From To-Do to Ta-Da: Transforming Task-Focused IR with Generative AI

Chirag Shah
University of Washington
Seattle, WA, USA
chirags@uw.edu

Ryen W. White
Microsoft Research
Redmond, WA, USA
ryenw@microsoft.com

Abstract

For decades, scholars have emphasized that *tasks* should be the central focus in Information Retrieval (IR). This point of view holds even more significance with the advent of Generative Artificial Intelligence (GenAI) models, which can, among other capabilities, understand natural language, engage in dialog with users, generate bespoke user interfaces, and power agents to help complete tasks. GenAI presents an unprecedented opportunity to finally realize the potential of tasks in IR, enhance task-focused retrieval and interaction, and create "magical" task completion moments for users. In this paper, we explore the rationale and methodology behind this argument. Traditional IR systems support mostly simple tasks. The emergence of GenAI creates an opportunity for IR systems to help users achieve *complex* tasks and for the IR community to rekindle its interest and demonstrate leadership in this sizable and significant problem space. We underscore the pivotal role of tasks in IR and introduce new evidence supporting the notion that task-centric approaches, abstracted from specific modalities, represent the future of IR. Building on this foundation, we envision the development, utilization, and evaluation of next-generation IR systems. We propose a promising future where IR *agents* prioritize users, their tasks, and their situations. However, despite their potential to address task-focused and modality-independent IR, agents alone are insufficient. We propose a robust ecosystem around these agents that transcends traditional queries, questions, prompts, and modalities to address users' fundamental needs, tasks, and goals.

CCS Concepts

• **Information systems → Task models**; **Personalization**; • **Computing methodologies → Planning and scheduling**.

Keywords

Tasks, Modalities, Agents, Generative AI

ACM Reference Format:
Chirag Shah and Ryen W. White. 2025. From To-Do to Ta-Da: Transforming Task-Focused IR with Generative AI. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25), July 13–18, 2025, Padua, Italy*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3726302.3730352

SIGIR '25, July 13–18, 2025, Padua, Italy

ACM ISBN 979-8-4007-1592-1/2025/07
https://doi.org/10.1145/3726302.3730352

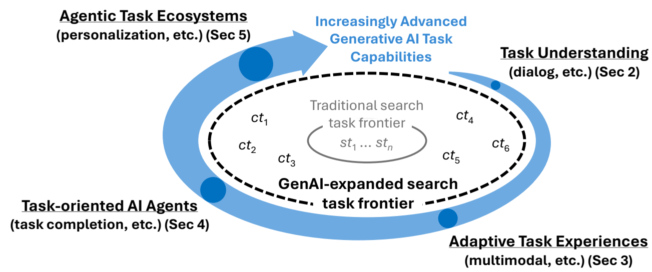

Figure 1: Range of GenAI-based task capabilities of increasing sophistication discussed in this paper, from Section 2 (task understanding) to Section 5 (agentic task ecosystems). Adding these capabilities to task-focused IR systems significantly expands the task frontier, depicting the range of tasks that the system can directly help users tackle. *st* and *ct* denote examples of simple and complex tasks respectively.

1 Introduction

People seek information to bridge gaps in their knowledge [8], solve problems [23], and satisfy their curiosity [95]. Generally, information seeking is purpose-driven. Many scholars have identified this purpose as a *task* and argued that the primary goal of an IR system should be to help people accomplish their tasks [78]. Despite such arguments and recommendations over several decades, IR systems still lack the core ability to fully address and support user tasks. Paradigms such as query-based retrieval, question-answering, and answer generation have served as substitutes for task-focused retrieval and interaction. While these approaches have performed well, recent scholarship has highlighted their limitations [66].

The emergence of GenAI creates new opportunities for task-focused IR. However, recent advances in integrating GenAI into the search process, e.g., via Retrieval Augmented Generation (RAG) [42] have still focused on the traditional query-response paradigm (e.g., the TREC RAG track[1]). There is a pressing need to explore creative ways to use transformational GenAI technologies to facilitate more efficient and effective search task completion.

Beyond addressing the limitations of existing IR paradigms, as argued by various scholars as above, we believe that there are bold new opportunities from focusing on tasks. However, our approach diverges from the traditional arguments presented in these works. Figure 1 illustrates the ongoing expansion of the task frontier (the set of tasks that an IR system can support) from predominantly simple tasks, to both simple and complex tasks that these new capabilities unlock, and the core GenAI advances of increasing sophistication (task understanding, etc.) that we cover in this paper.

[1] https://trec-rag.github.io/

We envision tasks as central elements, independent of any specific system or modality, and explore how they should be understood and supported. In this context, we present findings from a recent survey demonstrating that information seekers select interaction modalities based on their tasks and, more importantly, desire IR systems equipped with *adaptive task experiences* to seamlessly address their tasks across modalities (e.g., spanning search *and* chat). Ideally, users could rely on such a system to determine the most appropriate modality, granularity, and interaction fidelity.

Moving beyond search task interaction, users may want the system to help directly action some (or all) aspects of the task. This desire creates a compelling case for *intelligent agents*, as has been argued previously [86, 99]. However, merely building agents to support tasks is insufficient. Many informational tasks are complex, requiring personal and private information, and span multiple domains, skills, and modalities. Issues of trust, agency, and reliability arise, along with trade-offs between privacy and personalization, ability versus control, and bias versus customization. We must avoid solving one set of problems only to create another. Agentic IR systems must be designed from a holistic perspective. To this end, we propose a novel *agentic task ecosystem* for building, deploying, and evaluating agents. Ultimately, we emphasize the importance of returning to our foundational goal of addressing tasks, now equipped with new GenAI tools and a renewed vision.

2 Task Understanding

A task is typically viewed as a series of interconnected physical, cognitive, and emotional actions that individuals undertake to achieve specific goals in their professional or personal lives [3, 14, 79]. They involve various activities, constraints, and timeframes. Users of task management applications benefit from assistance with task planning [10] and prioritization [56]. Recent progress in task intelligence includes discovering digital assistant capabilities [85], estimating task durations [90], and auto-tracking task status [91].

Within the realm of IR, the notion of a task has gained particular significance, emphasizing the understanding and facilitation of information seeking and searching behaviors [66]. The focus of this prior work has largely been on *understanding* tasks and applying that understanding to improve IR system performance. We proceed with a brief overview, examine notable efforts to integrate task knowledge into IR, and conclude with our perspectives on how task understanding is transforming with GenAI (e.g., through multi-turn dialog with the IR system to help resolve uncertainty).

2.1 Tasks in IR

Early research in IR related to tasks can be traced back to the cognitive perspective in IR [9], which focused on understanding the motivations behind information seeking and searching. This perspective influenced works by Vakkari [78] and Ingwersen and Järvelin [32], who examined tasks in the design of IR systems to determine their purpose [64]. These studies provided insights for personalizing information search based on the task at hand. Vakkari [78] developed a framework of task-based information searching comprising three stages: *pre-focus*, *focus formulation*, and *post-focus*.

Tasks are often viewed as multi-level information seeking processes where individuals need information to achieve a goal [e.g., 16, 64, 65, 77]. Many task models [e.g., 19, 38, 43] have identified searchers' tasks as static, overarching goals that drive search actions. However, tasks evolve over time and with changing cognitive states. Different characteristics or facets of tasks [43] influence interactions with intelligent systems, such as search engines [47], and inform system design [76]. Search tasks are influenced by the work or everyday life tasks that drive information seeking or are associated with problematic situations [15].

Identifying tasks at various levels has been a key focus. Broder [12] proposed that search engine engagement could be categorized into three types: *informational*, *transactional*, and *navigational*. This taxonomy has been widely used to classify tasks according to search behaviors and support search studies. Rose and Levinson [63] extended Broder's scheme by specifying types of information goals and adding a new goal type (*resource*). They tested their scheme by classifying search engine queries.

Early works explored various aspects of tasks that influence search behaviors, including task complexity [e.g., 16], task difficulty [e.g., 39], and work context [e.g., 25]. Others examined the interactive and dynamic nature of search tasks [e.g., 6]. While performing tasks, searchers' actions are driven by intentions and can be well-defined or ill-defined [32]. These studies highlight the close association between task performance, information needs, search strategies, and document relevance and utility.

Not all task-related research in IR focuses explicitly on modeling or using tasks. Examples include research on task trails [44], personalized search [88], trail recommendation [72], cross-session tasks [83], task continuation [1], and cross-device tasks [84]. See Shah et al. [66] for a detailed overview of research on tasks in IR.

2.2 Applying Task Knowledge in IR

Applications of task knowledge in IR have shown that task representations can enhance query suggestions [4], build user models for improved personalized search [50, 88] and recommendations [100], and aid in satisfaction prediction [28, 82]. Mehrotra et al. [49] employed a tensor-based approach, representing each user as a combination of their topical interests and search task behaviors for personalization. Other studies have developed novel task context embeddings to represent queries via search logs, facilitating task-based personalization, query suggestion, and re-ranking [50, 51]. Tolomei et al. [75] explored task flows and analyzed query logs to generate task-based query suggestions, while Baraglia et al. [5] introduced search shortcuts to drive goal attainment.

Vu et al. [81] also utilized tasks to model user interests in search. Similarly, other researchers have leveraged task information to provide long-term support for task completion [e.g., 1, 35, 90]. Cai et al. [17] used task models to improve the ranking of retrieved search results, offering task-based support to users. Tasks help users achieve their search goals and assess a system's competency in assisting them. Hassan et al. [28] used search task constructs to predict satisfaction, while White and Kelly [92] employed them to enhance relevance feedback. Song and Guo [73] demonstrated that task information could automate tasks, reducing user burden.

Other researchers have focused on assistive systems, such as tours or trails, to guide users through their search process [29, 52, 59], predicting users' next search actions based on current actions, either by forecasting the next result click [18] or by predicting short-term interests based on task topic information [87].

2.3 Tasks in the Era of Generative AI

The search task frontier [86] (also see Figure 1), depicting the aspects of the task process that IR systems can support, continues to expand with GenAI, and is making significant inroads in task understanding, task completion, and automation in particular. This expansion is paving the way for more sophisticated and effective task management systems and reimagining task-focused IR.

We are moving towards a future where IR systems have the capability to work directly with tasks in ways that were previously challenging. In the past, we relied on approximations of tasks from users in the form of keyword queries or, at best, natural language questions. Now, with the advent of GenAI-based systems, users can express their tasks fully and freely and systems can reason over complex task descriptions and maintain state across multiple iterations of dialog with users. This iterative dialog can be freeform or driven by the system through AI-generated follow-up questions to improve the system's understanding of the user's task. A system component to do this could plug into any modular agentic AI workflow to first ensure that the task representations are accurate and fully-formed before proceeding with downstream processing.

In practice, the complexity of a task often makes it difficult for users to clearly express their needs, as they may not be familiar with the extensive problem space, let alone the solution space. Examples include significant purchasing decisions (such as buying a home or making financial investments), conducting market research in countries with different languages and cultural norms, and policymaking that involves multiple stakeholder groups. In these scenarios, a user's task goal is ambiguous and highly susceptible to external influences. It is important to note that GenAI systems may still struggle in handling such situations. In fact, the linguistic fluency and visual/interactive appeal of GenAI systems may increase the risk of steering users towards suboptimal goals or making premature decisions more quickly than traditional search engines. The likelihood of such outcomes must be taken into account when designing the interaction models for these systems.

3 Adaptive Task Experiences

As the focus shifts from user actions such as queries or questions sent to a specific interface or system such as a search engine or a voice-based personal assistant to tasks that are expressed in natural language, potentially spanning several dialog turns, we can also start abstracting out from those specific IR systems and modalities to envision a future where systems and users adapt the modalities that they offer and use depending on the task [94]. Therefore, as the next step in our perspective, we explore how modalities of interaction play a role in task execution, how users could work with and across multiple modalities to address their information needs, and how systems could leverage different modalities to assist.

An information interaction *modality* defines how users engage with a system, encompassing various interaction paradigms (e.g., query-response, multi-turn dialog, proactive suggestions) and various input mechanisms (e.g., text, speech, touch) across different device types. For decades, mainstream search interfaces in search engines have primarily offered a single modality: query-response, or simply search. This includes text queries, hyperlinked results, clicks to landing pages, and iterative query refinement to find relevant information. While search engines have been a primary information source for consumers, emerging modalities powered by advanced models such as Google's Gemini and OpenAI's GPT-4o, which can reason over text, audio, and visual inputs, and generate text and multimedia outputs, can complement search engine capabilities and create new possibilities for information interaction [86].

To understand how people use different modalities for their informational tasks, we surveyed a few hundred crowdworkers. For this exploratory study, we focused on two currently prevalent modalities for search: traditional keyword-based search, and AI-infused chat-based search. After data cleaning to remove bots and low-quality responses, we had 130 responses to analyze. Respondents' reported genders were male (60%), female (39%), and other (1%). They were all from the United States, English-speaking, and from various professions and educational backgrounds. Respondents reported using search and chat to accomplish a wide range of personal and professional tasks. Each worker with a valid response was paid $4 USD for 15 minutes of their time. This study was reviewed and approved by the Institutional Review Board at Microsoft Research. Future work needs to consider other modalities (e.g., proactive recommendations, summarization/synthesis per recent advances on in-depth, multi-step research[2,3]), employ a field study to collect data from users with real information needs, and also consider other task characteristics (e.g., single- vs. multi-user).

We divide interactions information interactions into four types—*monomodal*, *multimodal*, *crossmodal*, and *transmodal*—depending on the task support and the interactions they enable.

3.1 Monomodal Information Interaction

In monomodal interactions, individuals select a single modality to complete a task. To understand their choices among available modalities, we presented respondents with six specific tasks that varied in complexity and were inspired by our prior work that looked at information seeking trends in search and chat modalities [69]. Figure 2 shows these tasks with responses about the most appropriate modality between search and chat that the participants selected. Respondents overwhelmingly chose search engines for informational and planning tasks (T1, T3), while they preferred chat-based modalities for understanding/learning and creation tasks (T2, T4). Additionally, they favored chat-based tools for learning tasks (T6), although search remained an important modality for these tasks as well as for medical diagnosis (T5).

Respondents reported a preference for traditional search engines when seeking more options, detailed information, or diverse and comprehensive content. Familiarity also plays a significant role in users' preference for search engines over other modalities for their information needs. Conversely, respondents found chat-based systems more effective for obtaining direct and quick answers to specific questions, without the need to sift through unrelated information. Additionally, respondents preferred chat-based modalities for tasks involving learning, synthesis, or creation.

While these preferences highlight general tendencies, we aimed to understand how respondents leverage access to multiple modalities to accomplish real tasks. This deeper exploration seeks to uncover the nuanced ways in which searchers integrate different modalities to meet their diverse information needs.

[2]https://openai.com/index/introducing-deep-research/
[3]https://gemini.google/overview/deep-research/

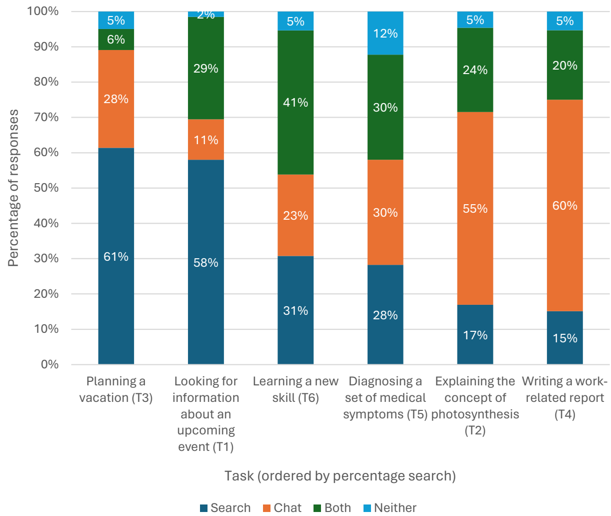

Figure 2: Distribution of modalities selected by respondents for various task examples. Tasks are ordered from left to right by the percentage of responses favoring search.

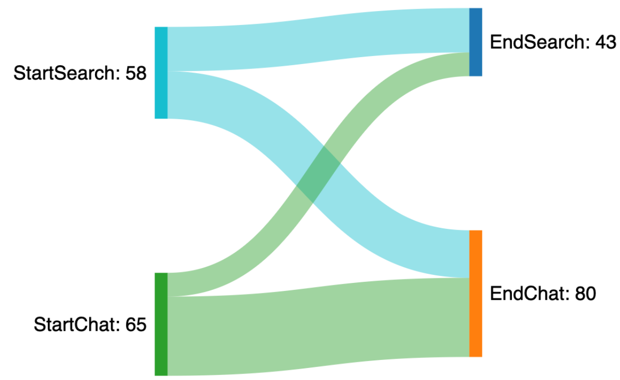

Figure 3: Evidence of crossmodality information interactions as people move from one modality to another to accomplish a task. The flow from search to chat is much greater than chat to search. The numbers are the number of respondents in each start and end state. A small number of ambiguous responses (n=7) were removed from the data for this analysis.

3.2 Multimodal Information Interaction

Multimodal interfaces have been studied for decades in the human factors community [60], including for information interactions.[4] Search systems have long provided various facilities and search strategies based on user needs and goals [21], and even suggested specific search engines based on the current query [93].

As mentioned earlier, search engines now offer multiple modalities, such as search and chat, juxtaposed in the search interface. This is available on Bing and in Google as *AI Overviews*.[5] Our survey findings indicate that the search mode is preferred for many personal tasks, such as finding recipes, planning vacations, and shopping. Conversely, chat-based modalities dominate professional tasks, such as report generation, coding, and researching relevant business topics. Insights from our survey shed light on how respondents choose each modality and what they should consider when making their selection decisions (Table 1).

For users, making informed selections about which modality to use also depends on sound mental models of system capabilities. People have developed mental models of IR systems over decades of use [11, 55]. However, for chat-based systems, users' mental models are still evolving, and they may experiment with different prompt formulations to tailor the system output to meet their expectations. Presenting both GenAI-powered chat experiences and traditional search results on a single SERP, can be imprecise and overwhelming, potentially impeding task progress.

3.3 Crossmodal Information Interactions

While users have preferences for a modality based on the task or their prior experiences, there are instances when their initial choice does not suffice, or they may require capabilities from a different modality, necessitating a switch. We refer to these as *crossmodal information interactions* and have investigated when and why users switch modalities. Crossmodal interactions have been explored in the context of search (e.g., text and voice-based input/output, different device types [53]), but transitions between search and chat are less well understood and supported by current IR systems.

Our survey responses revealed that respondents switch from search to chat when the search results are not sufficiently interactive or tailored to their needs. They also found switching from search to chat beneficial when they required detailed or highly specific information. Conversely, respondents switch from chat to search when they believe that search will provide a more comprehensive and diverse set of results than what they received through chat.

Figure 3 illustrates the frequency with which respondents switch between modalities in the tasks they reported in our survey. When users start with search, they are almost equally likely to continue with search or switch to chat. Conversely, when starting with chat, they are much less likely to switch to search. This disparity in crossmodality can be attributed to the strengths of the chat-based modality, such as its ability to provide direct answers that reduce the need for further exploration and its interactive nature, which allows its users to effectively reshape their requests across dialog turns to achieve their desired task outcomes.

We observe that current mainstream search experiences are multimodal but offer very limited crossmodal support (e.g., GenAI-powered chat on SERPs). An IR system could actively engage with users to assist with switching decisions, suggest a suitable modality for the task or task stage, or proactively enlist agents to try alternative modalities and share prospective outcomes in advance of a switch. Switching considerations extend beyond technological solutions; they also involve modality awareness, which can influence usage habits. Additionally, the time spent with the chat modality is crucial for users to learn how to best utilize its functionality.

[4] In the context of AI, "multimodal" refers to using different data types—such as text, images, video—while in human factors, it describes various modes of user interaction. Since this section focuses on user experience, we adopt the human factors definition.
[5] https://blog.google/products/search/generative-ai-google-search-may-2024/

Table 1: Summary of reasons to use and challenges in using the two primary modalities (search and chat) offered by current commercial search systems, as derived from qualitative responses to survey questions.

Search modality	
USE THIS MODALITY when need for: (1) **Broad and Unprompted Information:** Search engines often provide related or historical data that was not explicitly requested. (2) **Trustworthy and Verified Sources:** For academic or scholarly research, search engines provide direct access to authentic academic journals and books. (3) **Product Research and Shopping:** When deciding on a purchase, search engines provide detailed reviews, user experiences, specs, and multimedia content such as images and videos. (4) **Multiple Sources and Opinions:** Search engines are preferred to see multiple opinion-based pieces or reviews on a possible purchase or topic. (5) **Real-Time, Current, and Location-Based Information:** For real-time data such as movie times, weather, game scores, or election results, search engines are more effective as they can provide up-to-date information. (6) **Specific and Quick Information:** For quick and simple information, such as restaurant details, holiday dates, or trivia, search engine are fast.	**CHALLENGES** with using modality include: (1) **Irrelevant Results and Overwhelming Information:** Search engines could have results that are unrelated to specific queries or could overwhelm the user with too much information, irrelevant results, or excessive advertisements. (2) **Inadequate Detail or Specificity:** Search engines may not be good for a need for detailed or highly specific information. (3) **Lack of Interactivity, Personalization, and Creativity:** Search engines could be inadequate in cases when an interactive experience is needed, such as asking follow-up questions or having a real-time discussion to solve a problem. (4) **Difficulty in Finding Specific Items, Services, or Paywalled Content:** Search engines can be inadequate when users have difficulty finding specific items, services, or parts, or when they are unable to access paywalled or restricted content. (5) **Biased or Misleading Information:** Search engines may provide information that is biased, misleading, or inconsistent.
Chat modality	
USE THIS MODALITY when need for: (1) **Direct and Quick Answers:** Chat-based systems are more efficient for direct and quick answers to specific questions, without having to sift through unrelated information. (2) **Complex Queries and Specialized Knowledge:** For high-level topics or complex queries that require specialized knowledge, chat-based systems can be helpful in summarizing and clarifying points. (3) **Personalized Assistance and Decision Making:** Chat-based systems can provide dynamic and personalized assistance provided when making decisions or seeking advice tailored to their specific situations. (4) **Interactive Troubleshooting and Real-Time Engagement:** When dealing with technical issues or problems with devices, chat-based systems can be more effective in offering solutions and troubleshooting due to real-time interactions with ability to do back-and-forth. (5) **Creative and Original Content:** Chat-based systems are superior for help with creative tasks such as brainstorming ideas, writing essays, emails, reviews, or creating campaigns. (6) **Learning and Understanding:** Chat-based systems are beneficial for learning new concepts, languages, or understanding complex topics.	**CHALLENGES** with using modality include: (1) **Difficulty Following Instructions, Handling Complex Queries, and Providing In-Depth Information:** This includes instances where the chat-based system fails to adhere to specific instructions given by the user, leading to outputs that do not meet the user's requirements. (2) **Limited Creativity, Originality, Personalization, and Handling of Sensitive Topics:** This category encompasses scenarios where the chat-based system fails to generate creative, original, or personalized content. (3) **Inadequate for Quick Searches, General Information, and Real-Time Updates:** This category includes scenarios where chat-based systems could be slower or more complicated than traditional search engines for quick and simple queries. (4) **Inadequate for Specialized Support and Complex Scenarios:** This category includes scenarios where the chat-based system fails to provide specialized support for technical issues, complex software applications, or specific customer service needs.

3.4 Transmodal Information Interaction

As has been evident, no single modality of information interaction can satisfy users' full range of informational needs. More importantly, strategically using multiple modalities can help users accomplish their tasks more effectively and unlock possibilities for tasks not traditionally performed through monomodal, multimodal, or crossmodal approaches. Respondents reported several such incidents, which we term *transmodal information interactions.*

Several respondents described complex information needs requiring dynamic support, such as using multiple modalities for the same task, either sequentially [53] or concurrently [89]; representing and preserving task state; and using GenAI to decompose tasks, plan actions, and select and sequence modalities. These needs often related to tasks in healthcare (e.g., diagnosis and mitigation), legal (e.g., seeking legal advice), or research (e.g., creating summary reports for a business need). For example, one respondent mentioned how a search engine could find examples of high-quality reports while a chat-based system could generate a draft, helping them learn about and accomplish the task more effectively.

However, most current interfaces lack clear support for running concurrent modalities with seamless integration and focus on the task at hand, leading respondents to perform such transmodal work through ad hoc workflows. Given the importance of supporting transmodal information interactions and the current lack of such support, there is a need for platforms to facilitate the use of multiple modalities and create user experiences with the most appropriate constellation of modalities and connections between them.

In many of these adaptive task experiences, the task execution burden remains with the user. Given recent advances in AI agents powered by GenAI, IR systems could invoke agentic workflows to accomplish some or all of a task on the user's behalf. In this scenario, agents become the new IR systems, transforming users' to-do lists into "ta-da" completion moments, regardless of the set of modalities used in the task so far. We now consider such task-oriented agents.

4 Task-oriented AI Agents

Instead of expressing their knowledge gaps, problems, or curiosity through queries, questions, or prompts, users could present their high-level tasks in natural language for agents to directly address. While using agents to accomplish tasks is not a novel idea [96], GenAI has renewed interest in leveraging agents for this purpose.

4.1 AI Agents

In the context of AI, an agent is an autonomous entity or program that takes preferences, instructions, or other forms of inputs from a user to accomplish specific tasks on their behalf. Agents can range from simple systems, such as smart thermostats that adjust ambient temperature based on sensor readings, to complex systems, such as autonomous vehicles navigating through traffic. Key characteristics of agents include autonomy, programmability, reactivity, and proactivity. The area of agentic AI encompasses intelligent systems designed to operate with a high degree of autonomy, making decisions and taking actions independently of human intervention.

The evolution of AI agents can be traced through five distinct eras, each marked by unique architectures and specific challenges. In the 1950s, early AI agents such as the General Problem Solver [58] utilized symbolic reasoning but struggled with real-world complexity due to their reliance on predefined rules and lack of adaptability [31]. The 1980s introduced expert systems such as MYCIN and DENDRAL, which leveraged domain-specific knowledge for decision-making. Despite their effectiveness in narrow domains, these systems were brittle and could not generalize beyond their programmed expertise, limiting their broader applicability [37]. The 1990s saw the emergence of reactive agents, exemplified by Rodney Brooks' subsumption architecture, which prioritized real-time interaction over complex reasoning. However, these agents were limited by their inability to plan or learn from past experiences, reducing their effectiveness in dynamic environments [13, 48].

Multi-agent systems (MAS) introduced the concept of multiple interacting agents with specific roles, showing promise in distributed problem-solving [97]. Tasks are also at the heart of Google's new Agent-to-Agent (A2A) protocol.[6] However, MAS faced significant challenges in coordination, communication, and scalability, often leading to inefficiencies and unpredictable behaviors [98]. Cognitive architectures such as Soar and ACT-R attempted to model human cognition by integrating perception, memory, and reasoning. Despite their sophisticated designs, these architectures struggled with scalability and real-time performance, resulting in high computational costs and limited practical applications [41]. These historical efforts underscore the ongoing challenges in developing AI agents that are both adaptable and efficient in real-world scenarios.

[6]https://google.github.io/A2A/#/documentation?id=task

4.2 Agents in IR

In IR, agents are increasingly being leveraged to address complex search needs by integrating foundation models [86] and provide more tailored task support [99]. These agents enhance traditional IR systems by incorporating capabilities such as semantic search, context-aware retrieval, and real-time interaction. For instance, RAG systems combine the strengths of retrieval-based and generative models to provide more accurate and contextually relevant responses. By utilizing external knowledge bases and real-time data streams, RAG systems can dynamically adapt to user queries, significantly improving the relevance and precision of the retrieved information [71].

Agents are being deployed in IR systems to facilitate better coordination and communication among various components of the system, thereby enhancing the overall efficiency of IR processes [45, 101]. These agents can autonomously manage tasks such as document retrieval, summarization, and search result ranking, ensuring that the most pertinent information is presented to users [20, 26]. Recent advancements in dense retrieval techniques, such as Dense Passage Retrieval [46] and Sentence-BERT [62], have further improved the performance of AI agents in handling complex queries and multimodal content.

4.3 Limitations of Agents

By definition, agents remove agency from a user to perform tasks on their behalf, saving time and effort. However, this trade-off may be questioned if relinquishing control does not generate sufficient user value. Agents may also make mistakes, require intervention or supervision, and may be limited to performing only simple tasks. These shortcomings have been evident in agentic research and development over the past decade. Agents, often referred to as operators, skills, apps, extensions, and plugins, have been widely integrated into computers, smartphones, speakers, wearables, and automobiles. However, their utility has been severely limited [2, 40]. Beyond limited applications, there are persistent shortcomings that are not addressed by simply creating more capable systems.

AI agents have faced significant challenges [e.g., 27], including their inability to generalize across different domains due to reliance on predefined rules and lack of adaptive learning mechanisms, scalability issues where computational demands grow exponentially with task complexity, difficulties in coordination and communication within MAS and between agents and users, a lack of robustness in performance under unexpected conditions, and ethical and safety concerns that arise from biased decision-making and the trade-offs between agent control and user agency.

What, then, can we do to support user tasks with IR agents? Or as history has shown many times, will agents end up providing a limited use case scenario? We believe this time it is different for three reasons: (1) Advancements in various technologies, specifically encoder-only transformers, have made it possible to take natural language inputs in multiple situations and for a range of tasks (text classification, named entity recognition, coreference resolution, etc); (2) Foundation-model-powered systems are able to not only generate responses, but also reason to perform operations such as task decomposition; and (3) Many of the modalities have advanced and matured enough to allow for more seamless connections among them. Combined, these reasons have created an environment where we could develop and deploy agentic systems with much more applicability and reliability than before. However, as we recently argued [67], even such capable and advanced agents are not sufficient to transform how we tackle and complete tasks.

5 Agentic Task Ecosystems

Agents are compelling mechanisms to help us address user tasks, but as outlined in the previous section, they are not without their issues. We now answer the question: what will make agents successful? In addition to creating task-capable agents, we need to create a new ecosystem around them. This will not be a single research project, but a whole new direction for multiple generations of researchers and will require multidisciplinary collaboration.

5.1 Making Agents Successful

For agents to be successful, attention must be given to several critical aspects involving users, tasks, and contexts.

First, value generation is essential; agents must provide significant autonomous task execution benefits to users, outweighing the costs and risks involved. Frequent user intervention or privacy concerns can diminish the perceived value, making users reluctant to adopt these agents. Value is essentially the balance between the perceived benefits and the costs, such as time and privacy. Second, adaptable personalization is crucial. Agents must adapt to the unique needs and contexts of each user and situation. For example, if an agent needs to reset a password during an online transaction, it should be capable of handling this autonomously or seeking user input based on the task's context and the user's preferences. Third, trustworthiness is vital. Users must trust agents to handle sensitive tasks such as bank transactions and personal communications. This trust is developed gradually through consistent accuracy and transparency. Fourth, social acceptability is necessary for widespread adoption. Agents must be accepted across diverse populations, cultures, and customs, which may take time. Finally, standardization is needed to ensure compatibility, reliability, and security in a decentralized development environment. Standardizing deployment, connectivity, and service protocols will help create a sustainable ecosystem for agents, similar to networking protocols or app stores.

5.2 Going Beyond Agents

We posit that to overcome the challenges of agentic AI listed above, we need three specific mechanisms:

(1) A private and secure version of an agent to ensure user information is protected while making both private and public versions of agents tackle more meaningful/complex tasks.
(2) A representation of the user that can interact with an agent on a user's behalf so that the user does not have to keep intervening or providing frequent inputs.

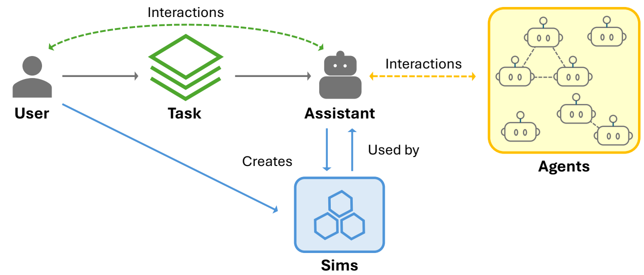

Figure 4: Envisioning a new agentic task ecosystem. Motivated by task, users work with their Assistant, which leverages Sims and Agents, and determines interaction modalities.

(3) Ability for an agent with intimate knowledge of users and their tasks to communicate and negotiate with other agents and/or tools/applications on a user's behalf to accomplish complex tasks without added burden on the user.

To put these recommendations into practice, we propose a new kind of ecosystem that is built around agents, but also includes other critical components that provide standardization, privacy, personalization, and increased trust. Specifically, we envision the ecosystem in Figure 4 that comprises Agents, Sims, and Assistants.

Agents are specialized, purpose-driven modules trained to perform specific tasks. Each agent operates autonomously but can interface with other agents with complementary capabilities as needed. Section 5.1 provides suggestions for improving agents.

Sims represent users, created from a combination of user profiles, preferences, and behaviors, capturing different aspects of the user. Unlike user personas or profiles, Sims carry information about the user's preferences, behaviors, privacy, and contexts. Sims are constantly updated as the user interacts with their environment, generating new information about their preferences and behaviors. In addition, Sims can be edited by the user's Assistant (a private agent) based on task execution and outcomes.

Assistants are programs that directly interact with the user, deeply understands them, and can call upon Sims and Agents as needed to reactively or proactively accomplish tasks and sub-tasks. Essentially, an Assistant is a private version of an agent, fine-tuned to the user, with access to personal information, allowing it to represent the user and to act on the user's behalf.

The interaction between Agents, Sims, and Assistants is highly synergistic. The Assistant, with its comprehensive understanding of the user, co-creates and manages Sims under the user's supervision, reflecting the user's multifaceted life. These Sims provide most accurate and current information about the user to the Assistant and task-oriented Agents to perform tasks efficiently. This layered approach ensures tasks are handled with more precision and with personalization, enhancing overall user satisfaction.

Also important is a representation of the user task that accurately and comprehensively depicts the user's needs, goals, and intents and is actionable by the agent(s). Figure 4 shows that the user interactions with the Assistant (two-way) are centered around the task. It all starts by the user first expressing a need, even if it is vague, to the Assistant, which in turn constructs a more detailed and personalized version of task description using relevant Sims(s). The

User: I would like to find some afterschool activities for Hannah.

Agent: I understand that you are looking for afterschool activities for your 9-year-old daughter Hannah and since you may need to take her to that activity, you would like to explore some options that meet your schedule and proximity. [REASONING]

User: Yes, but I'm not sure where to begin. Can you look into it?

Agent: Certainly. After doing some investigation, I found four classes in the area that Hannah would like, and you could manage within your schedule. [RESEARCH]. However, one of them is on Wednesday evening and I know some of those evenings you work late, so perhaps we should look for other day. [REASONING]

User: That's true. Can you plan out a full schedule for the other three activities and run it by Adam and also check for availability?

Agent: Sure. Here are three plans ... [GENERATION]. I will (1) share these plans with your husband Adam for his feedback, (2) contact the organizers of those classes on your behalf to find out about availability, and (3) bring back my final recommendation for your approval. [ACTION]

User: I want to make sure our beta is compliant with the latest regulations.

Agent: I researched [RESEARCH] and found that for the most part, your app follows the state and the federal regulations. However, there are a few areas where the legal guidelines are not very clear and while we are not in any violation, we should consider the implications. Here are my recommendations... [GENERATION, REASONING]

User: Can you run risk assessments on those and give me arguments for why we should take any actions and what it would cost, as well as what we could risk and save if we don't act?

Agent: Here are my analyses for costs, risks, and benefits of addressing those gray area compliance issues... [REASONING]

User: We need a second opinion on this. Can you figure out who to send this within our org and ensure that confidential information is only shared with them if they have the right permissions?

Agent: Sure. I will (1) find the appropriate people, (2) assess if they have the right clearance, (3) ask for their opinions, (4) synthesize those opinions and share with you. If needed, I can also set up meetings. [ACTION]

Figure 5: Examples of a user interacting with an agent for accomplishing informational tasks with appropriate incorporation of generation, research, reasoning, and action components (flagged in square brackets).

Assistant, then decides appropriate sets of Agents and information interaction modalities to use to accomplish that task.

5.3 Envisioning New Kinds of IR Agents

We envision agents that can understand not only the complex task, but also the user, and the context surrounding them. We imagine agents that could generate information, while also being able to reason, research, and take actions. Figure 5 provides two examples of what interactions with such agents could look like.

6 Task Futures

Beyond those already discussed, GenAI unlocks many opportunities for the future of task-focused IR. We cover some in more detail.

6.1 Task Support with Agents

As described above, agents that can complete tasks on behalf of users are rapidly becoming a reality. These agents include planning capabilities to decompose a complex task into a set of sub-tasks [61], some of which can be completed by the system and some that may require user input or actions. These agents can also be specialized and work together with each other in MAS and directly with humans (not via Assistants) to tackle complex tasks [86, 97].

Agents also signal an expansion of search engine capabilities and more recently answer engine capabilities. *Action engines* [57] are emerging, where some search tasks can be completed by agents operating on behalf of users. Tasks that are repetitive, menial, and/or low stakes may be especially suited to this type of support. Action engines could also draw on some of the principles of IR-system design, such as inverted indices mapping tasks/intents to action sequences as a caching solution to avoid high costs from running agents for every task at massive scale and transparent user interfaces that communicate task progress and collect user feedback.

Agent design principles from mixed-initiative systems [30] will help ensure that users maintain control over agents. We need to consider automation and control, and since it is not zero-sum, increase automation while maintaining human control [70]. Users need to stay involved in the process of oversight and to guide the system as needed; that will help build trust in autonomous operation but does not mean that systems should be entirely autonomous. Users need to engage with the retrieved information for learning purposes [80] and removing access to that information could be detrimental for their cognitive development. Also, pragmatically, with agents capable of task automation, there are many issues associated with safety, privacy, and reliability, including validation, robustness, and risk mitigation, that need to be considered in the design process.

Adaptation of the agents and any task support to individuals (personalization), their context (contextualization), and the task at hand (task adaptation) is crucial so that they factor in variables that should affect their actions and outputs. Personalized planning may also be valuable, reducing load on the user to collect already-provided task information and load on the system if elements of the task have already been completed previously. IR agents can also use tools, APIs, and other applications to help complete tasks. This information could be encoded in the foundation models that power the agent, but also in knowledge bases and search results that augment/ground agent responses and actions. To do that, the agent needs to know what else is available, how to use it, and when to send requests where. Different modalities in search engines offer users various capabilities to complete tasks, with search and chat being two prominent examples. Recommendations on the types of capabilities that could or should be used depending on the characteristics of the task are emerging [68]. The system could also pick the best modality or generate experiences that combine the available modalities in crossmodal and transmodal experiences, as noted above, to help users best tackle different facets of a task.

Long-term memory will enable agents to personalize their actions and support tasks that span sessions over days and weeks. The additional time between user interactions can afford the system greater opportunity to generate more complete answers, find the best resources, coordinate with humans, use additional tools, and so on. Lessons and ideas from slow search [74], which advocates for taking more time to find the most relevant search results, can be applied here too. These systems will be able to track their task progress over time and report back to users for feedback. Asynchronous task completion will also enable systems to operate when users may not be present, such as reserving class activities when a registration website opens at an inconvenient time. Doing this well opens up a new class of challenges around system *preparedness*, where the agent needs to collect all of the required information in advance of performing the task since it will be unable to engage the user at runtime. Task histories reflecting user preferences may also ensure the agent has the required data should questions arise.

6.2 Task Support in General

GenAI is expanding the searcher-system boundary [7] at both ends of the traditional IR search process: moving from queries and prompts to tasks and intents on the input and moving from results to first answers and now actions/decisions on the output. This expansion unlocks many opportunities to do more for users, but also to engage them more in the task completion process (e.g., via dialog to iteratively refine the task representation, to allow an agent to observe and henceforth repeat/optimize their task completion processes, to provide feedback on agent actions and output).

Task support that anticipates user needs and completes tasks on their behalf will become increasingly important. This support might be especially useful for tasks that are repeated often, leading to the emergence of standing tasks similar to how standing queries [54] have already been discussed in the context of search engines. We may also see the emergence of proactive experiences that engage the user without direct initiation and promote the need to complete a task given, say, an external prompt such as a deadline, and even offer to complete it on behalf of the user. These recommendations should be offered with explanations for greater transparency, and to increase trust and drive uptake. Continual learning of task preferences and efficient task solutions over users, tasks, and time will help task support to be more useful, as seen with search engines as they learn from aggregated user feedback [e.g., 34].

While the focus has been on tasks in the digital world, we are starting to see more support for tasks in the physical world through progress in embodied AI agents and robotics. Systems with world and human action models [e.g., 36] that can handle multimodal inputs are essential in making this transition from digital to physical and from text generation to actuation. Mixed reality systems are already supporting humans in completing physical tasks with overlays in, e.g., headsets, smartphones, and smart glasses. Soon, AI systems will include physical elements in their task completion processes, interacting with the physical environment to extend the task frontier even further and into the analog domain.

6.3 Evaluation

Evaluation in task-focused IR is an important area for innovation. Evaluation is one area of strength for the IR community and, given required datasets containing, say, tasks, activities, and outcomes, task-focused metrics and methodologies could be devised to help drive progress in this next wave of task-centric systems.

Task-based evaluation metrics have been proposed in IR for some time [33]. With the advent of GenAI, there is an opportunity/need to revisit this area with a special focus on task (e.g., task representations, task completion efficiency, task completion accuracy). The presence of a task completion signal from an agent provides, finally, when compared to the various proxies that the community has relied on (e.g., lengthy dwell times on documents [24]) a clear sign of task success, but completion on its own is also insufficient: different degrees of success need to be considered as well (e.g., booking any seat on an airplane versus booking one that aligns with user preferences, is well-positioned on the airplane, satisfies cost constraints, etc.). For cases where the task can be successfully completed with an AI-generated answer and no further actions are needed, the success metrics are less clear but prior work on estimating good abandonment in SERP interactions [e.g., 22] may provide some inspiration for how to estimate task success.

The generation of sets of simulated users and their synthetic task completion activities could also help drive progress in this area more quickly. Given the complexities in GenAI-powered task understanding, adaptive task experiences, and task agents and ecosystems, TREC-style batch evaluation tailored to the query-response paradigm may also be insufficient to measure system performance.

7 Concluding Remarks

The emergence of GenAI can revitalize the task movement within IR. Tasks have been discussed in the community for decades, but we have lacked the technological basis for deep task understanding and negotiation with users and an ability to translate that understanding into action through agents in a generalizable way across a range of tasks. As a call to action, we encourage the IR community to dive deeply into task-focused support once again given new GenAI capabilities, perhaps following a paced trajectory (e.g., from learning/research tasks to large-scale planning/analytics tasks).

In doing all this, and in carefully thinking through more complete task understanding and task automation, we must not lose sight of the central role that humans play in tasks. Humans provide the input and consume any output from task execution, so are the arbiters on whether the task has been successfully completed. While our role in the process is being reshaped by GenAI, humans are still central in guiding AI systems and learning from the resources accessed and actions taken to complete tasks. Systems need to adapt their support per task demands and we still must consider degrees of system automation and human control. There are many open questions on task support and the IR community is pivotal in shaping (complex) task completion futures in the GenAI era.

References

[1] Eugene Agichtein, Ryen W. White, Susan T. Dumais, and Paul N. Bennett. 2012. Search, interrupted: understanding and predicting search task continuation. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 315–324.

[2] Emily Couvillon Alagha and Rachel Renee Helbing. 2019. Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. *BMJ Health and Care Informatics* 26, 1 (2019).

[3] David Allen. 2015. *Getting Things Done: The Art of Stress-free Productivity*. Penguin.

[4] Ahmed Hassan Awadallah, Ryen W. White, Patrick Pantel, Susan T. Dumais, and Yi-Min Wang. 2014. Supporting complex search tasks. In *Proceedings of the ACM CIKM Conference on Information and Knowledge Management*. 829–838.

[5] Ranieri Baraglia, Fidel Cacheda, Victor Carneiro, Vreixo Formoso, Raffaele Perego, and Fabrizio Silvestri. 2009. Search shortcuts: Driving users towards their goals. In *Proceedings of the International Conference on the World Wide Web*. 1073–1074.

[6] Marcia J Bates. 1989. The design of browsing and berrypicking techniques for the online search interface. *Online Review* 13, 5 (1989), 407–424.

[7] Marcia J Bates. 1990. Where should the person stop and the information search interface start? *Information Processing and Management* 26, 5 (1990), 575–591.

[8] Nicholas J Belkin. 1980. Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science* 5, 1 (1980), 133–143.

[9] Nicholas J Belkin. 1990. The cognitive viewpoint in information science. *Journal of Information Science* 16, 1 (1990), 11–15.

[10] Victoria Bellotti, Brinda Dalal, Nathaniel Good, Peter Flynn, Daniel G Bobrow, and Nicolas Ducheneaut. 2004. What a to-do: Studies of task management towards the design of a personal task list manager. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. 735–742.

[11] Christine L Borgman. 1985. The user's mental model of an information retrieval system. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 268–273.

[12] Andrei Broder. 2002. A taxonomy of web search. *ACM SIGIR Forum* 36, 2 (2002), 3–10.

[13] Rodney Brooks. 1986. A robust layered control system for a mobile robot. *IEEE Journal on Robotics and Automation* 2, 1 (1986), 14–23.

[14] Katriina Byström. 2007. Approaches to "task" in contemporary information studies. In *Proceedings of the International Conference on Conceptions of Library and Information Science*, Vol. 12. 1–10.

[15] Katriina Byström and Preben Hansen. 2005. Conceptual framework for tasks in information studies. *Journal of the American Society for Information Science and Technology* 56, 10 (2005), 1050–1061.

[16] Katriina Byström and Kalervo Järvelin. 1995. Task complexity affects information seeking and use. *Information Processing and Management* 31, 2 (1995), 191–213.

[17] Fei Cai, Shangsong Liang, and Maarten De Rijke. 2014. Personalized document re-ranking based on bayesian probabilistic matrix factorization. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 835–838.

[18] Huanhuan Cao, Daxin Jiang, Jian Pei, Enhong Chen, and Hang Li. 2009. Towards context-aware search by learning a very large variable length hidden markov model from search logs. In *Proceedings of the International Conference on the World Wide Web*. 191–200.

[19] Robert Capra, Jaime Arguello, Heather O'Brien, Yuan Li, and Bogeum Choi. 2018. The effects of manipulating task determinability on search behaviors and outcomes. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 445–454.

[20] Vinay Chavan and Sanjay E Yedey. 2005. Coordination Model for Communication Among Agents in Multi-Agent Based Intelligent Systems. *Recent Advances in Information Technology* (2005), 210.

[21] W Bruce Croft and Roger H Thompson. 1987. I3R: A new approach to the design of document retrieval systems. *Journal of the American Society for Information Science* 38, 6 (1987), 389–404.

[22] Abdigani Diriye, Ryen White, Georg Buscher, and Susan Dumais. 2012. Leaving so soon? Understanding and predicting web search abandonment rationales. In *Proceedings of the ACM CIKM International Conference on Information and Knowledge Management*. 1025–1034.

[23] Joan C Durrance, Maria Souden, Dana Walker, and Karen E Fisher. 2006. Community problem-solving framed as a distributed information use environment: bridging research and practice. *Information Research: An International Electronic Journal* 11, 4 (2006), n4.

[24] Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan Dumais, and Thomas White. 2005. Evaluating implicit measures to improve web search. *ACM Transactions on Information Systems* 23, 2 (2005), 147–168.

[25] Luanne Freund, Elaine G. Toms, and Charles L.A. Clarke. 2005. Modeling task-genre relationships for IR in the workplace. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 441–448.

[26] Frans CA Groen, Matthijs TJ Spaan, Jelle R Kok, and Gregor Pavlin. 2007. Real world multi-agent systems: information sharing, coordination and planning. In *Proceedings of the 6th International Tbilisi Symposium on Logic, Language, and Computation*. Springer, 154–165.

[27] Shanshan Han, Qifan Zhang, Yuhang Yao, Weizhao Jin, Zhaozhuo Xu, and Chaoyang He. 2024. LLM multi-agent systems: Challenges and open problems. *arXiv preprint arXiv:2402.03578* (2024).

[28] Ahmed Hassan, Xiaolin Shi, Nick Craswell, and Bill Ramsey. 2013. Beyond clicks: query reformulation as a predictor of search satisfaction. In *Proceedings of the ACM CIKM Conference on Information and Knowledge Management*. 2019–2028.

[29] Ahmed Hassan and Ryen W White. 2012. Task tours: Helping users tackle complex search tasks. In *Proceedings of the ACM CIKM Conference on Information and Knowledge Management*. 1885–1889.

[30] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. 159–166.

[31] Yu Huang. 2024. Levels of AI agents: From rules to large language models. *arXiv preprint arXiv:2405.06643* (2024).

[32] Peter Ingwersen and Kalervo Järvelin. 2006. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer.

[33] Kalervo Järvelin, Pertti Vakkari, Paavo Arvola, Feza Baskaya, Anni Järvelin, Jaana Kekäläinen, Heikki Keskustalo, Sanna Kumpulainen, Miamaria Saastamoinen, Reijo Savolainen, et al. 2015. Task-based information interaction evaluation: The viewpoint of program theory. *ACM Transactions on Information Systems* 33, 1 (2015), 1–30.

[34] Thorsten Joachims. 2002. Optimizing search engines using clickthrough data. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 133–142.

[35] Rosie Jones and Kristina Lisa Klinkner. 2008. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *Proceedings of the ACM CIKM Conference on Information and Knowledge Management*. 699–708.

[36] Anssi Kanervisto, Dave Bignell, Linda Yilin Wen, Martin Grayson, Raluca Georgescu, Sergio Valcarcel Macua, Shan Zheng Tan, Tabish Rashid, Tim Pearce, Yuhan Cao, Abdelhak Lemkhenter, Chentian Jiang, Gavin Costello, Gunshi Gupta, Ryen W. White, Shun Ishida, Cecily Gault, Aidan Curtis, Marko Vukolic, Tarun Gupta, and Katja Hofmann. 2025. World and human action models towards gameplay ideation. *Nature* 638 (2025), 656—-663.

[37] Sayash Kapoor, Benedikt Stroebl, Zachary S Siegel, Nitya Nadgir, and Arvind Narayanan. 2024. AI agents that matter. *arXiv preprint arXiv:2407.01502* (2024).

[38] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-ching Wu. 2015. Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. In *Proceedings of the ACM ICTIR International Conference on the Theory of Information Retrieval*. 101–110.

[39] Jeonghyun Kim. 2006. Task difficulty as a predictor and indicator of web searching interaction. In *ACM SIGCHI Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 959–964.

[40] Miriam Kurz, Birgit Brüggemeier, and Michael Breiter. 2021. Success is not final; failure is not fatal–task success and user experience in interactions with Alexa, Google Assistant and Siri. In *Proceedings of the HCI International Conference*. 351–369.

[41] John E Laird. 2022. An analysis and comparison of ACT-R and Soar. In *Proceedings of the Cognitive Systems Conference*.

[42] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems* 33 (2020), 9459–9474.

[43] Yuelin Li and Nicholas J Belkin. 2008. A faceted approach to conceptualizing tasks in information seeking. *Information Processing and Management* 44, 6 (2008), 1822–1837.

[44] Zhen Liao, Yang Song, Yalou Huang, Li-wei He, and Qi He. 2014. Task trail: An effective segmentation of user search behavior. *IEEE Transactions on Knowledge and Data Engineering* 26, 12 (2014), 3090–3102.

[45] David Lillis, Rem Collier, Fergus Toolan, and John Dunnion. 2007. Evaluating communication strategies in a multi agent information retrieval system. In *Proceedings of the 18th Irish Conference on Artificial Intelligence and Cognitive Science*. 81–90.

[46] Sheng-Chieh Lin, Minghan Li, and Jimmy Lin. 2023. Aggretriever: A simple approach to aggregate textual representations for robust dense passage retrieval. *Transactions of the Association for Computational Linguistics* 11 (2023), 436–452.

[47] Jiqun Liu, Matthew Mitsui, Nicholas J Belkin, and Chirag Shah. 2019. Task, information seeking Intentions, and user behavior: Toward a multi-level understanding of web search. In *Proceedings of the ACM CHIIR Conference on Human Information Interaction and Retrieval*. 123–132.

[48] Tula Masterman, Sandi Besen, Mason Sawtell, and Alex Chao. 2024. The landscape of emerging AI agent architectures for reasoning, planning, and tool calling: A survey. *arXiv preprint arXiv:2404.11584* (2024).

[49] Rishabh Mehrotra, Prasanta Bhattacharya, and Emine Yilmaz. 2016. Characterizing users' multi-tasking behavior in web search. In *Proceedings of the ACM CHIIR Conference on Human Information Interaction and Retrieval*. 297–300.

[50] Rishabh Mehrotra and Emine Yilmaz. 2015. Terms, topics & tasks: Enhanced user modelling for better personalization. In *Proceedings of the ACM ICTIR International Conference on The Theory of Information Retrieval*. 131–140.

[51] Rishabh Mehrotra, Emine Yilmaz, and Manisha Verma. 2014. Task-based user modelling for personalization via probabilistic matrix factorization. In *Proceedings of RecSys Posters*.

[52] Bhaskar Mitra. 2015. Exploring session context using distributed representations of queries and reformulations. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3–12.

[53] George D Montanez, Ryen W White, and Xiao Huang. 2014. Cross-device search. In *Proceedings of the ACM CIKM Conference on Information and Knowledge Management*. 1669–1678.

[54] Meredith Ringel Morris and Eric Horvitz. 2007. S3: Storable, shareable search. In *IFIP Conference on Human-Computer Interaction*. Springer, 120–123.

[55] Jack Muramatsu and Wanda Pratt. 2001. Transparent Queries: Investigation users' mental models of search engines. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 217–224.

[56] Karen Myers, Pauline Berry, Jim Blythe, Ken Conley, Melinda Gervasio, Deborah L McGuinness, David Morley, Avi Pfeffer, Martha Pollack, and Milind Tambe. 2007. An intelligent personal assistant for task and time management. *AI Magazine* 28, 2 (2007), 47–47.

[57] Suman Nath, Ryen W. White, Fazle E. Faisal, Morris E. Sharp, Robert Gruen, and Lenin Ravindranath Sivalingam. 2025. From search engines to action engines. *IEEE Computer* 58, 6 (2025).

[58] Allen Newell, John C. Shaw, and Herbert A. Simon. 1959. Report on a General Problem Solving Program. In *Proceedings of the International Conference on Information Processing*. UNESCO, 256–264.

[59] Brendan O'Connor, Michel Krieger, and David Ahn. 2010. Tweetmotif: Exploratory search and topic summarization for twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 4. 384–385.

[60] Sharon Oviatt. 2007. Multimodal interfaces. *The Human-Computer Interaction Handbook* (2007), 439–458.

[61] Archiki Prasad, Alexander Koller, Mareike Hartmann, Peter Clark, Ashish Sabharwal, Mohit Bansal, and Tushar Khot. 2023. Adapt: As-needed decomposition and planning with language models. *arXiv preprint arXiv:2311.05772* (2023).

[62] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).

[63] Daniel E Rose and Danny Levinson. 2004. Understanding user goals in web search. In *Proceedings of the International Conference on the World Wide Web*. 13–19.

[64] Miamaria Saastamoinen and Kalervo Järvelin. 2017. Search task features in work tasks of varying types and complexity. *Journal of the Association for Information Science and Technology* 68, 5 (2017), 1111–1123.

[65] Reijo Savolainen. 2012. Expectancy-value beliefs and information needs as motivators for task-based information seeking. *Journal of Documentation* 68, 4 (2012), 492–511.

[66] Chirag Shah, Ryen White, Paul Thomas, Bhaskar Mitra, Shawon Sarkar, and Nicholas Belkin. 2023. Taking search to task. In *Proceedings of the ACM CHIIR Conference on Human Information Interaction and Retrieval*. 1–13.

[67] Chirag Shah and Ryen W White. 2024. Agents are not enough. *arXiv preprint arXiv:2412.16241* (2024).

[68] Chirag Shah and Ryen W White. 2025. Panmodal information interaction. *Commun. ACM* 68, 4 (2025), 33–36.

[69] Chirag Shah, Ryen W White, Reid Andersen, Georg Buscher, Scott Counts, Sarkar Snigdha Sarathi Das, Ali Montazer, Sathish Manivannan, Jennifer Neville, Xiaochuan Ni, et al. 2025. Using large language models to generate, validate, and apply user intent taxonomies. *ACM Transactions on the Web* (2025).

[70] Ben Shneiderman. 2022. *Human-Centered AI*. Oxford University Press.

[71] Aditi Singh, Abul Ehtesham, Saket Kumar, and Tala Talaei Khoei. 2025. Agentic retrieval-augmented generation: A survey on agentic RAG. *arXiv preprint arXiv:2501.09136* (2025).

[72] Adish Singla, Ryen White, and Jeff Huang. 2010. Studying trailfinding algorithms for enhanced web search. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 443–450.

[73] Yang Song and Qi Guo. 2016. Query-less: Predicting task repetition for nextgen proactive search and recommendation engines. In *Proceedings of the International Conference on the World Wide Web*. 543–553.

[74] Jaime Teevan, Kevyn Collins-Thompson, Ryen W White, and Susan Dumais. 2014. Slow search. *Commun. ACM* 57, 8 (2014), 36–38.

[75] Gabriele Tolomei, Salvatore Orlando, and Fabrizio Silvestri. 2010. Towards task-based search and recommender systems. In *Proceedings of the IEEE International Conference on Data Engineering Workshops*. 333–336.

[76] 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. 2019. Learning about work tasks to inform intelligent assistant design. In *Proceedings of the ACM CHIIR Conference on Human Information Interaction and Retrieval*. 5–14.

[77] Pertti Vakkari. 1999. Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval. *Information Processing and Management* 35, 6 (1999), 819–837.

[78] Pertti Vakkari. 2001. A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of Documentation* 57, 1 (2001).

[79] Pertti Vakkari. 2003. Task-based information searching. *Annual Review of Information Science and Technology* 37, 1 (2003), 413–464.

[80] Pertti Vakkari. 2016. Searching as learning: A systematization based on literature. *Journal of Information Science* 42, 1 (2016), 7–18.

[81] Thanh Vu, Alistair Willis, Son N Tran, and Dawei Song. 2015. Temporal latent topic user profiles for search personalisation. In *Proceedings of the European Conference on Information Retrieval*. 605–616.

[82] Hongning Wang, Yang Song, Ming-Wei Chang, Xiaodong He, Ahmed Hassan Awadallah, and Ryen W. White. 2014. Modeling action-level satisfaction for search task satisfaction prediction. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 123–132.

[83] Hongning Wang, Yang Song, Ming-Wei Chang, Xiaodong He, Ryen W White, and Wei Chu. 2013. Learning to extract cross-session search tasks. In *Proceedings of the International Conference on the World Wide Web*. 1353–1364.

[84] Yu Wang, Xiao Huang, and Ryen W White. 2013. Characterizing and supporting cross-device search tasks. In *Proceedings of the ACM WSDM Conference on Web Search and Data Mining*. 707–716.

[85] Ryen W White. 2018. Skill discovery in virtual assistants. *Commun. ACM* 61, 11 (2018), 106–113.

[86] Ryen W White. 2024. Advancing the search frontier with AI agents. *Commun. ACM* 67, 9 (2024), 54–65.

[87] Ryen W White, Paul N Bennett, and Susan T Dumais. 2010. Predicting short-term interests using activity-based search context. In *Proceedings of the ACM CIKM Conference on Information and Knowledge Management*. 1009–1018.

[88] Ryen W White, Wei Chu, Ahmed Hassan, Xiaodong He, Yang Song, and Hongning Wang. 2013. Enhancing personalized search by mining and modeling task behavior. In *Proceedings of the International Conference on World Wide Web*. 1411–1420.

[89] Ryen W White, Adam Fourney, Allen Herring, Paul N Bennett, Nirupama Chandrasekaran, Robert Sim, Elnaz Nouri, and Mark J Encarnación. 2019. Multi-device digital assistance. *Commun. ACM* 62, 10 (2019), 28–31.

[90] Ryen W White and Ahmed Hassan Awadallah. 2019. Task duration estimation. In *Proceedings of the ACM WSDM Conference on Web Search and Data Mining*. 636–644.

[91] Ryen W White, Ahmed Hassan Awadallah, and Robert Sim. 2019. Task completion detection: A study in the context of intelligent systems. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 405–414.

[92] Ryen W White and Diane Kelly. 2006. A study on the effects of personalization and task information on implicit feedback performance. In *Proceedings of the ACM CIKM Conference on Information and Knowledge Management*. 297–306.

[93] Ryen W White, Matthew Richardson, Mikhail Bilenko, and Allison P Heath. 2008. Enhancing web search by promoting multiple search engine use. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 43–50.

[94] Ryen W White and Chirag Shah. 2025. Panmodal information interaction. *Commun. ACM* 68, 4 (2025), 33–36.

[95] Thomas D Wilson. 2024. Curiosity and information-seeking behaviour: a review of psychological research and a comparison with the information science literature. *Journal of Documentation* 80, 7 (2024), 43–59.

[96] Michael Wooldridge. 1999. Intelligent agents. *Multiagent systems: A modern approach to distributed artificial intelligence* 1 (1999), 27–73.

[97] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Awadallah, Ryen W. White, Doug Burger, and Chi Wang. 2024. AutoGen: Enabling next-gen LLM applications via multi-agent conversations. In *Proceedings of the Conference on Language Modeling*.

[98] Jing Xie and Chen-Ching Liu. 2017. Multi-agent systems and their applications. *Journal of International Council on Electrical Engineering* 7, 1 (2017), 188–197.

[99] ChengXiang Zhai. 2024. Large language models and future of information retrieval: Opportunities and challenges. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 481–490.

[100] Yongfeng Zhang, Min Zhang, Yiqun Liu, Chua Tat-Seng, Yi Zhang, and Shaoping Ma. 2015. Task-based recommendation on a web-scale. In *Proceedings of the IEEE International Conference on Big Data*. 827–836.

[101] Yan-ping Zhang, Quan-xin Tian, Li Liu, and Zheng-ying Wang. 2012. Research on multi-agents information retrieval system based on intelligent evolution. In *Proceedings of the International Conference on Wavelet Active Media Technology and Information Processing*. 308–311.