Advanced Computational Techniques in AI: A Survey on Deep Research, Search Agents, AI Writing, Information Retrieval, and Natural Language Processing

Abstract

This survey paper provides a comprehensive examination of advanced computational techniques in artificial intelligence (AI), emphasizing their transformative potential across various domains. The integration of machine learning (ML) and natural language processing (NLP) is pivotal in enhancing systematic reviews, improving document retrieval, and fostering personalized interactions with large language models (LLMs). The paper explores the role of AI in mitigating information overload during crises, such as the COVID-19 pandemic, and highlights innovations like SynAsk, which enhances LLM capabilities in organic synthesis. The survey also addresses the ethical considerations of AI writing tools in educational settings and the differentiation between AI-generated and human-written texts. Furthermore, it examines the synergies between AI technologies, particularly in voice and language processing, and their applications in clinical and educational settings. The paper concludes by discussing future directions, including refining document retrieval techniques, expanding user studies for LLMs, and integrating AI with environmental sustainability initiatives. Through these insights, the survey underscores the critical role of AI in driving technological innovation and addressing complex challenges across diverse fields.

1 Introduction

1.1 Significance and Relevance of AI Techniques

Artificial intelligence (AI) techniques are transforming various fields by enhancing efficiency, fostering innovation, and managing complex tasks. In information retrieval, AI addresses challenges presented by large language models (LLMs), particularly in multilingual contexts, necessitating innovative approaches to improve retrieval accuracy and relevance [1]. The integration of machine learning (ML) and natural language processing (NLP) broadens the scope of systematic reviews in the social sciences, significantly enhancing information synthesis efficiency [2].

In legal systems, such as India's, the complexity of judicial processes underscores the need for automated solutions that assist legal professionals in swiftly identifying pertinent legal information [3]. The combination of information retrieval (IR) and neural methods enhances AI's capability to generate natural language descriptions for source code, highlighting the importance of these techniques in software development [4].

AI's ability to mitigate information overload is especially crucial during crises like the COVID-19 pandemic, where NLP techniques are utilized to manage vast data volumes and satisfy urgent information needs [5]. The evolution of NLP, particularly with the rise of LLMs, has significantly advanced predictive tasks in specialized fields such as organic chemistry [6].

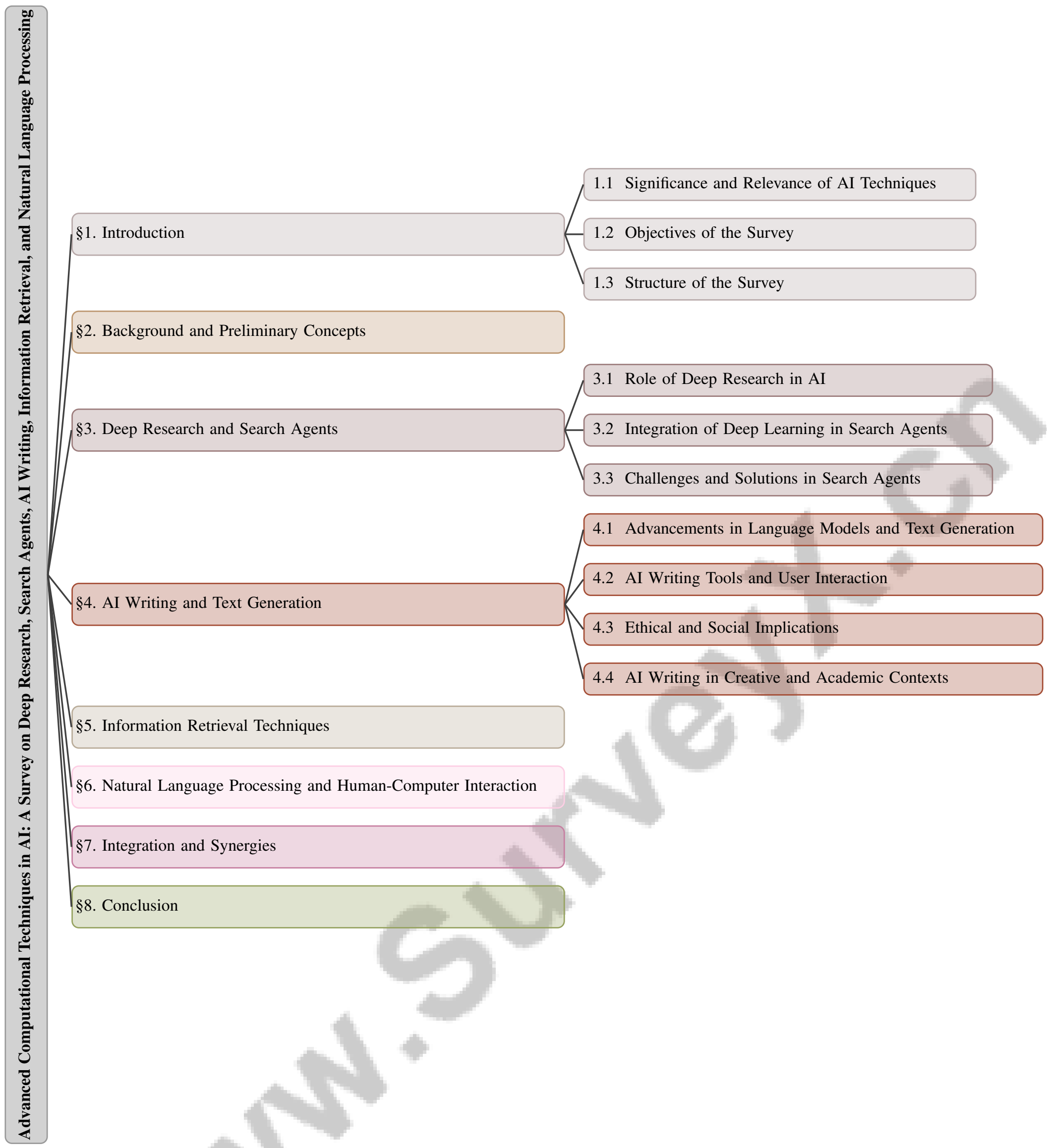

Figure 1: chapter structure

The educational sector is undergoing significant changes due to AI, with text generative technologies impacting academic integrity and the teaching-learning dynamic. Careful integration is necessary to uphold educational standards [7]. AI writing tools also enhance the writing experience for non-native English speakers by providing accessible and explainable support [8].

The distinction between AI-generated and human-written scientific texts is gaining traction, emphasizing the need for quality and reliability in AI-generated outputs [9]. As AI technologies like ChatGPT proliferate, ethical considerations in scholarly publishing require thorough evaluation to address emerging concerns [7].

AI techniques are reshaping technological innovation across various sectors, presenting opportunities and challenges that necessitate ongoing scrutiny and adaptation. The advancement of AI in expert identification on platforms such as Question Answer forums further illustrates the significance of these techniques in accurately categorizing user contributions [10]. Additionally, improvements in document retrieval methods address limitations in traditional approaches, particularly for East Asian languages [11].

Moreover, analyzing social phenomena through discourse in large document corpora reveals the limitations of existing keyword-based computational techniques, necessitating a more nuanced semantic approach [12]. These advancements underscore the transformative impact of AI techniques across modern technology and various fields.

1.2 Objectives of the Survey

This survey aims to comprehensively examine advanced computational techniques in artificial intelligence, focusing on their transformative potential across diverse domains. A primary objective is to automate time-intensive stages of the systematic reviewing process using machine learning (ML) and natural language processing (NLP) techniques, thereby enhancing the efficiency and scope of systematic reviews [2]. Another goal is to explore algorithmic aspects of document retrieval, addressing challenges posed by general sequence collections across various fields [11].

Enhancing user experience through the integration of task context and user perceptions in interactions with large language models (LLMs) is also a significant aim, potentially allowing for more personalized and task-supportive interactions [13]. The survey seeks to fill knowledge gaps by reviewing a substantial body of NLP studies and systems, particularly in response to crises such as the COVID-19 pandemic, thereby contributing to a more informed understanding of AI's role in emergency information management [5].

Furthermore, the introduction of SynAsk aims to enhance LLM capabilities in organic synthesis through the integration of domain-specific data and advanced chemistry tools [6]. The survey also intends to evaluate models' performance in generating lay language summaries and background explanations, which are crucial for improving the interpretability of biomedical literature [14].

Finally, the survey emphasizes the innovation of tools like Doris, which combine semantic understanding with traditional keyword approaches, facilitating a more insightful exploration of document trends over time [12]. Through these objectives, the survey aims to provide a comprehensive overview of the challenges and opportunities presented by AI technologies in modern research, education, and various professional domains.

1.3 Structure of the Survey

This survey is meticulously structured to offer a comprehensive examination of advanced computational techniques in artificial intelligence, organized into distinct sections reflecting the interdisciplinary nature of the field. The survey begins with an **Introduction**, establishing the significance and relevance of AI techniques across various domains, outlining the survey's objectives, and providing an overview of the paper's structure.

The subsequent section, **Background and Preliminary Concepts**, delves into foundational concepts essential for understanding the survey's scope, including deep research, search agents, AI writing, information retrieval, artificial intelligence, and natural language processing. This section also traces the historical development and evolution of these core technologies, offering a contextual framework for the discussions that follow [15].

Following this, the **Deep Research and Search Agents** section explores the role of deep research in AI and the utilization of search agents for conducting in-depth investigations. This includes an analysis of the integration of deep learning techniques to enhance search agent functionality and a discussion of the challenges and solutions in search agents.

The **AI Writing and Text Generation** section examines advancements in language models and text generation technologies, exploring user interactions with AI writing tools and analyzing the ethical and social implications of AI writing. This section considers the challenges posed by technologies like ChatGPT and the need for updated methods to assess originality in AI-generated content [16].

In the **Information Retrieval Techniques** section, the survey focuses on techniques employed to enhance the retrieval process through AI and machine learning, examining innovative retrieval methods, the integration of machine learning, and domain-specific retrieval applications.

The **Natural Language Processing and Human-Computer Interaction** section discusses the role of natural language processing in understanding and processing human language. This section addresses improvements in voice recognition, modular NLP frameworks for research and development, and the application of NLP in conversational agents and virtual assistants.

The survey further explores **Integration and Synergies** between these technologies, highlighting their interconnections and synergies, particularly in voice and language processing, and the application of AI synergies in clinical and educational settings. The section also considers the Task Supportive and Personalized Human-Large Language Model Interaction (TSPLM) method, which integrates user perceptions and task context into LLM interactions [13].

Finally, the **Conclusion** section synthesizes the key findings and insights from the survey, discussing future directions and potential developments in the field of advanced computational techniques in AI. Through this structured approach, the survey aims to provide a comprehensive roadmap of the current landscape and future potential of AI technologies.The following sections are organized as shown in Figure 1.

2 Background and Preliminary Concepts

2.1 Importance of Core Concepts

The foundational concepts of artificial intelligence (AI) and computational techniques are integral to technological advancement and innovation across multiple domains. Ensuring originality in AI-generated content remains a primary challenge, complicating traditional plagiarism detection and necessitating novel evaluation methods [16]. In natural language processing (NLP), prompt engineering significantly influences language model performance, affecting the coherence and relevance of outputs [17]. Standardized benchmarks, such as the Project Gutenberg corpus, are vital for addressing inconsistent data handling and limited sample sizes, providing a robust foundation for AI research [18]. In multilingual information retrieval, issues like suboptimal data feeding and the need for timely updates to mitigate hallucinations underscore the necessity for accelerated information delivery to improve retrieval accuracy [1].

Traditional textual representation methods, such as term-frequency-based approaches, often fall short in NLP tasks, highlighting the need for more sophisticated representation techniques [19]. The increasing volume of scientific literature challenges conventional systematic review methods, prompting the use of machine learning to enhance efficiency [2]. Existing benchmarks often lack sufficient annotated examples, complicating the development of effective supervised models [3]. This challenge is exacerbated by processing languages without clear word boundaries and effectively indexing diverse sequence collections [11]. In specialized domains like organic chemistry, large language models (LLMs) face difficulties in generative tasks due to insufficient domain-specific tuning and understanding of complex molecular structures [6].

Utilizing domain-specific data can significantly enhance model performance, as shown by methods generating labeled data through large language models [20]. The semi-supervised approach to identifying expert users, leveraging both labeled and unlabeled comments, further illustrates the critical role of core concepts in enhancing AI capabilities [10]. These foundational concepts underscore AI and computational techniques' essential role in addressing complex challenges, fostering innovation, and supporting diverse applications across various fields.

2.2 Historical Development and Evolution

The historical development of artificial intelligence (AI) and computational techniques is marked by significant advancements and challenges that shape contemporary applications. Early AI systems relied on rule-based approaches, utilizing expert knowledge for query formulation, a process that was time-consuming and complex, especially for novices [21]. Dimensionality reduction techniques emerged to streamline data complexity while preserving essential features, though maintaining data integrity during reduction remained a challenge [22].

The transition from traditional rule-based systems to neural network models marked a pivotal shift, enhancing AI's ability to process and retrieve information through end-to-end neural architectures, which improved information retrieval (IR) systems' capacity to meet diverse needs [23]. Retrieval-augmented generation (RAG) techniques further advanced AI's content generation capabilities by integrating retrieval processes with generation models [24]. The evolution of information retrieval models, such as the Information Search Process (ISP), reflects a dynamic adaptation to the digital age, underscoring the evolving nature of IR systems [25]. This evolution is mirrored in the diversification of retrieval methods, including keyword-based, semantic-based, interactive, and metadata-based approaches, accommodating modern information demands [26].

Despite these advancements, challenges persist in dynamically optimizing resource allocation based on workload variations, limiting AI applications' full potential [27]. However, integrating exploratory search with learning theories presents a promising avenue for enhancing AI's educational applications, illustrating the potential for continuous learning and adaptation in the digital era [28].

The historical advancements in AI and NLP technologies, such as ChatGPT, reveal a profound transformation across various fields. This progress emphasizes significant enhancements in academic writing and research efficiency, while also highlighting persistent challenges related to authenticity, credibility, and ethical considerations in AI-generated content. As these technologies evolve, they prompt critical discussions about their implications for human intelligence and critical thinking within the academic process, underscoring the need for careful integration and transparent usage to leverage AI's potential while addressing risks of factual inaccuracies and biases inherent in AI outputs [9, 29, 30].

3 Deep Research and Search Agents

Deep research methods are integral to the progress of artificial intelligence (AI), particularly in refining information retrieval and text generation. This section explores the essential role of deep research in shaping AI technologies and enhancing search agents' performance, with a focus on developing sophisticated algorithms and innovative approaches that propel intelligent systems forward.

Figure 2 illustrates the hierarchical structure of deep research and search agents, highlighting the role of deep research in advancing AI capabilities. The figure categorizes the advancements, applications, technological enhancements, and innovative approaches in AI and search agents, while also outlining the challenges and proposed solutions to improve search agents' effectiveness and accuracy. This visual representation underscores the integration of deep learning in search agents and the complexities involved, thereby enriching our understanding of the current landscape in AI research.

3.1 Role of Deep Research in AI

Deep research methods significantly advance AI capabilities, particularly in enhancing the precision and efficiency of information retrieval and text generation. Efficient algorithms and data structures are vital for improving document retrieval systems, especially for non-traditional sequences, enabling AI systems to manage complex datasets and enhance search effectiveness [11, 31]. The integration of retrieval-augmented generation (RAG) techniques, which blend retrieval processes with generation models, has notably improved information retrieval by providing additional context and background information [14, 20].

In pandemic-specific applications, deep research has addressed urgent information needs by categorizing research into traditional NLP tasks and pandemic-specific applications, demonstrating NLP's significance in crisis management [5]. These methods have also facilitated novel categorization systems on QA platforms, enhancing user engagement and task completion by effectively identifying expert, non-expert, and out-of-scope comments [10].

The evolution of deep research is evident in personalized human-AI interactions, exemplified by the Task Supportive and Personalized Human-Large Language Model Interaction (TSPLM) method, which enhances user engagement by addressing specific needs and contexts [13]. Furthermore, the need for models capturing semantic nuances in discourse underscores deep research's importance in enabling comprehensive analysis beyond keyword searches [12]. These methods ensure AI technologies evolve innovatively and ethically, addressing technical challenges while expanding AI's frontiers.

3.2 Integration of Deep Learning in Search Agents

The integration of deep learning techniques into search agents has significantly enhanced their functionality through sophisticated data processing, improved user interaction, and more accurate

AI-generated, for reference only.

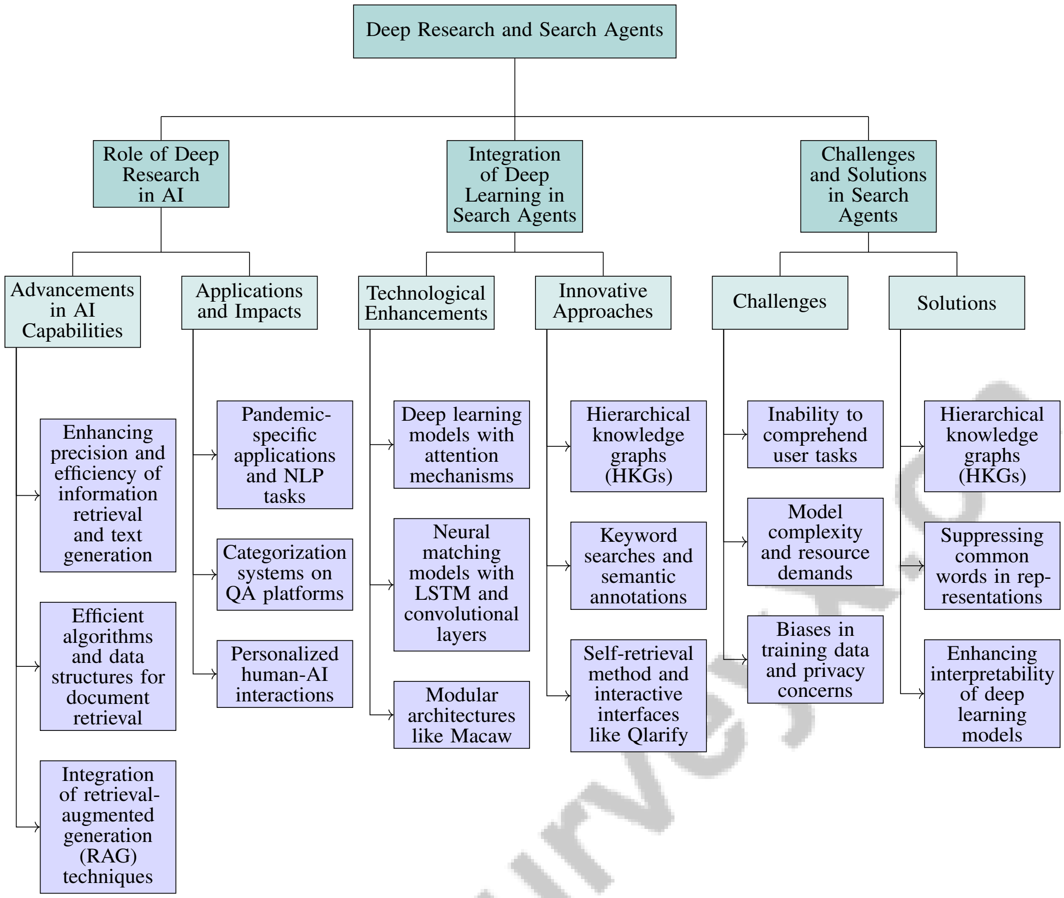

Figure 2: This figure illustrates the hierarchical structure of deep research and search agents, highlighting the role of deep research in advancing AI capabilities, the integration of deep learning in search agents, and the challenges and solutions faced by search agents. It categorizes the advancements, applications, technological enhancements, and innovative approaches in AI and search agents, while also outlining the challenges and proposed solutions to improve search agents' effectiveness and accuracy.

information retrieval. Deep learning models with attention mechanisms have transformed search agents by focusing on the most relevant input data parts, crucial in machine translation and information retrieval [32]. Neural matching models incorporating long short-term memory (LSTM) networks and convolutional layers exemplify deep learning's impact on search capabilities, facilitating more accurate and context-aware retrieval [33].

Systems like Macaw, with their modular architecture, support the development of conversational information-seeking systems, allowing the implementation of deep learning techniques to enhance search agent functionality [34]. This modularity enables integrating various components and techniques, such as hierarchical knowledge graphs (HKGs), which combine low-level entity relationships with high-level central concepts to facilitate exploratory search [35].

Innovative approaches like Doris leverage keyword searches and semantic annotations to provide deeper insights into document corpora, enhancing search agents' ability to explore social phenomena [12]. Additionally, methods such as KG-RAG, which construct knowledge graphs from unstructured text and employ the Chain of Explorations (CoE) retrieval algorithm, enhance the accuracy and relevance of information retrieved by search agents [36]. The self-retrieval method, integrating indexing, retrieval, and reranking functionalities into a single large language model (LLM), streamlines the information retrieval process [37]. Interfaces like Qlarify enable users to interactively expand abstracts with clarifying information from full texts, enhancing engagement and comprehension

[38]. Additionally, a two-stage approach in machine translation reduces the need for full document translation, lowering computational costs while improving retrieval accuracy [39].

3.3 Challenges and Solutions in Search Agents

Search agents face numerous challenges that hinder their effectiveness and accuracy. A primary issue is the inability of current systems to fully comprehend user tasks, leading to ineffective outcomes, compounded by difficulties in task representation, user engagement, and the complexity of user needs [40]. Integrating heterogeneous features from users and items into post-ranking models presents another challenge, as large language models (LLMs) are primarily designed for semantic text input and struggle with non-semantic feature embeddings [41].

Model complexity and resource demands are significant obstacles, particularly in domain-driven applications. The phenomenon of hallucination in LLMs complicates keyword extraction, impacting search result accuracy and reliability [42]. Systems like LPar also face limitations in managing multiple agents and ensuring effective communication and context retention [43]. Furthermore, managing noise in retrieval results, increased system complexity, and overhead costs associated with retrieval processes pose additional challenges [24].

Biases in training data, limited understanding of concepts by AI, and privacy concerns further exacerbate the challenges faced by search agents [44]. Existing benchmarks often fail to measure systems' ability to personalize and ground responses accurately, leading to issues with hallucinations and biases in generated content [45].

To address these challenges, several solutions have been proposed. Hierarchical knowledge graphs (HKGs) preserve the advantages of both knowledge graphs and hierarchical representations, reducing the need for extensive document reading while providing effective navigation support [35]. Suppressing common words in representations is crucial for improving retrieval effectiveness, as it helps focus on more relevant and distinctive terms [31]. Enhancing the interpretability of deep learning models and optimizing hyperparameter tuning are essential for improving recommender systems' performance [46]. By addressing these challenges through innovative solutions, search agents can become more effective, user-centric, and capable of delivering accurate and relevant results across diverse applications.

4 AI Writing and Text Generation

4.1 Advancements in Language Models and Text Generation

Recent advancements in large language models (LLMs) have significantly transformed text generation, enhancing both efficiency and contextual relevance. A notable development is the establishment of benchmarks for evaluating LLM performance in text summarization, using diverse metrics for comprehensive assessment [47]. These benchmarks highlight the increasing sophistication of LLMs in producing concise and coherent summaries, a critical aspect of text generation.

Innovations such as integrating user perceptions and task context into interaction processes, as proposed by Wang et al., represent a shift towards personalized AI systems, tailoring interactions to users' specific needs and improving the relevance of generated text [13]. This personalization is crucial for applications ranging from customer service to educational tools.

In specialized domains like organic synthesis, the SynAsk model illustrates the effectiveness of combining fine-tuning with external resources, demonstrating the adaptability of modern language models for complex tasks [6]. Incorporating domain-specific knowledge into language models is a pivotal advancement, enabling more accurate and contextually aware text generation.

Sophisticated AI systems like ChatGPT signify a transformative shift in text generation technologies, offering substantial benefits for academic writing efficiency while raising ethical concerns about authenticity, bias, and credibility [29, 48]. These advancements enhance AI's ability to generate relevant, coherent, and contextually appropriate text, paving the way for sophisticated applications across diverse domains.

4.2 AI Writing Tools and User Interaction

The interaction between users and AI writing tools is increasingly characterized by advanced AI technologies that promote personalization and enhance engagement. A significant challenge is the lack of established design frameworks for developing user-friendly AI writing interfaces, as noted by Buschek et al. [49]. Scraft addresses this by providing recursive feedback through Socratic questioning, fostering deeper engagement and reflection [50].

AI writing tools are adopting frameworks that categorize interactions into push and pull paradigms, influencing user satisfaction and dynamics [30]. Tools like GhostWriter and Effidit allow users to personalize their writing experience through editable style and context profiles, offering comprehensive support for diverse tasks.

Additionally, models like SS-LSTM illustrate AI's capability to process user inputs, enhancing emotion classification and interaction [51]. This functionality is crucial for developing responsive and context-aware writing tools that adapt to the emotional tone of user inputs.

Datasets tailored to specific language learners, such as the Gazelle instruction dataset for Arabic, emphasize the need for culturally relevant AI tools [52]. These datasets provide targeted writing advice, enhancing the educational value of AI writing tools.

AI writing tools are evolving to offer more personalized, interactive, and transparent user experiences. By leveraging advanced AI capabilities and addressing design challenges, these tools significantly enhance the writing process across various contexts, promoting a collaborative and user-centric approach to text generation. Users report enjoying the collaborative experience without losing ownership, although concerns regarding bias and AI feedback accuracy remain critical areas for future development [50, 53, 30].

4.3 Ethical and Social Implications

The ethical and social implications of AI writing technologies encompass concerns including bias, authorship, transparency, and environmental impact. A primary ethical challenge is the potential for bias in AI-generated text, which current studies often inadequately address, necessitating ongoing research into ethical AI practices [17]. Such biases can significantly influence user outputs, especially in educational contexts [7].

Integrating AI in educational settings raises concerns about academic integrity. Establishing clear guidelines for AI use is vital to ensure ethical employment without undermining independent writing skills [8]. Reliance on AI for content generation poses risks related to the accuracy and reliability of outputs, as studies often fall short in addressing differences in depth and factual accuracy between AI-generated and human-written texts [9].

Transparency in AI-generated content is essential for fostering trust and understanding among users. Explainable AI systems can enhance user confidence by enabling critical engagement with AI-generated suggestions, ensuring reliability and comprehensibility [7]. However, the risks of plagiarism and inaccuracies necessitate a balanced approach, where human oversight remains integral to content creation to uphold quality and originality.

The environmental impact of AI writing technologies is another critical consideration. While AI can potentially reduce carbon emissions compared to traditional human efforts, its widespread use raises questions about long-term sustainability. Addressing these concerns is essential to ensure that AI benefits do not come at the expense of ecological health [7].

4.4 AI Writing in Creative and Academic Contexts

The integration of AI writing tools in both creative and academic domains has initiated a paradigm shift in content creation, transforming how users generate, interact with, and refine their work. In creative writing, the Controlled Text Generation System (CTGS) exemplifies how AI can enhance control and creativity, enabling users to produce coherent outputs under various constraints [62]. This capability is crucial for fostering creativity while maintaining user agency, as demonstrated by AI systems like Effidit, which significantly improve writing quality and efficiency [63].

Benchmark	Size	Domain	Task Format	Metric
COVID-IR[54]	1,000,000	Biomedical Information Retrieval	Information Retrieval	bpref, P@5
Sy-SE-PQA[55]	100,000	Community Question Answering	Question Answering	P@1, NDCG@10
LAP[56]	206,339	Cognitive Science	Similarity Judgment	Pearson correlation
Ericson[57]	230,000	Conversational AI	Dialogue Management	User Ratings, Engagement
UTS/UTSB[58]	12,000	Information Retrieval	Relevance Estimation	RBP
NLP-MR[59]	1,000,000	Multilingual Information Retrieval	Cross-Language Information Retrieval	Precision, Recall
LAMP[60]	1,057	Literary Fiction	Text Editing	Initial Writing Quality Score, Final Writing Quality Score
CSN[61]	6,000,000	Software Engineering	Code Retrieval	NDCG

Table 1: This table presents a comprehensive overview of various benchmarks utilized in different domains, highlighting their size, domain of application, task format, and evaluation metrics. The benchmarks range from biomedical information retrieval to software engineering, illustrating the diverse applications of AI writing tools. Each benchmark is associated with specific metrics, providing a quantitative basis for performance assessment.

In academic contexts, AI writing tools enhance outcomes for non-native English speakers (NNESs) by providing accessible support, thereby contributing to both creative and academic endeavors [64]. The concept of collage, discussed by Buschek et al., offers a valuable framework for understanding and designing AI writing tools, emphasizing fragmentation and user interaction in modern writing practices [49]. This approach highlights AI's potential to facilitate diverse writing styles and genres, enriching users' creative and academic experiences.

Tools like Scraft, which serve as thought-provoking writing tutors, demonstrate AI's potential to engage users in deeper reflection and revision processes. However, improvements are needed to enhance the relevance and accuracy of the Socratic questions generated by such tools [50]. In scientific writing, AI tools like ChatGPT can augment the writing process, but they should not supplant human judgment and expertise, underscoring the need for regulatory frameworks to ensure ethical use [65].

Table 1 provides a detailed overview of representative benchmarks used to evaluate AI writing tools across multiple domains, illustrating the breadth of applications and the metrics used for performance assessment. The future of AI-assisted writing lies in refining user experiences and exploring various writing contexts, as suggested by Pereira [30]. This involves conducting independent evaluations of AI-assisted writing outputs to ensure that these tools effectively complement human creativity and academic integrity. By focusing on ethical guidelines, human oversight, and sustainable practices, AI writing tools can continue to support innovative applications across diverse fields, balancing the benefits of automation with the nuances of human creativity and critical thinking.

5 Information Retrieval Techniques

5.1 Innovative Retrieval Methods

Innovative retrieval methods have significantly advanced information retrieval (IR) by improving data processing efficiency and accuracy. The integration of neural index and hybrid search schemes, which combine symbolic and neural indices, exemplifies advancements that enhance recall during initial retrieval stages [66]. The Self-Retrieval method, which unifies indexing, retrieval, and reranking with large language models, streamlines the retrieval process and boosts overall efficiency [20]. Similarly, the RegulatoryBERT model's domain-specific capabilities achieve state-of-the-art results in text classification and named entity recognition, enhancing retrieval accuracy [67].

In specialized domains, methods like SynAsk leverage domain-specific knowledge and external tools to enhance contextual understanding and problem-solving capabilities [6]. The Expert Identification Method (EIM) improves expert identification precision by categorizing user contributions [10]. Tools such as Doris facilitate interactive exploration, allowing users to search through documents by specific words or topics while visualizing result distributions, thus enhancing user engagement [12].

The Contrastive Pre-training a Discriminative AutoEncoders (CPDAE) method enhances representation discrimination by generating word distributions from input texts and applying a contrastive loss, refining model performance and ensuring relevant information retrieval [31]. Established datasets like MS MARCO and LoCoV0 provide a foundation for benchmarking retrieval methods [20].

These innovative methods illustrate the ongoing evolution of IR, driven by advanced computational techniques that enhance efficiency and accuracy. By integrating interactive, probabilistic, and neural methodologies, these IR systems address traditional limitations, enabling richer user interactions through natural language and multimodal inputs while leveraging deep learning to improve query and document representations. This evolution paves the way for sophisticated solutions that adapt to diverse information needs [22, 68, 23, 69].

5.2 Integration of Machine Learning and AI

Integrating machine learning (ML) into information retrieval (IR) systems has significantly enhanced precision, efficiency, and adaptability to complex data environments. Large language models (LLMs) redefine traditional IR processes through advanced query rewriting, retrieval, reranking, and reading capabilities, facilitating more nuanced and context-aware data processing [70].

The Mamba Retriever, employing a bi-encoder architecture and a selective state space model, achieves linear time scaling, setting a new standard for retrieval efficiency and accuracy [71]. Retrieval-augmented generation (RAG) methods, which adapt responses based on linguistic and contextual factors, exemplify the synergy between ML and IR, improving accuracy and relevance [1].

ML's role in IR is further underscored by novel evaluation frameworks tailored for conversational search, distinguishing these approaches from traditional ad-hoc retrieval benchmarks [66]. The application of contrastive pre-training techniques, such as CPDAE, enhances representation discrimination by focusing on salient words while suppressing common ones, refining precision in IR systems [31].

In data augmentation, models fine-tuned on synthetic data exhibit superior performance compared to strong baselines, demonstrating synthetic data's potential to enhance IR systems' adaptability and robustness [20].

Advancements in ML underscore its pivotal role in revolutionizing IR systems, leading to more sophisticated, efficient, and user-centric solutions. These innovations address the intricate challenges faced by users in diverse information environments, particularly in managing complex search tasks beyond basic fact-finding. By leveraging deep learning and large pretrained transformer models, researchers enhance IR systems' capabilities to understand user intent, personalize search experiences, and improve task completion outcomes. This evolution highlights ML's transformative potential in IR and points toward a future where intelligent systems effectively cater to the dynamic needs of users navigating vast information landscapes [40, 68, 23, 20].

5.3 Applications and Domain-Specific Retrieval

Advanced information retrieval (IR) techniques in domain-specific contexts enhance query result precision and relevance, tailoring retrieval systems to meet various fields' unique demands. In the biomedical domain, integrating semantic knowledge and advanced neural architectures has proven effective. Methods incorporating large language models (LLMs) into IR systems demonstrate superior performance in understanding and generating relevant content compared to traditional approaches, as evidenced by benchmarks like SciFact+AIGC and NQ320K+AIGC, which include both human-written and LLM-generated documents with relevancy labels [72].

In healthcare, models enhancing retrieval precision and relevance are critical for ensuring healthcare professionals efficiently access pertinent data, improving patient care and decision-making. The incorporation of structured information, as exemplified by frameworks like WikiFormer, enhances retrieval performance across multiple benchmark datasets, including MS MARCO and TREC DL 2019, using metrics such as MRR and nDCG [73].

Probabilistic models in security and intelligence domains enhance domain-specific information retrieval, particularly for critical topics like terrorism, underscoring probabilistic approaches' importance. In scientific research, addressing the vocabulary gap is essential, as demonstrated by methods emphasizing semantic understanding of queries to improve recall and relevance [74].

Integrating large language models into IR systems has been transformative, particularly where LLMs excel in content understanding and generation. This is evident in the Mamba Retriever, which performs comparably or better than Transformer models on both short and long-text retrieval tasks, highlighting innovative architectural designs' efficacy in enhancing retrieval outcomes [71]. Furthermore, using LLMs for generating training data enhances retrieval model performance, as demonstrated by benchmarks that efficiently utilize these models in data augmentation [20].

6 Natural Language Processing and Human-Computer Interaction

The convergence of Natural Language Processing (NLP) and human interaction is increasingly pivotal as technology advances. This section delves into key NLP areas, highlighting both advancements and challenges, with a focus on enhancing voice recognition systems and addressing biases affecting diverse user demographics.

6.1 Improving Voice Recognition and Addressing Bias

Enhancing voice recognition systems while addressing biases in NLP is a critical challenge requiring comprehensive strategies. Selection bias, due to the underrepresentation of women and minorities in training datasets, leads to performance disparities across demographic groups, necessitating inclusive datasets for equitable system accuracy [75]. Voice recognition improvement also depends on user input quality, as misunderstandings can arise from poor input [76]. Integrating memory into conversational systems is vital for retaining user context, significantly enhancing response accuracy and relevance [77, 78].

The complexity of multi-user interactions in collaborative search scenarios presents additional challenges. Solutions like CoSearchAgent enhance user experience by facilitating efficient voice and text-based interactions, emphasizing the need for systems capable of navigating collaborative environments with diverse user inputs [79]. Furthermore, understanding and interpreting emotions from text without visual or auditory cues remains challenging due to context understanding, sarcasm, and natural language ambiguity [51]. These issues underscore the necessity for advanced NLP techniques to accurately capture human language nuances and address biases from imbalanced class sizes or ambiguous constructs.

6.2 Modular NLP Frameworks for Research and Development

Modular NLP frameworks are crucial for advancing research and development, offering flexible architectures that accommodate diverse linguistic tasks. These frameworks facilitate integrating various NLP components, allowing researchers to experiment with different configurations and optimize performance across multiple domains [77]. They often incorporate tokenization, parsing, and semantic analysis modules, enabling seamless adaptation to specific research needs.

A notable advancement in these frameworks is the integration of memory mechanisms, enhancing user context retention and improving interaction quality [78]. This capability is particularly beneficial in conversational AI, where maintaining context over extended interactions is crucial for coherent responses. Moreover, modular frameworks support advanced neural architectures, such as attention mechanisms and transformer models, revolutionizing NLP by enabling more accurate processing of large-scale text data [75].

The modularity of NLP frameworks also allows for incorporating domain-specific knowledge, enhancing performance in specialized fields like healthcare, finance, and legal research. By integrating domain-specific modules, researchers can tailor NLP systems to meet unique requirements, improving output relevance and accuracy [79].

6.3 NLP in Conversational Agents and Virtual Assistants

NLP is fundamental to developing conversational agents and virtual assistants, enabling effective understanding, processing, and response to human language. Advanced NLP techniques facilitate natural human-computer interactions, essential for personal and professional applications [75]. A core application of NLP in these systems is maintaining context over extended interactions through sophisticated memory mechanisms, allowing virtual assistants to retain user preferences and previous interactions, enhancing personalization and response relevance [78]. The use of attention mechanisms and transformer-based models further advances conversational agents' capabilities, enabling improved handling of complex queries and accurate language understanding [77].

Moreover, NLP techniques are crucial for addressing emotion detection and sentiment analysis challenges, vital for creating empathetic conversational agents. By leveraging sentiment and semantic embeddings, these systems can better interpret the emotional tone of user inputs, enhancing interaction quality and user satisfaction [51]. The development of modular NLP frameworks has accelerated advancements in conversational agents by providing flexible architectures supporting diverse NLP components, allowing for the customization of virtual assistants to meet varied user needs and preferences [79].

6.4 Ethical and Practical Implications of NLP in AI

The ethical and practical implications of NLP in AI encompass considerations such as bias, transparency, interpretability, and societal impact. Algorithmic bias in NLP systems can lead to inequitable outcomes across demographic groups, particularly in legal applications where biased algorithms may influence decision-making processes, highlighting the necessity for human oversight to ensure fairness [15]. The theoretical properties of conversational information retrieval systems further emphasize the importance of ethical considerations for effective performance across tasks [77].

Transparency and interpretability are vital for building trust in AI systems. The black-box nature of many deep learning models complicates understanding decision-making processes, crucial for fostering user confidence [80]. The MatchZoo framework, for example, simplifies the learning and application of neural text matching models, addressing interpretability concerns while advancing NLP [80].

In education, the ethical implications of AI-generated content relate to academic integrity and the potential influence on educational outcomes. The integration of NLP technologies necessitates ongoing evaluation of their effectiveness and ethical use, as highlighted by Chan [7], with significant implications for academic integrity requiring careful consideration of AI-generated content in educational settings [16].

Practically, NLP technologies provide substantial benefits in applications such as systematic reviews and legal document retrieval. The integration of machine learning tools in systematic reviews raises questions about usability and the effectiveness of explainability techniques, crucial for ensuring practical and ethical application [2]. In legal contexts, the relevance of factual information correlates positively with retrieval effectiveness, underscoring the importance of accuracy in NLP applications [81].

The development of domain-specific language models, such as those for the architecture, engineering, and construction (AEC) domain, illustrates the practical utility of NLP in specialized fields. These models provide tailored evaluation frameworks that address unique challenges, facilitating better model training and application [67].

7 Integration and Synergies

7.1 Interconnections and Synergies

The integration of various AI technologies significantly enhances system capabilities across diverse domains. A notable instance is CodeMatcher, which combines information retrieval (IR) and deep learning to improve code search accuracy and response times, showcasing the complementary strengths of these approaches [82]. Similarly, conversational agents benefit from frameworks merging conversational retrieval systems with traditional IR techniques, improving user interaction and retrieval performance [77]. This synergy provides intuitive access to information, enhancing user experience by integrating conversational and traditional search technologies.

Advancements in semantic encoding and fusion techniques have bolstered multimodal data processing capabilities [83]. Frameworks accommodating multimodal queries expand AI systems' scope, enabling comprehensive data processing and analysis. The integration of factuality and objectivity detection into IR systems systematically enhances document retrieval effectiveness, crucial in high-accuracy domains like legal and academic research [81].

Ma's research categorizes AI-generated and human-written texts into syntax, semantics, and pragmatics, facilitating a nuanced understanding of AI-generated content [9]. This framework aids in distinguishing AI-generated texts from human-authored ones, enhancing interpretability and credibility. In education, Artiana's framework categorizes students' perceptions of AI tools like ChatGPT into advantages, challenges, and impacts on writing, highlighting AI's potential benefits and challenges in education [8].

Integrating advanced natural language generation models and retrieval-augmented generation techniques is driving innovation in AI-generated content. These synergies enhance content creation quality and efficiency across domains, addressing challenges like factual accuracy and bias [24, 9, 30, 29]. Leveraging these interconnections, AI systems can achieve greater efficiency, accuracy, and ethical alignment, enhancing their applicability and impact in real-world scenarios.

7.2 Synergies in Voice and Language Processing

Voice and language processing technologies have advanced significantly, enhancing human-computer interaction systems. In voice recognition, advanced NLP techniques improve accuracy and contextual understanding, addressing selection bias by including diverse demographic groups for equitable performance [75]. Memory mechanisms in NLP frameworks are crucial for maintaining context, improving voice processing systems' ability to generate contextually relevant responses, essential for applications like virtual assistants [78].

Semantic encoding techniques further enhance multimodal data processing, enabling systems to handle complex queries and deliver nuanced responses [83]. Frameworks integrating voice and language processing technologies underscore the potential for creating sophisticated AI systems, incorporating AI components like attention mechanisms and transformer models to enhance large-scale text data processing [75].

Advancements in voice and language processing technologies are critical for AI, enhancing solution effectiveness and user-centricity while addressing challenges like bias and authenticity of generated content. Recent developments have led to sophisticated AI models that improve communication and interaction, raising ethical considerations regarding diverse voices and the credibility of AI-generated text in academic and professional contexts. By leveraging these synergies, AI systems can achieve greater accuracy, contextual understanding, and adaptability, enhancing their applicability and impact across various domains [30, 29, 9, 75].

7.3 Frameworks and Synergies in AI Task Execution

Integrated frameworks are crucial for efficient and adaptable AI task execution. These frameworks provide the structural foundation for integrating diverse AI components, facilitating coherent and efficient task execution. Modular architectures supporting advanced neural models and attention mechanisms are essential for processing complex data and improving task outcomes [77].

Frameworks combining NLP and IR techniques exemplify the potential for creating sophisticated AI systems, enhancing information retrieval accuracy and relevance in applications requiring nuanced understanding and contextual analysis [81]. Incorporating memory mechanisms into AI frameworks enhances the system's ability to maintain context over extended interactions, crucial for personalized and contextually relevant responses in conversational AI systems [78].

Integrated AI frameworks also support multimodal data fusion, enabling effective processing and analysis of diverse data types. This capability is vital for applications requiring integration of voice, text, and visual data, allowing comprehensive and accurate task execution [83]. Leveraging these synergies, AI frameworks enhance systems' ability to handle complex queries and deliver nuanced, informed responses.

7.4 Synergies in Clinical and Educational AI Applications

AI synergies in clinical and educational settings have the potential to revolutionize both fields by enhancing service efficiency, accuracy, and personalization. In clinical environments, AI technologies improve diagnostic accuracy and patient care by analyzing large volumes of medical data to identify crucial patterns for early diagnosis and treatment planning. Integrating NLP with electronic health records facilitates extracting relevant patient information, improving clinical decision-making and patient management [15].

AI-driven tools automate routine tasks like data retrieval and information processing, allowing healthcare professionals to focus on complex patient care activities. These tools leverage advanced technologies like Large Language Models (LLMs) to enhance medical information searches' efficiency and accuracy, fostering a more informed healthcare environment. AI technologies in telemedicine enhance remote patient monitoring and consultation, providing access to healthcare services regardless of geographical barriers, benefiting resource-limited settings [9, 84, 25, 29, 30].

In education, AI synergies transform learning experiences by offering personalized learning pathways and improving educational outcomes. AI-powered adaptive learning systems analyze student performance data to tailor educational content to individual learning needs, optimizing the learning process. These systems identify struggling areas and provide targeted interventions, facilitating a more personalized and effective learning experience [8].

AI technologies enhance academic integrity by detecting plagiarism and ensuring originality in student submissions. The development of AI tools assisting non-native English speakers in improving writing skills demonstrates AI's potential to support diverse learning needs and promote inclusivity in education [8]. AI in educational settings also extends to administrative tasks, streamlining processes like enrollment, scheduling, and resource allocation, increasing institutional efficiency.

AI synergies in clinical and educational applications highlight the transformative potential of AI in enhancing service delivery, personalization, and operational efficiency. By harnessing advanced information retrieval techniques and large language models, both biomedical research and educational initiatives can achieve substantial progress. This collaboration enhances the clarity and accessibility of complex biomedical literature for the public and optimizes information search for healthcare professionals, improving health outcomes and informed decision-making [25, 14, 84].

8 Conclusion

8.1 Future Directions and Innovations

The trajectory of artificial intelligence (AI) technologies is set to redefine their capabilities and applications across multiple domains. A primary focus for future research will be enhancing document retrieval techniques to improve scalability and address computational challenges associated with large-scale sequence data. This aligns with the broader aim of enhancing AI interpretability and addressing algorithmic biases, which are crucial for fostering trust and accountability in AI systems.

Expanding user studies to include diverse populations and refining interfaces to improve interactions with large language models (LLMs) are critical objectives. Enhancements in these areas should focus on the model's ability to interpret complex data structures and integrate advanced natural language processing techniques. Additionally, innovative fusion methods, such as dynamic and adaptive fusion, will be essential for managing multimodal data complexity.

The integration of AI in scientific writing and education is another promising research avenue. Future initiatives will aim to expand model capabilities through additional data integration and explore autonomous functionalities in laboratory settings. Furthermore, research should focus on effective strategies for incorporating AI tools in educational environments while promoting independent writing skills among students.

Efforts to enhance the quality of background explanations, explore additional knowledge resources, and develop tailored evaluation metrics for lay language generation will be pivotal. This involves expanding datasets, investigating transfer learning techniques, and validating benchmarks on diverse datasets to ensure the robustness and generalizability of AI models.

In multilingual information retrieval, future research could explore data augmentation techniques and the application of contrastive pre-training methods to other retrieval scenarios. Additionally, developing larger, annotated datasets for specific tasks and improving the interoperability of NLP tools will be essential. Addressing these considerations will facilitate the responsible evolution of AI technologies, supporting ethical innovation and practical applications across diverse fields.

8.2 Technological Innovations and Future Directions

The future of artificial intelligence (AI) is poised for significant transformation, driven by ongoing technological innovations and a deeper understanding of foundational theories. Insights into the historical context of deep learning models provide critical perspectives on their evolution, highlighting the foundational theories that shape their development. This understanding is vital for guiding future advancements in AI technologies, enabling the creation of more sophisticated and efficient models.

A key area of innovation lies in the integration of AI systems with environmental sustainability initiatives. While AI systems have environmental costs, they generally exhibit a lower carbon footprint compared to human counterparts performing similar tasks. This suggests that AI can significantly contribute to reducing overall emissions, aligning technological advancements with global sustainability goals. Developing AI technologies that prioritize energy efficiency and environmental impact will be crucial in shaping the future landscape of AI.

Advancements in large language models (LLMs) and their applications across various domains will remain a focal point of future research. Expanding user studies to include diverse populations and refining interfaces for enhanced user interaction will be critical in maximizing LLM utility across different applications. Moreover, exploring dynamic and adaptive fusion methods will enhance multimodal data processing capabilities, further broadening AI application scopes.

In scientific research and education, AI's role is expected to expand, with future enhancements focusing on integrating additional data and exploring autonomous functionalities in laboratory settings. The potential of AI to revolutionize educational practices while fostering independent learning and writing skills will be a key exploration area. This includes developing tailored evaluation metrics and expanding datasets to support more robust and generalizable AI models.

Technological innovations that emphasize sustainability, efficiency, and inclusivity will shape the future of AI. By leveraging historical insights and addressing environmental considerations, AI technologies can evolve responsibly, supporting ethical innovation and practical applications across diverse fields.

References

[1] Syed Rameel Ahmad. Enhancing multilingual information retrieval in mixed human resources environments: A rag model implementation for multicultural enterprise, 2024.

[2] Iman Munire Bilal, Zheng Fang, Miguel Arana-Catania, Felix-Anselm van Lier, Juliana Outes Velarde, Harry Bregazzi, Eleanor Carter, Mara Airoldi, and Rob Procter. Machine learning information retrieval and summarisation to support systematic review on outcomes based contracting, 2024.

[3] Baban Gain, Dibyanayan Bandyopadhyay, Arkadipta De, Tanik Saikh, and Asif Ekbal. Iitp at aila 2019: System report for artificial intelligence for legal assistance shared task, 2021.

[4] Huang Yuchao, Wei Moshi, Wang Song, Wang Junjie, and Wang Qing. Yet another combination of ir- and neural-based comment generation, 2021.

[5] Qingyu Chen, Robert Leaman, Alexis Allot, Ling Luo, Chih-Hsuan Wei, Shankai Yan, and Zhiyong Lu. Artificial intelligence (ai) in action: Addressing the covid-19 pandemic with natural language processing (nlp), 2021.

[6] Chonghuan Zhang, Qianghua Lin, Biwei Zhu, Haopeng Yang, Xiao Lian, Hao Deng, Jiajun Zheng, and Kuangbiao Liao. Synask: Unleashing the power of large language models in organic synthesis, 2024.

[7] 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.

[8] Nisa Artiana and Ria Fakhrurriana. Efl undergraduate students'perspective on using ai-based chatgpt in academic writing. *Language and Education Journal*, 9(1):1–11, 2024.

[9] Yongqiang Ma, Jiawei Liu, Fan Yi, Qikai Cheng, Yong Huang, Wei Lu, and Xiaozhong Liu. Ai vs. human – differentiation analysis of scientific content generation, 2023.

[10] Sofia Strukova, José A. Ruipérez-Valiente, and Félix Gómez Mármol. Identifying experts in question answer portals: A case study on data science competencies in reddit, 2022.

[11] Gonzalo Navarro. Spaces, trees and colors: The algorithmic landscape of document retrieval on sequences, 2013.

[12] Sreya Guha. Doris: A tool for interactive exploration of historic corpora (extended version), 2017.

[13] Ben Wang, Jiqun Liu, Jamshed Karimnazarov, and Nicolas Thompson. Task supportive and personalized human-large language model interaction: A user study, 2024.

[14] Yue Guo, Wei Qiu, Gondy Leroy, Sheng Wang, and Trevor Cohen. Retrieval augmentation of large language models for lay language generation, 2024.

[15] Harry Surden. Artificial intelligence and law: An overview. *Ga. St. UL Rev.*, 35:1305, 2018.

[16] Mohammad Khalil and Erkan Er. Will chatgpt g et you caught? rethinking of plagiarism detection. In *International conference on human-computer interaction*, pages 475–487. Springer, 2023.

[17] Golam Md Muktadir. A brief history of prompt: Leveraging language models. (through advanced prompting), 2023.

[18] Martin Gerlach and Francesc Font-Clos. A standardized project gutenberg corpus for statistical analysis of natural language and quantitative linguistics, 2018.

[19] Mohamed Morchid, Juan-Manuel Torres-Moreno, Richard Dufour, Javier Ramírez-Rodríguez, and Georges Linarès. Automatic text summarization approaches to speed up topic model learning process, 2017.

[20] Luiz Bonifacio, Hugo Abonizio, Marzieh Fadaee, and Rodrigo Nogueira. Inpars: Data augmentation for information retrieval using large language models, 2022.

[21] Hemanth Kandula, Damianos Karakos, Haoling Qiu, Benjamin Rozonoyer, Ian Soboroff, Lee Tarlin, and Bonan Min. Querybuilder: Human-in-the-loop query development for information retrieval, 2024.

[22] Benjamin Piwowarski, Ingo Frommholz, Mounia Lalmas, and Keith van Rijsbergen. Exploring a multidimensional representation of documents and queries (extended version), 2010.

[23] Ye Zhang, Md Mustafizur Rahman, Alex Braylan, Brandon Dang, Heng-Lu Chang, Henna Kim, Quinten McNamara, Aaron Angert, Edward Banner, Vivek Khetan, Tyler McDonnell, An Thanh Nguyen, Dan Xu, Byron C. Wallace, and Matthew Lease. Neural information retrieval: A literature review, 2017.

[24] Penghao Zhao, Hailin Zhang, Qinhan Yu, Zhengren Wang, Yunteng Geng, Fangcheng Fu, Ling Yang, Wentao Zhang, Jie Jiang, and Bin Cui. Retrieval-augmented generation for ai-generated content: A survey, 2024.

[25] Forhan Bin Emdad and Mohammad Ishtiaque Rahman. Potential renovation of information search process with the power of large language model for healthcare, 2024.

[26] Walid Shalaby and Wlodek Zadrozny. Patent retrieval: A literature review, 2018.

[27] Carlos M. Lorenzetti and Ana G. Maguitman. Learning better context characterizations: An intelligent information retrieval approach, 2010.

[28] Yiming Luo, Patrick Cheong-Iao Pang, and Shanton Chang. Enhancing exploratory learning through exploratory search with the emergence of large language models, 2025.

[29] Ismail Dergaa, Karim Chamari, Piotr Zmijewski, and Helmi Ben Saad. From human writing to artificial intelligence generated text: examining the prospects and potential threats of chatgpt in academic writing. *Biology of sport*, 40(2):615–622, 2023.

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

[31] Xinyu Ma, Ruqing Zhang, Jiafeng Guo, Yixing Fan, and Xueqi Cheng. A contrastive pre-training approach to learn discriminative autoencoder for dense retrieval, 2022.

[32] Jian-wei LIU, Jun-wen LIU, and Xiong-lin LUO. Research progress in attention mechanism in deep learning. *Chinese Journal of Engineering*, 43(11):1499–1511, 2021.

[33] Liu Yang, Hamed Zamani, Yongfeng Zhang, Jiafeng Guo, and W. Bruce Croft. Neural matching models for question retrieval and next question prediction in conversation, 2017.

[34] Hamed Zamani and Nick Craswell. Macaw: An extensible conversational information seeking platform, 2019.

[35] Bahareh Sarrafzadeh, Adam Roegiest, and Edward Lank. Hierarchical knowledge graphs: A novel information representation for exploratory search tasks, 2020.

[36] Diego Sanmartin. Kg-rag: Bridging the gap between knowledge and creativity, 2024.

[37] Qiaoyu Tang, Jiawei Chen, Zhuoqun Li, Bowen Yu, Yaojie Lu, Cheng Fu, Haiyang Yu, Hongyu Lin, Fei Huang, Ben He, Xianpei Han, Le Sun, and Yongbin Li. Self-retrieval: End-to-end information retrieval with one large language model, 2024.

[38] Raymond Fok, Joseph Chee Chang, Tal August, Amy X. Zhang, and Daniel S. Weld. Qlarify: Recursively expandable abstracts for directed information retrieval over scientific papers, 2024.

[39] Atsushi Fujii and Tetsuya Ishikawa. Applying machine translation to two-stage cross-language information retrieval, 2000.

[40] Ryen W. White. Advancing the search frontier with ai agents, 2024.

[41] Yang Yan, Yihao Wang, Chi Zhang, Wenyuan Hou, Kang Pan, Xingkai Ren, Zelun Wu, Zhixin Zhai, Enyun Yu, Wenwu Ou, and Yang Song. Llm4pr: Improving post-ranking in search engine with large language models, 2024.

[42] Sandeep Chataut, Tuyen Do, Bichar Dip Shrestha Gurung, Shiva Aryal, Anup Khanal, Carol Lushbough, and Etienne Gnimpieba. Comparative study of domain driven terms extraction using large language models, 2024.

[43] Pranav Sharma. Lpar – a distributed multi agent platform for building polyglot, omni channel and industrial grade natural language interfaces, 2020.

[44] David Baidoo-Anu and Leticia Owusu Ansah. Education in the era of generative artificial intelligence (ai): Understanding the potential benefits of chatgpt in promoting teaching and learning. *Journal of AI*, 7(1):52–62, 2023.

[45] Quinn Patwardhan and Grace Hui Yang. Sequencing matters: A generate-retrieve-generate model for building conversational agents, 2023.

[46] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. Deep learning based recommender system: A survey and new perspectives, 2019.

[47] 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.

[48] Jesse G Meyer, Ryan J Urbanowicz, Patrick CN Martin, Karen O'Connor, Ruowang Li, Pei-Chen Peng, Tiffani J Bright, Nicholas Tatonetti, Kyoung Jae Won, Graciela Gonzalez-Hernandez, et al. Chatgpt and large language models in academia: opportunities and challenges. *BioData Mining*, 16(1):20, 2023.

[49] Daniel Buschek. Collage is the new writing: Exploring the fragmentation of text and user interfaces in ai tools, 2024.

[50] Tae Wook Kim and Quan Tan. Repurposing text-generating ai into a thought-provoking writing tutor, 2023.

[51] Umang Gupta, Ankush Chatterjee, Radhakrishnan Srikanth, and Puneet Agrawal. A sentiment-and-semantics-based approach for emotion detection in textual conversations, 2018.

[52] Samar M. Magdy, Fakhraddin Alwajih, Sang Yun Kwon, Reem Abdel-Salam, and Muhammad Abdul-Mageed. Gazelle: An instruction dataset for arabic writing assistance, 2024.

[53] David Zhou and Sarah Sterman. Ai.llude: Encouraging rewriting ai-generated text to support creative expression, 2024.

[54] Sarvesh Soni and Kirk Roberts. An evaluation of two commercial deep learning-based information retrieval systems for covid-19 literature, 2020.

[55] Marco Braga, Pranav Kasela, Alessandro Raganato, and Gabriella Pasi. Synthetic data generation with large language models for personalized community question answering, 2024.

[56] Raja Marjieh, Pol van Rijn, Ilia Sucholutsky, Theodore R. Sumers, Harin Lee, Thomas L. Griffiths, and Nori Jacoby. Words are all you need? language as an approximation for human similarity judgments, 2023.

[57] Zihao Wang, Ali Ahmadvand, Jason Choi, Payam Karisani, and Eugene Agichtein. Ericson: An interactive open-domain conversational search agent, 2023.

[58] Gabriella Kazai, Bhaskar Mitra, Anlei Dong, Nick Craswell, and Linjun Yang. Less is less: When are snippets insufficient for human vs machine relevance estimation?, 2022.

[59] KR1442 Chowdhary and KR Chowdhary. Natural language processing. *Fundamentals of artificial intelligence*, pages 603–649, 2020.

[60] Tuhin Chakrabarty, Philippe Laban, and Chien-Sheng Wu. Can ai writing be salvaged? mitigating idiosyncrasies and improving human-ai alignment in the writing process through edits, 2025.

[61] Hamel Husain, Ho-Hsiang Wu, Tiferet Gazit, Miltiadis Allamanis, and Marc Brockschmidt. Codesearchnet challenge: Evaluating the state of semantic code search. *arXiv preprint arXiv:1909.09436*, 2019.

[62] Allen Roush, Sanjay Basu, Akshay Moorthy, and Dmitry Dubovoy. Most language models can be poets too: An ai writing assistant and constrained text generation studio, 2023.

[63] Shuming Shi, Enbo Zhao, Duyu Tang, Yan Wang, Piji Li, Wei Bi, Haiyun Jiang, Guoping Huang, Leyang Cui, Xinting Huang, Cong Zhou, Yong Dai, and Dongyang Ma. Effidit: Your ai writing assistant, 2022.

[64] Yewon Kim, Mina Lee, Donghwi Kim, and Sung-Ju Lee. Towards explainable ai writing assistants for non-native english speakers, 2023.

[65] Michele Salvagno, Fabio Silvio Taccone, and Alberto Giovanni Gerli. Can artificial intelligence help for scientific writing? *Critical care*, 27(1):75, 2023.

[66] Chuan Meng, Negar Arabzadeh, Mohammad Aliannejadi, and Maarten de Rijke. Query performance prediction: From ad-hoc to conversational search, 2023.

[67] Zhe Zheng, Xin-Zheng Lu, Ke-Yin Chen, Yu-Cheng Zhou, and Jia-Rui Lin. Pretrained domain-specific language model for general information retrieval tasks in the aec domain, 2022.

[68] Tom Kenter, Alexey Borisov, Christophe Van Gysel, Mostafa Dehghani, Maarten de Rijke, and Bhaskar Mitra. Neural networks for information retrieval, 2017.

[69] Mohammad Aliannejadi, Jacek Gwizdka, and Hamed Zamani. Interactions with generative information retrieval systems, 2024.

[70] Yutao Zhu, Huaying Yuan, Shuting Wang, Jiongnan Liu, Wenhan Liu, Chenlong Deng, Haonan Chen, Zheng Liu, Zhicheng Dou, and Ji-Rong Wen. Large language models for information retrieval: A survey, 2024.

[71] Hanqi Zhang, Chong Chen, Lang Mei, Qi Liu, and Jiaxin Mao. Mamba retriever: Utilizing mamba for effective and efficient dense retrieval, 2024.

[72] Sunhao Dai, Yuqi Zhou, Liang Pang, Weihao Liu, Xiaolin Hu, Yong Liu, Xiao Zhang, Gang Wang, and Jun Xu. Neural retrievers are biased towards llm-generated content, 2024.

[73] Weihang Su, Qingyao Ai, Xiangsheng Li, Jia Chen, Yiqun Liu, Xiaolong Wu, and Shengluan Hou. Wikiformer: Pre-training with structured information of wikipedia for ad-hoc retrieval, 2024.

[74] Christophe Van Gysel. Remedies against the vocabulary gap in information retrieval, 2017.

[75] Kashav Piya, Srijal Shrestha, Cameran Frank, Estephanos Jebessa, and Tauheed Khan Mohd. Addressing the selection bias in voice assistance: Training voice assistance model in python with equal data selection, 2022.

[76] Bai Li, Nanyi Jiang, Joey Sham, Henry Shi, and Hussein Fazal. Real-world conversational ai for hotel bookings, 2019.

[77] Filip Radlinski and Nick Craswell. A theoretical framework for conversational search. In *Proceedings of the 2017 conference on conference human information interaction and retrieval*, pages 117–126, 2017.

[78] Pei-Hung Chung, Kuan Tung, Ching-Lun Tai, and Hung-Yi Lee. Joint learning of interactive spoken content retrieval and trainable user simulator, 2018.

[79] Peiyuan Gong, Jiamian Li, and Jiaxin Mao. Cosearchagent: A lightweight collaborative search agent with large language models, 2024.

[80] Jiafeng Guo, Yixing Fan, Xiang Ji, and Xueqi Cheng. Matchzoo: A learning, practicing, and developing system for neural text matching, 2019.

[81] Christina Lioma, Birger Larsen, Wei Lu, and Yong Huang. A study of factuality, objectivity and relevance: Three desiderata in large-scale information retrieval?, 2016.

[82] Chao Liu, Xin Xia, David Lo, Zhiwei Liu, Ahmed E. Hassan, and Shanping Li. Codematcher: Searching code based on sequential semantics of important query words, 2022.

[83] Pierre Lamart, Yinan Yu, and Christian Berger. Semantic-aware representation of multi-modal data for data ingress: A literature review, 2024.

[84] Alexandros Ioannidis. An analysis of indexing and querying strategies on a technologically assisted review task, 2021.