpretability. This is particularly evident in education and the deployment of LLMs, where understanding AI decision-making is essential for effective interaction and ethical application [9, 10, 11, 12, 13]. Moreover, the exploration of AI reliance behaviors and alignment with human moral preferences emphasizes the need for AI systems to conform to societal norms.

This survey offers a comprehensive overview of these advanced AI fields, examining their architectures, applications, and transformative impacts while addressing the ethical and policy considerations that accompany their development. The historical trajectory of AI, from its inception in the 1950s to the rise of deep learning by 2021, provides context for understanding these advancements [14]. Additionally, the exploration of epistemological tensions in AI-mediated knowledge, such as bridging modern language processing and traditional dance knowledge, highlights the interdisciplinary challenges and opportunities in AI research [15].

1.2 Significance in AI Development

The integration of advanced AI technologies, including LLMs, multimodal models, and cognitive computing, is crucial in reshaping AI development. LLMs, as seen in Valley, enhance video comprehension and instruction-following capabilities, underscoring their significance in AI's evolution [3]. Their application in domains like psychotherapy addresses accessibility and cost challenges, exemplifying transformative potential [16].

These technologies also counter skepticism about full automation, particularly in fact-checking, where human judgment remains essential [17]. This illustrates the need for a balanced integration of AI and human oversight. Furthermore, the shift from diverse AI approaches to a singular focus on deep learning, driven by benchmarking, marks a critical transition in AI development [14].

Neuro-Symbolic AI, combining symbolic reasoning with neural networks, addresses challenges in effective reasoning within AI systems, particularly in natural language processing [18]. This integration is vital for enhancing reasoning capabilities and ensuring reliable outcomes. Additionally, understanding deceptive UI/UX patterns is essential for improving user interaction and writing outcomes, which are crucial for the broader acceptance and effectiveness of AI technologies [19].

An interdisciplinary approach that incorporates ecological and evolutionary theories into AI opens new research avenues, expanding the scope of AI development [20]. The potential emergence of high-level machine intelligence (HLMI) and the automation of specific jobs represent significant milestones in AI's trajectory, emphasizing the profound impact of these technologies on the future of work and society [21].

To foster cultural dialogue and innovation, new approaches in AI development are necessary, particularly in bridging modern AI capabilities with traditional knowledge systems [15]. This highlights the importance of these technologies in promoting cultural inclusivity and innovation. Collectively, these advancements demonstrate the transformative impact of AI technologies in enhancing decision-making, improving human-AI interaction, and democratizing access to AI tools, underscoring their significance in the broader context of AI development.

1.3 Structure of the Survey

This survey is organized to provide a comprehensive exploration of advanced AI models, structured into distinct sections that reflect the multifaceted nature of AI research and development. The initial section introduces the overarching theme, offering a high-level overview of advanced AI fields such as LLMs, multimodal models, and cognitive computing. Following this, the survey delves into the significance of these technologies in AI development, highlighting their transformative potential across various domains.

The second section provides detailed definitions and historical context for core AI concepts, setting the stage for understanding subsequent discussions. The historical development and evolution of AI, including Turing's objections and AI capabilities' progression, contextualize current advancements [22].

The third section focuses on LLMs, examining their architecture, training processes, applications, ethical considerations, and computational demands. This is complemented by an analysis of LLMs' impact on computing education, encompassing literature reviews, student and instructor attitudes, instructional approaches, ethical issues, and performance benchmarking [23].

The fourth section investigates multimodal models and cross-modal learning, emphasizing the integration of multimodal data and knowledge transfer between modalities. This section also provides examples of applications in AI systems, illustrating the practical implications of these technologies.

The fifth section discusses the foundational role of neural networks in AI, detailing their architecture, recent advancements, and evaluation methods. This is followed by a section on cognitive computing, which analyzes its aim to simulate human thought processes and its potential to enhance decision-making and problem-solving.

The concluding section synthesizes key findings and insights, reflecting on future directions and emphasizing the importance of interdisciplinary collaboration in advancing AI research. The survey's organization into fields such as text generation, sentiment analysis, question-answering, and commonsense reasoning provides a structured overview of the paper's content, ensuring a coherent narrative that aligns with the current landscape of AI research [24].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Definitions and Core Concepts

Foundational understanding of artificial intelligence (AI) concepts is pivotal for comprehending the capabilities and limitations of advanced models. Large Language Models (LLMs) are integral to natural language processing (NLP), generating coherent, contextually relevant text and facilitating mental health applications like Cognitive Behavioral Therapy (CBT) [8]. They address moral dilemmas across languages and cultures, highlighting their role in multilingual contexts [25].

The emergence of multimodal models represents a significant advancement, integrating diverse data types such as text, images, and audio to enhance complex information synthesis. The Valley model exemplifies this, demonstrating capabilities in video comprehension and instruction-following [3]. The translation of user-generated prompts into culturally specific sequences, like Thai classical dance, underscores the interdisciplinary nature of multimodal AI applications [15].

Neural networks, inspired by the human brain, form the backbone of AI systems, facilitating advanced pattern recognition and problem-solving. Parameter sharing, linked to analogy-making and cognitive metaphor, is essential for enhancing AI's cognitive functions [26]. Despite progress, replicating human-like performance remains challenging, necessitating AI agents with a broad skill set akin to human cognition [20].

Cognitive computing endeavors to replicate human thought processes, particularly in decision-making and problem-solving. AI's influence on text message composition can alter recipients' perceptions, emphasizing the need to understand AI's communicative role [27]. Concepts like 'confirmation bias' are significant in cognitive computing, especially in healthcare decision-making [28].

Understanding AI reliance in sociotechnical systems is crucial for human-AI interaction [29]. Challenges in managing misinformation highlight the need for technological support in fact-checking [17]. Ethical concerns arise from potential deceptive patterns in AI writing assistants, affecting user influence and writing integrity [19]. These definitions and concepts provide a framework for exploring advanced AI fields, facilitating a detailed examination of their architectures, applications, and ethical implications.

2.2 Historical Development and Evolution

The historical evolution of artificial intelligence (AI) is marked by significant milestones, transitioning from Symbolic AI (1955-1987) to the Benchmarking Era (1987-2011), and culminating in the Deep Learning Era (2012-2021) [14]. This evolution reflects a shift from rule-based systems to sophisticated models utilizing extensive data and computational power, moving from specific algorithms to integrated approaches combining various AI methodologies.

Large Language Models (LLMs) exemplify AI advancements, tackling complex linguistic challenges. However, effective generalization with limited training data remains an obstacle, underscoring the reliance on large datasets [26]. Models like GPT-4 have set new benchmarks in natural language processing through extensive data use.

Integrating AI with traditional scientific approaches has enhanced fields like climate modeling [30]. This collaboration shows AI's potential to complement traditional methods. Neuro-symbolic AI architectures, merging symbolic reasoning with neural networks, represent a crucial advancement in achieving effective AI reasoning [18].

In multimodal models, historical challenges include bridging discrete and continuous computation, as seen in neurosymbolic learning systems [31]. Performance variances among architectures, especially when pretrained and fine-tuned, highlight the complexities of developing multimodal AI systems [32].

Cognitive computing has evolved to replicate human thought processes, focusing on decision-making and problem-solving. The historical reliance on AI for data preparation and summarization tasks illustrates AI's gradual integration into cognitive processes [33]. The exploration of AI reliance in sociotechnical systems highlights the intricacies of human-AI interactions as a significant factor in AI reliance research development [29].

AI benchmark evolution has shifted from narrow task-focused evaluations to comprehensive criteria reflecting real-world complexities [34]. This shift is essential for capturing AI's full capabilities, particularly in non-language domains where traditional benchmarks fall short [35]. AI's exploration in creativity and reasoning requires novel evaluation frameworks beyond conventional human judgment [36].

The historical evolution of AI narrates a story of technological innovation and interdisciplinary collaboration. Continuous efforts to address ethics, governance, and societal impact challenges are reshaping AI research and development's future. This transformation is evident in large language models (LLMs), where transparency, privacy, fairness, and accountability issues are crucial. Prioritizing a human-centered approach to transparency and developing ethical frameworks tailored to specific domains fosters robust, adaptable, and inclusive AI systems that responsibly influence information dissemination and societal interactions [37, 10].

3 Large Language Models

Category	Feature	Method
Architecture and Training Processes	Sequential Data Processing Ongoing Learning Processes User Interaction Systems	READ[27] MGPT[5] PF[38]
Applications and Impacts	Decision and Optimization Strategies Data Integration and Processing Feedback and Iterative Improvement	LLM-AI[6] BAI[39], MMLLM[40], NOF[31], VAL[3] SCM[41]
Ethical Considerations and Challenges	Cultural Awareness	T2T[15]
Computational Demands and Limitations	Efficiency Enhancement	PM[42]

Table 1: This table provides a comprehensive overview of various methods categorized under architecture and training processes, applications and impacts, ethical considerations, and computational demands related to Large Language Models (LLMs). Each category delineates specific features and corresponding methods, highlighting the multifaceted nature and broad implications of LLMs across different domains. The table serves as a foundational reference for understanding the diverse methodologies employed in LLM development and deployment.

Exploring the foundational aspects of Large Language Models (LLMs) is crucial for understanding their functionality and effectiveness. By examining the architecture and training processes, we gain insights into their high performance across various applications, setting the stage for appreciating their broader implications in real-world contexts. As illustrated in **??**, the hierarchical structure of key concepts related to LLMs encompasses not only their architecture and training processes but also their applications and impacts, ethical considerations and challenges, as well as computational demands and limitations. Table 1 presents a detailed categorization of methods associated with Large Language Models (LLMs), focusing on architecture and training processes, applications and impacts, ethical considerations, and computational demands. Each primary category is further divided into specific subcategories and detailed points, emphasizing the multifaceted nature of LLMs and their influence across various domains. Additionally, Table 2 provides a comparative analysis of the architectural and training methodologies of various Large Language Models, emphasizing their distinct features and ethical considerations. The subsequent subsection will delve into these architectural elements and training methodologies, highlighting their role in shaping LLM capabilities.

3.1 Architecture and Training Processes

The architecture and training methodologies of LLMs are pivotal to their precision and adaptability across tasks. Central to these models is the transformer architecture, which uses self-attention mechanisms for efficient data processing, enabling coherent and contextually relevant outputs. This is vital for sequential data processing in fields like NLP and decision-making [3]. The Valley model, for instance, employs a structured architecture with a temporal modeling module and visual encoder, showcasing innovative LLM development [3].

LLMs typically undergo a dual-phase training process: pre-training and fine-tuning. Pre-training involves large corpora to learn linguistic patterns and build a knowledge base, as seen in MarineGPT, which uses a dataset of over 5 million marine image-text pairs for continuous pre-training [5]. Fine-tuning adapts models to specific datasets for optimized task performance, exemplified by the HELPERT model, a CBT-prompted LLM tested against CBT-trained peer counselors [8].

Frameworks like Prompt Middleware enhance LLM integration into user interfaces by mapping user options to generate prompts, improving AI system usability [42]. GigaCheck introduces a detection architecture based on DETR-like models, enabling the identification of LLM-generated content [43].

Incorporating human-centric design principles into LLM architecture is vital for user understanding, transparency, and ethical considerations [20]. Experiments with models like Llama 3.2 1B, fine-tuned with LoRA on structured memories, highlight the potential for developing sophisticated cognitive capabilities [27].

Surveys of deceptive patterns in AI writing assistants emphasize the importance of ethical considerations in LLM deployment [19]. The categorization of neuro-symbolic architectures provides insights into techniques for knowledge representation and reasoning, crucial for advancing LLM capabilities [18]. In schizophrenia rehabilitation, AI technologies integration illustrates the diverse applications of LLMs [16].

The dynamic landscape of LLM architecture and training processes, characterized by innovative methodologies and frameworks, enhances their performance across applications. This environment facilitates the creation of sophisticated LLM-generated content and necessitates effective detection methods to mitigate misuse. Advanced optimization algorithms integrated with LLMs improve decision-making in complex scenarios, while ethical considerations and transparency discussions highlight responsible deployment importance [44, 37, 43, 10, 45]. These developments address technical challenges and underscore ethical considerations and human-centric design in LLM deployment.

In **??**, understanding LLM architecture and training processes is crucial for appreciating their capabilities and functionality. The examples provide insights into different facets of these models. "The Standard Model of the Mind" visually represents a cognitive model, emphasizing the parallels with language models. "PlanFitting" showcases a system for managing dialogues and planning, highlighting adaptive learning processes in language models. "Human-Model Interaction in NLP" underscores human intervention's role in refining model outputs, reflecting collaborative human-AI interactions in NLP tasks [46, 38, 33].

3.2 Applications and Impacts

LLMs have revolutionized various domains, enhancing capabilities in healthcare, education, and legal systems. In healthcare, LLMs automate processes like dental diagnosis and treatment planning, managing multi-source data to streamline workflows [40]. Technologies like EEGChat enable users with severe language impairments to communicate by interpreting intentions and generating responses [39].

In the legal field, LLMs enhance judicial processes by processing complex legal texts, demonstrated in question-answering systems and information retrieval [47]. In education, LLMs foster innovative learning experiences, particularly in visual-based AI applications [2]. Research categorization into stages of coding generation, data analysis, and educational enhancement highlights generative AI tools streamlining educational processes [48].

LLMs enhance decision-making by analyzing key decision variables and trade-offs, optimizing solutions in complex scenarios [6]. In neurosymbolic tasks, LLMs optimize mixed computation, addressing methodological shortcomings [31].

In AI-assisted writing, LLMs guide model output, select or rate output, post-edit, and write with model assistance [33]. Fine-tuning frameworks improve content generation quality by critiquing model-generated and human-generated summaries [41]. AI's influence on mundane tasks enhances information retrieval efficiency [49].

Applications of Valley, an LLM evaluated on diverse datasets for video and image understanding, demonstrate significant advancements in visual information processing [3]. FeedbackBuffet, a successful Prompt Middleware, bridges user interfaces and LLMs, enhancing user interaction [42].

These applications underscore LLMs' transformative impact across domains, driving innovation and expanding AI capabilities. As LLMs evolve, their role in advancing AI systems and expanding AI boundaries remains pivotal, emphasizing the need for continued research and ethical considerations. While participants perform above chance using AI, explainable AI has not significantly enhanced decision-making accuracy compared to AI alone, highlighting areas for improvement [11].

In Figure 3, LLMs are transformative tools with diverse applications and impacts across domains such as healthcare, legal systems, and education. The figure illustrates examples highlighting LLMs' applications and challenges. The "Transformer Recurrent Decoder Architecture" demonstrates intricate LLM design, while the "Snowballing Process" underscores methodological approaches in AI and computing education. Furthermore, "Prompts and Techniques for Hallucination Mitigation" presents strategies for reducing LLM output hallucinations, ensuring reliable model performance. Collectively, these examples encapsulate LLMs' multifaceted applications and efforts to address limitations, underscoring their profound influence on technology and education [24, 23, 50].

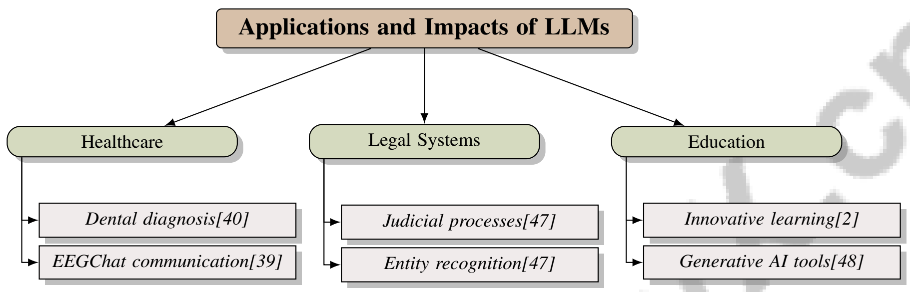


Figure 3: This figure illustrates the diverse applications and impacts of Large Language Models (LLMs) across domains such as healthcare, legal systems, and education. In healthcare, LLMs automate dental diagnosis and facilitate communication for language-impaired users. Legal systems benefit from enhanced judicial processes and entity recognition. Education experiences innovative learning through generative AI tools, emphasizing LLMs' transformative role in expanding AI capabilities and driving innovation.

3.3 Ethical Considerations and Challenges

LLMs' deployment introduces ethical challenges requiring careful consideration for responsible use. A major issue is the potential for misinformation and deceptive content, manipulating user behavior due to system transparency lack [19]. LLMs' limited ability to generate culturally specific content, like dance movements, highlights the need for models respecting cultural nuances [15].

Bias in LLMs is a critical concern, with standardized prompting potentially leading to outputs not representing diverse perspectives, problematic in educational and therapeutic applications [16]. Integrating neural and symbolic systems enhances reasoning and interpretability, yet balancing effective reasoning and ethical considerations remains challenging [18].

Aligning LLMs with human moral judgments is crucial for ethical AI interactions, especially in decision-making processes. However, human-AI interaction complexity and lack of external validity in studies challenge AI reliance understanding and improvement [29]. In fact-checking, AI tools often fall short in addressing fact-checkers' specific needs, emphasizing tailored solutions [17].

Transparency and explainability are essential for fostering trust in AI systems, yet current benchmarks and studies often inadequately address these aspects and the environmental costs of large-scale models. Addressing AI systems' ethical challenges, particularly LLMs, requires a comprehensive approach emphasizing transparency, inclusivity, and adaptability. Implementing tailored ethical frameworks, enhancing accountability through dynamic auditing, and fostering interdisciplinary collaboration mitigate challenges like hallucination and bias. A human-centered perspective ensures transparency supports diverse stakeholder needs and promotes responsible AI development and deployment [37, 51, 10]. Focusing on these areas aligns LLMs with societal values and ethical standards, ensuring safe and effective deployment across applications.

3.4 Computational Demands and Limitations

LLMs' computational demands are significant, driven by extensive hardware resources and energy consumption. These demands raise concerns about operational costs and environmental impacts, necessitating innovative optimization strategies for efficiency. Methods like Prompt Middleware abstract prompt creation complexity and embed domain expertise, improving user interaction with LLMs while managing computational demands [42].

A critical challenge in LLM deployment is aligning with human moral and ethical preferences, as most models do not align well, indicating significant computational demands for accurate moral alignment [25]. This misalignment underscores ongoing research to refine models to better reflect human values. Benchmarks like GigaCheck address LLM-generated content detection, highlighting the computational complexity in distinguishing human-written and machine-generated text [43].

Computational requirements extend to domains like healthcare, where LLMs analyze and interpret multi-source data, including electronic health records and patient audio inputs. However, accurately modeling human emotional dynamics complexity remains a limitation, leading to inconsistencies in synthesized expressions when models lack robustness. This limitation is exacerbated by sensory data quality reliance, as LLMs may misinterpret commands with inadequate environmental context representation, leading to potential hallucinations—seemingly factual content lacking reality grounding. This issue is concerning in sensitive applications, like medical record summarization or legal advice, where misinterpretations can have serious consequences. Addressing this challenge is essential for LLMs' safe deployment in real-world scenarios, emphasizing improved evaluation mechanisms considering automated and human assessments to ensure LLM output reliability [45, 50].

In legal applications, models like Mistral and Gemma exemplify LLMs' computational demands and limitations in specialized tasks like legal entity recognition, evaluated for accurately identifying and classifying domain-specific entities within complex legal texts. Recent studies show these models demonstrate balanced precision and recall, crucial for effective entity extraction, enhancing legal applications like question-answering systems and information retrieval from case law documents [43, 47, 52, 45]. Despite advancements, LLMs face challenges in modeling complex relationships and trade-offs crucial for informed decision-making. The necessity for customizable AI interactions and transparency in AI decision-making underscores addressing these computational limitations.

While LLMs present groundbreaking capabilities across fields, their significant computational requirements and unique challenges—such as hallucination, accountability, and transparency—highlight the urgent need for continued research and innovation to enhance efficiency and reliability. This ongoing effort is essential for fostering sustainable progress and ensuring LLM technologies' ethical deployment, addressing critical issues like bias reduction and responsible information dissemination [37, 43, 10, 45].

Feature	Valley Model	MarineGPT	HELPERT Model
Architecture Type	Structured Architecture	Transformer-based	Cbt-prompted
Training Process	Temporal Modeling	Continuous Pre-training	Fine-tuning
Ethical Considerations	Not Specified	Not Specified	Not Specified

Table 2: Table comparing the features of three Large Language Models (LLMs): Valley Model, MarineGPT, and HELPERT Model. The table outlines differences in their architecture types, training processes, and ethical considerations, highlighting the diverse approaches in LLM development and deployment.

4 Multimodal Models and Cross-Modal Learning

4.1 Integration of Multimodal Data

Integrating multimodal data is essential for advanced AI systems, enabling the synthesis and processing of diverse sources like text, images, and audio, which enhances contextual understanding and functionality in complex environments. MarineGPT exemplifies this by merging marine images with text to improve object recognition [5], while the Valley model demonstrates the potential of combining video and language understanding to enhance human-AI interactions [3].

Multimodal models leverage structured data integration to boost predictive accuracy. The Recursive Attention Model (RAM) uses recursive reasoning to retrieve relevant information, adapting memory based on user feedback for enhanced adaptability [53]. In dynamic audio-visual scenarios, the CAT model improves MLLMs by aggregating question-related clues and employing an innovative training strategy, emphasizing adaptive learning's importance [54].

The TaskMatrix.AI framework organizes methods into stages, including a multimodal conversational foundation model, illustrating a systematic approach to multimodal data integration [55]. Exploring vulnerabilities in MLLMs underscores the need for robust security measures in multimodal systems [56].

Recent surveys introduce a unified mathematical framework encapsulating single-stream and dual-stream architectures, providing a basis for understanding multimodal pretraining [32]. This aids in developing models capable of managing multiple data streams, enhancing performance.

In educational contexts, AI tutors powered by LLMs engage Deaf and Hard of Hearing learners through culturally relevant interactions, showcasing multimodal models' potential to enrich educational experiences [57].

Integrating multimodal data is crucial for advancing machine learning systems, significantly enhancing model performance and applicability across real-world scenarios. Advancements in Natural Language Generation have led to AI-assisted writing tools that predict user needs, supporting seamless collaboration. Detection methods like GigaCheck ensure responsible generative AI use by distinguishing between human and AI-generated content [12, 43, 58, 59].

4.2 Cross-Modal Knowledge Transfer

Cross-modal knowledge transfer is crucial in AI, facilitating knowledge transfer across modalities like text, images, and audio, enhancing understanding and synthesis of complex information for improved performance. Integrating neuroimaging data with symbolic clinical features enhances predictive accuracy in medical contexts [60]. This holistic approach leads to improved outcomes.

In multimodal emotional understanding, datasets containing images, videos, and textual instructions related to emotional tasks highlight cross-modal knowledge transfer's significance in enhancing AI's emotional intelligence [61]. This capability is essential for developing AI systems that naturally interact with humans.

Integrating multimodal graphs poses challenges like managing varying graph topologies and ensuring effective data fusion while preserving relationships [62]. Addressing these challenges is crucial for seamless cross-modal knowledge transfer.

Enhancing MLLMs through a clue aggregator captures question-related clues and employs a mixed training strategy, improving understanding in audio-visual contexts [54]. This approach highlights cross-modal knowledge transfer's potential in refining AI models' capabilities.

In robotic systems, integrating foundation models is challenged by the need for improved planning and reasoning capabilities, critical for effective cross-modal knowledge transfer [63]. Enhancing these capabilities will enable robots to better interpret and respond to multimodal inputs.

Cross-modal knowledge transfer enhances AI systems by bridging diverse forms of knowledge, such as cultural practices and modern methodologies. The Text2Tradition system exemplifies this by translating prompts into Thai classical dance sequences, preserving cultural heritage while fostering innovation [15]. Generative AI transforms computational social sciences, enabling researchers to analyze multimodal data effectively, enriching cultural dialogue and educational practices [48].

4.3 Applications in AI Systems

The application of multimodal models in AI systems spans sectors like healthcare, social media analysis, recommendation systems, and creative industries. In healthcare, integrating multimodal data enhances diagnostic and treatment planning, improving predictive accuracy and patient outcomes [60]. Multimodal graph learning techniques advance healthcare applications by enabling comprehensive data analysis [62].

In social media analysis, multimodal models capture and analyze data from multiple sources, offering insights into user behavior and preferences. Techniques like Multimodal Graph Convolutional Networks enhance social media analytics [62]. However, challenges like data imbalance and alignment remain critical [62].

AI-generated, for reference only.

In creative industries, text-to-image generation models provide tools for artists, enabling textual descriptions' translation into visual content, fostering innovation [64]. Models like GPT-3 generate multilingual content, presenting opportunities and challenges for spreading narratives [65].

In education, multimodal models enhance summarization methods, supporting extractive and abstractive approaches to improve accessibility [58]. In medical domains, multilingual tasks are supported by new evaluation datasets, underscoring language diversity's importance [66].

The CAT model excels in audio-visual question answering, accurately localizing specific objects and providing clear responses, enhancing user interactions [54]. AI systems' adaptability and collaborative learning among agents improve through innovative approaches, fostering robust applications [67]. Diverse training data is crucial to mitigate bias, ensuring equitable outcomes [68].

The application of multimodal models in AI systems drives innovation across sectors, enhancing AI technologies' capabilities in addressing complex challenges. Exploring attack methodologies against MLLMs underscores the necessity for robust security measures [56].

Evaluation Setting	Description
Loebner Prize[6]	General Turing Test implementation.
U. of Reading TT 2014[7]	General Turing Test implementation.
BotPrize[8]	Contest about bot believability in videogames. [114, 84]
Robo Chat Challenge[9]	Chattering bots competition.
CAPTCHAs[10]	Spotting bots in applications requiring humans. [172, 173]
Humies awards[11]	Human-competitive results using genetic and evolutionary computation. [103]
Graphics Turing Test	Tell between a computer-generated virtual world and a real camera. [126, 16]

(a) Evaluation Setting Description[69]

Dataset	Usage	Size	Mean Num. of Hypotheses
$Train_1$	Model training	10.4MM	5.9
$Train_2$	Model training with data augmentation	10.4MM	12.5
$Test_1$	Offline evaluation	9.7MM	5.4
$Test_2$	Online evaluation	9.5MM	7.5

(b) Dataset Usage and Mean Number of Hypotheses for Different Datasets[70]

Figure 4: Examples of Applications in AI Systems

As illustrated in Figure 4, the integration of multimodal models and cross-modal learning is pivotal in AI, offering applications across systems. The first figure presents an "Evaluation Setting Description," detailing evaluation environments like the Loebner Prize, assessing AI systems through unique criteria. The second figure, "Dataset Usage and Mean Number of Hypotheses for Different Datasets," provides insight into these models' practical application, outlining how different datasets are employed in training. These visuals underscore multimodal models' complexity and versatility in AI, emphasizing their role in enhancing robustness and adaptability through cross-modal learning techniques [69, 70].

5 Neural Networks

5.1 Foundational Role and Architecture

Neural networks are pivotal in artificial intelligence (AI), providing the architecture essential for diverse applications. Mimicking the human brain's neocortex, responsible for complex cognitive functions, these networks consist of interconnected layers of nodes or neurons that process and transmit information, enabling learning and generalization from data [71]. Their architecture benefits from parameter sharing, enhancing learning efficiency and adaptability, which is crucial for developing AI models capable of performing sophisticated tasks with minimal human intervention. For instance, neural networks underpin chatbot systems that automate customer support by analyzing inquiries and generating contextually relevant responses [72].

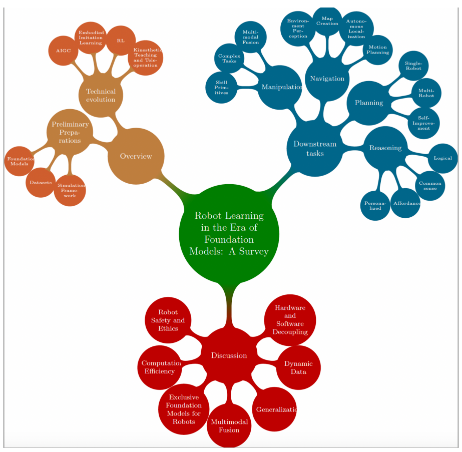
Convolutional neural networks (CNNs), a specialized category of neural networks, excel in processing image data. The ZigZag ResNet illustrates the versatility of CNNs by classifying text through image representations of word embeddings, demonstrating their capability to handle varied data modalities [73]. Additionally, the Bilateral Artificial Neural Network (BANN) architecture, which features two hemispheres trained with distinct objectives, specializes in local and global feature extraction, mirroring the bilateral structure of the human brain [74].

Domain-specific models like MarineGPT exemplify the application of neural networks in specialized fields, combining continuous pre-training with instruction-following fine-tuning to provide targeted knowledge for marine research [5]. Similarly, the Valley model integrates a large language model with a temporal modeling module, highlighting the foundational role of neural networks in processing sequential data and enhancing video comprehension [3].

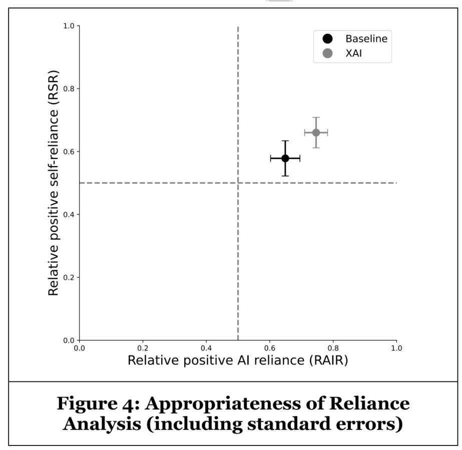
Innovative concepts such as 'neural erosion'—which manipulates weight matrices in neural networks to simulate cognitive decline—provide insights into the limitations and potential of AI technologies [7]. This underscores the importance of understanding neural architectures in emulating human cognitive processes and addressing challenges related to aging and neurodegeneration.

Moreover, the integration of structured knowledge with AI capabilities, such as interpreting and translating cultural narratives into movement, illustrates the potential of neural networks to bridge traditional knowledge systems with contemporary AI techniques [15]. This synthesis highlights the role of neural networks in connecting diverse knowledge domains.

Neural networks are integral to AI advancement, providing the architectural framework necessary for developing intelligent systems capable of a wide range of tasks. Their ability to learn, generalize, and adapt is pivotal across applications, as evidenced by progress in Natural Language Processing (NLP) and Generative AI technologies. These systems enhance user experiences in writing and data analysis while facilitating the automation of complex processes, thus improving productivity in fields such as education and social sciences. By enabling machines to interpret and generate human-like text, these technologies underscore their foundational importance in driving innovation and addressing challenges like academic integrity and user engagement in AI-assisted tools [73, 59, 75, 48, 12].



(a) Robot Learning in the Era of Foundation Models: A Survey[63]



(b) The figure represents the appropriateness of Reliance Analysis, including standard errors.[76]

Figure 5: Examples of Foundational Role and Architecture

As shown in Figure 5, the exploration of the foundational role and architecture of neural networks presents two illustrative examples. The first, titled "Robot Learning in the Era of Foundation Models: A Survey," is depicted through a mind map that delineates essential topics pertinent to robot learning amidst the influence of foundation models. This representation serves as a guide, emphasizing the core theme of robot learning and its interconnected aspects. The second example examines the appropriateness of reliance analysis through a scatter plot, juxtaposing "Baseline" and "XAI" data against "Relative positive AI reliance (RAIR)" and "Relative positive self-reliance (RSR)," with error bars illustrating standard errors. This visual analysis provides insights into the dynamics of AI reliance versus self-reliance, contributing to a nuanced understanding of decision-making processes in human-AI interactions. Together, these examples encapsulate the foundational principles and architectural considerations pivotal in advancing neural network technologies.

5.2 Advancements in Neural Network Technologies

Recent advancements in neural network technologies have significantly enhanced their capabilities and broadened applicability across various domains. Key improvements include the integration of large datasets and rapid computing, revitalizing neural networks for greater effectiveness and efficiency [71]. This progress has facilitated the development of lightweight solutions, exemplified by the ConvNLP method, achieving an average detection rate of 88.35% [73].

Innovations in neural architectures, such as the Bilateral Artificial Neural Network (BANN), have demonstrated improved performance in classifying both fine and coarse classes compared to baseline models [74]. This underscores the potential of bilateral architectures to enhance classification capabilities. Furthermore, integrating parameter sharing with cognitive metaphor theory presents novel perspectives on learning dynamic skills, paving the way for adaptive neural networks [26].

In healthcare, the effective merging of multimodal data—specifically, neuroimaging and clinical features—using Convolutional Neural Network (CNN) architectures represents a significant innovation, enhancing predictive accuracy in healthcare applications [60]. However, challenges remain, including high computational demands and the need for methodologies that optimize neural architectures efficiently without excessive resource consumption [77].

Despite advancements, research is constrained by the absence of standardized methods for verifying and validating AI outputs, leading to inconsistent quality and reliability [78]. Moreover, some Neuro-Symbolic AI systems face high computational costs and reliance on domain-specific rules, limiting scalability and applicability [18].

The advancements in neural network technologies reflect a dynamic landscape characterized by innovative architectures and methodologies that enhance AI systems' performance and adaptability across diverse applications. These developments address significant technical challenges in AI text detection and performance prediction, facilitating the creation of efficient and reliable neural network implementations, as evidenced by models like ZigZag ResNet and GigaCheck, which enhance AI-generated content detection and improve generalization capabilities across datasets [12, 73, 59, 43].

5.3 Evaluation and Benchmarking

Benchmark	Size	Domain	Task Format	Metric
GAIE[79]	905	Question Answering	Evaluation	Accuracy, F1
OffTheRails[80]	400	Moral Reasoning	Moral Dilemma Evaluation	Permissibility, Intention
RELAX[81]	1,000	Language Model Editing	Knowledge Retrieval	Accuracy, Reversion
LLM-SUM[58]	300,000	Text Summarization	Abstractive Summarization	BLEU, ROUGE
EmoBench[61]	287,000	Emotional Recognition	Emotion Classification	Accuracy, F1-score
LLM-IS[82]	1,500	Conversational AI	Language Style Imitation	Human Evaluation, Automatic Evaluation
AIEB[69]	10,000	Cognitive Robotics	Cognitive Ability Assessment	Accuracy, F1-score
AraSum[83]	4,000	Clinical Documentation	Summarization	F1 Score, ROUGE-1

Table 3: Table ef presents an overview of representative benchmarks used for evaluating neural networks across various domains. It includes details on benchmark size, domain, task format, and the metrics employed for performance assessment, providing a comprehensive resource for understanding different evaluation criteria in neural network research.

Evaluation and benchmarking of neural networks are critical for assessing performance and ensuring reliability across applications. These processes involve comparing neural networks' capabilities on language tasks with human cognitive functions, utilizing established datasets and benchmarks to draw meaningful comparisons [71]. Such evaluations are essential for understanding how closely neural networks mimic human cognitive processes and for identifying areas for improvement. Table 3 provides a detailed overview of representative benchmarks utilized in the evaluation and benchmarking of neural networks, highlighting their application across different domains and task formats.

Performance assessment typically employs metrics such as accuracy, balanced accuracy, the area under the Receiver Operating Characteristic (ROC) curve, and the F1-score, providing a comprehensive overview of a model's effectiveness. Cross-validation techniques ensure robustness and generalizability, mitigating overfitting and confirming models perform well on unseen data, crucial for real-world deployment [60].

In neural architecture search (NAS), comparative analyses of methodologies highlight the effectiveness of evolutionary algorithms versus reinforcement learning, with evolutionary algorithms offering superior computational efficiency and performance outcomes, making them preferred for optimizing neural network architectures [77]. This analysis guides the selection of appropriate NAS methodologies, ensuring efficient and effective neural networks.

Evaluation and benchmarking are vital in neural network development, providing insights into performance metrics and identifying enhancement areas in design and implementation. This is particularly important in advanced applications, such as AI-generated text detection, where models like ZigZag ResNet demonstrate significant improvements in generalization and efficiency. The integration of modularized approaches, such as Modularized Adaptive Neural Architecture Search (MANAS), emphasizes the importance of adaptive learning in creating personalized architectures for specific tasks. Additionally, exploring the combination of deep learning with symbolic reasoning in Neuro-Symbolic AI aims to enhance capabilities in natural language processing, improving reasoning and generalization. These evaluations guide iterative improvements and contribute to safeguarding academic integrity while enhancing neural networks' applicability across diverse domains [73, 18, 59, 84]. By employing rigorous evaluation methods and leveraging advanced NAS techniques, researchers can enhance neural networks' capabilities, ensuring their continued advancement and applicability across diverse fields.

6 Cognitive Computing

Cognitive computing aims to simulate human cognitive processes through diverse technologies and methodologies. This simulation is crucial for enhancing AI-assisted writing tools, detecting LLM-generated content, and integrating AI within educational frameworks [12, 43, 85]. By emulating human thought processes, cognitive computing enhances decision-making and problem-solving capabilities. This section explores the mechanisms and models that facilitate these simulations, emphasizing the principles guiding cognitive computing.

6.1 Simulating Human Thought Processes

Cognitive computing seeks to replicate human cognition, thereby augmenting decision-making and problem-solving through advanced AI technologies. Models like Valley exemplify this by processing video and language inputs to generate nuanced outputs, showcasing AI's potential to mimic human cognition [3]. In rehabilitation management for schizophrenia, AI enhances symptom tracking, personalizes interventions, and develops engaging therapeutic tools, underscoring its role in improving therapeutic outcomes [16].

AI models such as MarineGPT demonstrate cognitive computing's adaptability by generating contextually relevant responses based on specialized knowledge, highlighting the importance of ecological complexity in cognitive skill development [5, 20]. Parameter sharing, informed by cognitive theories of analogy making, enhances AI's learning and adaptability, fostering systems that simulate human-like reasoning [26]. Despite these advancements, achieving true understanding and reasoning remains challenging, as AI systems often lack comprehension of underlying concepts. Ongoing research aims to replicate human decision-making processes, revealing complexities in cognitive emulation [29].

Cognitive computing's quest to replicate human thought processes drives AI innovation while exposing significant challenges. Advanced systems, such as generative AI (GenAI), improve usability by integrating metacognitive support strategies. AI chatbots like ReMe are explored for personalized cognitive training, addressing cognitive disorders through tailored interventions leveraging large language models. Research shows AI interactions can stimulate human creativity, illustrating cognitive computing's dual nature: expanding AI capabilities while revealing the complexities of mimicking human cognition [86, 87, 88]. By integrating insights from cognitive science and leveraging advanced neural architectures, cognitive computing seeks to bridge the gap between human and machine intelligence, enhancing AI technologies across diverse domains.

6.2 Enhancing Decision-Making and Problem-Solving

Cognitive computing significantly enhances decision-making and problem-solving by simulating human cognitive processes and providing sophisticated tools for complex scenarios. The BRAINSTORM method, for instance, assists radiologists by generating a broader spectrum of diagnostic hypotheses, improving medical decision-making [28]. In automated recruitment systems, removing sensitive information from training data promotes fairer decision-making processes, addressing biases and enhancing AI-driven decision transparency [89].

The application of Large Language Models (LLMs) in therapy illustrates cognitive computing's potential, providing structured Cognitive Behavioral Therapy (CBT) techniques. However, LLMs may lack the emotional connection inherent in human interactions, highlighting the need for further development in AI systems to emulate human empathy [8]. In educational contexts, AI-powered tutors emphasize transparency in persona design and ethical considerations, fostering trust in AI interactions [57]. The Text2Tradition framework demonstrates cognitive computing's potential to enhance decision-making in cultural interpretation and co-creation, integrating traditional knowledge with modern technology [15].

Despite advancements, concerns about job displacement in creative fields and the ethical implications of AI-generated content persist [64]. Addressing these concerns is vital for ensuring cognitive computing positively impacts decision-making and problem-solving across various domains.

Cognitive computing enhances decision-making and problem-solving by providing tools and frameworks that mimic human cognition, support personalized interactions, and ensure ethical considerations in AI deployments. Recent advancements highlight its transformative potential across sectors, including business, education, healthcare, and social sciences, through predictive models like the READ model, which integrates cognitive biases with natural language processing. These innovations drive progress, improve decision-making processes, and facilitate personalized cognitive training through AI chatbots, ultimately making complex data analysis and content creation more accessible and efficient [27, 86, 90, 48, 12].

6.3 Current Implementations

Cognitive computing has led to innovative implementations across various sectors, enhancing decision-making and problem-solving capabilities. Integrating self-modeling and theory of mind into agent architectures significantly improves decision-making in dynamic environments [46]. In healthcare, GatorTron, a clinical language model, outperforms existing transformer models on multiple clinical NLP tasks, demonstrating cognitive computing's potential in processing complex clinical data [91]. Similarly, the MOTOR framework shows promise in advancing medical artificial general intelligence through its exceptional performance in various medical tasks [92].

Robotic systems benefit from cognitive computing, with architectures enhancing multi-agent learning and long-term autonomy, improving coordination and decision-making in collaborative environments [67]. The ReMe framework exemplifies personalized cognitive training through tailored tasks leveraging personal experiences, enhancing user engagement and effectiveness [86]. The MRKL system, represented by Jurassic-X, showcases a modular neuro-symbolic architecture, integrating symbolic reasoning with neural networks to enhance AI capabilities [93].

In sustainable infrastructure planning, cognitive computing supports complex decision-making through a multi-objective optimization framework addressing cost, environmental impact, and efficiency [6]. Current implementations of cognitive computing demonstrate its transformative potential across sectors, enhancing productivity and accessibility in computational social science and personalized cognitive training. Generative AI automates complex coding tasks, enabling researchers with limited programming skills to analyze multimodal data more effectively. AI chatbots like ReMe revolutionize personalized cognitive training by providing interactive and tailored experiences targeting memory tasks, addressing cognitive disorders like Alzheimer's. These advancements significantly improve AI systems' decision-making and problem-solving capabilities, underscoring the importance of continued research and development [86, 48].

6.4 Future Directions and Opportunities

The future of cognitive computing holds significant promise, with numerous research directions poised to enhance AI systems' capabilities and applicability. Developing more interpretable AI models is critical to address ethical and environmental challenges associated with current practices [14]. Expanding the cultural scope of AI projects and developing ethical frameworks will ensure cultural authenticity, particularly in applications integrating traditional knowledge with modern AI capabilities [15].

Future research should enhance the multidimensional nature of AI reliance, particularly in high-stakes decision-making scenarios, and validate findings in practical applications [29]. This includes exploring the scalability and adaptability of models like the READ model across various domains and languages [27]. In combating misinformation, developing multilingual AI tools to assist fact-checkers in real-time is crucial [17]. Integrating audio inputs and expanding multilingual capabilities in models like Valley can enhance decision-making, further advancing cognitive computing's potential [3].

Exploring ethical implications and improving methods for predicting technological progress are vital for addressing challenges posed by AI advancements [21]. Future research should investigate user awareness of deceptive practices in AI interactions, examining the psychological impact and developing guidelines for ethical design in AI writing assistants [19]. Incorporating human-like reasoning capabilities in AI systems through diverse benchmarks and exploring various types of reasoning will enhance cognitive processes' integration [18]. Leveraging social media data for monitoring, developing serious games for rehabilitation, and enhancing multimodal data integration will advance personalized patient interactions, particularly in mental health applications [16].

The future of cognitive computing is set to catalyze substantial advancements in AI systems, addressing existing challenges and capitalizing on emerging opportunities. This evolution will enhance AI capabilities across sectors, including transportation, healthcare, and education, while improving user experiences in applications like AI-assisted writing tools that predict user needs and provide real-time suggestions. As AI systems become more sophisticated, they are expected to outperform humans in various tasks, leading to transformative changes in job markets and necessitating proactive policy adaptations. Innovations in large language models (LLMs) will further democratize access to advanced analytical tools, empowering researchers in computational social sciences and beyond. The trajectory of cognitive computing promises to reshape our interaction with technology, fostering collaboration and enhancing productivity while raising important considerations regarding ethics and bias [12, 24, 48, 21]. By focusing on these research directions, cognitive computing can continue to evolve and contribute to developing more intelligent, adaptable, and ethically aligned AI technologies.

7 Conclusion

This survey explores the transformative potential of advanced artificial intelligence (AI) models, particularly large language models (LLMs), multimodal models, and cognitive computing. The findings emphasize the need for dynamic auditing frameworks and tailored ethical guidelines to address the unique challenges posed by LLMs, ensuring responsible deployment and mitigating societal inequities. Notably, advancements in LLMs, such as ChatGPT, suggest pathways to general intelligence and innovative applications [24].

Interoperability among AI and language technology platforms is essential for maximizing technological potential through collaboration and standardized practices [94]. The integration of personalized, guideline-informed frameworks like PlanFitting illustrates AI's adaptability and usability in real-world scenarios, paving the way for tailored solutions that enhance user engagement. Future research should focus on developing ability-oriented assessments reflecting AI's cognitive capabilities, ensuring evaluations align with human cognitive benchmarks [69].

Examining AI's moral discourse across cultural contexts can reveal the complex interactions between perceived quality and source attribution, offering deeper insights into AI's societal role [95]. Enhancing frameworks' robustness against conversation deviations and investigating applications in multi-agent systems and cooperative AI are crucial future research areas [96]. The RAM model's effectiveness in dynamic knowledge acquisition underscores AI systems' potential for continuous learning and adaptation, enhancing capabilities in real-time applications [53].

Interdisciplinary collaboration is vital for advancing AI technologies, fostering innovation, and addressing the multifaceted challenges these systems present. By adopting a collaborative approach, stakeholders can ensure that AI technologies are developed and deployed equitably and effectively, aligning with societal values and enhancing human experiences across diverse applications [97]. Research into human-AI interaction frameworks that promote user-centered explanations can bridge the gap between explainable AI and human-robot interaction, further underscoring the importance of interdisciplinary inquiry [98].

The conclusion highlights that MarineGPT significantly enhances the generation of accurate, informative responses for marine-related inquiries, indicating future research directions to further improve the model's capabilities [5]. Additionally, findings suggest that leveraging parameter sharing through analogy-making can enhance AI models' abilities, particularly in generalization and dynamic task learning [26]. Key insights indicate a 50

AI-generated, for reference only.

This survey underscores the transformative potential of advanced AI models and the critical role of interdisciplinary collaboration in shaping AI's future, ensuring that these technologies contribute positively to society while addressing ethical and practical challenges.

AI-generated, for reference only.

References

[1] Poorvesh Dongre, Majid Behravan, Kunal Gupta, Mark Billinghurst, and Denis Gračanin. Integrating physiological data with large language models for empathic human-ai interaction, 2024.

[2] Eliza Kosoy, Soojin Jeong, Anoop Sinha, Alison Gopnik, and Tanya Kraljic. Children's mental models of generative visual and text based ai models, 2024.

[3] Ruipu Luo, Ziwang Zhao, Min Yang, Junwei Dong, Da Li, Pengcheng Lu, Tao Wang, Linmei Hu, Minghui Qiu, and Zhongyu Wei. Valley: Video assistant with large language model enhanced ability, 2023.

[4] Suyi Zhang, Ekram Alam, Jack Baber, Francesca Bianco, Edward Turner, Maysam Chamanzar, and Hamid Dehghani. Mindspeech: Continuous imagined speech decoding using high-density fnirs and prompt tuning for advanced human-ai interaction, 2024.

[5] Ziqiang Zheng, Jipeng Zhang, Tuan-Anh Vu, Shizhe Diao, Yue Him Wong Tim, and Sai-Kit Yeung. Marinegpt: Unlocking secrets of ocean to the public, 2023.

[6] Gaurav Singh and Kavitesh Kumar Bali. Enhancing decision-making in optimization through llm-assisted inference: A neural networks perspective, 2024.

[7] Antonios Alexos, Yu-Dai Tsai, Ian Domingo, Maryam Pishgar, and Pierre Baldi. Neural erosion: Emulating controlled neurodegeneration and aging in ai systems, 2024.

[8] Zainab Iftikhar, Sean Ransom, Amy Xiao, and Jeff Huang. Therapy as an nlp task: Psychologists' comparison of llms and human peers in cbt, 2024.

[9] Subhankar Maity and Aniket Deroy. Human-centric explainable ai in education, 2024.

[10] Q. Vera Liao and Jennifer Wortman Vaughan. Ai transparency in the age of llms: A human-centered research roadmap, 2023.

[11] Yasmeen Alufaisan, Laura R. Marusich, Jonathan Z. Bakdash, Yan Zhou, and Murat Kantarcioglu. Does explainable artificial intelligence improve human decision-making?, 2020.

[12] Carlos Alves Pereira, Tanay Komarlu, and Wael Mobeirek. The future of ai-assisted writing, 2023.

[13] Vijay Arya, Rachel K. E. Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C. Hoffman, Stephanie Houde, Q. Vera Liao, Ronny Luss, Aleksandra Mojsilović, Sami Mourad, Pablo Pedemonte, Ramya Raghavendra, John Richards, Prasanna Sattigeri, Karthikeyan Shanmugam, Moninder Singh, Kush R. Varshney, Dennis Wei, and Yunfeng Zhang. One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques, 2019.

[14] Bernard J. Koch and David Peterson. From protoscience to epistemic monoculture: How benchmarking set the stage for the deep learning revolution, 2024.

[15] Pat Pataranutaporn, Chayapatr Archiwaranguprok, Phoomparin Mano, Piyaporn Bhongse-tong, Pattie Maes, and Pichet Klunchun. Text2tradition: From epistemological tensions to ai-mediated cross-cultural co-creation, 2024.

[16] Hongyi Yang, Fangyuan Chang, Dian Zhu, Muroi Fumie, and Zhao Liu. Application of artificial intelligence in schizophrenia rehabilitation management: A systematic scoping review, 2025.

[17] Andrea Hrckova, Robert Moro, Ivan Srba, Jakub Simko, and Maria Bielikova. Autonomation, not automation: Activities and needs of fact-checkers as a basis for designing human-centered ai systems, 2024.

[18] Kyle Hamilton, Aparna Nayak, Bojan Božić, and Luca Longo. Is neuro-symbolic ai meeting its promise in natural language processing? a structured review, 2022.

[19] Karim Benharrak, Tim Zindulka, and Daniel Buschek. Deceptive patterns of intelligent and interactive writing assistants, 2024.

[20] Eleni Nisioti and Clément Moulin-Frier. Grounding artificial intelligence in the origins of human behavior, 2020.

[21] Katja Grace, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. When will ai exceed human performance? evidence from ai experts. *Journal of Artificial Intelligence Research*, 62:729–754, 2018.

[22] Nitisha Aggarwal, Geetika Jain Saxena, Sanjeev Singh, and Amit Pundir. Can i say, now machines can think?, 2023.

[23] James Prather, Paul Denny, Juho Leinonen, Brett A. Becker, Ibrahim Albluwi, Michelle Craig, Hieke Keuning, Natalie Kiesler, Tobias Kohn, Andrew Luxton-Reilly, Stephen MacNeil, Andrew Peterson, Raymond Pettit, Brent N. Reeves, and Jaromir Savelka. The robots are here: Navigating the generative ai revolution in computing education, 2023.

[24] Frank Joublin, Antonello Ceravola, Joerg Deigmoeller, Michael Gienger, Mathias Franzius, and Julian Eggert. A glimpse in chatgpt capabilities and its impact for ai research, 2023.

[25] Zhijing Jin, Max Kleiman-Weiner, Giorgio Piatti, Sydney Levine, Jiarui Liu, Fernando Gonzalez, Francesco Ortu, András Strausz, Mrinmaya Sachan, Rada Mihalcea, Yejin Choi, and Bernhard Schölkopf. Language model alignment in multilingual trolley problems, 2024.

[26] Roland Memisevic. Metaphors we learn by, 2022.

[27] Nimrod Dvir, Elaine Friedman, Suraj Commuri, Fan yang, and Jennifer Romano. A predictive model of digital information engagement: Forecasting user engagement with english words by incorporating cognitive biases, computational linguistics and natural language processing, 2023.

[28] Liyan Tang, Yifan Peng, Yanshan Wang, Ying Ding, Greg Durrett, and Justin F. Rousseau. Less likely brainstorming: Using language models to generate alternative hypotheses, 2023.

[29] Sven Eckhardt, Niklas Kühl, Mateusz Dolata, and Gerhard Schwabe. A survey of ai reliance, 2024.

[30] Christopher Irrgang, Niklas Boers, Maike Sonnewald, Elizabeth A. Barnes, Christopher Kadow, Joanna Staneva, and Jan Saynisch-Wagner. Will artificial intelligence supersede earth system and climate models?, 2021.

[31] Emile van Krieken. Optimisation in neurosymbolic learning systems, 2024.

[32] Emanuele Bugliarello, Ryan Cotterell, Naoaki Okazaki, and Desmond Elliott. Multimodal pretraining unmasked: A meta-analysis and a unified framework of vision-and-language berts, 2021.

[33] Ruijia Cheng, Alison Smith-Renner, Ke Zhang, Joel R. Tetreault, and Alejandro Jaimes. Mapping the design space of human-ai interaction in text summarization, 2022.

[34] Mingyu Derek Ma, Chenchen Ye, Yu Yan, Xiaoxuan Wang, Peipei Ping, Timothy S Chang, and Wei Wang. Clibench: A multifaceted and multigranular evaluation of large language models for clinical decision making, 2024.

[35] Mohamad Ballout, Ulf Krumnack, Gunther Heidemann, and Kai-Uwe Kühnberger. Investigating pre-trained language models on cross-domain datasets, a step closer to general ai, 2023.

[36] Ximing Lu, Melanie Sclar, Skyler Hallinan, Niloofar Mireshghallah, Jiacheng Liu, Seungju Han, Allyson Ettinger, Liwei Jiang, Khyathi Chandu, Nouha Dziri, and Yejin Choi. Ai as humanity's salieri: Quantifying linguistic creativity of language models via systematic attribution of machine text against web text, 2025.

[37] Junfeng Jiao, Saleh Afroogh, Yiming Xu, and Connor Phillips. Navigating llm ethics: Advancements, challenges, and future directions, 2024.

[38] Donghoon Shin, Gary Hsieh, and Young-Ho Kim. Planfitting: Tailoring personalized exercise plans with large language models, 2023.

[39] Anja Meunier, Michal Robert Žák, Lucas Munz, Sofiya Garkot, Manuel Eder, Jiachen Xu, and Moritz Grosse-Wentrup. A conversational brain-artificial intelligence interface, 2024.

[40] Hanyao Huang, Ou Zheng, Dongdong Wang, Jiayi Yin, Zijin Wang, Shengxuan Ding, Heng Yin, Chuan Xu, Renjie Yang, Qian Zheng, and Bing Shi. Chatgpt for shaping the future of dentistry: The potential of multi-modal large language model, 2023.

[41] Self-critiquing models for assisting human evaluators.

[42] Stephen MacNeil, Andrew Tran, Joanne Kim, Ziheng Huang, Seth Bernstein, and Dan Mogil. Prompt middleware: Mapping prompts for large language models to ui affordances, 2023.

[43] Irina Tolstykh, Aleksandra Tsybina, Sergey Yakubson, Aleksandr Gordeev, Vladimir Dokholyan, and Maksim Kuprashevich. Gigacheck: Detecting llm-generated content, 2024.

[44] Sen Huang, Kaixiang Yang, Sheng Qi, and Rui Wang. When large language model meets optimization, 2024.

[45] Bhashithe Abeysinghe and Ruhan Circi. The challenges of evaluating llm applications: An analysis of automated, human, and llm-based approaches, 2024.

[46] Antonio Chella, Francesco Lanza, and Valeria Seidita. Decision process in human-agent interaction: Extending jason reasoning cycle. In *Engineering Multi-Agent Systems: 6th International Workshop, EMAS 2018, Stockholm, Sweden, July 14-15, 2018, Revised Selected Papers 6*, pages 320–339. Springer, 2019.

[47] Atin Sakkeer Hussain and Anu Thomas. Large language models for judicial entity extraction: A comparative study, 2024.

[48] Yongjun Zhang. Generative ai has lowered the barriers to computational social sciences, 2023.

[49] Quentin Gallea. From mundane to meaningful: Ai's influence on work dynamics – evidence from chatgpt and stack overflow, 2023.

[50] S. M Towhidul Islam Tonmoy, S M Mehedi Zaman, Vinija Jain, Anku Rani, Vipula Rawte, Aman Chadha, and Amitava Das. A comprehensive survey of hallucination mitigation techniques in large language models, 2024.

[51] Kevin R. McKee. Human participants in ai research: Ethics and transparency in practice, 2024.

[52] Da Song, Xuan Xie, Jiayang Song, Derui Zhu, Yuheng Huang, Felix Juefei-Xu, and Lei Ma. Luna: A model-based universal analysis framework for large language models, 2024.

[53] Jiaqi Li, Xiaobo Wang, Wentao Ding, Zihao Wang, Yipeng Kang, Zixia Jia, and Zilong Zheng. Ram: Towards an ever-improving memory system by learning from communications, 2024.

[54] Qilang Ye, Zitong Yu, Rui Shao, Xinyu Xie, Philip Torr, and Xiaochun Cao. Cat: Enhancing multimodal large language model to answer questions in dynamic audio-visual scenarios, 2024.

[55] Yaobo Liang, Chenfei Wu, Ting Song, Wenshan Wu, Yan Xia, Yu Liu, Yang Ou, Shuai Lu, Lei Ji, Shaoguang Mao, Yun Wang, Linjun Shou, Ming Gong, and Nan Duan. Taskmatrix.ai: Completing tasks by connecting foundation models with millions of apis, 2023.

[56] Yihe Fan, Yuxin Cao, Ziyu Zhao, Ziyao Liu, and Shaofeng Li. Unbridled icarus: A survey of the potential perils of image inputs in multimodal large language model security, 2024.

[57] Haocong Cheng, Si Chen, Christopher Perdriau, and Yun Huang. Llm-powered ai tutors with personas for d/deaf and hard-of-hearing online learners, 2024.

[58] Lochan Basyal and Mihir Sanghvi. Text summarization using large language models: A comparative study of mpt-7b-instruct, falcon-7b-instruct, and openai chat-gpt models, 2023.

[59] Akshansh Mishra, Vijaykumar S Jatti, Vaishnavi More, Anish Dasgupta, Devarrishi Dixit, and Eyob Messele Sefene. Performance prediction of data-driven knowledge summarization of high entropy alloys (heas) literature implementing natural language processing algorithms, 2023.

[60] Adam White, Margarita Saranti, Artur d'Avila Garcez, Thomas M. H. Hope, Cathy J. Price, and Howard Bowman. Predicting recovery following stroke: deep learning, multimodal data and feature selection using explainable ai, 2023.

[61] Qu Yang, Mang Ye, and Bo Du. Emollm: Multimodal emotional understanding meets large language models, 2024.

[62] Ciyuan Peng, Jiayuan He, and Feng Xia. Learning on multimodal graphs: A survey, 2024.

[63] Xuan Xiao, Jiahang Liu, Zhipeng Wang, Yanmin Zhou, Yong Qi, Qian Cheng, Bin He, and Shuo Jiang. Robot learning in the era of foundation models: A survey, 2023.

[64] Jonas Oppenlaender, Aku Visuri, Ville Paananen, Rhema Linder, and Johanna Silvennoinen. Text-to-image generation: Perceptions and realities, 2023.

[65] Kris McGuffie and Alex Newhouse. The radicalization risks of gpt-3 and advanced neural language models, 2020.

[66] Iker García-Ferrero, Rodrigo Agerri, Aitziber Atutxa Salazar, Elena Cabrio, Iker de la Iglesia, Alberto Lavelli, Bernardo Magnini, Benjamin Molinet, Johana Ramirez-Romero, German Rigau, Jose Maria Villa-Gonzalez, Serena Villata, and Andrea Zaninello. Medical mt5: An open-source multilingual text-to-text llm for the medical domain, 2024.

[67] Georgios Th. Papadopoulos, Margherita Antona, and Constantine Stephanidis. Towards open and expandable cognitive ai architectures for large-scale multi-agent human-robot collaborative learning, 2021.

[68] Francisco Castro, Jian Gao, and Sébastien Martin. Human-ai interactions and societal pitfalls, 2023.

[69] Jose Hernandez-Orallo. Ai evaluation: past, present and future, 2016.

[70] Han Li, Sunghyun Park, Aswarth Dara, Jinseok Nam, Sungjin Lee, Young-Bum Kim, Spyros Matsoukas, and Ruhi Sarikaya. Neural model robustness for skill routing in large-scale conversational ai systems: A design choice exploration, 2021.

[71] David Mumford. The convergence of ai code and cortical functioning – a commentary, 2020.

[72] André Barbosa and Alan Godoy. Augmenting customer support with an nlp-based receptionist, 2021.

[73] Suriya Prakash Jambunathan, Ashwath Shankarnarayan, and Parijat Dube. Convnlp: Image-based ai text detection, 2024.

[74] Chandramouli Rajagopalan, David Rawlinson, Elkhonon Goldberg, and Gideon Kowadlo. Deep learning in a bilateral brain with hemispheric specialization, 2024.

[75] Hao Wu. What is learning? a primary discussion about information and representation, 2015.

[76] Max Schemmer, Andrea Bartos, Philipp Spitzer, Patrick Hemmer, Niklas Kühl, Jonas Liebschner, and Gerhard Satzger. Towards effective human-ai decision-making: The role of human learning in appropriate reliance on ai advice, 2023.

[77] Fanfei Meng, Chen-Ao Wang, and Lele Zhang. Evolution and efficiency in neural architecture search: Bridging the gap between expert design and automated optimization, 2024.

[78] Erik Blasch, James Sung, Tao Nguyen, Chandra P. Daniel, and Alisa P. Mason. Artificial intelligence strategies for national security and safety standards, 2019.

[79] Juhyun Oh, Eunsu Kim, Inha Cha, and Alice Oh. The generative ai paradox on evaluation: What it can solve, it may not evaluate, 2024.

[80] Jan-Philipp Fränken, Kanishk Gandhi, Tori Qiu, Ayesha Khawaja, Noah D. Goodman, and Tobias Gerstenberg. Procedural dilemma generation for evaluating moral reasoning in humans and language models, 2024.

[81] Xinbei Ma, Tianjie Ju, Jiyang Qiu, Zhuosheng Zhang, Hai Zhao, Lifeng Liu, and Yulong Wang. On the robustness of editing large language models, 2024.

[82] Ziyang Chen and Stylios Moscholios. Using prompts to guide large language models in imitating a real person's language style, 2024.

[83] Chanseo Lee, Sonu Kumar, Kimon A. Vogt, Sam Meraj, and Antonia Vogt. Advancing complex medical communication in arabic with sporo arasum: Surpassing existing large language models, 2024.

[84] Hanxiong Chen, Yunqi Li, He Zhu, and Yongfeng Zhang. Learn basic skills and reuse: Modularized adaptive neural architecture search (manas), 2022.

[85] Cecilia Ka Yuk Chan. A comprehensive ai policy education framework for university teaching and learning. *International journal of educational technology in higher education*, 20(1):38, 2023.

[86] Zilong Wang, Nan Chen, Luna K. Qiu, Ling Yue, Geli Guo, Yang Ou, Shiqi Jiang, Yuqing Yang, and Lili Qiu. The potential and value of ai chatbot in personalized cognitive training, 2024.

[87] Ara Vartanian, Xiaoxi Sun, Yun-Shiuan Chuang, Siddharth Suresh, Xiaojin Zhu, and Timothy T. Rogers. Learning interactions to boost human creativity with bandits and gpt-4, 2023.

[88] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. The metacognitive demands and opportunities of generative ai, 2024.

[89] Alejandro Peña, Ignacio Serna, Aythami Morales, Julian Fierrez, Alfonso Ortega, Ainhoa Herrarte, Manuel Alcantara, and Javier Ortega-Garcia. Human-centric multimodal machine learning: Recent advances and testbed on ai-based recruitment, 2023.

[90] Yixin Zhu, Tao Gao, Lifeng Fan, Siyuan Huang, Mark Edmonds, Hangxin Liu, Feng Gao, Chi Zhang, Siyuan Qi, Ying Nian Wu, Joshua B. Tenenbaum, and Song-Chun Zhu. Dark, beyond deep: A paradigm shift to cognitive ai with humanlike common sense, 2020.

[91] Xi Yang, Aokun Chen, Nima PourNejatian, Hoo Chang Shin, Kaleb E Smith, Christopher Parisien, Colin Compas, Cheryl Martin, Mona G Flores, Ying Zhang, Tanja Magoc, Christopher A Harle, Gloria Lipori, Duane A Mitchell, William R Hogan, Elizabeth A Shenkman, Jiang Bian, and Yonghui Wu. Gatortron: A large clinical language model to unlock patient information from unstructured electronic health records, 2022.

[92] Bingqian Lin, Zicong Chen, Mingjie Li, Haokun Lin, Hang Xu, Yi Zhu, Jianzhuang Liu, Wenjia Cai, Lei Yang, Shen Zhao, Chenfei Wu, Ling Chen, Xiaojun Chang, Yi Yang, Lei Xing, and Xiaodan Liang. Towards medical artificial general intelligence via knowledge-enhanced multimodal pretraining, 2023.

[93] Ehud Karpas, Omri Abend, Yonatan Belinkov, Barak Lenz, Opher Lieber, Nir Ratner, Yoav Shoham, Hofit Bata, Yoav Levine, Kevin Leyton-Brown, Dor Muhlgay, Noam Rozen, Erez Schwartz, Gal Shachaf, Shai Shalev-Shwartz, Amnon Shashua, and Moshe Tenenholtz. Mrkl systems: A modular, neuro-symbolic architecture that combines large language models, external knowledge sources and discrete reasoning, 2022.

[94] Georg Rehm, Dimitrios Galanis, Penny Labropoulou, Stelios Piperidis, Martin Welß, Ricardo Usbeck, Joachim Köhler, Miltos Deligiannis, Katerina Gkirtzou, Johannes Fischer, Christian Chiarcos, Nils Feldhus, Julián Moreno-Schneider, Florian Kintzel, Elena Montiel, Víctor Rodríguez Doncel, John P. McCrae, David Laqua, Irina Patricia Theile, Christian Dittmar, Kalina Bontcheva, Ian Roberts, Andrejs Vasiljevs, and Andis Lagzdiņš. Towards an interoperable ecosystem of ai and lt platforms: A roadmap for the implementation of different levels of interoperability, 2020.

[95] Eyal Aharoni, Sharlene Fernandes, Daniel J. Brady, Caelan Alexander, Michael Criner, Kara Queen, Javier Rando, Eddy Nahmias, and Victor Crespo. Attributions toward artificial agents in a modified moral turing test, 2024.

[96] Guohao Li, Hasan Abed Al Kader Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. Camel: Communicative agents for "mind" exploration of large language model society, 2023.

[97] Amon Rapp, Arianna Boldi, Lorenzo Curti, Alessandro Perrucci, and Rossana Simeoni. How do people ascribe humanness to chatbots? an analysis of real-world human-agent interactions and a theoretical model of humanness. *International Journal of Human–Computer Interaction*, 40(19):6027–6050, 2024.

[98] Marco Matarese, Francesco Rea, and Alessandra Sciutti. A user-centred framework for explainable artificial intelligence in human-robot interaction, 2021.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

AI-generated, for reference only.