

Adapting Generative Information Retrieval Systems to Users, Tasks, and Scenarios

Johanne R. Trippas^[0000–0002–7801–0239], Damiano Spina^[0000–0001–9913–433X],
and Falk Scholer^[0000–0001–9094–0810]

RMIT University, Melbourne, Australia
{j.trippas,damiano.spina,falk.scholer}@rmit.edu.au

Draft of a chapter intended to appear in a forthcoming book on generative information retrieval, co-edited by Chirag Shah and Ryen White.

Abstract. *Generative Information Retrieval* (GenIR) signifies an advancement in *Information Retrieval* (IR). GenIR encourages more sophisticated, conversational responses to search queries by integrating generative models and chat-like interfaces. However, this approach retains core principles of traditional IR and conversational information seeking, illustrating its capacity to augment current IR frameworks.

In this chapter, we propose that introducing GenIR enhances traditional information retrieval tasks and expands their scope. This allows systems to manage more complex queries, including generative, critiquing, and extractive tasks. These advancements surpass traditional systems, handling queries with greater depth and flexibility. This sometimes speculative chapter suggests *Generative Information Access* (GenIA), a term that more accurately encapsulates the widened scope and enhanced functionalities of GenIR, particularly in how this relates to tasks. By investigating the impact of GenIR, this discussion aims to reiterate that generative research should not abandon traditional interactive information retrieval research but rather incorporate it into future research and development efforts.

Keywords: Generative AI · Information retrieval · Personalisation · Information seeking tasks.

1 Introduction

In an era where the volume of digital information expands more rapidly than ever, the ability of IR systems to sift through data to understand and anticipate users' information needs becomes even more essential. Additionally, in *Information Retrieval* (IR), the emergence of *Generative Information Retrieval* (GenIR) systems represents a paradigm shift in how we search for – and use found – information. The next-generation information access systems not only retrieve documents that may be relevant to a user's query but ideally can combine, synthesise, or abstract information, making the information directly applicable.

This chapter explores what GenIR systems may mean for users. We argue that even though *Generative Artificial Intelligence* (GenAI) could help us towards genuine interactive information retrieval, building on existing research is crucial. We, therefore, contextualise the broader GenIA and stress the importance of keeping the user central. We aim to bridge the gap between established IR principles and new generative technologies, ensuring that user needs, tasks, and contexts drive innovation in IR. We explore the dynamic interaction between advanced GenIR technologies and the user’s information needs, tasks, and queries. We examine users’ tasks in IR systems, from simple fact-finding to complex, exploratory searches and transfer these tasks to generative information-seeking.

We consider how the context (e.g., academic research, learning and teaching, or personalised personal information management) affects the requirements and expectations of a GenIR system. We discuss the integration of context-aware technologies that adapt the retrieval process to fit the user’s current environment, device, or application, thereby enhancing the relevance and utility of retrieved information.

Through theoretical insights and practical examples, this chapter aims to provide an overview of current strategies and emerging trends in GenIR. This chapter emphasises the need for an approach considering the dynamic interaction between users, tasks, and scenarios. Such an approach helps researchers and practitioners develop more efficient, user-friendly information access systems. The significance of this chapter lies in demonstrating how GenIR systems can enhance the IR process by providing more flexible, advanced, and user-centric approaches. The integration of GenIR within the broader context of GenIA offers the potential for dynamic personalisation and improved task understanding. Additionally, the chapter highlights the human’s role in ensuring the relevance and reliability of GenIR outputs and the importance of ethical considerations and user privacy in evaluating these systems.

1.1 Chapter Overview

In Section 2, we conceptualise that GenIR affects IR information needs, tasks and queries. We suggest that the generative systems’ flexibility enables more advanced tasks than traditional IR. We argue that even though the flexibility of GenIR systems introduces new capabilities to the search process, the core structure of traditional IR and *Interactive Information Retrieval* (IIR) remains. We introduce the parallel to conversational information seeking and suggest that we leverage prior research and apply it to the context of GenIR.

In Section 3, we reemphasise the importance of the user’s centrality for GenAI. We highlight the potential GenIR with more sophisticated user adaptation techniques, enabling dynamic personalisation. We emphasise that even though GenIR is driven by advanced algorithms, humans-in-the-loop is indispensable for curating and refining the system’s output to ensure relevance and reliability.

In Section 4, we extend the importance of tasks within IR. We then discuss how tasks are expanding in *Artificial Intelligence* (AI) while pulling this through to GenIR. Next, we distinguish between using GenIR to enhance system tasks versus user tasks. We then map commonly used information-seeking process stages from IR to GenIR. We conceptualise task complexity for GenIR systems and provide tasks that are suitable and less suitable for GenIR.

Next, Section 5 discusses how different scenarios and applications can use generative technology, including work, knowledge base access via customised conversational agents, learning and teaching, research, and personal information management.

Given the human-centred nature of this chapter, we discuss user evaluation in relation to GenIR in Section 6. We briefly overview commonly used user-based evaluations in IR, such as user studies, online evaluation, and implicit measures. We then propose challenges and considerations for evaluating GenIR systems, including its ethical considerations and user privacy.

Lastly, we conclude with an overview of the chapter in Section 7 and discuss the future proactivity of generative systems.

1.2 Chapter Approach and Definitions

Our approach is the following. We study past work on IR and IIR, and suggest how future information access systems can leverage prior research and what may be different when GenIR is further developed.

Generative Information Retrieval (GenIR) and *Generative Interactive Information Retrieval* (GenIIR) represent emerging concepts within *Information Retrieval* (IR), which traditionally focuses on retrieving relevant information from a large corpus of documents based on a user’s query. The new approaches incorporate generative models, especially those based on deep learning, to enhance the search process. We begin by defining the key concepts used in this chapter.

Generative Information Retrieval (GenIR). GenIR is a subset of IR technologies that leverage GenAI to enhance the search process. Unlike traditional IR systems, which focus primarily on matching keywords and returning pre-existing documents, GenIR systems can synthesise, critique, or create new content for user queries. GenIR systems aim to move beyond the limitations of keyword-based searches and static document retrieval, offering users more nuanced, conversational, and interactive search experiences. This approach opens new possibilities for automated content creation, question-answering systems, and personalised information delivery.

While traditional IR systems focus on efficiently finding and presenting existing information, GenIR systems extend this by, for example, by creating or synthesising new information in response to user queries. This fundamental difference in output (i.e., retrieving existing documents versus generating new content) represents a shift in how these systems address user information needs.

Generative Interactive Information Retrieval (GenIIR). GenIIR extends the GenIR concept by emphasising the interactive nature of – and user centrality in – the search process. Similar to IIR, GenIIR keeps the user central. *Interactive* implies the involvement of humans, in contrast to GenIR, which is system-oriented. GenIIR is a system that incorporates the capabilities of GenIR within an interactive framework that prioritises user engagement and feedback throughout the search process. Unlike GenIR, which primarily focuses on the system’s ability to generate and retrieve information, GenIIR emphasises a collaborative search process where the user’s inputs, queries, and feedback directly influence the generation and refinement of information. This approach leverages generative models to synthesise, adapt, and present information in response to user–system interactions (also seen as conversations).

The key to GenIIR is its dynamic, user-driven approach to IR, where the system understands and generates content based on initial queries, and evolves its responses through continuous interaction. These interactions ensure the generated information aligns with the user’s changing information needs and contexts. GenIIR fundamentally transforms the nature of retrieval by dynamically generating information tailored to the user’s evolving needs during the interactions.

Generative Information Access (GenIA). GenIA represents a holistic approach to how users discover, interact with, and utilise information across multiple platforms and formats. The emphasis of GenIA is on the breadth of access and the innovative generation of information, rather than on the depth of the user–system interaction as is the case for GenIIR. Interaction is one component of GenIA, but not its defining feature. It leverages GenAI to retrieve existing information and create, synthesise, or enhance the content in real-time. This includes transforming raw data into understandable narratives, generating visualisations from complex datasets, or creating new textual content that fills the gaps in existing information.

The relationships between *Generative Information Access*, *Generative Information Retrieval*, and *Generative Interactive Information Retrieval* are illustrated in Figure 1, and for convenient reference the definitions are summarised in Table 1.

1.3 Information Needs, Tasks, and Queries

Information needs, tasks, and queries are foundational to understanding how users interact with IR systems. Thus, we define these concepts. IR and GenIR information needs, tasks, and queries share fundamental similarities, as they all revolve around the user’s need to find information. We will explore the differences between traditional IR and GenIR systems in the subsequent section, reflecting the evolving capabilities of GenAI technologies.

- **Information need:** The genesis of the IR process. An information need arises when users recognise a gap in their knowledge or require information

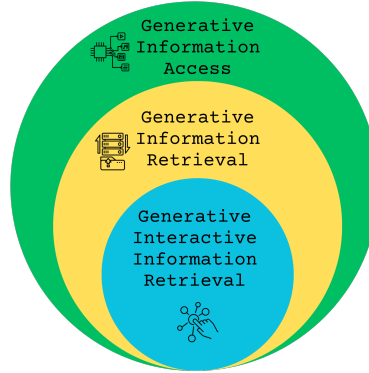


Fig. 1. Relationships of *Generative Information Access*, *Generative Information Retrieval*, and *Generative Interactive Information Retrieval*

on a particular topic or question [12]. It is the intent or requirement for information the user seeks to fulfill [75]. Information needs are often complex and may not be fully formed or explicitly understood by the user initially [20].

- **Tasks:** The notion of *task* has been widely studied in both the IR and *Information Science* (IS) fields, with two broad perspectives. First, task may refer to a “goal”, incorporating a specific scenario providing context for the need [84, 69]. This context elucidates the breadth and depth of the user’s information requirements, and can influence the search approach. Tasks, such as planning a trip, making a meal, or fixing a car, directly influence the search execution and the type of information deemed necessary, be it detailed explanations, quick facts, or comprehensive overviews [40]. The concept of *search task* focuses more specifically on the actions and activities carried out by a user to resolve their information need, such as when interacting with an IR system. For example, Broder’s taxonomy specifies search tasks as informational, navigational, or transactional [14].
- **Queries:** Queries are explicit expressions of information needs, formulated by users to interact with an IR system. A query translates the user’s information need into a system-processable format, such as a set of keywords or a question. This translation is influenced by the user’s understanding of their information need and their perception of the system’s capability to meet that need. Crafting an effective query requires users to distil their information needs and task context into a concise and precise information request.¹

The search for information, therefore, follows a logical flow, beginning with an underlying *information need* – a gap in our knowledge we aim to fill in a specific scenario. This need propels us to define a *task* – how we will acquire

¹ In this text, we use the term “queries” broadly to encompass any system-oriented specification of an information need. It therefore includes things such as “keyword queries”, “questions”, “prompts”, and “Boolean queries”.

Table 1. Relationship and short description of *Generative Information Access*, *Generative Information Retrieval*, and *Generative Interactive Information Retrieval*.

Technology	Description
<i>Generative Information Access</i> (GenIA)	Broadest concept, encompassing generative AI techniques for information access
<i>Generative Information Retrieval</i> (GenIR)	Focuses on retrieving and generating content, enhancing search beyond keyword matching
<i>Generative Interactive Information Retrieval</i> (GenIIR)	Adds human interactivity (i.e., conversational elements) to GenIR, allowing dynamic user engagement

that information. Finally, we *translate this task into a query* – employing specific words or phrases to search for information in a system like a search engine. The success of this search hinges on how accurately the query represents our initial information requirement. For instance, the need to prepare an evening meal leads to a requirement of information for a recipe; the task involves using an appropriate web search engine to seek a suitable recipe, and translating translates into the task of finding a nutritious meal and may manifest through queries such as “healthy dinner recipes” or “easy recipes with vegetables”.

2 Does *Generative Information Retrieval* (GenIR) Change *Information Retrieval* (IR) Information Needs, Tasks, and Queries?

Searching for information in GenIR can be more dynamic and interactive than traditional IR. The process still begins with an *information need*, but GenIR allows for more interactions, such as conversational engagement, to better refine and understand the user’s question (also referred to as *prompt*). Additionally, GenIR can generate new, synthesised information relevant to the user’s scenario rather than simply returning existing documents.

Tasks in GenIR extends beyond traditional search and retrieval, incorporating direct question-answering, content summarisation, and content creation based on the user’s needs. This aspect of GenIR can adapt and respond to the nuances of the user’s requirements in real time. Users might still need to craft keyword-based queries carefully, but they can also express their needs naturally. This ability to interpret and respond to conversational input improves the feedback loop between user input and information output, making responses more immediate and relevant to the user’s context.

For example, consider a user interested in starting an urban garden and seeking information on sustainable practices. In a traditional IR scenario, the user might input several keyword-based queries such as “urban gardening tips”,

Table 2. Comparison of Components in Traditional IR vs. GenIR

Component	Traditional IR	GenIR
Information needs	Defined by user’s desire to find specific information within existing documents.	Interpreted flexibly, generating new content that fulfils the user’s need.
Tasks	Involves searching, browsing, filtering, and sorting through existing information.	Extends to content creation, summarisation, and question answering through content generation.
Queries	Typically keyword-based queries, relying on precise user articulation.	Can be more natural or conversational, with the system interpreting the query’s intent.

“sustainable urban gardening”, or “how to start an urban garden”, with the system providing relevant articles in response.

In contrast, with GenIR, the user could ask, “Can you guide me through starting a sustainable urban garden?” The GenIR system could then generate a step-by-step guide from multiple documents, including multimedia, web pages, or personal documents. This guide could include selecting the right location, choosing plants based on the local climate, and implementing sustainable water drainage, all synthesised into a coherent, personalised response. This example highlights the transformative potential of GenIR in making the information retrieval process more aligned with natural human inquiry (similar to what is known from conversational information seeking), and potentially more efficient in addressing complex, multifaceted information needs.

2.1 Fulfilling Information Needs with GenIR

While the GenIR search process introduces more dynamic interactions and content generation capabilities to broaden the search process, it retains the core structure of the traditional IR and IIR processes [24]. In essence, both approaches navigate from an information need, through a task, to formulating the information need (i.e., query or prompt), aiming to fulfil the user’s search intent.

However, GenIR encapsulates a broader concept by integrating these foundational steps into a more fluid and conversation-like model. Rather than fundamentally altering the process, this development adds new interaction layers, understanding, and response generation to the established framework. GenIR’s relationship with conversational information-seeking highlights this progression.

Conversational information seeking focuses on natural language processing and understanding to facilitate a dialogue-based interaction between the user and the system [65, 95, 71, 81, 30, 80]. This approach, for example, enables the system to ask clarifying questions [4], to refine search parameters based on user responses [62, 25], and to present information in a more conversational and accessible format [82, 9]. By building on the principles of conversational information

seeking, GenIR should leverage prior research in the field, applying it within a generative context to produce synthesised information that directly addresses the user’s needs. This connection to conversational information seeking enables researchers to draw upon existing studies and methodologies, to further develop and refine GenIR systems. The accumulated knowledge in understanding user intent, processing natural language queries, and generating relevant responses, forms a solid foundation for advancing GenIR. This continuity ensures that innovations in GenIR are grounded in established IR and IIR research, while at the same time expanding the boundaries of what information retrieval systems can achieve.

While GenIR introduces novel capabilities and a broader conceptual scope, its search process remains ingrained in the traditional IR framework, enriched by the advancements in conversational information seeking. This relationship validates the effectiveness of GenIR in meeting contemporary information needs and encourages a seamless integration of new technologies with existing IR research to enhance information seeking.

Figure 2 depicts the progression of tasks within information retrieval settings, highlighting the extension of capabilities by GenIR systems. The “Critique and evaluate” layer represents an advanced function where the system generates content and provides feedback, broadening the task’s scope from mere creation to critical assessment. The diagram captures the concept of GenIR expanding the frontier of tasks beyond what was traditionally possible with search alone.

3 User-Centric Generative AI

People are at the heart of *Information Retrieval* (IR), as information-seekers and as “system component” as part of Human-in-the-Loop. Since information needs are inherently personal and unique to each individual, and therefore *adapting* systems to users has been an important goal of much IR research. The adaptation of IR systems as the potential to support users in many ways, including

- providing more relevant search results by tailoring the search to account for individual preferences;
- reducing cognitive load by aligning information with the user’s abilities and experience;
- providing context-sensitive adaptations based on a user’s location, time and device; and,
- continuously evolving to match the user’s changing preferences.

In addition, in trying to keep the user central, we also acknowledge people involved on the system side, through Human-in-the-Loop (HITL) approaches. This concept emphasises the necessity of integrating human insights within the system development and operation processes. HITL that the systems benefit from continuous human oversight and expertise. This approach enhances the system’s adaptability, reliability, and overall effectiveness. By incorporating HITL methodologies, the aim is to create more robust systems integrating human judgment with advanced technological capabilities.

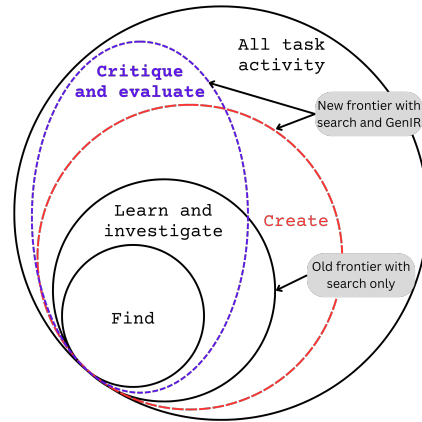


Fig. 2. The diagram is an adapted visualisation of possible tasks [88], demonstrating different information retrieval and activity levels. The diagram highlights the progression from simple information finding to more complex tasks like learning, investigating, critiquing, evaluating, and creating. The new frontier with GenIR (marked by dashes) indicates that these systems can enable advanced tasks such as *critiquing* and *evaluating*, expanding beyond the traditional search frontier. This conceptualisation shows how GenIR systems are pushing the boundaries of what can be achieved with information retrieval, making it possible to engage in higher-order cognitive tasks (e.g., create and critique).

3.1 User Adaptation

User adaptation is crucial because there is a difference in effectiveness between a search engine designed for everyone and one personalised for an individual, as highlighted by Teevan et al. as the *potential for personalisation* [76]. With the advent of GenIR, existing approaches for user adaptation are enhanced, and as the technology continues to develop, increasingly nuanced approaches are likely to become available. In addition,

It is important to note that there is a wide variety of literature on adaptivity [57, 45, 58, 44]. Therefore, the examples provided should be understood as illustrative rather than exhaustive.

User Characteristics and Individual Differences. User factors (personal characteristics and individual differences) can substantially impact how people interact with IR systems. For instance, *Cognitive abilities*, such as working memory and processing speed, have been found to have a significant impact on how effectively users search and make use of an IR system [57]. Research has shown that users with more *search experience* tend to make better and more effective use of IR systems. This is because they employ more efficient search strategies based on their understanding of system features [45]. Additionally, users with greater domain knowledge or expertise exhibit different search behaviours,

including the sites they visit, query length, and vocabulary breadth. These variations significantly impact overall search success [89].

User Adaptation Techniques. The individual difference factors of searchers have direct implications for the design of IR systems, and the techniques that can be deployed to make them adaptable to different preferences, needs, and experiences. Adaptability aims to enhance the user experience by enabling the delivery of content that is more relevant, engaging, and accessible, furthering the mission to help the user resolve their information need.

User profiling and personalisation involves collecting and analysing data about individual users (such as their behaviour, preferences, and interaction history) to tailor search results specifically to them, typically by re-ranking or filtering. Research in this area has explored various methods for creating dynamic user profiles, including machine learning algorithms that adapt to changes in user behaviour over time. Depending on the data a system can collect, user profiles could be short- or long-term. For example, past browsing behaviours have been used to create user profiles, which are then applied to personalise search results by re-ranking items [58]. Other research has demonstrated that the Big Five personality traits can predict visual search performance [64]. Incorporating such individual differences into user profiles can allow IR systems to provide personalised recommendations, and content that aligns with the characteristics of users.

Context-Aware Search considers the user’s current context, such as location, device, time, and other situational factors, to provide more relevant search results [36]. For example, a user searching for “restaurants” on a mobile device would likely expect results tailored to their current location.

User Feedback Techniques aims to integrate information directly from the user to improve search results. Widely explored approaches are the use of relevance feedback, where a user may provide *explicit* information (e.g. by marking items in an initial search results list as being relevant or not) [44], or the system makes use of *implicit* information (e.g. aggregated historical click behaviour for the same query) [42], or even simply assumes the top items that were initially retrieved to be useful and uses this information to re-write the query [19].

GenIR systems using techniques such as language models have the potential to further *personalise* responses, tailor language style, and adjust information complexity based on user profiles. These systems can leverage user profile data to generate content that matches the user’s reading level, interests, and conversational preferences. Recent advances in language models enable them to perform tasks with little task-specific data [15], suggesting strong potential for personalisation even based on limited user input. Moreover, transfer learning and fine-tuning on user-specific data will allow these models to adapt their output even further to suit individual users better.

3.2 The Role of Humans In The Loop for GenIR

In addition to GenIR users for accessing information, we acknowledge that people are integral to developing these systems. While GenIR models will continue to improve, humans still play a crucial role. They provide critical thinking to ensure that the information is useful, accurate, and ethical. HITL AI refers to a methodology where humans are actively involved in some or all stages of an AI system’s training, testing, and deployment. This approach combines the efficiency of algorithms with the nuanced understanding and decision-making capabilities of humans. Mosqueira-Rey et al. [60] identifies three broad categories for HITL machine-learning approaches. These categories are differentiated by the degree of control machines have over the learning process. From highest to lowest machine control are Active Learning, Interactive Machine Learning, and Machine Teaching. In addition, HITL is now also used more broadly across various AI applications.

For GenIR, human-AI collaboration is crucial for curating and validating information. Generative models may be good at finding information but often struggle with understanding its nuance and accuracy. Humans act as fact-checkers, evaluating information for relevance, credibility, and potential bias. Generative models might misinterpret the true intent behind a search query. Humans refine searches by providing context or reformulating queries to meet users’ needs better. Human domain expertise is invaluable for interpreting and evaluating information in specific fields. Expert knowledge helps distinguish relevant and irrelevant results, especially in complex or high-risk domains like healthcare or legal information retrieval.

Human input will continue to be key to address *ethical concerns*. Generative models can potentially surface harmful or offensive content. Therefore, humans need to be responsible for setting ethical guidelines and ensuring retrieved information is appropriate and unbiased. This aligns with the growing focus on Fairness, Accountability, Transparency and Ethics (FATE) in AI systems, where human oversight is crucial for mitigating bias and ensuring ethical outcomes. In addition, human feedback on the retrieved information is also crucial for improving generative models. This feedback can be used to train the models to better understand user needs and return more accurate and relevant results in the future. This aspect aligns with the core principles of HITL ML, where human feedback forms a continuous loop for improving the machine learning system [61].

4 Tasks and Information-seeking Processes

4.1 Tasks in Information Retrieval

The concept of tasks (also referred to as work tasks, information-seeking tasks, search tasks, or information retrieval tasks [72]) is central to the design and effectiveness of IR systems [69, 90, 10]. Tasks represent the goals or objectives

that users aim to achieve, ranging from simple queries to complex information-seeking behaviours. Identifying and understanding these tasks are crucial for developing IR systems that align with user intentions and contextual needs. These IR systems leverage computational models to provide responses relevant to users’ tasks. By tailoring the retrieval process to the characteristics of individual tasks, IR systems can provide more relevant, accurate, and useful results, thereby enhancing user satisfaction and improving the overall effectiveness of the search process. This task-centric approach to IR highlights the need for systems to understand beyond the content they index and the context and purpose behind user queries [53]. This enables a more nuanced and effective retrieval experience that aligns with the specific demands of different tasks.

Tasks are essential to users’ search strategies, the type of information they seek, and how they engage with retrieval systems [54]. For instance, a well-defined task, such as looking up a specific fact (i.e., factoid information need), typically leads to direct and focused search behaviour, with users employing precise queries and expecting quick, accurate answers. Conversely, more complex tasks (i.e., non-factoid information needs), such as conducting research for an academic paper, involve iterative search processes, refinement of information needs, and extensive interaction with the IR system to explore, contrast, and evaluate diverse information sources.

The influence of tasks extends to the design and functionality of IR systems themselves. Systems need to adapt to accommodate the varying requirements of different tasks, offering functionalities like query suggestion, personalised filtering, and context-aware retrieval to enhance user satisfaction and search efficiency [93, 72]. Understanding the task-driven nature of search behaviours and information needs is already essential for current systems; as discussed in the following sections, it may become even more crucial for developing generative IR systems that dynamically adapt to user contexts, anticipate information needs, and provide tailored responses.

4.2 Expanding IR Tasks into GenIR

Generative AI tasks include a range of “generative” tasks with varying levels of human-AI interaction, from fully automated content creation to collaborative co-creation where humans and AI work together to produce novel outcomes [23, 26]. These models are not only capable of generating text [50], but can be used for image creation [94], music composition [74], data augmentation [13], simulation [3], classification [70], or predictions [22]. As generative AI continues to evolve, the interaction between humans and AI systems is becoming more nuanced and sophisticated [6]. With AI evolving, the applications and tasks are expanding too, from enhancing artistic output to scientific research. It has been suggested that the key in successful human-AI interaction lies in finding the right balance between leveraging the AI’s capabilities and maintaining human oversight [23]. Ultimately, the goal is to harness the strengths of both humans and AI for improved outcomes.

Generative AI is also impacting the field of information retrieval. By leveraging natural language processing and machine learning techniques, generative AI models can understand and interpret complex queries, providing more accurate and relevant search results [8]. In addition, these models can generate summaries, answer questions, and create content tailored to users' needs. These AI abilities enhance information retrieval efficiency and open new possibilities for personalised and interactive search experiences.

4.3 Using GenIR to Enhance System Tasks

Tasks can be categorised as *system tasks*, given their execution by the IR system autonomously rather than by the end user. These tasks are integral to the system's enhanced capability to comprehend, process, and retrieve information relevant to the user's query. For example, a GenIR system can incorporate an automated *query expansion* technique. The automated query expansion process within a GenIR system can autonomously enhance a user's query to improve the relevance of results. This can be achieved with classic synonym additions or semantic enrichments but now completed by the underlying generative system. The outcome are search results that match the exact terms of the original query and include information linked to synonymous terms and related concepts.

An example of a system task query expansion is seen in Figure ??.

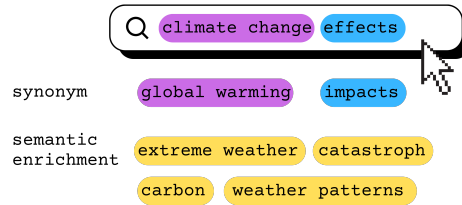


Fig. 3. Example of a typical GenIR system enhancement task for query expansion.

These tasks encapsulate many functionalities within NLP, ML, and AI domains, aimed at improving the system's performance and effectiveness. By incorporating these advanced computational techniques, the IR system can better deal with the complexities of language and patterns within vast datasets. Consequently, these system-centric tasks are key in refining the system's responsiveness and reliability in delivering relevant search results, thereby contributing to the advancement of information retrieval technologies.

4.4 Using GenIR to Support User Tasks: Mapping GenIR to Existing Information-seeking Processes

We discuss possible GenIR actions and behaviours for three commonly used stages of the information-seeking process: *query formulation*, *search result exploration*, and *query reformulation* [68, 81], search stages equivalent *express*,

examine, and *reformulate* [56]. The information-seeking process model provides broad stages for possible actions and behaviours while providing a structure.

Information Need (Query/Prompt) Formulation. The initial stage of *query formulation*, or what can also be referred to as *Information Need (Query/Prompt) Formulation*, is critical in the information-seeking process. It involves the user identifying and articulating their need for information into a query or prompt that the GenIR system can understand. This stage is critical because the entire search process’s effectiveness is based on the user’s ability to accurately express their information needs and the system’s ability to interpret them correctly.

In traditional IR systems, query formulation often relies on the user’s ability to distil their information need into a set of keywords or phrases. However, in GenIR, this process takes on a more dynamic and interactive character. GenIR systems, with their conversational capabilities, allow users to formulate their queries more naturally. This can include posing questions, making statements, or engaging in a dialogue with the system to refine the query.

Ideally, GenIR systems enhance the query formulation stage through interactive information needs refinement and prompt generation. The systems should engage with users to clarify ambiguities, request more context, or suggest different ways to phrase information needs, ensuring a deeper understanding of the user’s intent. Additionally, for users unsure of how to express their information needs, GenIR can generate guiding information needs or prompts, aiding them in refining their search objectives. This approach fosters a more intuitive and user-centric search experience.

In a recent interaction log study, it was indicated that Generative AI prompts are often verbose and structured, encapsulating a broader range of information needs and *imperative* (e.g., directive) tasks distinct from traditional search queries [79]. The study showed that LLMs can support users in tasks beyond the three main types based on user intent: informational, navigational, and transactional [14]. Prompts also included instances where entire documents are copied and pasted into the “prompt box”. These “document” prompts were often used to extract or summarise a user’s personal data, indicative of pseudo-navigational tasks or personal information management. They identified unique tasks, such as text formatting and information extraction, that extend beyond traditional search queries and uncover a range of user intents, predominantly commands to the system.

An example session from this log analysis is shown in Table 3. The session starts with a generic question from the user wanting the system to explain the stages of a waterfall model in bullet points (Turn 1). Throughout the session, the user changes the way they are formulating their information need. For example, in Turn 2, the user specifies their initial need in more details.

Examine Generated Information. In the context of GenIR-enhanced information-seeking processes, the search result exploration phase expands to include the examination of traditional search results and newly generated information. This

Table 3. Interaction session with user input to Google Bard [79]. The user input is verbatim. For example, turns 5, 8, 9, and 10 include bullet points directly copy-pasted by the user. from one user looking for information about project management.

Turn no.	User Input
1	Explain the stages of the waterfall method in bullet points
2	ok again but explain the distinct stages in teh waterfall methodology
3	give me a reference for this
4	ok can you give me a web page reference
5	<ul style="list-style-type: none"> •Evaluate the advantages and disadvantages of using Waterfall for the project, considering the specific project context.
6	ok project considerations for Agile and the PRINCE2
7	no like you did for Waterfall, project considerations for Agile and then PRINCE2
8	<ul style="list-style-type: none"> •Explain the significance of gathering user requirements in project management. •Describe how user requirements were gathered in each of the evaluated methodologies. Agile, PRINCE2 and Waterfall
9	<ul style="list-style-type: none"> •Define the various job roles and responsibilities within project management. •Compare the job roles and responsibilities in the three methodologies. •Highlight any differences or similarities among them.
10	<ul style="list-style-type: none"> •Describe the role of quality assurance, testing, and deployment in IT projects. •Explain how each methodology handled quality assurance, testing, and deployment. •Evaluate the effectiveness of these processes within the context of each methodology.
11	m,uch more concise
12	Summarize the key findings of your investigation. Discuss the factors that could influence the choice of project management methodology in the new software development company.
13	Give me a short conclusion:
14	no in paragrpahs
15	ok again, just a conclusion to theis assignemtn
16	rewrite in paragrpahs
17	I need a conclusion for my assignment. Can you give me one?
18–19	[<i>complete copy-paste document inputs</i>]
20	You missed out Agile
21	please put Agile first
22	ok rewrite that in paragrpahs

phase involves evaluating the relevance and usefulness of the initial search results and assessing the quality, novelty, and relevance of information generated by the AI system.

In the search process’ “examine generated information” stage, personalisation is crucial in tailoring the generated information to the individual’s preferences and needs. This stage involves the AI system leveraging user profiles, search history, and contextual information to create personalised summaries, answers, or content directly relevant to the user’s query. By doing so, the system must ensure that the generated information is grounded in accurate information and aligned with the user’s interests and requirements, thereby enhancing the overall effectiveness of the information-seeking process. A recent paper investigated the *readability* of generative information systems’ output and their accessibility barriers, especially for people with literacy difficulties [67]. This paper showed that responses from from widespread large language models may not be accessible to people with cognitive and literacy challenges. The authors stress that generative systems have potential accessibility issues for users with low literacy or reading impairments. To address this, it is imperative for GenIA systems to incorporate mechanisms that can adapt the complexity of language and presentation of information according to the user’s. This further highlights the important role of personalisation for generative systems. Furthermore, the design of such systems should be informed by inclusive user studies, ensuring that language and information are accessible to everyone, thereby mitigating the risk of exacerbating existing information access inequalities.

When we investigate the example interaction session again in Table 3, we see that the user is trying to specify *how* the information is presented. In turn 11, the user specifies that they want the information presented in a much more concisely.

Information Need (Query/Prompt) Reformulation. As for all interactive search systems, the reformulation stage is critical in GenIR systems. This reformulation acts as the dynamic interface where users refine their queries in response to generated content and initial search outcomes. This iterative process is integral to GenIR, enabling users to adjust their information requests based on the presented information. By continuously refining their queries, users can further investigate their topic, leveraging the generative capabilities of the system to explore complex ideas and uncover connections. This feedback loop enhances the precision of search results and enriches the users engagement with the information, demonstrating the unique interactivity and adaptability of GenIR systems.

Re-investigating the example interaction in Table 3, we can see many different reformulations. The search interaction excerpts highlight the iterative nature of the reformulation stage in the context of the user’s search process. Each step, from initial, often imprecise requests (“Give me a short conclusion” in Turn 13) to more specific demands (“no in paragraphs” in Turn 14), illustrates how user queries evolve as they refine their need. This dynamic is crucial in both tra-

ditional and GenIR systems, where the capacity to adapt responses based on user feedback can enhance the relevance and utility of the information provided. For instance, requests like “You missed out Agile” and “please put Agile first” (Turns 20–21) emphasise the importance of adaptability and specificity in search queries, including the need for systems that can flexibly accommodate changing user priorities and insights. In GenIR, this is particularly important, as the system must not only search but also generate content, demonstrating a sophisticated generative model for content creation. These reformulation interactions are practical examples of continuous feedback and essential for refining search outputs and accuracy. This capability to iterate and evolve search queries and responses is foundational in delivering a more personalised and effective search experience.

4.5 Conceptualising Task Complexity for GenIR Systems

The field of information retrieval has long recognised the diverse nature of information-seeking tasks and acknowledges that tasks vary in their complexity [17, 41, 90, 93]. Understanding this variation is crucial for developing information systems that effectively support users across a spectrum of needs. This section introduces a conceptual framework for categorising information-seeking tasks by two critical dimensions: task complexity and *generative involvement*. For simplicity, we refer to task complexity as the number of steps, the intricacy of these steps, and the level of decision-making needed to complete a task. Figure 4 illustrates a continuum of task complexity and the level of generative intervention from AI as discussed below:

Basic information retrieval (low task complexity, minimal generative involvement). It involves direct queries with precise answers, like looking up straightforward facts. Interaction with GenIA systems is transactional, the user requests specific information, and the AI retrieves it with little to no additional generative contribution.

Guided topic expansion (medium task complexity, moderate generative involvement). It entails broadening the scope of an inquiry to include related topics or concepts, requiring users to navigate through and select relevant information. The GenIA system aids this process by suggesting related areas and generating ancillary information that users can incorporate into their search.

In-depth analysis and synthesis (high task complexity, substantial generative involvement). Requires comprehensive research and the integration of multiple information sources to construct detailed knowledge or insights. The GenIA system plays a significant role by generating complex outputs like summaries of extensive literature, which the user then critically evaluates and refines for their purposes.

Intelligent research design (very high task complexity, interactive generative involvement). Involves the generation of new research frameworks, theoretical models, or innovative problem-solving approaches. The GenIA system

and the user interact together, with the GenIA system proposing novel ideas and designs that the user iteratively refines, leading to sophisticated outcomes that may not have been achievable individually.

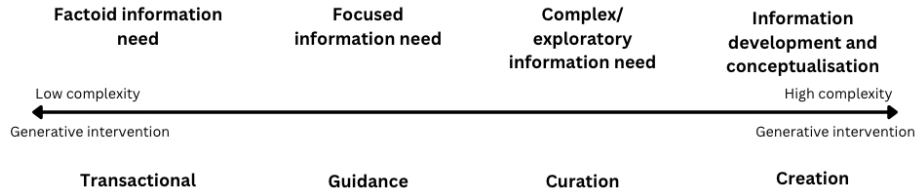


Fig. 4. Possible visualisation of how generative systems interact depending on task complexity.

Tasks less suitable for GenIR. Based on the above conceptualisation, we can see that not all search tasks are suited for a GenIR approach. The effectiveness of GenIR systems largely depends on the query’s nature, the user’s information need, and the context in which the information will be used. There are several scenarios where traditional IR systems might be more appropriate or where GenIR systems may need to be carefully designed to meet specific requirements.

- **Factoid information need.** Consider someone who wants to have an answer to a very concrete information need, “first person on South Pole”. This query seeks a factual answer about a historical event related to world exploration. A traditional IR system would look for information from historical records, exploration archives, or authoritative history websites to provide the name of the explorer who first reached the South Pole. It makes sense for a factoid task to embed the information in current authoritative information. In contrast, even though a generative system may contain the information, the user may not have to conduct extra fact-checking.
- **Legal or medical information.** In domains where the accuracy of information can have serious implications, such as legal and medical research, the conservative approach of traditional IR systems may be preferred. The potential for GenIR systems to synthesise information in ways that misinterpret complex legal statutes or medical guidelines necessitates a cautious application [47].
- **(Re)finding an original online document.** Imagine someone is working on a paper about tasks in information retrieval. They want to retrieve the topics of a previous TREC Track. Even though the user can ask for the topics in a GenIR system, users may prefer to access original documents directly from the sources, rather than receiving synthesised or generated content. In such cases, traditional IR systems that provide direct links to original

sources would be more appropriate. In addition, many documents are not online and reside in physical archives or within proprietary databases. The researcher may need to consult these offline materials for academic rigour, necessitating a hybrid approach combining digital searches with traditional library methods.

- **Niche topics.** GenIR systems are typically trained on broad datasets, which may not cover highly specialised or niche topics sufficiently. For nice queries, traditional IR systems that index specialised databases or pay-walled articles might provide more comprehensive and relevant results.
- **Complex topics with critical and high-level reasoning.** While advances in AI and natural language processing have enabled GenIR systems to handle complex queries, there are still limitations in their ability to perform multi-step reasoning or to understand queries that require deep domain-specific knowledge. Complemented by human expertise, traditional IR systems may be better suited for these scenarios.

Tasks suitable for GenIR. Next, we present example *Information Access* (IA) tasks that are suitable for GenIR:

- **Content creation.** GenIR systems are proficient at creating new, original content tailored to specific needs. This includes writing articles, generating reports, or producing creative pieces like short stories and poetry. The strength of GenIR in content creation lies in its ability to analyse vast amounts of data, understand context, and generate coherent, relevant text based on the user’s input or prompts. GenIR can help streamline the content creation, offering efficiency and creativity while reducing the time and resources traditionally required for these tasks.
- **Content summarisation.** With the overwhelming amount of online information, there is a growing need for concise summaries that capture the essence of longer texts. GenIR systems can automatically generate accurate, coherent summaries of articles, research papers, books, or reports, making information more accessible and digestible for readers.
- **Content extraction.** GenIR systems can help with content extraction, where specific information, data points, or insights need to be identified and extracted from large volumes of text or complex datasets. GenIR systems can parse through documents, identify relevant pieces of information based on the criteria set by the user, and generate summaries or reports highlighting the extracted content. This could save people time, and GenIR systems may identify patterns that the user may initially overlook.
- **Personalisation.** Leveraging the strengths of GenIR, systems can craft personalised information. By analysing a user’s past interactions, search behaviours, or preferences, GenIR systems could curate content, increasing user satisfaction and engagement. The strategic deployment of GenIR for personalised suggestions enriches the user’s experience by ensuring relevancy and is crucial in enhancing loyalty and improving conversion rates. GenIR may offer a more personalised, engaging, and user-centric service.

5 Scenarios and Applications

5.1 Work

Information access systems’ role in working environments and work tasks have been studied for decades [51, 52, 55, 83]. There are a variety of tasks on which information access systems are used, including communication, documentation, planning, problem-solving, admin and management, among others. In 2019, Trippas et al. [83] asked participants in an online survey about work tasks and how digital assistants can support them.

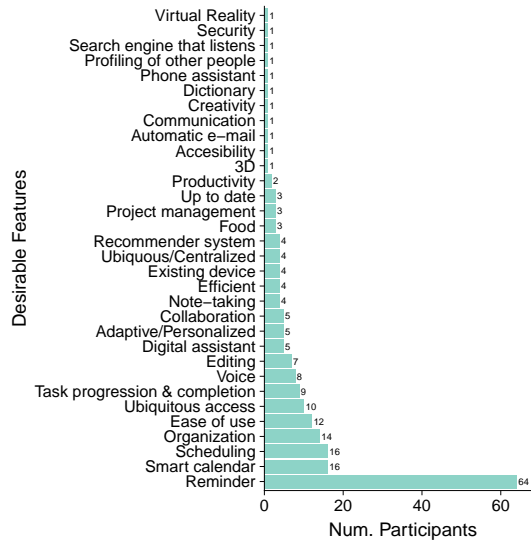


Fig. 5. Desirable features for an imaginary new piece of technology as reported by participants from [83].

The survey was conducted from May 17 to July 2 2018, with 410 respondents. One of the questions in the survey asked participants to describe the features or capabilities they would want to have in a hypothetical new piece of technology. Figure 5 shows an aggregation of the relevant responses – 46.9% of the participants do not report a particular feature, or report they do not need any. Features such as automatic “reminders”, “scheduling”, or “ubiquitous use” are easily recognisable in current applications such as e-mail clients or personal information managers. Yet, other features such as “automatic e-mail”, “profiling of other people”, “note taking”, or keeping oneself “up to date” were less obvious to foresee before the uptake of generative AI solutions such as ChatGPT.

A report by Microsoft [16] discusses the opportunities LLM-based technology can create to assist in work tasks. The report suggests that we are witnessing

an expansion in the range of tasks automated by human-AI cooperation with assistants or copilots. For instance, the task of *finding* similar snippets of code implementing a particular functionality may shift to *critically analyse* a generated code that satisfies a given functional requirement. Synthesis tasks (e.g., summarising a set of relevant documents or the discussion in a particular meeting) are also likely to be more automated in working settings, particularly for retrospective and real-time feedback in collaborative scenarios.

As generative IR enable automating more complex information-related tasks in work settings, it is important to consider the ethical implications that automation has in terms of the workforce. As every technological advancement reduces the need for manual labour, it is crucial to have measures that enable everybody to benefit from shared prosperity [43, 91].

5.2 Knowledge Base Access via Customised Conversational Agents

Recent developments in GenIR, including RAG, have sparked interest in both industry and academia in exploring the role of customised LLM-based conversational agents as a viable solution to provide access to information that is typically stored in private data and knowledge bases (e.g., information used by customer services, internal policies and procedures, project management data).

Pathiyam Cherumanal et al. [63] described *Walert*, an LLM-powered conversational agent customised to answer frequently asked questions about enrolment in computer science programs. Walert aims to demonstrate how common practices in conversational information seeking and IR can be adopted by practitioners when designing and evaluating customised LLM-based conversational agents, and identify aspects in the evaluation process that may help practitioners better understand the suitability of GenIR in the context of customised information seeking systems, such as the importance of evaluating Retrieval-Augmented Generation (RAG) at the retrieval phase – and the role of truncated rankings [7] –, the inclusion of out of knowledge base (i.e., unanswerable) questions in the test collection, and the need of more informative evaluation measures to assess correctness of the generated responses at the end of the pipeline [63].

Another common problem in organisations and companies is that the vast amount of information generated and collected by different teams within the organisation is not easily accessible. Information is typically stored in different data and knowledge bases, and organised in different formats. Enterprise search [48] is a structured search system that enables stakeholders in an organisation to find relevant information in decision-making processes with organisational or personal goals [87]. Reliable access to relevant information is arguably one of the most essential needs of an organisation. And yet, the nature of the problem – including data fragmentation, lack of redundancy in information, vocabulary gap, and secure access for different roles within the organisation, among others – makes enterprise search one of the most challenging problems in IR [48]. Considering the capability of LLMs in characterising language and semantic relationships, new GenIR techniques could considerably reduce the vocabulary gap present in enterprise search. In the past, enterprise search has attracted little

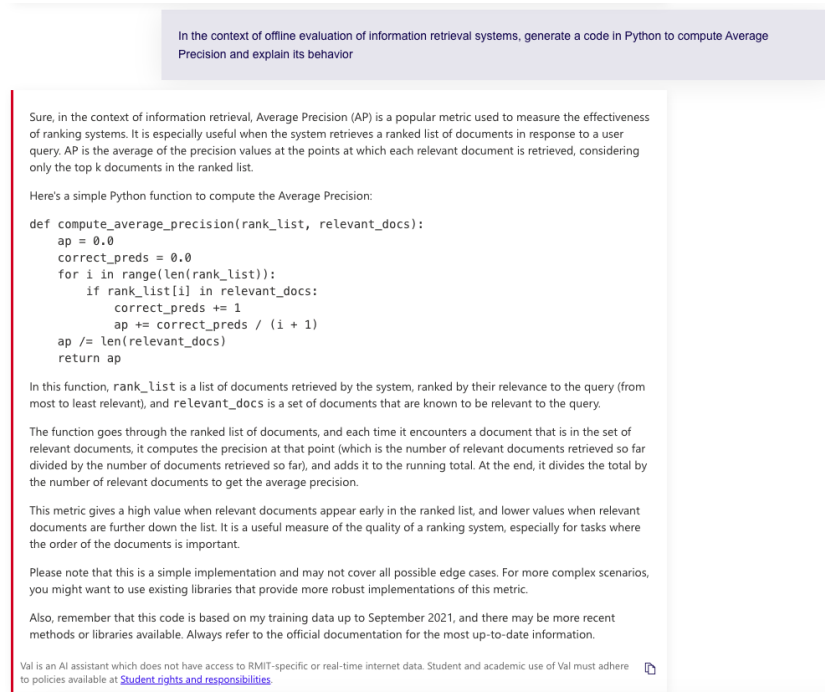


Fig. 6. Example of code generation for implementing Average Precision (AP) in Python by RMIT Val using GPT-4.

attention in the academic community [24]. The increasing interest (both in industry and academia) of customised LLM-based conversational agents and RAG may push the state of the art in enterprise search.

5.3 Learning and Teaching

The arrival of ChatGPT has provoked a series of debates around the implications of generative AI in the education domain [2, 1, 21]. Recent studies bring empirical results to the discussion on ways students have been using GenAI tools [1]. Code and text generation for completing assignments, essays, and academic projects are some of the most common applications in learning and teaching environments. Although these applications may enhance productivity, they also raise some ethical concerns and implications to pedagogy and academic integrity [21].

Figure 6 shows the interaction with RMIT's Val learning assistant,² using GPT-4 to generate a snippet of code in Python to compute the Average Precision (AP) effectiveness measure.

² <https://www.rmit.edu.au/students/support-services/study-support/val> [Accessed: 12 Apr 2024]

Besides the code, the response includes an explanation of the code, as well as a disclaimer to warn the user that the code does not cover edge cases (e.g., topics without relevant documents in the qrels would give a divided-by-zero runtime error). An Information Retrieval course student may prefer to interact with a conversational agent to learn about fundamental concepts such as AP instead of running a search in a commercial search engine – or reading the evaluation chapter of a classic IR book.

A promising direction is using GenAI for personalised and adaptive learning experiences [21, 33]. In particular, novel ways for providing personalised feedback or instruction tailored to individual student needs [33, 21].

The effective use of this type of technology in learning and teaching environments heavily relies on creating awareness of its limitations and critical engagement. There is also a demand for more established guidelines and policies to ensure the responsible use of GenAI in educational scenarios.

5.4 Research

Researchers perform a variety of tasks during the life-cycle of a research project. Some common tasks across different study fields are planning, literature review, experimental design, data collection and analysis, writing and publishing, and collaboration. While many of these tasks overlap with the work tasks described in Section 5.1 (e.g., planning or collaboration), it is important to consider how GenIR can assist practitioners with research tasks.

The literature review phase is where most of the finding tasks occur. Although Retrieval-Augmented Generation (RAG) systems [49] – i.e., systems that generate an answer from a set of passages retrieved from a knowledge base – specifically designed for scientific repositories may provide a complementary way to find relevant work, GenIR can still present unreliable information to researchers. Recent work explored the use of LLMs to make systematic reviews more cost-effective. Wang et al. [86] investigated the effectiveness of Boolean queries designed for systematic review literature search generated by transformed-based systems such as ChatGPT, showing a promising ground for research directions but also important caveats related to incorrect terms in the queries and non-determinism of prompts. Another work in the context of systematic reviews is reducing the number of retrieved documents that need to be manually screened by experts/researchers performing systematic reviews. Recent work in automatic document screening has explored the use of ChatGPT [5], fine-tuning [66] and zero-shot open-sourced LLMs [85]. The results indicate that techniques based on LLM, particularly fine-tuned, can automatically be developed to screen documents for systematic reviews.

GenAI tools are instead becoming commonly available to assist researchers in refining their writing, e.g., by recommending alternative ways to formulate titles, abstracts, or sentences. Researchers may also benefit from using tools for synthesis or translation tasks [59].

Mittelstadt, Wachter, and Russell [59] discuss how other tasks, such as data formatting and conversion, are likely to get more automated with the assistance

of GenAI. However, the use of GenAI in other research tasks could compromise research integrity [27, 59], by increasing the risk of lack of reproducibility and transparency, especially if used without robust quality assurance protocols in the data collection and analysis phases.

As in other research fields, information retrieval has also started to include GenAI approaches in research. In addition to automatic relevance assessment [78, 32] and simulation of user’s interactions [11, 29] [(see Chapter 6 for more details)], recent work explore the use of GenAI to characterise tasks and information needs. Zendel et al. [96] explores the effectiveness of instruction-based LLMs to automatically classify the cognitive complexity of information needs described as backstories [96]. Alaofi et al. [3] explores the role of LLMs in generating new query variants for a given information need. Pathiyan Cherumanal et al. [63] use open-source LLMs as a data augmentation approach to generate training phrases to build the conversational model of a customised intent-based conversational agent.

5.5 Personalised Personal Information Management

Personal Information Management (PIM) is a set of practices to manage personal information ecosystems [18, 77]. This ecosystem includes various physical and digital information formats like emails, documents, web content, and social media interactions. PIM enables users to control their information environment, enhancing their productivity, decision-making, and learning [38].

Incorporating GenAI and GenIR into PIM could potentially enhance personalised information access. GenAI can extrapolate new insights, link diverse data sources, and propose novel viewpoints, aiding knowledge integration from personal information sources. In addition, GenIR can improve how information is retrieved and presented to the user. GenIR can provide more relevant and digestible information by understanding the user’s context and preferences. This can save time and effort in information retrieval and make using information more efficient and enjoyable.

In this context, PIM is not just about managing information, but also about effectively utilising this information to achieve tasks and fulfil roles within individual contexts. This could include professional roles where specific information is needed to make decisions or personal roles where information could help plan activities or learn new skills. Thus enhancing current PIM techniques making it even more personalised. The combination of PIM, GenAI, and GenIR aims to create a more personalised, efficient, and insightful way of managing and utilising information. This integrated approach can empower individuals to control their information environment and use information more effectively to achieve their goals. It represents a significant step forward in the evolution of personal information management.

6 User Evaluation of *Generative Information Retrieval*

New approaches enabled through *Generative Information Retrieval* offer extensive opportunities to support users to resolve their information needs. To understand these new approaches, and to support ongoing development and improvements, the ability to measure and evaluate system performance is a key requirement. The acir field has a strong history of evaluation. This includes approaches based on offline evaluation using test collections (often called the “Cranfield” methodology) [34] and instantiated through shared evaluation campaigns such as TREC, NTCIR, CLEF, and FIRE; online evaluation through techniques such as A/B testing [35]; and, user-based evaluation [39].

While the established approaches for evaluating IR systems provide a good foundation, they are not always directly usable in the context of new features that GenIR systems support. For example, Gienapp et al. [31] have recently proposed a framework for generative *ad hoc* retrieval – the task of ranking documents by their expected relevance in response to a single search query – that defines *utility*, *reading* and *accumulation* components for an effectiveness metric. In line with traditional ad hoc retrieval evaluation based on test collections, this framework offers a promising direction to enable GenIR systems to be evaluated for ad hoc search, offering a clearly defined, repeatable, and cost-effective way to quantify effectiveness. However, similar to the use of test collections for traditional IR systems, this comes at the cost of simplification by essentially abstracting out the user and their interactions with the system.

As was highlighted in the previous sections of this chapter, *Generative Information Retrieval* in particular offers substantial new opportunities at the level of users, tasks and scenarios: key opportunities arise in the interaction between users and systems, working to resolve an information need that is situated in the scope of a particular task – *Generative Interactive Information Retrieval*. Evaluation here typically required user-focused approaches, rather than the use of test collections which typically abstract out the variability that users and interactions introduce. We therefore provide an overview of the key methodologies and associated considerations that arise in the context of user-based evaluation of these systems.

6.1 Current IR Approaches to User Evaluation

User Studies. Evaluating the effectiveness of interactive systems can require careful study of the interactions between users and systems. It can be helpful to consider different approaches based on the goals of the research: exploratory, descriptive, and explanatory [39]. User studies can vary widely depending on the phenomenon being studied. Generally, they fall into exploratory, descriptive, and explanatory categories, reflecting the level of researcher control. Exploratory studies involve minimal intervention, while explanatory studies often require extensive intervention for formal experimental inference.

- *Exploratory studies.* In situations where relatively little is known about the phenomenon, exploratory studies are useful to enable better understand-

ing [73]. The aim is often to learn more about the phenomenon, which means that the research questions may be broad or open-ended. As a result, exploratory studies are typically less structured. Exploratory studies often inform subsequent descriptive and explanatory studies.

- *Descriptive studies.* Descriptive studies aim to describe a phenomenon by careful observation and documentation [28]. Such studies can provide benchmarks of interactive systems, and serve as taxonomies related to the phenomenon of interest.
- *Explanatory studies.* When variables of interest have been identified, explanatory studies offer a framework to determine relationships between them. This includes formal experiments to establish causality. Explanatory studies are sometimes termed “laboratory experiments”, since they often take place in controlled conditions, with the aim of isolating the key variables of interest from possible confounding conditions [92].

Online Evaluation and Implicit Measures. Online evaluation aims to measure the effectiveness of IR systems by considering implicit indicators of user behaviour as they interact with a live system. Indicators may be any measurable signals that reflect user activity, and can range from low-level events such as the number of clicks on a hyperlink and the dwell time on particular web page, to higher-level events such as decisions to purchase items in an online store [35].

To establish the relative effectiveness of two systems, online evaluation typically makes use of AB Testing, a between-subjects experiment where users are randomly exposed to either system A or system B (the independent variable) to establish the presence or absence of an effect on the chosen implicit indicator (the dependent variable) [46].

6.2 Challenges and Considerations for Evaluating *Generative Information Retrieval Systems*

The evolution of IR to include conversational and generative aspects necessitates a deeper understanding of user needs and behaviors, especially since these systems may substantially change user expectations, and interaction approaches. GenIR systems require rigorous user evaluation methodologies to ensure their effectiveness and relevance. User studies, incorporating both quantitative and qualitative methodologies, will be essential in identifying the effectiveness of such systems in addressing complex user information needs.

A key open challenge for GenIR evaluation will be to establish realistic approaches for evaluating system output with users when we cannot control the system’s output. For generative systems, a key consideration regarding user studies is the extent to which the system output needs to be controlled. Generative systems, by their nature, create “new” responses, and it may be difficult to ensure that such a system generates identical output even in response to the same input query.

The impact of this factor will vary depending on the type of study being conducted. In the context of *user studies*, this is e.g. unlikely to be problematic

for an open-ended exploratory study aiming to learn about interactions between users and chatbots, but may present new complications into the design of an exploratory study in which the system output needs to be a controlled variable.

Since the indicators used in *online evaluation* rely on signals of user behavior, rather than the specific output of a GenIR system, this evaluation approach can be used directly to evaluate systems that include new generative components. It is however important to bear in mind the usual limitations of online evaluation, namely that the implicit indicators are very likely to only be a proxy for variables that are actually of interest, such as whether the system actually conveyed useful information to the user, or whether the user was ultimately satisfied.

Other challenges in evaluating GenIR systems include accounting for the natural language understanding (NLU) and generation (NLG) components, managing context and state across conversational turns, and ensuring the relevance and coherence of system responses. User evaluation methods will therefore need to be tailored to address these challenges, e.g. by incorporating scenario-based testing, user satisfaction surveys, and task completion rates as part of the evaluation criteria.

Beyond individual studies that focus on particular aspects of evaluation, the development and ongoing evaluation of GenIR systems will benefit from the use of *user-centered design principles*, involving users early and throughout the design process of such systems. This includes understanding user preferences for conversational interactions, personalization, and response generation. Design decisions should be informed by user feedback, ensuring that the system aligns with user expectations and information seeking behaviors.

Ethical Considerations and User Privacy. The ethical landscape of user evaluation in GenIR systems is complex, underscored by the importance of the ethical use of data and privacy considerations. In this context, the methodologies employed to gather, analyse, and store user data should be carefully designed to uphold the highest standards of privacy and ethics. In most countries, regulatory requirements around GenIR are developing; but perhaps even more importantly, ethical considerations and practices are a fundamental aspect of building trust and ensuring the integrity of the interaction between users and systems.

Transparency in the collection, usage, and storage of user data forms the cornerstone of ethical user evaluation. Users should be fully informed about what data is being collected, how it is being used, and where it is stored. This transparency is crucial not just for compliance with privacy laws and regulations, such as the General Data Protection Regulation (GDPR) in the European Union, but also for fostering a relationship of trust with users. When users understand how their data contributes to the improvement and effectiveness of generative IR systems, they are more likely to participate willingly in the evaluation process.

Informed consent is another critical element, ensuring that users are not just aware of how their data is used, but have explicitly agreed to it. This consent should be obtained through clear, understandable language that avoids technical jargon, so that it is accessible to all users regardless of their background in tech-

nology. Moreover, informed consent should not be treated as a one-time process. Users should have the ability to withdraw their consent at any time, necessitating systems that can accommodate such requests without compromising the integrity of the data or the user experience.

A key technique in preserving user privacy is the *de-identification* or *anonymisation* of user data. By removing or obfuscating identifiers that can link data back to an individual, researchers and developers can analyse patterns, behaviors, and feedback while minimising risks around compromising user anonymity in situations such as data breaches or unauthorized access. Privacy-preserving methodologies extend beyond anonymisation principles to include techniques such as differential privacy, which adds noise to the data to prevent the identification of individuals while still allowing for the aggregate data to be useful for analysis and system improvement. These methodologies ensure that the evaluation of generative IR systems can proceed without exposing sensitive user information or compromising the privacy of individual users.

7 Conclusion and the Future of Generative Systems

This chapter demonstrates that research on highly interactive information retrieval is not new. However, with the new developments around generative technology, interactivity has become central again in information retrieval research. This *generative interactive information retrieval* resurgence may have the potential to make IA systems true assistants. However, questions such as *how much generation* is really needed for particular tasks, and *what kind of interactivity best enhances user experience*, remain open. The challenge lies in finding the right balance between generative capabilities and user control, ensuring that the systems are powerful but also intuitive and user-friendly. As we progress, we must continue exploring these questions, testing and refining generative interactive approaches to realise their potential in transforming information retrieval into a more dynamic and collaborative process.

7.1 Proactivity in GenIR

Generative systems, despite their advancements, largely are still not proactive. The concept of “search engines that listen” aimed to introduce a more interactive dimension to information retrieval systems [37, 81]. This vision sought to transform search engines from “responders” to active participants in the search process, capable of understanding and adapting to the user’s context in real-time.

With GenIR, we have progressed in expanding the task types these systems can handle, moving beyond traditional search queries to include content generation, summarisation, critiquing, and even dialogues-like user interactions. However, this leap has yet to realise the proactive potential of GenIR systems fully. The envisioned “search engines that listen” imply initiative and anticipation, actively engaging with users, seeking clarification, and offering suggestions

even before a query is fully articulated. Perhaps even imply that a system is part of an agent that performs tasks on behalf of users.

To achieve genuinely proactive systems, enhancements in GenIR should focus on better interpreting user interactions, effectively using context, and applying predictive analytics to anticipate user needs. This shift towards proactive participation, making the system a co-navigator rather than just a responder, may improve the user experience, making information retrieval more intuitive and aligning more with natural human information-seeking behaviours.

References

- [1] Muhammad Abbas, Farooq Ahmed Jam, and Tariq Iqbal Khan. “Is It Harmful or Helpful? Examining the Causes and Consequences of Generative AI Usage Among University Students”. In: *International Journal of Educational Technology in Higher Education* 21.1 (Feb. 2024), p. 10. ISSN: 2365-9440. DOI: [10.1186/s41239-024-00444-7](https://doi.org/10.1186/s41239-024-00444-7). (Visited on 04/08/2024).
- [2] Tufan Adiguzel, Mehmet Haldun Kaya, and Fatih Kürşat Cansu. “Revolutionizing Education with AI: Exploring the Transformative Potential of ChatGPT”. In: *Contemporary Educational Technology* 15.3 (July 1, 2023), ep429. ISSN: 1309-517X. DOI: [10.30935/cedtech/13152](https://doi.org/10.30935/cedtech/13152). URL: <https://www.cedtech.net/article/revolutionizing-education-with-ai-exploring-the-transformative-potential-of-chatgpt-13152> (visited on 03/05/2024).
- [3] Marwah Alaofi, Luke Gallagher, Mark Sanderson, Falk Scholer, and Paul Thomas. “Can Generative LLMs Create Query Variants for Test Collections? An Exploratory Study”. In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '23. New York, NY, USA: Association for Computing Machinery, July 18, 2023, pp. 1869–1873. ISBN: 978-1-4503-9408-6. DOI: [10.1145/3539618.3591960](https://doi.org/10.1145/3539618.3591960). URL: <https://dl.acm.org/doi/10.1145/3539618.3591960> (visited on 02/26/2024).
- [4] Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W. Bruce Croft. “Asking Clarifying Questions in Open-Domain Information-Seeking Conversations”. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '19: The 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Paris France: ACM, July 18, 2019, pp. 475–484. ISBN: 978-1-4503-6172-9. DOI: [10.1145/3331184.3331265](https://doi.org/10.1145/3331184.3331265). URL: <https://dl.acm.org/doi/10.1145/3331184.3331265> (visited on 03/06/2024).
- [5] Ahmad Alshami, Moustafa Elsayed, Eslam Ali, Abdelrahman E. E. El-toukhy, and Tarek Zayed. “Harnessing the Power of ChatGPT for Automating Systematic Review Process: Methodology, Case Study, Limitations, and Future Directions”. In: *Systems* 11.7 (July 2023), p. 351. ISSN: 2079-8954. DOI: [10.3390/systems11070351](https://doi.org/10.3390/systems11070351). (Visited on 03/25/2024).

- [6] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. “Guidelines for Human-AI Interaction”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI ’19: CHI Conference on Human Factors in Computing Systems. Glasgow Scotland Uk: ACM, May 2, 2019, pp. 1–13. ISBN: 978-1-4503-5970-2. DOI: [10.1145/3290605.3300233](https://doi.org/10.1145/3290605.3300233). URL: <https://dl.acm.org/doi/10.1145/3290605.3300233> (visited on 02/27/2024).
- [7] Enrique Amigó, Stefano Mizzaro, and Damiano Spina. “Ranking Interruptus: When Truncated Rankings Are Better and How to Measure That”. In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Madrid Spain: ACM, July 2022, pp. 588–598. ISBN: 978-1-4503-8732-3. DOI: [10.1145/3477495.3532051](https://doi.org/10.1145/3477495.3532051). (Visited on 03/28/2024).
- [8] Avishek Anand, Abhijit Anand, and Vinay Setty. “Query Understanding in the Age of Large Language Models”. 2023. arXiv: [2306.16004](https://arxiv.org/abs/2306.16004). URL: <https://arxiv.org/abs/2306.16004> (visited on 02/27/2024).
- [9] Sandeep Avula, Gordon Chadwick, Jaime Arguello, and Robert Capra. “SearchBots: User Engagement with ChatBots during Collaborative Search”. In: *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval - CHIIR ’18*. The 2018 Conference. New Brunswick, NJ, USA: ACM Press, 2018, pp. 52–61. ISBN: 978-1-4503-4925-3. DOI: [10.1145/3176349.3176380](https://doi.org/10.1145/3176349.3176380). URL: <http://dl.acm.org/citation.cfm?doid=3176349.3176380> (visited on 03/06/2024).
- [10] Jing Bai and Jian-Yun Nie. “Adapting Information Retrieval to Query Contexts”. In: *Information Processing & Management* 44.6 (2008), pp. 1901–1922. URL: <https://www.sciencedirect.com/science/article/pii/S0306457308000824> (visited on 02/29/2024).
- [11] Krisztian Balog and ChengXiang Zhai. *User Simulation for Evaluating Information Access Systems*. June 2023. arXiv: [2306.08550 \[cs\]](https://arxiv.org/abs/2306.08550). (Visited on 03/28/2024).
- [12] Nicholas J. Belkin, Robert N. Oddy, and Helen M. Brooks. “ASK for Information Retrieval: Part I. Background and Theory”. In: *Journal of documentation* 38.2 (1982), pp. 61–71. URL: <https://www.emerald.com/insight/content/doi/10.1108/eb026722/full/html> (visited on 03/05/2024).
- [13] Angona Biswas, Nasim Md Abdullah Al, Al Imran, Anika Tabassum Sejuty, Fableha Fairouz, Sai Puppala, and Sajedul Talukder. “Generative Adversarial Networks for Data Augmentation”. In: *Data Driven Approaches on Medical Imaging*. Ed. by Bin Zheng, Stefan Andrei, Md Kamruzzaman Sarker, and Kishor Datta Gupta. Cham: Springer Nature Switzerland, 2023, pp. 159–177. ISBN: 978-3-031-47771-3 978-3-031-47772-0. DOI: [10.1007/978-3-031-47772-0_8](https://doi.org/10.1007/978-3-031-47772-0_8). URL: https://link.springer.com/10.1007/978-3-031-47772-0_8 (visited on 02/27/2024).

- [14] Andrei Broder. “A Taxonomy of Web Search”. In: *ACM SIGIR Forum* 36.2 (Sept. 2002), pp. 3–10. ISSN: 0163-5840. DOI: [10.1145/792550.792552](https://doi.org/10.1145/792550.792552). URL: <https://dl.acm.org/doi/10.1145/792550.792552> (visited on 02/10/2024).
- [15] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, and Tom Henighan. “Language Models Are Few-Shot Learners”. In: ().
- [16] Jenna Buttler, Sonia Jaffe, Nancy Baym, Mary Czerwinski, Shamsi Iqbal, Kate Nowak, Sean Rintel, Abigail Sellen, Mihaela Vorvoreanu, Brent Hecht, and Jaime Teevan. *Microsoft New Future of Work Report 2023*. Tech Report MSRTR-2023-34. Microsoft Research, 2023. URL: <https://aka.ms/nfw2023>.
- [17] Katriina Byström and Kalervo Järvelin. “Task complexity affects information seeking and use”. In: *Information processing & management* 31.2 (1995), pp. 191–213.
- [18] Robert Capra. “A survey of personal information management practices”. In: *Proceedings of Personal Information Management: PIM* (2009).
- [19] Claudio Carpineto and Giovanni Romano. “A Survey of Automatic Query Expansion in Information Retrieval”. In: *ACM Computing Surveys* 44.1 (Jan. 2012), pp. 1–50. ISSN: 0360-0300, 1557-7341. DOI: [10.1145/2071389.2071390](https://doi.org/10.1145/2071389.2071390). URL: <https://dl.acm.org/doi/10.1145/2071389.2071390> (visited on 03/16/2024).
- [20] Donald O. Case and Lisa M. Given. “Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior”. In: (2016). URL: https://books.google.com.au/books?hl=en&lr=&id=IAYvDAAAQBAJ&oi=fnd&pg=PP1&dq=scenarios+information+seeking&ots=Y-OTmzrOC3&sig=cAp_H91RVR5yXT65hvUM5tuUByA (visited on 03/05/2024).
- [21] Cecilia Ka Yuk Chan and Katherine K. W. Lee. “The AI Generation Gap: Are Gen Z Students More Interested in Adopting Generative AI Such as ChatGPT in Teaching and Learning than Their Gen X and Millennial Generation Teachers?” In: *Smart Learning Environments* 10.1 (2023), pp. 60–23. ISSN: 2196-7091. DOI: [10.1186/s40561-023-00269-3](https://doi.org/10.1186/s40561-023-00269-3).
- [22] Boyang Chen, Zongxiao Wu, and Ruoran Zhao. “From Fiction to Fact: The Growing Role of Generative AI in Business and Finance”. In: *Journal of Chinese Economic and Business Studies* 21.4 (Oct. 2, 2023), pp. 471–496. ISSN: 1476-5284, 1476-5292. DOI: [10.1080/14765284.2023.2245279](https://doi.org/10.1080/14765284.2023.2245279). URL: <https://www.tandfonline.com/doi/full/10.1080/14765284.2023.2245279> (visited on 02/27/2024).
- [23] Ruijia Cheng, Alison Smith-Renner, Ke Zhang, Joel Tetreault, and Alejandro Jaimes-Larrarte. “Mapping the Design Space of Human-AI Interaction in Text Summarization”. In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Proceedings of the 2022 Conference of the

- North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Seattle, United States: Association for Computational Linguistics, 2022, pp. 431–455. DOI: [10.18653/v1/2022.naacl-main.33](https://doi.org/10.18653/v1/2022.naacl-main.33). URL: <https://aclanthology.org/2022.naacl-main.33> (visited on 02/27/2024).
- [24] W. Bruce Croft, Donald Metzler, and Trevor Strohman. *Search Engines: Information Retrieval in Practice*. Vol. 520. Addison-Wesley Reading, 2010. URL: https://www.academia.edu/download/30740463/z2009_2465.pdf (visited on 03/06/2024).
- [25] Jeffrey Dalton, Sophie Fischer, Paul Owoicho, Filip Radlinski, Federico Rossetto, Johanne R. Trippas, and Hamed Zamani. “Conversational Information Seeking: Theory and Application”. In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR ’22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. Madrid Spain: ACM, July 6, 2022, pp. 3455–3458. ISBN: 978-1-4503-8732-3. DOI: [10.1145/3477495.3532678](https://doi.org/10.1145/3477495.3532678). URL: <https://dl.acm.org/doi/10.1145/3477495.3532678> (visited on 03/06/2024).
- [26] Zijian Ding and Joel Chan. *Mapping the Design Space of Interactions in Human-AI Text Co-creation Tasks*. Mar. 14, 2023. DOI: [10.48550/arXiv.2303.06430](https://doi.org/10.48550/arXiv.2303.06430). arXiv: [2303.06430 \[cs\]](https://arxiv.org/abs/2303.06430). URL: <http://arxiv.org/abs/2303.06430> (visited on 02/27/2024). preprint.
- [27] Matt Duckham, Falk Scholer, Daniel Barr, David Blades, Chi-Tsun (Ben) Cheng, Brian Falzon, Anthony Forsyth, Lisa Given, Dana McKay, Anne-Laure Mention, Xavier Mulet, Magdalena Plebanski, Jason Potts, Mark Sanderson, John Thangarajah, Julian Thomas, Karin Verspoor, Charlie Xue, Irene Yarovsky, Xinghuo Yu, and Fabio Zambetta. *Research Integrity and Generative AI*. [object Object], Nov. 8, 2023. DOI: [10.5281/ZENODO.10081201](https://doi.org/10.5281/ZENODO.10081201). URL: <https://zenodo.org/doi/10.5281/zenodo.10081201> (visited on 03/05/2024).
- [28] Carsten Eickhoff, Jaime Teevan, Ryen White, and Susan Dumais. “Lessons from the Journey: A Query Log Analysis of Within-Session Learning”. In: *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*. New York New York USA: ACM, Feb. 2014, pp. 223–232. ISBN: 978-1-4503-2351-2. DOI: [10.1145/2556195.2556217](https://doi.org/10.1145/2556195.2556217). (Visited on 04/07/2024).
- [29] Björn Engelmann, Timo Breuer, Jana Isabelle Friese, Philipp Schaer, and Norbert Fuhr. “Context-Driven Interactive Query Simulations Based on Generative Large Language Models”. In: *Advances in Information Retrieval*. Ed. by Nazli Goharian, Nicola Tonello, Yulan He, Aldo Lipani, Graham McDonald, Craig Macdonald, and Iadh Ounis. Vol. 14609. Cham: Springer Nature Switzerland, 2024, pp. 173–188. ISBN: 978-3-031-56059-0 978-3-031-56060-6. DOI: [10.1007/978-3-031-56060-6_12](https://doi.org/10.1007/978-3-031-56060-6_12). (Visited on 03/28/2024).

- [30] Jianfeng Gao, Michel Galley, and Lihong Li. “Neural Approaches to Conversational AI”. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. SIGIR ’18: The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval. Ann Arbor MI USA: ACM, June 27, 2018, pp. 1371–1374. ISBN: 978-1-4503-5657-2. DOI: [10.1145/3209978.3210183](https://doi.org/10.1145/3209978.3210183). URL: <https://dl.acm.org/doi/10.1145/3209978.3210183> (visited on 03/06/2024).
- [31] Lukas Gienapp, Harrison Scells, Niklas Deckers, Janek Bevendorff, Shuai Wang, Johannes Kiesel, Shahbaz Syed, Maik Fröbe, Guido Zuccon, Benno Stein, Matthias Hagen, and Martin Potthast. *Evaluating Generative Ad Hoc Information Retrieval*. Nov. 2023. arXiv: [2311.04694 \[cs\]](https://arxiv.org/abs/2311.04694). (Visited on 03/28/2024).
- [32] Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. “ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks”. In: *Proceedings of the National Academy of Sciences* 120.30 (July 2023), e2305016120. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.2305016120](https://doi.org/10.1073/pnas.2305016120). (Visited on 03/28/2024).
- [33] Shalin Hai-Jew. *Generative AI in Teaching and Learning*. 1st ed. Hershey: IGI Global, 2023. ISBN: 9798369300749. DOI: [10.4018/979-8-3693-0074-9](https://doi.org/10.4018/979-8-3693-0074-9).
- [34] Donna K. Harman. “TREC: Experiment and Evaluation in Information Retrieval”. In: ed. by Ellen M. Voorhees and Donna K. Harman. MIT Press, 2005.
- [35] Katja Hofmann, Lihong Li, and Filip Radlinski. “Online Evaluation for Information Retrieval”. In: *Foundations and Trends® in Information Retrieval* 10.1 (2016), pp. 1–117. ISSN: 1554-0669, 1554-0677. DOI: [10.1561/15000000051](https://doi.org/10.1561/15000000051). (Visited on 04/07/2024).
- [36] Yoshihide Hosokawa, Yuichiro Nakazawa, and Takanori Yamamoto. “Location-Aware Information Retrieval for Identifying Local and Distant Landmark”. In: *Proceedings of the 29th Annual ACM Symposium on Applied Computing*. SAC 2014: Symposium on Applied Computing. Gyeongju Republic of Korea: ACM, Mar. 24, 2014, pp. 428–435. ISBN: 978-1-4503-2469-4. DOI: [10.1145/2554850.2554974](https://doi.org/10.1145/2554850.2554974). URL: <https://dl.acm.org/doi/10.1145/2554850.2554974> (visited on 03/16/2024).
- [37] Hideo Joho. “Taciturn Search, Loquacious Search”. In: *IEICE Technical Report; IEICE Tech. Rep.* 115.184 (2015), pp. 19–19. URL: <https://ken.ieice.org/ken/paper/20150821Tb1s/eng/> (visited on 03/06/2024).
- [38] William Jones, Rob Capra, Mary Czerwinski, Jesse David Dinneen, Jacek Gwizdka, and Unmil Karadkar. “PIM 2024: The Information We Need, When We Need It...: As We Get Ever Closer, Is this Ideal Still Ideal?” In: *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*. CHIIR ’24. Sheffield, United Kingdom: Association for Computing Machinery, 2024, pp. 438–440. ISBN: 9798400704345. DOI: [10.1145/3627508.3638333](https://doi.org/10.1145/3627508.3638333). URL: <https://doi.org/10.1145/3627508.3638333>.

- [39] Diane Kelly. “Methods for Evaluating Interactive Information Retrieval Systems with Users”. In: *Foundations and Trends® in Information Retrieval* 3.1–2 (2009), pp. 1–224. URL: <https://www.nowpublishers.com/article/Details/INR-012> (visited on 02/10/2024).
- [40] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-ching Wu. “Development and Evaluation of Search Tasks for IIR Experiments using a Cognitive Complexity Framework”. en. In: *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*. Northampton Massachusetts USA: ACM, Sept. 2015, pp. 101–110. ISBN: 978-1-4503-3833-2. DOI: [10.1145/2808194.2809465](https://doi.org/10.1145/2808194.2809465). URL: <https://dl.acm.org/doi/10.1145/2808194.2809465> (visited on 03/28/2024).
- [41] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-ching Wu. “Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework”. In: *Proceedings of the 2015 international conference on the theory of information retrieval*. 2015, pp. 101–110.
- [42] Diane Kelly and Jaime Teevan. “Implicit Feedback for Inferring User Preference: A Bibliography”. In: *ACM SIGIR Forum* 37.2 (Sept. 2003), pp. 18–28. ISSN: 0163-5840. DOI: [10.1145/959258.959260](https://doi.org/10.1145/959258.959260). URL: <https://dl.acm.org/doi/10.1145/959258.959260> (visited on 03/16/2024).
- [43] Hisham O. Khogali and Samir Mekid. “The Blended Future of Automation and Ai: Examining Some Long-Term Societal and Ethical Impact Features”. In: *Technology in Society* 73 (May 2023), p. 102232. ISSN: 0160791X. DOI: [10.1016/j.techsoc.2023.102232](https://doi.org/10.1016/j.techsoc.2023.102232). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0160791X23000374> (visited on 03/06/2024).
- [44] Jin Young Kim, Jaime Teevan, and Nick Craswell. “Explicit In Situ User Feedback for Web Search Results”. In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR ’16: The 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. Pisa Italy: ACM, July 7, 2016, pp. 829–832. ISBN: 978-1-4503-4069-4. DOI: [10.1145/2911451.2914754](https://doi.org/10.1145/2911451.2914754). URL: <https://dl.acm.org/doi/10.1145/2911451.2914754> (visited on 03/16/2024).
- [45] Kyung-Sun Kim and Bryce Allen. “Cognitive and Task Influences on Web Searching Behavior”. In: *Journal of the American Society for Information Science and Technology* 53.2 (Jan. 2002), pp. 109–119. ISSN: 1532-2882, 1532-2890. DOI: [10.1002/asi.10014](https://doi.org/10.1002/asi.10014). URL: <https://onlinelibrary.wiley.com/doi/10.1002/asi.10014> (visited on 03/15/2024).
- [46] Ron Kohavi, Roger Longbotham, Dan Sommerfield, and Randal M. Henne. “Controlled Experiments on the Web: Survey and Practical Guide”. In: *Data Mining and Knowledge Discovery* 18.1 (Feb. 2009), pp. 140–181. ISSN: 1384-5810, 1573-756X. DOI: [10.1007/s10618-008-0114-1](https://doi.org/10.1007/s10618-008-0114-1). (Visited on 04/07/2024).
- [47] Bevan Koopman and Guido Zuccon. “Dr Chatgpt Tell Me What i Want to Hear: How Different Prompts Impact Health Answer Correctness”. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language*

- Processing*. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Singapore: Association for Computational Linguistics, 2023, pp. 15012–15022. DOI: [10.18653/v1/2023.emnlp-main.928](https://doi.org/10.18653/v1/2023.emnlp-main.928). URL: <https://aclanthology.org/2023.emnlp-main.928> (visited on 03/04/2024).
- [48] Udo Kruschwitz and Charlie Hull. “Searching the Enterprise”. In: *Foundations and Trends® in Information Retrieval* 11.1 (2017), pp. 1–142. ISSN: 1554-0669, 1554-0677. DOI: [10.1561/15000000053](https://doi.org/10.1561/15000000053). (Visited on 03/28/2024).
- [49] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”. In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin. Vol. 33. Curran Associates, Inc., 2020, pp. 9459–9474.
- [50] Yang Li, Quan Pan, Suhang Wang, Tao Yang, and Erik Cambria. “A Generative Model for Category Text Generation”. In: *Information Sciences* 450 (2018), pp. 301–315. URL: <https://www.sciencedirect.com/science/article/pii/S0020025518302366> (visited on 02/27/2024).
- [51] Yuelin Li and Nicholas J. Belkin. “A Faceted Approach to Conceptualizing Tasks in Information Seeking”. In: *Information Processing & Management* 44.6 (Nov. 2008), pp. 1822–1837. ISSN: 03064573. DOI: [10.1016/j.ipm.2008.07.005](https://doi.org/10.1016/j.ipm.2008.07.005). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0306457308000836> (visited on 03/06/2024).
- [52] Yuelin Li and Nicholas J. Belkin. “An Exploration of the Relationships Between Work Task and Interactive Information Search Behavior”. In: *Journal of the American Society for Information Science and Technology* 61.9 (Sept. 2010), pp. 1771–1789. ISSN: 1532-2882, 1532-2890. DOI: [10.1002/asi.21359](https://doi.org/10.1002/asi.21359). URL: <https://onlinelibrary.wiley.com/doi/10.1002/asi.21359> (visited on 03/06/2024).
- [53] Jiqun Liu. “Deconstructing Search Tasks in Interactive Information Retrieval: A Systematic Review of Task Dimensions and Predictors”. In: *Information Processing & Management* 58.3 (May 1, 2021), p. 102522. ISSN: 0306-4573. DOI: [10.1016/j.ipm.2021.102522](https://doi.org/10.1016/j.ipm.2021.102522). URL: <https://www.sciencedirect.com/science/article/pii/S0306457321000315> (visited on 02/29/2024).
- [54] Jiqun Liu, Matthew Mitsui, Nicholas J. Belkin, and Chirag Shah. “Task, Information Seeking Intentions, and User Behavior: Toward A Multi-level Understanding of Web Search”. In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. CHIIR ’19. New York, NY, USA: Association for Computing Machinery, Mar. 8, 2019, pp. 123–132. ISBN: 978-1-4503-6025-8. DOI: [10.1145/3295750.3298922](https://doi.org/10.1145/3295750.3298922). URL: <https://dl.acm.org/doi/10.1145/3295750.3298922> (visited on 02/28/2024).
- [55] Jiqun Liu, Shawon Sarkar, and Chirag Shah. “Identifying and Predicting the States of Complex Search Tasks”. In: *Proceedings of the 2020 Confer-*

- ence on Human Information Interaction and Retrieval*. CHIIR '20: Conference on Human Information Interaction and Retrieval. Vancouver BC Canada: ACM, Mar. 14, 2020, pp. 193–202. ISBN: 978-1-4503-6892-6. DOI: [10.1145/3343413.3377976](https://doi.org/10.1145/3343413.3377976). URL: <https://dl.acm.org/doi/10.1145/3343413.3377976> (visited on 03/06/2024).
- [56] Gary Marchionini. *Information Seeking in Electronic Environments*. 9. Cambridge university press, 1995. URL: <https://books.google.com.au/books?hl=en&lr=&id=cYOHgr18DSQC&oi=fnd&pg=PR8&dq=G.+Marchionini.+Information+seeking+in+electronic+environments.&ots=e8oLNGTwZU&sig=7wd1IpxzmQYJgbyudhD3bpnEEJc> (visited on 03/06/2024).
- [57] Azzah Al-Maskari and Mark Sanderson. “The Effect of User Characteristics on Search Effectiveness in Information Retrieval”. In: *Information Processing & Management* 47.5 (Sept. 2011), pp. 719–729. ISSN: 03064573. DOI: [10.1016/j.ipm.2011.03.002](https://doi.org/10.1016/j.ipm.2011.03.002). URL: <https://linkinghub.elsevier.com/retrieve/pii/S030645731100029X> (visited on 03/15/2024).
- [58] Nicolaas Matthijs and Filip Radlinski. “Personalizing Web Search Using Long Term Browsing History”. In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. WSDM'11: Fourth ACM International Conference on Web Search and Data Mining. Hong Kong China: ACM, Feb. 9, 2011, pp. 25–34. ISBN: 978-1-4503-0493-1. DOI: [10.1145/1935826.1935840](https://doi.org/10.1145/1935826.1935840). URL: <https://dl.acm.org/doi/10.1145/1935826.1935840> (visited on 03/16/2024).
- [59] Brent Mittelstadt, Sandra Wachter, and Chris Russell. “To Protect Science, We Must Use LLMs as Zero-Shot Translators”. In: *Nature Human Behaviour* 7.11 (Nov. 20, 2023), pp. 1830–1832. ISSN: 2397-3374. DOI: [10.1038/s41562-023-01744-0](https://doi.org/10.1038/s41562-023-01744-0). URL: <https://www.nature.com/articles/s41562-023-01744-0> (visited on 03/19/2024).
- [60] Eduardo Mosqueira-Rey, Elena Hernández-Pereira, David Alonso-Ríos, José Bobes-Bascarán, and Ángel Fernández-Leal. “Human-in-the-Loop Machine Learning: A State of the Art”. In: *Artificial Intelligence Review* 56.4 (Apr. 2023), pp. 3005–3054. ISSN: 0269-2821, 1573-7462. DOI: [10.1007/s10462-022-10246-w](https://doi.org/10.1007/s10462-022-10246-w). (Visited on 07/02/2024).
- [61] Eduardo Mosqueira-Rey, Elena Hernández-Pereira, David Alonso-Ríos, José Bobes-Bascarán, and Ángel Fernández-Leal. “Human-in-the-Loop Machine Learning: A State of the Art”. In: *Artificial Intelligence Review* 56.4 (Apr. 2023), pp. 3005–3054. ISSN: 0269-2821, 1573-7462. DOI: [10.1007/s10462-022-10246-w](https://doi.org/10.1007/s10462-022-10246-w). URL: <https://link.springer.com/10.1007/s10462-022-10246-w> (visited on 03/22/2024).
- [62] Paul Owoicho, Jeffrey Dalton, Mohammad Aliannejadi, Leif Azzopardi, Johanne R. Trippas, and Svitlana Vakulenko. “TREC CAsT 2022: Going beyond User Ask and System Retrieve with Initiative and Response Generation”. In: *NIST Special Publication* (2022), pp. 500–338. URL: https://trec.nist.gov/pubs/trec31/papers/Overview_cast.pdf (visited on 03/06/2024).

- [63] Sachin Pathiyen Cherumanal, Lin Tian, Futoon M. Abushaqra, Angel Felipe Magnossao de Paula, Kaixin Ji, Danula Hettiachchi, Johanne R. Trippas, Halil Ali, Falk Scholer, and Damiano Spina. “Walert: Putting Conversational Search Knowledge into Action by Building and Evaluating a Large Language Model-Powered Chatbot”. In: *Proceedings of the ACM Conference on Information Interaction and Retrieval (CHIIR’24)* (2024), arXiv-2401. URL: <https://ui.adsabs.harvard.edu/abs/2024arXiv240107216P/abstract> (visited on 02/10/2024).
- [64] Chad Peltier and Mark W. Becker. “Individual Differences Predict Low Prevalence Visual Search Performance”. In: *Cognitive Research: Principles and Implications* 2.1 (Dec. 2017), p. 5. ISSN: 2365-7464. DOI: [10.1186/s41235-016-0042-3](https://doi.org/10.1186/s41235-016-0042-3). URL: <http://cognitiveresearchjournal.springeropen.com/articles/10.1186/s41235-016-0042-3> (visited on 03/16/2024).
- [65] Filip Radlinski and Nick Craswell. “A Theoretical Framework for Conversational Search”. In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval. CHIIR ’17: Conference on Human Information Interaction and Retrieval*. Oslo Norway: ACM, Mar. 7, 2017, pp. 117–126. ISBN: 978-1-4503-4677-1. DOI: [10.1145/3020165.3020183](https://doi.org/10.1145/3020165.3020183). URL: <https://dl.acm.org/doi/10.1145/3020165.3020183> (visited on 03/06/2024).
- [66] Ambrose Robinson, William Thorne, Ben P. Wu, Abdullah Pandor, Munira Essat, Mark Stevenson, and Xingyi Song. *Bio-SIEVE: Exploring Instruction Tuning Large Language Models for Systematic Review Automation*. Aug. 2023. arXiv: [2308.06610 \[cs\]](https://arxiv.org/abs/2308.06610). (Visited on 03/25/2024).
- [67] Adam Roegiest and Zuzana Pinkosova. “Generative Information Systems Are Great If You Can Read”. In: *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval. CHIIR ’24*. Sheffield, United Kingdom: Association for Computing Machinery, 2024, pp. 165–177. ISBN: 9798400704345. DOI: [10.1145/3627508.3638345](https://doi.org/10.1145/3627508.3638345). URL: <https://doi.org/10.1145/3627508.3638345>.
- [68] Nuzhah Gooda Sahib, Anastasios Tombros, and Tony Stockman. “A Comparative Analysis of the Information-Seeking Behavior of Visually Impaired and Sighted Searchers”. In: *Journal of the American Society for Information Science and Technology* 63.2 (2012), pp. 377–391. ISSN: 1532-2890. DOI: [10.1002/asi.21696](https://doi.org/10.1002/asi.21696). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21696> (visited on 03/06/2024).
- [69] Chirag Shah, Ryen White, Paul Thomas, Bhaskar Mitra, Shawon Sarkar, and Nicholas Belkin. “Taking Search to Task”. In: *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval. CHIIR ’23: ACM SIGIR Conference on Human Information Interaction and Retrieval*. Austin TX USA: ACM, Mar. 19, 2023, pp. 1–13. ISBN: 9798400700354. DOI: [10.1145/3576840.3578288](https://doi.org/10.1145/3576840.3578288). URL: <https://dl.acm.org/doi/10.1145/3576840.3578288> (visited on 01/18/2024).

- [70] Chirag Shah, Ryen W. White, Reid Andersen, Georg Buscher, Scott Counts, Sarkar Snigdha Sarathi Das, Ali Montazer, Sathish Manivannan, Jennifer Neville, Xiaochuan Ni, Nagu Rangan, Tara Safavi, Siddharth Suri, Mengting Wan, Leijie Wang, and Longqi Yang. *Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies*. Jan. 31, 2024. arXiv: [2309.13063 \[cs\]](https://arxiv.org/abs/2309.13063). URL: <http://arxiv.org/abs/2309.13063> (visited on 02/16/2024). preprint.
- [71] Stefan Sitter and Adelheit Stein. “Modeling Information-Seeking Dialogues: The Conversational Roles (COR) Model”. In: *RIS: Review of Information Science (online journal)* 1.1 (1996), pp. 165–180. URL: <http://www.fb10.uni-bremen.de/anglistik/langpro/webospace/jb/info-pages/misc/sds/cor.pdf> (visited on 03/06/2024).
- [72] Ayah Soufan, Ian Ruthven, and Leif Azzopardi. “Untangling the Concept of Task in Information Seeking and Retrieval”. In: *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*. ICTIR ’21: The 2021 ACM SIGIR International Conference on the Theory of Information Retrieval. Virtual Event Canada: ACM, July 11, 2021, pp. 73–81. ISBN: 978-1-4503-8611-1. DOI: [10.1145/3471158.3472259](https://doi.org/10.1145/3471158.3472259). URL: <https://dl.acm.org/doi/10.1145/3471158.3472259> (visited on 02/29/2024).
- [73] Amanda Spink. “A User-Centered Approach to Evaluating Human Interaction with Web Search Engines: An Exploratory Study”. In: *Information Processing & Management* 38.3 (May 2002), pp. 401–426. ISSN: 03064573. DOI: [10.1016/S0306-4573\(01\)00036-X](https://doi.org/10.1016/S0306-4573(01)00036-X). (Visited on 04/07/2024).
- [74] Minhyang (Mia) Suh, Emily Youngblom, Michael Terry, and Carrie J Cai. “AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI ’21. New York, NY, USA: Association for Computing Machinery, May 7, 2021, pp. 1–11. ISBN: 978-1-4503-8096-6. DOI: [10.1145/3411764.3445219](https://doi.org/10.1145/3411764.3445219). URL: <https://dl.acm.org/doi/10.1145/3411764.3445219> (visited on 02/26/2024).
- [75] Robert S. Taylor. “The Process of Asking Questions”. In: *American Documentation* 13.4 (Oct. 1962), pp. 391–396. ISSN: 0096-946X, 1936-6108. DOI: [10.1002/asi.5090130405](https://doi.org/10.1002/asi.5090130405). URL: <https://onlinelibrary.wiley.com/doi/10.1002/asi.5090130405> (visited on 03/05/2024).
- [76] Jaime Teevan, Susan T. Dumais, and Eric Horvitz. “Potential for Personalization”. In: *ACM Transactions on Computer-Human Interaction* 17.1 (Mar. 2010), pp. 1–31. ISSN: 1073-0516, 1557-7325. DOI: [10.1145/1721831.1721835](https://doi.org/10.1145/1721831.1721835). URL: <https://dl.acm.org/doi/10.1145/1721831.1721835> (visited on 03/12/2024).
- [77] Jaime Teevan, William Jones, and Benjamin B Bederson. “Personal information management”. In: *Communications of the ACM* 49.1 (2006), pp. 40–43.
- [78] Paul Thomas, Seth Spielman, Nick Craswell, and Bhaskar Mitra. *Large Language Models Can Accurately Predict Searcher Preferences*. Sept. 19,

2023. arXiv: [2309.10621 \[cs\]](https://arxiv.org/abs/2309.10621). URL: <http://arxiv.org/abs/2309.10621> (visited on 03/05/2024). preprint.
- [79] Johanne R. Trippas, Sara Fahad Dawood Al Lawati, Joel Mackenzie, and Luke Gallagher. “What Do Users Really Ask Large Language Models? An Initial Log Analysis of Google Bard Interactions in the Wild”. In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery. New York, NY, USA, 2024, pp. 1–5.
- [80] Johanne R. Trippas, Luke Gallagher, and Joel Mackenzie. “Exploring Conversational Search and Re-evaluating the Command-and-Control Paradigm with Conversational Agents”. In: *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24)*. Association for Computing Machinery. New York, NY, USA, 2024.
- [81] Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, Hideo Joho, and Mark Sanderson. “Informing the Design of Spoken Conversational Search: Perspective Paper”. In: *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval - CHIIR '18*. The 2018 Conference. New Brunswick, NJ, USA: ACM Press, 2018, pp. 32–41. ISBN: 978-1-4503-4925-3. DOI: [10.1145/3176349.3176387](https://doi.org/10.1145/3176349.3176387). URL: <http://dl.acm.org/citation.cfm?doid=3176349.3176387> (visited on 03/06/2024).
- [82] Johanne R. Trippas, Damiano Spina, Mark Sanderson, and Lawrence Cavedon. “Results Presentation Methods for a Spoken Conversational Search System”. In: *Proceedings of the First International Workshop on Novel Web Search Interfaces and Systems*. CIKM'15: 24th ACM International Conference on Information and Knowledge Management. Melbourne Australia: ACM, Oct. 22, 2015, pp. 13–15. ISBN: 978-1-4503-3789-2. DOI: [10.1145/2810355.2810356](https://doi.org/10.1145/2810355.2810356). URL: <https://dl.acm.org/doi/10.1145/2810355.2810356> (visited on 03/06/2024).
- [83] Johanne R. Trippas, Damiano Spina, Falk Scholer, Ahmed Hassan Awadallah, Peter Bailey, Paul N. Bennett, Ryen W. White, Jonathan Liono, Yongli Ren, Flora D. Salim, and Mark Sanderson. “Learning About Work Tasks to Inform Intelligent Assistant Design”. In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. CHIIR '19: Conference on Human Information Interaction and Retrieval. Glasgow Scotland UK: ACM, Mar. 8, 2019, pp. 5–14. ISBN: 978-1-4503-6025-8. DOI: [10.1145/3295750.3298934](https://doi.org/10.1145/3295750.3298934). URL: <https://dl.acm.org/doi/10.1145/3295750.3298934> (visited on 03/04/2024).
- [84] Pertti Vakkari. “Task-based Information Searching”. In: *Annual Review of Information Science and Technology* 37.1 (Jan. 2003), pp. 413–464. ISSN: 0066-4200, 1550-8382. DOI: [10.1002/aris.1440370110](https://doi.org/10.1002/aris.1440370110). (Visited on 03/28/2024).
- [85] Jiayin Wang, Weizhi Ma, Peijie Sun, Min Zhang, and Jian-Yun Nie. *Understanding User Experience in Large Language Model Interactions*. Jan. 16, 2024. arXiv: [2401.08329 \[cs\]](https://arxiv.org/abs/2401.08329). URL: <http://arxiv.org/abs/2401.08329> (visited on 02/05/2024). preprint.

- [86] Shuai Wang, Harrison Scells, Bevan Koopman, and Guido Zuccon. “Can ChatGPT Write a Good Boolean Query for Systematic Review Literature Search?” In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Taipei Taiwan: ACM, July 2023, pp. 1426–1436. ISBN: 978-1-4503-9408-6. DOI: [10.1145/3539618.3591703](https://doi.org/10.1145/3539618.3591703). (Visited on 03/25/2024).
- [87] Martin White. *Enterprise Search: Enhancing Business Performance*. Enhancing Business Performance. O’Reilly Media, 2015. ISBN: 978-1-4919-1551-6.
- [88] Ryen W. White. “Tasks, Copilots, and the Future of Search: A Keynote at SIGIR 2023”. In: *ACM SIGIR Forum* 57.2 (Jan. 22, 2024), 4:1–4:8. ISSN: 0163-5840. DOI: [10.1145/3642979.3642985](https://doi.org/10.1145/3642979.3642985). URL: <https://dl.acm.org/doi/10.1145/3642979.3642985> (visited on 02/10/2024).
- [89] Ryen W. White, Susan T. Dumais, and Jaime Teevan. “Characterizing the Influence of Domain Expertise on Web Search Behavior”. In: *Proceedings of the Second ACM International Conference on Web Search and Data Mining*. WSDM’09: Second ACM International Conference on Web Search and Web Data Mining. Barcelona Spain: ACM, Feb. 9, 2009, pp. 132–141. ISBN: 978-1-60558-390-7. DOI: [10.1145/1498759.1498819](https://doi.org/10.1145/1498759.1498819). URL: <https://dl.acm.org/doi/10.1145/1498759.1498819> (visited on 03/15/2024).
- [90] Barbara Wildemuth, Luanne Freund, and Elaine G. Toms. “Untangling Search Task Complexity and Difficulty in the Context of Interactive Information Retrieval Studies”. In: *Journal of Documentation* 70.6 (2014), pp. 1118–1140. URL: <https://www.emerald.com/insight/content/doi/10.1108/JD-03-2014-0056/full/html> (visited on 02/29/2024).
- [91] Scott A. Wright and Ainslie E. Schultz. “The Rising Tide of Artificial Intelligence and Business Automation: Developing an Ethical Framework”. In: *Business Horizons* 61.6 (Nov. 2018), pp. 823–832. ISSN: 00076813. DOI: [10.1016/j.bushor.2018.07.001](https://doi.org/10.1016/j.bushor.2018.07.001). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0007681318301046> (visited on 03/06/2024).
- [92] Wan-Ching Wu and Diane Kelly. “Online Search Stopping Behaviors: An Investigation of Query Abandonment and Task Stopping”. In: *Proceedings of the American Society for Information Science and Technology* 51.1 (Jan. 2014), pp. 1–10. ISSN: 0044-7870, 1550-8390. DOI: [10.1002/meet.2014.14505101030](https://doi.org/10.1002/meet.2014.14505101030). (Visited on 04/07/2024).
- [93] Iris Xie. “Dimensions of Tasks: Influences on Information-seeking and Retrieving Process”. In: *Journal of Documentation* 65.3 (Jan. 1, 2009), pp. 339–366. ISSN: 0022-0418. DOI: [10.1108/00220410910952384](https://doi.org/10.1108/00220410910952384). URL: <https://doi.org/10.1108/00220410910952384> (visited on 02/29/2024).
- [94] Yutong Xie, Zhaoying Pan, Jinge Ma, Luo Jie, and Qiaozhu Mei. “A Prompt Log Analysis of Text-to-Image Generation Systems”. In: *Proceedings of the ACM Web Conference 2023*. WWW ’23: The ACM Web Conference 2023. Austin TX USA: ACM, Apr. 30, 2023, pp. 3892–3902. ISBN: 978-1-4503-9416-1. DOI: [10.1145/3543507.3587430](https://doi.org/10.1145/3543507.3587430). URL: <https://dl.acm.org/doi/10.1145/3543507.3587430> (visited on 02/10/2024).

- [95] Hamed Zamani, Johanne R. Trippas, Jeff Dalton, and Filip Radlinski. “Conversational Information Seeking”. In: *Foundations and Trends® in Information Retrieval* 17.3-4 (2023), pp. 244–456. URL: <https://www.nowpublishers.com/article/Details/INR-081> (visited on 03/06/2024).
- [96] Oleg Zendel, J. Shane Culpepper, Falk Scholer, and Paul Thomas. “Enhancing Human Annotation: Leveraging Large Language Models and Efficient Batch Processing”. In: *Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval*. Sheffield United Kingdom: ACM, Mar. 2024, pp. 340–345. ISBN: 9798400704345. DOI: [10.1145/3627508.3638322](https://doi.org/10.1145/3627508.3638322). (Visited on 03/28/2024).