ABSTRACT

Title of Thesis: SERENDIPITY IN THE AGE OF GENERATIVE AI: HOW SEARCH SYSTEMS FOSTER IT, AND WHY IT MATTERS

Subin Seo
Master of Library and Information, 2026

Thesis Directed by: Professor Yiwei Xu
College of Information

Since generative AI-powered systems have emerged as search tools, they are influencing how people seek information online. In particular, this study focuses on serendipity, defined as "a memorable experience of unexpectedly discovering useful or interesting information that becomes the occasion for a new personal insight." Serendipity has recently been discussed as a critical dimension that information systems must address in order to satisfy users' underlying needs; however, it largely remains unexplored. In this study, I conducted a randomized controlled experiment by comparing three different search systems: traditional web search engine (WebSearch), generative AI powered search system (GenSearch), and retrieval augmented generation based search system (RAGSearch). The findings reveal that GenSearch and RAGSearch facilitate numerically higher levels of serendipity than WebSearch, though there were no statistically significant differences in the levels of serendipity across the three search system conditions. However, a critical distinction emerged in terms of user effort: AI-based systems (GenSearch and RAGSearch) required significantly fewer queries and clicks

compared to WebSearch to achieve comparable levels of serendipity. This suggests that conversational interfaces can effectively catalyze unexpected discoveries while minimizing the cognitive and physical effort required from the user, leading to a more efficient discovery process. Furthermore, WebSearch significantly outperformed both AI-assisted conditions in information novelty, providing empirical evidence that objective novelty and subjective serendipity are distinct theoretical constructs. The study also identifies serendipity as a correlate of overall user satisfaction, although it did not significantly relate to search self-efficacy. By highlighting the role of AI in fostering "effort-efficient discovery," this research provides critical implications for the design of next-generation search systems, emphasizing the need to facilitate perceived informational value and serendipitous moments to enhance user experience in the AI era.

SERENDIPITY IN THE AGE OF GENERATIVE AI: HOW SEARCH SYSTEMS FOSTER IT, WHY IT MATTERS

by

Subin Seo

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Library and Information Science
2026

Advisory Committee:
Professor Yiwei Xu, Chair/Advisor
Professor Ziang Xiao
Professor Joel Chan
Professor Julia Mendelsohn

Table of Contents

Table of Contents ii

List of Tables iv

List of Figures v

List of Abbreviations vi

Chapter 1: Introduction 1
1.0.1 Conceptual Model . . . 3
1.1 Outline of Thesis . . . 3

Chapter 2: Research Backgrounds 5
2.1 Defining Serendipity . . . 5
2.2 Exploratory Search and Search Task for Serendipity . . . 6
2.3 Search system for Serendipity . . . 6
2.3.1 Conversational Search vs. Navigational Search . . . 7
2.3.2 Parametric knowledge vs. Retrieved Knowledge . . . 8
2.4 Serendipity and its related psychological outcomes . . . 10
2.5 Research Questions and Hypotheses . . . 12
2.5.1 Which search system most effectively supports serendipity? . . . 12
2.5.2 How system characteristics drive serendipity? . . . 12
2.5.3 Other psychological outcomes related to users' serendipity in search . . . 14

Chapter 3: Research Method 16
3.1 Experimental Design . . . 17
3.1.1 Participants . . . 17
3.1.2 Procedure . . . 17
3.1.3 Search Tasks and Case . . . 20
3.1.4 Experiment Apparatus . . . 21
3.2 Measurement . . . 29
3.2.1 Serendipity . . . 29
3.2.2 Search Engagement . . . 30
3.2.3 Information Novelty . . . 30
3.2.4 User satisfaction . . . 32
3.2.5 Self-efficacy . . . 32
3.2.6 Topic Familiarity . . . 34
3.3 Analysis Plan . . . 35

Chapter 4: Results 37
4.1 GenSearch and RAGSearch showed numerically but not significantly higher serendipity than WebSearch . . . 37

4.2 WebSearch users showed more search engagement than RAGSearch users 37
4.3 WebSearch exhibited more novel information than RAGSearch and GenSearch 39
4.4 Search engagement and information novelty were not related to serendipity 41
4.5 The effect of Serendipity on user satisfaction and self-efficacy 42

Chapter 5: Conclusion 44
5.1 AI-powered search systems enable effort-efficient serendipitous discovery 44
5.2 Higher search engagement, higher novelty in output information, but lower serendipity in WebSearch users 44
5.3 Serendipity leads to user satisfaction but does not impact self-efficacy 45
5.4 Discussion 46

Bibliography 49

List of Tables

3.1 Demographic characteristics and search system usage frequency 18
3.2 Survey items for measuring serendipity 29
3.3 Survey items for measuring user satisfaction 33
3.4 Survey items for measuring self-efficacy 34
3.5 Survey items for measuring pre-task topic familiarity 35

4.1 Descriptive statistics of comparison across search systems 39
4.2 Pearson correlations between search engagement and serendipity, and between information novelty and serendipity 42
4.3 ANOVA and pairwise comparisons across search systems 43

List of Figures

1.1 Proposed conceptual model 4

2.1 Venn diagram of system characteristics across search mechanism and knowledge base . 13

3.1 Experimental procedure 19
3.2 Examples of search task and search case 21
3.3 Interface design of the WebSearch system 23
3.4 Interface design of the GenSearch system 25
3.5 Interface design of the RAGSearch system 27
3.6 Interface design of the RAGSearch system: source preview popup 28

4.1 RQ1: Serendipity differences across search systems 38
4.2 RQ2: Search engagement differences across search systems 40
4.3 RQ3: Information novelty differences across search systems 41

List of Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
GMO	Genetically Modified Organism
HCI	Human-Computer Interaction
LLM	Large Language Model
RAG	Retrieval-Augmented Generation
RQ	Research Question
SERP	Search Engine Result Page
α-nDCG	Alpha-normalized Discounted Cumulative Gain
WebSearch	Traditional Web Search Engine
GenSearch	Generative AI-based Search
RAGSearch	Retrieval-Augmented Generation-based Search

Chapter 1: Introduction

Serendipity, a memorable experience of unexpectedly discovering useful or interesting information that becomes the occasion for a new personal insight, extends beyond mere knowledge expansion, holding the potential to be a game changer in both personal life and society. For instance, at a personal level, a graffiti inscription glimpsed on the street may spark inspiration to write a novel, while at a societal level, a chance insight may lead to groundbreaking discoveries such as penicillin and X-rays [1,2]. However, research on serendipity has been comparatively underexplored within information behavior fields [3]. This is because serendipity is an 'event' of information acquisition that happens at a specific moment [4]. Moreover, this moment arrives through incidental pathways and cannot be planned prior to encounter [5], which makes it difficult to study in a controlled research setting [6]. People often find greater value from what they discover by accident than from what they seek [7]. Indeed, people frequently do not know what they do not know, a condition referred to as 'unknown unknowns,' where they are unaware of what they need [8]. Accordingly, understanding serendipity is essential to fully capturing the breadth of human information behavior, extending beyond what people consciously seek.

In this regard, information systems play a crucial role in helping people experiencing serendipity, by serving as a medium that connects humans with information. While little research has directly examined how search systems, particularly those incorporating generative AI, influence serendipity, the advance of search systems holds the potential to cut both ways in shaping serendipity. On the one hand, the advent of the internet and the world wide web has created more favorable conditions for experiencing serendipity compared to the era when information was sought primarily in libraries [6]. Taking this a step further, the emergence of large language models (LLMs) has made generative AI another viable tool for information search. More recently, a technique known as retrieval-augmented generation (RAG) has emerged, which retrieves up-to-date external data and reorganizes it through

generative AI before presenting it to the user. One of their key advantages is enabling conversational search without the need for search operators or keywords, offering greater convenience in the search process. Such advanced information systems may increase search engagement among users, which in turn could create more favorable conditions for serendipity.

On the other hand, advanced search systems may function in the opposite direction, putting the experience of serendipity at risk by promoting the homogenization of information. In fact, search engines have continued to suffer from the 'filter bubble' issue, which narrows users' exposure to information [9], and more recently, generative AI has also been shown to exhibit the echo chamber effect [10]. Furthermore, AI systems are inclined to converge toward similar opinions, raising concerns over the homogenization of information [11]. These issues persist despite researchers having advocated for "beyond-accuracy" metrics, such as serendipity, diversity, novelty, and relevance, as the new standards for search systems over the past decade [12–15]. Thus, from the perspective of knowledge base, the impact on serendipity may remain unchanged or even become more negative despite the advancement of search systems.

This thesis primarily contributes to the literature by providing some of the first empirical evidence regarding whether and how various online search systems influence users' serendipity and other related psychological outcomes. I examined three search systems—traditional search engines (WebSearch), generative AI (GenSearch), and generative AI augmented with retrieval-augmented generation (RAGSearch). I first compared them based on two key dimensions, search mechanism and knowledge base, to explain how each system supports serendipity (see the Figure 2.1). These conceptual dimensions were operationalized to produce differences in users' search engagement and the novelty and diversity of output information, and this relationship was also validated in this study (see the Figure 1.1). Lastly, I examined other related psychological outcomes of serendipity to suggest

why it should be considered in the design and evaluation of search systems. I conducted a randomized controlled online experiment to explore these three overarching questions:

- Which search system most effectively supports serendipity?
- How system characteristics drive serendipity?
  - (Search Behavior) conversational search vs. navigation-based search
  - (Knowledge Base) parametric knowledge vs. non-parametric knowledge
- What psychological outcomes are associated with users' serendipity in search?

1.0.1 Conceptual Model

Taken together, the guiding questions outlined above reflect a conceptual model in which search system characteristics shape the probability of serendipity in searching, which in turn influences downstream outcomes including user experience and self-efficacy. In particular, this study conceptualizes the characteristics of search systems along two dimensions: search behavior and knowledge base, which operate through search engagement and quality of output information, respectively. Serendipity, once experienced, is further examined in relation to two downstream outcomes: user experience and self-efficacy. (The specific Research Questions (RQs) and hypotheses derived from this model were introduced in Chapter 2.)

1.1 Outline of Thesis

In Chapter 2, I reviewed the theoretical and empirical literature underpinning this study, establishing the research background for each RQ and hypothesis. Next, Chapter 3 detailed the research

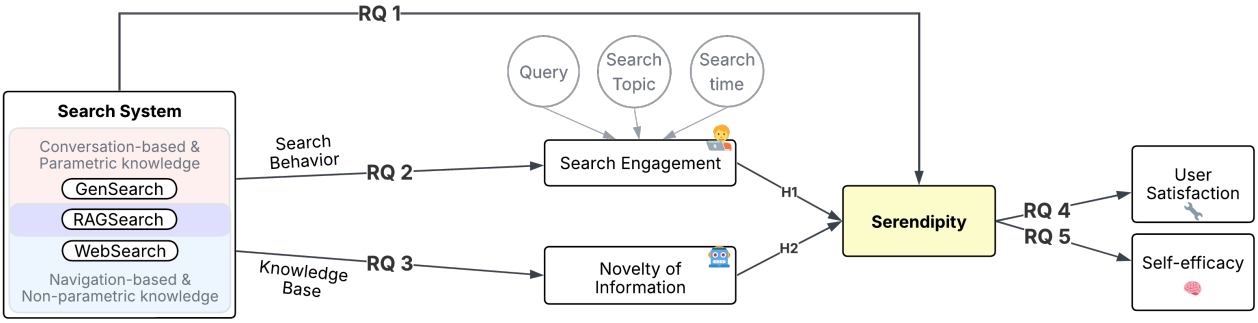


Figure 1.1: Proposed conceptual model illustrating the structural relationships between Search System types (GenSearch, RAGSearch, WebSearch), Serendipity, and its downstream effects. The model defines: (1) the direct link from systems to serendipity (RQ1); (2) the pathways through Search Engagement (RQ2, H1) and Novelty of Information (RQ3, H2); and (3) the subsequent downstream effects on User Satisfaction (RQ4) and Self-efficacy (RQ5).

methodology, describing how each construct was operationalized and measured, how the hypotheses and RQs were tested, and how the relationships among nodes in the conceptual model were estimated. Chapter 4 presented the results of these analyses. Finally, Chapter 5 discussed the theoretical and practical implications of the findings, and provided concluding remarks for the thesis.

Chapter 2: Research Backgrounds

In this chapter, I first defined serendipity based on previous studies (Section 2.1). Next, I presented the experimental design with consideration of the pathways through which serendipity is experienced (Section 2.2). In the following two sections, I reviewed related work to understand how prior studies had addressed the RQs proposed above (Section 2.3 and Section 2.4). Lastly, I described how this study builds on and contributes to the existing literature in the concluding paragraph of this chapter.

2.1 Defining Serendipity

Serendipity is commonly referred to as an “aha” moment when people unexpectedly discover valuable things [3, 16, 17]. In this study, I defined serendipity as "a memorable experience of unexpectedly discovering useful or interesting information that becomes the occasion for a new personal insight," primarily drawing from Erdelez [4] while referencing Merton (1948) [18]. In the field of information behavior, Erdelez first introduced the concept of serendipity as information encountering [19], which refers to "a memorable experience of unexpectedly discovering useful or interesting information" [4, p. 25]. This term reframes the phenomenon of serendipity through the lens of information behavior. On top of this definition, I incorporated an emphasis on the cognitive process of connecting encountered information to personal insight by referencing Merton’s definition. Merton defined serendipity as "the fairly common experience of observing an unanticipated, anomalous and strategic datum which becomes the occasion for developing a new theory or for extending an existing theory" [18, p. 506]. Through this, by adopting the term serendipity, the definition in this study encompasses not only the moment when a user bumps into information but also the subsequent process of recognizing it and connecting it to one’s own experiences and insights to derive meaning.

2.2 Exploratory Search and Search Task for Serendipity

Serendipity can arise through a wide variety of pathways. Since any unexpected and beneficial discovery may be considered serendipitous, the routes vary across individuals and situations [2]. Most typically, serendipity occurs when users are not looking for the specific information they happen to encounter [19]. Others suggest that it can happen regardless of whether users are purposive or non-purposive [3]. For example, while conducting a targeted search with a purpose, one can solve an unexpected problem or a targeted problem via an unexpected route. On the other hand, untargeted search may also facilitate solving a current problem or a future one [2]. Among these diverse pathways to serendipity, this study adopted a goal-directed search design by providing the search task itself as the external goal. Accordingly, the search task in this study was designed to enable participants to engage in exploratory search.

Exploratory search is a process encompassing learning and investigation, characterized by the combination of querying and browsing strategies [20]. It can involve serendipity within problem contexts that are open-ended, persistent, and multi-faceted [21]. Therefore, the study initiated the search task by providing only a search topic (e.g., "perform a search to explore evidence about the given topic"). Through this exploratory search task, I intended for participants not only to engage in the search process but also to have the opportunity to experience serendipity.

2.3 Search system for Serendipity

Search systems can be closely related to search outcomes, particularly in serendipity. While research to date has rarely tested which search systems influence the likelihood of serendipity, exist-

ing studies suggest that there are search environments that better facilitate serendipity [22, 23]. These environments exhibit serendipity-conducive characteristics, such as enabling exploration and being trigger-rich [23, 24]. First, an environment that supports exploration provides direct access to information, allows users to pause or stop searching, encourages exploratory behavior, and offers multiple pathways for navigation [23]. Additionally, a trigger-rich environment offers a variety of information or resources [24]. Based on these two characteristics, I conceptualized search systems along two dimensions: search mechanisms and knowledge base. The type of search mechanism, whether conversational or navigational, is expected to affect the degree of exploration, while the type of knowledge base—whether retrieved knowledge or the parametric knowledge of generative AI—is expected to influence how "trigger-rich" the information is. Thus, focusing on these two dimensions, I investigate how search systems influence serendipity.

2.3.1 Conversational Search vs. Navigational Search

Before people have generative AI as a search tool, 'search' has referred ad-hoc retrieval using search engines, where the system must find the most relevant documents for a information need represented by a query [25]. On the web, users typically iterate through the processes of formulating queries, selecting results, navigating, and modifying queries [26, 27]. In this process, the key difference from generative AI lies in 'navigating' behavior, which involves clicking links and browsing the displayed pages. Navigational behavior can foster serendipity by allowing users to browse information in multiple chunks rather than relying on a single result [28].

Now 'search' has been evolved to 'conversational search' which allows systems to understand users' needs through multiple rounds of conversations using natural languages [29]. Due to their natu-

ral and flexible conversational capabilities, systems can support users in expressing complex information needs, making them more applicable to complex search tasks [30]. While engaging in complex tasks that cannot be resolved with one single query, conversational retrieval models can suggest the most relevant results through re-ranking based on conversations, without requiring users to explicitly rewrite their queries [31].

Building on these competing perspectives, this study compared search systems to investigate which search mechanism better supports serendipity through increased exploration. To measure how much exploration each search system's mechanism facilitates, I used participants' level of search engagement in the search process as an indicator. In doing so, this study offers a new perspective on search mechanisms in the era of generative AI, extending beyond the focus of prior research on traditional web-based search systems.

2.3.2 Parametric knowledge vs. Retrieved Knowledge

The novelty and diversity of the information provided by a system can also influence serendipity. As illustrated, a trigger-rich environment for serendipity requires a diverse range of information, as serendipity is associated with the experience of discovering unexpected information. In addition to information diversity, evaluating the novelty of information can provide a nuanced understanding of the qualitative value of the results [32]. In other words, this novelty and diversity do not simply involve presenting unpopular results; rather, they reflect the extent to which the presented content is varied and new, with minimal redundancy [32,33]. These qualitative characteristics of information are largely determined by the type of knowledge base a system uses and each search system references a different knowledge base. In this study, I considered WebSearch as a system using retrieved knowl-

edge, GenSearch as a system using parametric knowledge (to be explained below), and RAGSearch as a hybrid system which incorporates both parametric and retrieved knowledge [34]. This approach allowed me to examine which of these two characteristics is more conducive to serendipity.

Generative AI uses parametric knowledge, the vast amount of intrinsic knowledge embedded within a model's parameters, to generate responses [35]. This knowledge base is cutoff at a specific point in time, meaning that new information is not incorporated into the model's parametric knowledge after the cutoff date [34]. This means, the novelty of information could be lower compared to a web search environment. Moreover, generative AI tends to align with the user's perspective, providing selectively biased information and reducing the diversity of information compared with web search [10]. Generative AI, which relies solely on parametric knowledge for language generation, performs poorly in terms of text diversity even when compared to RAG-based search systems that use both parametric and external data [34].

Even though a web search system using retrieved knowledge can outperform generative AI's parametric knowledge in terms of information novelty and diversity, search engine is also not inherently positive in these aspects as well. In fact, Eli Pariser introduced the concept of the filter bubble, arguing that personalized filtering in search engines can isolate users from diverse and novel information [36]. Similarly, long-tail results are often overlooked and excluded from top results due to ranking algorithms in search engines, even though they may contain valuable information that contributes to information diversity [37]. This suggested that, while the web result itself can provide users with broad, diverse, and novel information, search engine's recommendation systems may negatively affect serendipity.

Based on this evidence, this study compared search systems to investigate which knowledge base better supports serendipity through the novelty of output information. Therefore, I calculated the score

of novelty based on the output information for each search system.

2.4 Serendipity and its related psychological outcomes

Serendipity is an important metric for evaluating search systems in its own right because it reflects how much users have subjectively benefited from the system. However, unlike conventional metrics for web search systems, such as accuracy, precision, and recall, or common standards for evaluating generative AI, such as factual output, question answering, or summarization, serendipity is not directly captured by these measures. Therefore, I argue for its importance by examining the downstream outcomes it produces: user satisfaction and self-efficacy.

First, serendipity can influence user satisfaction by affecting how users feel while or after using a system and how they evaluate the information provided. In fact, serendipity is commonly accompanied by emotional responses such as happiness or a sense of being benefited [3]. More specifically, higher unexpectedness increases arousal and attention, which in turn raises the probability of experiencing positive emotions and satisfaction. However, at the same time, serendipity can also lead to disappointment when unexpected information contradicts users' existing beliefs [3]. In addition, the effect of serendipity on user satisfaction depends on the problem context [38]. For example, in social media use, which is largely characterized as a hedonic system focused on entertainment, users tend to report higher satisfaction with serendipitous experiences. In contrast, during information-seeking activities, the effect of serendipity on satisfaction is ambiguous. This is because users typically regard serendipity experience as disruptive, as they adopt a goal-oriented approach. However, this prior study measured satisfaction using a single-item question ("All together, how satisfied are you with the service you chose?") [38]. Accordingly, in this study, I employed a reliable measurement instrument to

assess user satisfaction with both the information system and the information provided (to be discussed in Chapter 3).

Second, I suggested that serendipity can also influence users' self-efficacy, which refers to "beliefs in one's power to produce the given levels of attainment" [39, p.382]. To the best of my knowledge, no study has empirically examined how the level of serendipity in search is associated with post-search self-efficacy. However, according to Erdelez's concept of the super-encounterer, some users do not passively experience accidental information encounters but instead develop them into a unique ability to discover valuable information [19]. In addition, research on serendipity beyond information behavior suggests that such experiences can evolve into recognizable patterns [40]. These experiences allow users to internalize successful search outcomes, fostering a more proactive approach to search and increasing their confidence in finding interesting or valuable information. In this sense, serendipity experiences in the search process may lead users to believe that they can discover serendipitous information, thereby enhancing their self-efficacy. Building on these prior studies, this study empirically tested how serendipity can lead to self-efficacy in searching.

To conclude, this study made a contribution to the literature on the "beyond-accuracy" perspective [12], arguing that it is necessary to evaluate search systems beyond merely assessing the accuracy of results, toward considering user's perceived value of information. In particular, I addressed a research gap by examining how different search systems vary to facilitate serendipity and by explaining the underlying mechanisms that drive such differences. Furthermore, this study represented a genuinely HCI-oriented investigation of serendipity by evaluating both users' search behavior and output information, addressing a limitation of most prior studies that have solely focused on one dimension. Finally, this study empirically demonstrated why serendipity matters by investigating its influence not only on direct satisfaction but also on users' self-efficacy at a more fundamental level.

2.5 Research Questions and Hypotheses

Based on these foundations, I propose five RQs and two hypotheses.

2.5.1 Which search system most effectively supports serendipity?

As the current literature provides limited guidance on which search system is most conducive to serendipity, I directly compare three different search systems to address this gap. Accordingly, this study poses this as an open RQ:

RQ 1: Which of the three search systems—WebSearch, GenSearch, or RAGSearch—exhibits the highest level of serendipity?

2.5.2 How system characteristics drive serendipity?

To unpack the system-level differences in serendipity, I compare WebSearch and RAGSearch as well as GenSearch and RAGSearch in terms of two key dimensions: (1) search mechanism, specifically whether the system supports conversational or navigational search, and (2) knowledge base, specifically whether the system draws on parametric or retrieved knowledge. At this point, I use RAGSearch as the reference system for comparison against both WebSearch and GenSearch, as RAGSearch serves as a hybrid system that incorporates characteristics of both systems along each dimension (see Figure 2.1).

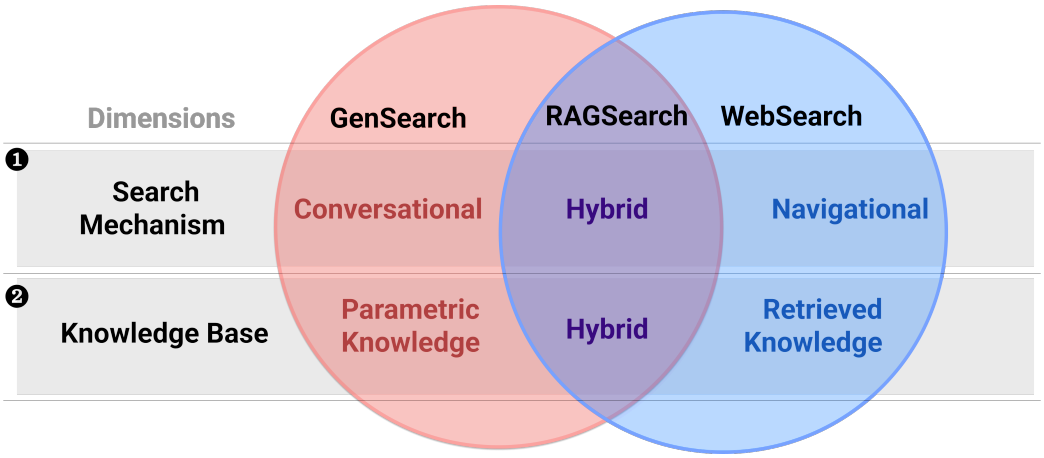


Figure 2.1: A Venn diagram illustrating the comparison of system characteristics (GenSearch, RAGSearch, WebSearch) across two dimensions: (1) Search Mechanism, ranging from conversational to navigational, and (2) Knowledge Base, ranging from parametric to retrieved knowledge. RAGSearch is positioned as a hybrid system incorporating both dimensions.

2.5.2.1 Search Mechanism: Conversational vs. Navigational Search

First, I compared conversational search with navigation-based search systems by analyzing participants' search behaviors, such as queries, search time, and click behavior. Specifically, I compared (1) WebSearch and RAGSearch, followed by (2) GenSearch and RAGSearch. RAGSearch is a hybrid search system that incorporates both conversational and navigation-based search characteristics by presenting responses in a chat interface with in-text citations. Thus, RAGSearch was used as an intermediate condition for comparison, allowing us to examine differences in users' search behaviors between WebSearch and GenSearch.

RQ 2: How does the type of search mechanism (conversational vs. navigational) affect the search engagement in the interaction with the system?

2.5.2.2 Knowledge Base: Parametric vs. Retrieved Knowledge

Second, I also compared systems that refer to parametric knowledge and retrieved knowledge by analyzing systems' outputs, such as AI responses or retrieved web pages. Specifically, I compare (1) WebSearch and RAGSearch, followed by (2) GenSearch and RAGSearch. In this comparison, RAGSearch is a hybrid search system that incorporates both parametric and retrieved knowledge. It first retrieves relevant results and then generates responses based on parametric knowledge by integrating the retrieved information. Thus, RAGSearch was used as an intermediate condition for comparison, allowing us to examine differences in users' search behaviors between WebSearch and GenSearch.

RQ 3: How does the type of knowledge base (parametric vs. retrieved) affect the diversity and novelty of information provided by the system?

I also tested the relationship between search engagement and serendipity as well as the novelty of output information and serendipity.

H1 (Serendipity; Search Engagement): Participants who exhibit more search engagement will report higher levels of serendipity than those with lower search engagement.

H2 (Serendipity;novelty of information): Participants of systems returning more diverse and novel output information will report higher levels of serendipity than those of systems returning less diverse, and novel output information.

2.5.3 Other psychological outcomes related to users' serendipity in search

So far, the preceding RQs and hypotheses investigated which system better facilitates serendipity and how its characteristics influence serendipity. In the following section, I examined how serendipity is associated with related psychological outcomes, including user satisfaction and self-efficacy, carry-

ing both theoretical and practical implications.

RQ 4: How does serendipity relate to user satisfaction?

RQ 5: How does serendipity relate to self-efficacy?

Chapter 3: Research Method

This study employed a randomized controlled online experiment with a three-condition between-subjects design, where participants were randomly assigned to one of three custom-built search systems: WebSearch, GenSearch, or RAGSearch. To mitigate item effects, each system was evaluated across three distinct search contexts, with each participant being randomly assigned to one of the topics. This design allowed this study to examine the effect of search systems on serendipity (RQ 1) and its downstream outcomes (RQ 4: user satisfaction, RQ 5: self-efficacy) while controlling for variance attributable to the search task topic. In addition, I compared characteristics of search system in terms of users' search engagement (RQ 2) and novelty of system's output information (RQ 3). To better understand how search engagement and information novelty contribute to serendipitous search experiences, I added two hypotheses examining their relationships with serendipity, respectively (H1: search engagement, H2: novelty of output information).

In the following section, I described the experimental design, including participants, the experimental procedure, the search task, and the experimental apparatus. Next, I explained the measurements for the key variables, including serendipity, search engagement, novelty of information, self-efficacy, and user satisfaction. Finally, I detailed the analytical strategy for addressing the RQs and proposing the conceptual model using statistical modeling.

3.1 Experimental Design

3.1.1 Participants

I recruited a sample representative of age and sex via Prolific[1], an online crowdsourcing platform, using their prescreening criteria. Participants were required to be United States residents and at least 18 years of age. Each participant received $2 as compensation for completing the experiment, which took approximately 13 minutes on average. There were 75 participants in total, randomly assigned to one of three conditions: WebSearch (n = 24), RAGSearch (n = 25), and GenSearch (n = 26). The demographic composition of the sample is summarized in Table 3.1. I confirmed that random assignment was successful, as no significant demographic differences were observed across the three search system conditions.

3.1.2 Procedure

The study employed a three-condition between-subjects design, with each participant randomly assigned to one of three search tasks (GMOs, nanotechnology, or cultivated meat) to reduce item effects. Following the procedure illustrated in Figure 3.1, after submitting the consent form, participants were assigned a search task before rating their familiarity with the topic and their uncertainty regarding the search goal and plan. Next, participants were instructed: "Now you will start a search! Perform a search to explore evidence about topic." Before beginning the search, participants were presented with two pop-up notifications: first, they were informed that they would need to answer open-ended questions upon completing the task; second, they were notified that a minimum of three queries was

[1] https://www.prolific.com

Table 3.1: Demographic characteristics and search system usage frequency

Variable	N or Mean	% or SD	χ^2 or F, p-value
Age	39.69	15.31	$F(2, 72) = 0.42, p = .659$
Gender			$\chi^2(4) = 3.01, p = .651$
Female	36	48.0%	
Male	38	50.7%	
Non-binary	1	1.3%	
Race			$\chi^2(10) = 5.40, p = .959$
White	59	78.7%	
Asian	5	6.7%	
Black	6	8.0%	
Amer. Indian/Alaska Native	1	1.3%	
Other (Mixed)	4	5.3%	
Education			$\chi^2(10) = 14.14, p = .215$
Bachelor's Degree	25	33.3%	
Master's Degree	16	21.3%	
Some College/Tech School	20	26.7%	
High School or Equivalent	9	12.0%	
Prof./Doctorate Degree	4	5.3%	
Other/Not specified	1	1.3%	
Politics (Party ID)			$\chi^2(6) = 4.48, p = .653$
Democrat	32	42.7%	
Republican	23	30.7%	
Independent	19	25.3%	
No preference	1	1.3%	
Ideology			$\chi^2(12) = 16.01, p = .239$
Liberal	38	50.7%	
Moderate	10	13.3%	
Conservative	27	36.0%	
AI Usage Frequency	4.41	1.87	$F(2, 72) = 0.34, p = .714$
Web Usage Frequency	6.05	1.04	$F(2, 72) = 0.63, p = .537$

Note. Participants ($N = 75$) had a mean age of 39.69 ($SD = 15.31$). The sample included 38 men, 36 women, and 1 non-binary individual. Racial composition was predominantly White (78.7%), followed by Black (8.0%) and Asian (6.7%), with small proportions of mixed-race (5.3%) and American Indian or Alaska Native (1.3%) participants. Educational backgrounds ranged from high school to doctoral degrees, with 33.3% holding a Bachelor's degree. Search system usage was measured on a 7-point scale, where participants reported frequent use of web search engines ($M = 6.05, SD = 1.04$) and moderate use of AI tools ($M = 4.41, SD = 1.87$).

required to proceed. They were then directed to conduct their search interactively using the provided system.

After completing the search, participants were asked to assess their perceived level of serendipity, user satisfaction, and self-efficacy based on their search experience in the experiment (to be more explained in section 3.2). Additionally, participants answered two open-ended questions designed to capture latent experiences by prompting participants to share advice for a friend and list valuable information discovered during the task. At the end of the experiment, participants answered demographic questions, including their prior frequency of using Generative AI models and web search engine. The experiment flow is presented in the Figure 3.1.

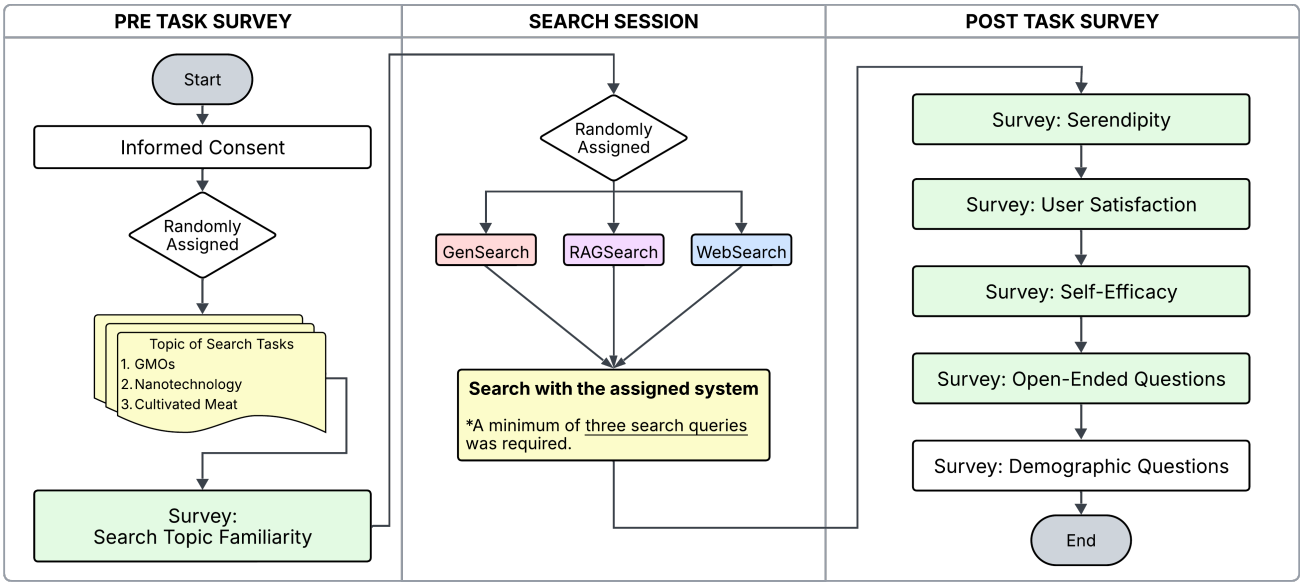


Figure 3.1: Overall experimental procedure. The process begins with participants providing informed consent, followed by the random assignment of a search task from topics including GMO, nanotechnology, and cultivated meat. In the pre-task survey, participants' topic familiarity is assessed through three items. During the search session, participants use one of the three systems (GenSearch, RAGSearch, or WebSearch) to gather information. Upon task completion, levels of serendipity, user satisfaction, and self-efficacy are measured. The procedure concludes with participants writing short essays for two open-ended questions and completing a demographic survey that includes their prior experience with generative AI and web search engines.

3.1.3 Search Tasks and Case

In this study, the search tasks focused on scientific topics such as GMOs, nanotechnology, and cultivated meat. To design the search tasks, I considered three criteria. First, the topics should enable participants to explore sufficient information. Second, the topic should be unfamiliar to the general public to evoke curiosity, prompting participants to engage with the search task [21]. Third, the search task should be still complex, requiring multiple searches and varied queries for resolution [41]. Accordingly, I selected these three scientific topics because they are widely debated and significantly impact daily life, yet remain relatively unfamiliar to the general public [42–44]. By selecting these topics, this study aimed to motivate participants to engage in exploratory search, as the topics' complexity and relevance necessitate deeper search.

In addition, the study provided the detailed scenarios with the search tasks (see Figure 3.2). The scenarios are designed by assuming simulated situations, providing a standardized environment for all participants while fostering a sense of realism individually [45]. This idea, presenting search tasks with simulated situations, was proposed by Borlund (2000) [46] and has been widely applied in many search studies. Based on Borlund's guidelines, simulated situations should adhere to three criteria. First, the situation should be a realistic scenario that participants can relate to and identify with. Second, the topic should be interesting to participants. Third, the situation should offer enough contexts to enable participants to apply themselves to the scenario. To meet these criteria, this study introduced a fictitious friend to allow participants to project their own relationships onto the task, while framing the scenario within a daily life context (e.g. grocery shopping) to enhance familiarity and interest.

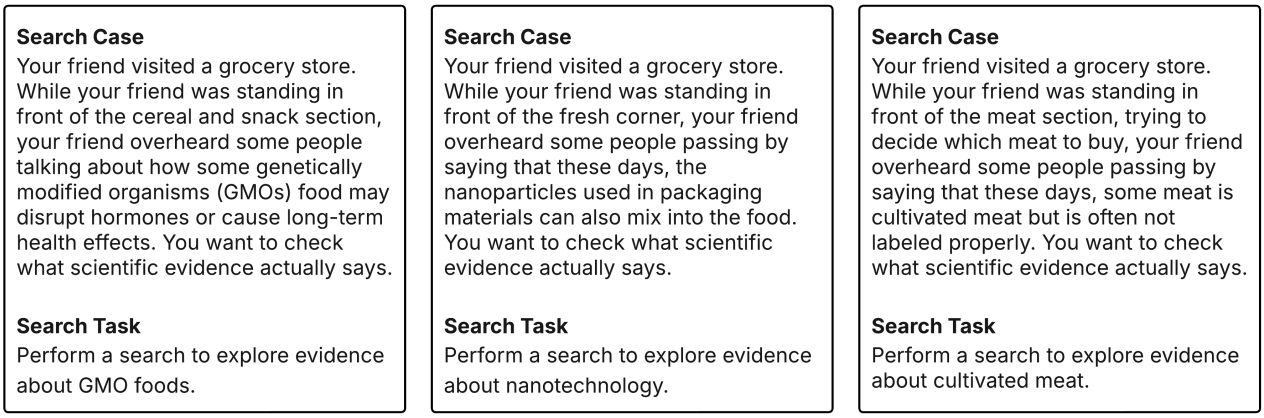


Figure 3.2: Design of search tasks and case scenarios. The experiment utilizes simulated information-seeking scenarios across three scientific domains: GMOs, nanotechnology, and cultivated meat. Each scenario is structured to provide participants with a specific context and a goal-oriented task, ensuring a standardized environment for evaluating search outcomes.

3.1.4 Experiment Apparatus

This study utilized three distinct search systems—WebSearch, GenSearch, and RAGSearch—which serve as interactive tools allowing participants to conduct direct search queries throughout the sessions. These custom-designed search systems featured a three-column interface, as illustrated in the following paragraph and Figures 3.3, 3.4, and 3.5.

1. The left panel includes a foldable section where users can revisit the search scenario and task at any time;

2. The middle pane displays the search system with the search bar for query entry, and this space varies in its characteristics across WebSearch, GenSearch, and RAGSearch. In WebSearch, results are displayed as a list with titles, snippets, and favicons, while in GenSearch and RAGSearch, responses are presented in a conversational format. RAGSearch additionally includes in-text citations displayed as clickable source buttons (see details below);

3. The right panel supports a space for note-taking and scrapping the links or AI responses. Participants can use this scrapbook for saving or taking notes on any interesting and valuable information.

Therefore, the three search systems incorporated intended differences in the middle pane, whereas the left and right panels were kept consistent across all systems to control for confounding variables. In the following sections, I described the design architecture of each system developed to facilitate an interactive search session.

3.1.4.1 WebSearch

WebSearch is a system that retrieved and displayed a list of websites in responses to queries via search engine. In the WebSearch condition, participants mainly interact with the Search Engine Result Page (SERP) to search for information, which displays results retrieved from the search engine. To implement this system, this study used the Google Custom Search Engine API [2], which is programmable search engine allowing researchers to custom retrieve results from Google [3]. In this study, I manipulated the results to improve participants' search efficiency by filtering them based on two criteria. First, I removed homepage URLs with one or fewer path segments, as these pages typically serve as main sites for companies or organizations rather than providing relevant information. For instance, I removed links such as "https://www.gmo.com/americas/" from the SERP because they contain only one path segment ("americas"). Second, I excluded URLs containing the term "product" to eliminate commercially oriented pages. As a result, the final set of results presented to participants included sources such as Wikipedia, NIH, FDA, and EU websites, as well as academic articles, blog posts, and

[2] `https://programmablesearchengine.google.com`
[3] `https://www.google.com`

social media webpage, including LinkedIn and Reddit. After applying these filters, the SERP in the WebSearch system consistently presents 20 results for each query to participants.

For the design rationale, I followed the typical SERP format, composed of titles, favicons, and snippets, as illustrated in Figure 3.3. Through this design, this study aims to enable participants to intuitively adapt to the system and perform the search task without requiring an additional practice session. After viewing the SERP, participants can click on individual web pages to access detailed information.

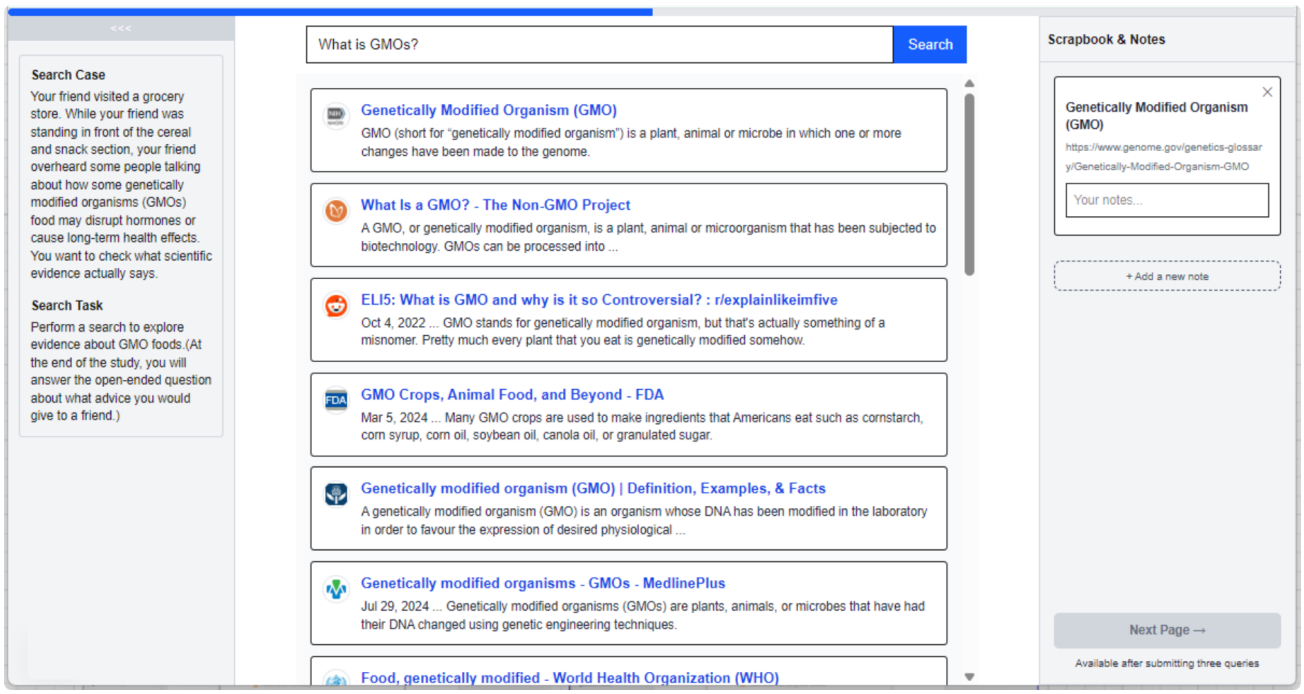


Figure 3.3: Interface design of the Search Engine Results Page (SERP) in the WebSearch system. The layout follows a typical SERP format, incorporating titles, favicons, and snippets to ensure participants can intuitively navigate the system. The interface includes: (1) the search case and task instructions on the left sidebar; (2) the central results area with interactive web page snippets; and (3) a "Scrapbook & Notes" section on the right for information gathering during the search session.

3.1.4.2 GenSearch

GenSearch is a system that generates text responses using LLM based on pre-trained data with the chatting style interface. In the GenSearch condition, participants can ask questions conversationally to search for information, with the system generating responses in a chat interface when users submit queries, as shown in Figure 3.4. To generate answers, the system employed the Gemini 2.5 Flash model, a hybrid reasoning model released in 2025 that supports a "thinking" capability while maintaining low cost and latency [47]. This means the GenSearch system solely relies on the LLM, which generates responses based on parametric knowledge. I configured the generation settings with a maximum of 400 output tokens, a temperature of 0.3, and a top-p value of 0.9 to promote consistent responses by prioritizing high-probability tokens. Additionally, I adapted system prompt to specify format of responses and length constraint of approximately 250–300 words, as shown below.

```
const systemInstruction = `
    Write a concise, well-organized answer in Markdown.
    Answering style requirements:
    - Provide a clear, multi-sentence explanation.
    - Use clear headings, bullet points for listed items, and formatting
     ↪ to organize the information.
    - Treat the interaction as a continuous conversation rather than
     ↪ isolated questions.
    - Use previous turns to maintain topic continuity and provide more
     ↪ relevant answers.`
```

In addition to employing the LLM to the search system, I adjusted the responses by preserving conversational continuity through structuring prior interactions into a formatted chat history and guid-

ing the model to maintain context across turns. For example, if a participant asks “What is GMOs?” followed by “Do I have to worry?”, the system can maintain the conversational context and interpret the follow-up question within the same context, even without the explicit mention of ’GMOs.’

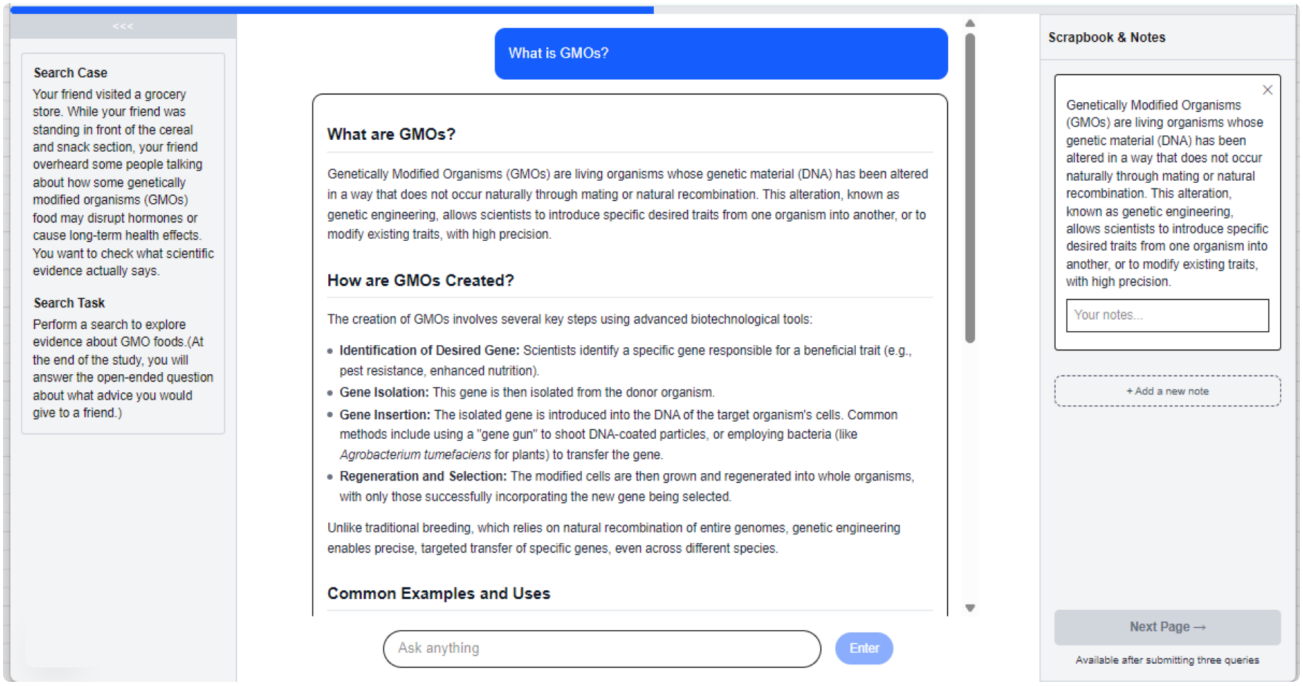


Figure 3.4: Interface of the GenSearch system. This interface is characterized by its conversation-based search. The layout features: (1) the search case and task instructions on the left sidebar; (2) a chat-style display in the central panel where the system provides generative answers to user queries, with an “Ask anything” input field at the bottom to support continuous dialogue; and (3) a “Scrapbook & Notes” section on the right sidebar for information gathering during the search session.

3.1.4.3 RAGSearch

RAGSearch refers to the retrieval of relevant information from external knowledge bases before answering questions with LLMs [48, p. 1]. In this study, it was positioned as a hybrid system, incorporating characteristics of both search behavior types and knowledge base types. First, it adopted a chat interface with in-text citation buttons that allow participants to browse the original sources (Figure 3.5). When the button is clicked, a pop-up window appears (Figure 3.6), displaying titles, links, fav-

icons, and snippets, with direct access to the corresponding web pages. In addition, to leverage both the search engine and LLM knowledge bases, the system followed an underlying process in which, when participants enter a query, RAGSearch first retrieves relevant information from a search engine and then generates a response by integrating the retrieved results with the model's output. The filtering of search results followed the same criteria as in the WebSearch condition, and the LLM settings for multi-turn conversation were kept consistent. The system prompt was also aligned with that used in the WebSearch condition, with the addition of citation formatting, as shown below.

```
const systemInstruction = `
  You are a helpful assistant engaged in an ongoing conversation.
  Answer the current question using the provided sources as your primary evidence.
  Use inline citations like [1], [2] to reference sources.
  Write in clear Markdown with headings and bullet points where appropriate.
  Important conversation rules:
  - If the current question is a follow-up, interpret it in the context of the previous conversation turns.Use previous turns to understand what topic the user is referring to, even if the sources don't explicitly repeat the topic.
  - For claims drawn from the provided sources, always cite with [n].
  - Do NOT invent sources or fabricate citations.`;
```

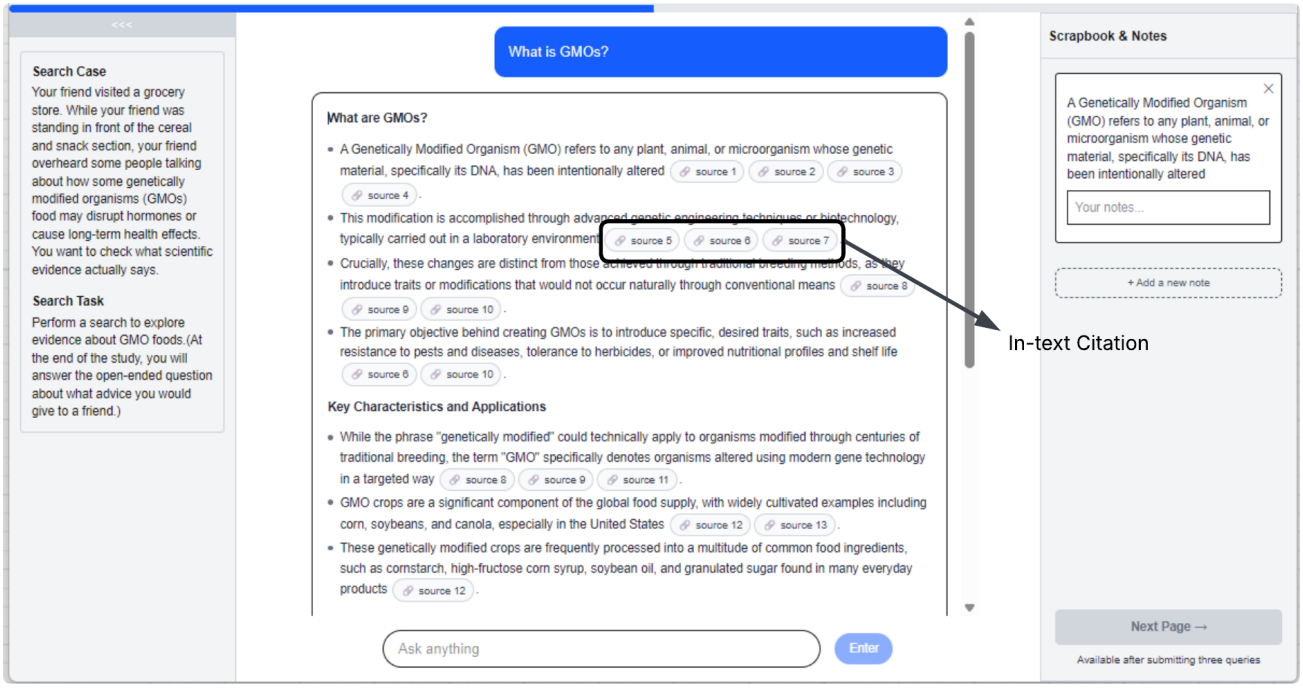


Figure 3.5: Interface of the RAGSearch system. This interface combines retrieval-augmented generation with a conversation-based search experience. The layout features: (1) the search case and task instructions on the left sidebar; (2) a chat-style display in the central panel where the system provides generative answers grounded in retrieved sources, with in-text citations linking to external web pages and an "Ask anything" input field at the bottom to support continuous dialogue; and (3) a "Scrapbook & Notes" section on the right sidebar for information gathering during the search session.

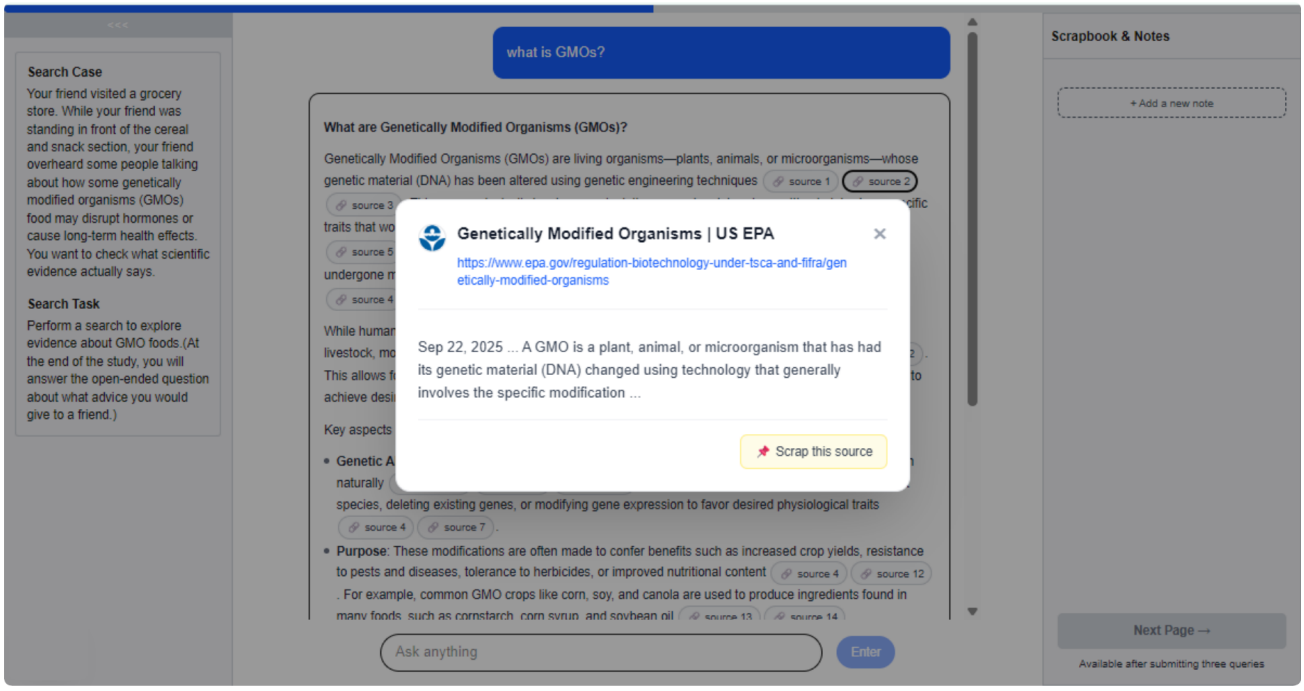


Figure 3.6: Source preview popup in the RAGSearch system, triggered by clicking an in-text citation button in the chat bubble. The popup displays the title, URL, publication date, and a brief excerpt of the retrieved external source, allowing users to evaluate its relevance without leaving the interface. Users may click the URL to access the full external webpage in a new tab, or immediately save the source to their Scrapbook via the "Scrap this source" button.

3.2 Measurement

To address our RQs, this study mainly measured serendipity, self-efficacy, and user satisfaction after the search session as outcomes of the search. In addition, the study evaluated participants' search engagement and novelty of systems' output information by analyzing participants' search log collected from the search session. Lastly, the study also accounted for participants' pre-task topic familiarity to control for individual differences in search task performance.

3.2.1 Serendipity

To evaluate serendipity, I applied the five items suggested by [16]. Through these five questions, this study assessed the various dimensions that conceptually define serendipity (see Table 3.2). This allows me to empirically measure the key components of the definition of serendipity, such as unexpectedness, interestingness, usefulness, and relevance. Participants reported their responses to these five items on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree), indicating the degree to which they perceived serendipity in the information they found (Cronbach's $\alpha = 0.94$).

Table 3.2: Survey items for measuring serendipity

No.	**Survey Items**
Q1	I had unexpected revelations about {Topic}.
Q2	I was able to see ordinary knowledge about {Topic} in new ways.
Q3	I obtained unexpected insights about {Topic}.
Q4	I found things about {Topic} that surprised me.
Q5	I made connections about {Topic} that I had not thought of before.

Note. All items were measured on a 7-point Likert scale, where {Topic} was replaced with the specific assigned search task (e.g., GMOs, nanotechnology, and cultivated meat).

3.2.2 Search Engagement

To measure search engagement, I adopted search behavioral metrics from [49], comparing WebSearch, GenSearch, and RAGSearch across several dimensions. Specifically, this study recorded participants' search behaviors, including issued queries, clicked link details, search duration, scrapped information, and notes. Based on these search behavioral data, three metrics were compared across all three systems: 1) the number of queries, 2) the number of topics in queries, and 3) search time. Search time was calculated by subtracting the time spent taking notes from the total completion time of search task. Additionally, 4.1) the length of queries was compared only between GenSearch and RAGSearch, and 4.2) the number of links clicked was compared only between WebSearch and RAGSearch due to its nature of search mechanism. Thus, these metrics collectively captured how actively and extensively participants engage with each search condition (see the Table 4.1).

3.2.3 Information Novelty

To evaluate the novelty of systems' output information, I employed the α-$nDCG$ framework introduced by [50]. This framework systematically rewards results that contribute new information beyond what has already been seen, while applying a diminishing-return penalty for redundant content. The unit of information was defined as the lemmatized word token. Each document (SERP snippets and clicked webpages for WebSearch, AI-generated responses for GenSearch, and both for RAGSearch) was tokenized, lowercased, filtered of stopwords, and lemmatized. I then constructed a judgment matrix $\mathbf{J}$ for each participant's search session, where each row corresponds to a query turn (ordered by the sequence in which the participant issued queries) and each column corresponds to a unique query token used across participants in the same task. In this matrix, a cell value of 1 indicates

that the token appeared in the document retrieved at that turn, and 0 indicates its absence. From this matrix, the redundancy count $r_{i,k-1}$ was obtained by counting how many times token i had already appeared in documents from earlier query turns — that is, by summing the 1s in column i from the first turn up to the turn immediately before the current one. To illustrate, suppose the token *"gmo"* appeared in the documents retrieved at turns 1, 2, and 3 within a participant's session. If it appeared again at turn 4, its redundancy count would be $r_{gmo,3} = 3$, and its contribution to the gain at turn 4 would therefore be discounted by $(1 - \alpha)^3$.

Formally, a gain function $G[k]$ accumulates judgments over retrieved documents weighted by the novelty penalty $(1 - \alpha)^{r_{i,k-1}}$, and the Discounted Cumulative Gain is computed as:

$$DCG[k] = \sum_{j=1}^{k} \frac{G[j]}{\log_2(1 + j)} \tag{3.1}$$

The final score is normalized against an ideal diversified ranking, yielding:

$$\alpha\text{-}nDCG[k] = \frac{DCG[k]}{DCG'[k]} \tag{3.2}$$

where $DCG'[k]$ denotes the ideal diversified ranking score, producing a final metric on a scale of 0 to 1. The α-$nDCG$ score was aggregated over matrix **J**, with $\alpha = 0.5$ (the strength of the redundancy penalty, where 0 applies no penalty and values closer to 1 impose stronger penalties; 0.5 is the standard choice in Clarke et al. [50]) and evaluation depth $k = 5$ (covering the full search session of most participants, who typically issued three to four queries).

3.2.4 User satisfaction

Next, the study measured participants' satisfaction with the given search system and its output information. Drawing on established measures from prior work [51, 52], I assessed how participants evaluate information quality through the first five items and how well the system meets their expectations through the remaining five items (see Table 3.3).Participants rated these ten items on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). First, to measure how participants are satisfied with the contents of information, I evaluated the reliability, relevancy, accuracy, precision, and completeness of output information provided by the system. Second, to measure how well the system meets the expectations of using it, I examined their understanding of the system, confidence in the system, and feeling of participating in the search, control over the system, and assurance in the system (Cronbach's $\alpha = 0.86$).

3.2.5 Self-efficacy

I also measured participants' post-search self-efficacy to evaluate the downstream outcomes of serendipity. Following Bandura's guide for constructing self-efficacy scales [53], I adapted the instrument to the context of exploratory search. Bandura emphasizes that an effective self-efficacy scale should assess confidence through concrete items tailored to each specific domain, rather than relying on a general-purpose instrument [53]. Guided by this principle, I developed a seven-item questionnaire targeting search ability (see Table 3.4), drawing on prior work by Brennan et al.'s factor analysis of search self-efficacy [54] and White and Roth's suggested characteristics of exploratory search [21]. First, from the search self-efficacy scale validated by Brennan et al. [54], I retained only the items applicable across all three search systems and suited to an exploratory search context—for example,

Table 3.3: Survey items for measuring user satisfaction

No.	Survey Item	Low Anchor	High Anchor
Information Satisfaction			
Q1	The information (___) what I was looking for.	unmatched	matched
Q2	The information provided was (___) to understand the topic.	not enough	enough
Q3	The information provided by the system was (___).	inconsistent	consistent
Q4	The search results contained (___) unnecessary information.	little	much
Q5	The information provided was (___).	inaccurate	accurate
Overall Search Experience			
Q6	Your understanding of the provided search system	low	high
Q7	The degree of your involvement in the search	uninvolved	involved
Q8	Your feelings of participating in search	negative	positive
Q9	Your feeling of the personal power to regulate the system	weak	strong
Q10	Your feelings of assurance about the search systems	low	high

Note. All items were rated on a 7-point Likert scale. Items for Information Satisfaction and Overall Search Experience were averaged to create composite scores representing user satisfaction. For Information Satisfaction (Q1–Q5), participants completed each sentence by selecting a point on a 7-point scale between two bipolar anchors (Low/High), where the chosen point reflects the degree to which the descriptor fits the blank.For Overall Search Experience (Q6-Q10), participants rated their subjective experience of the search system across five dimensions: understanding of the system, degree of involvement, feelings of participation, sense of personal control, and feelings of assurance.

items assessing the ability to formulate queries, distinguish relevant results, and locate sufficient information. Next, I incorporated four items derived from the exploratory search abilities identified by White and Roth [21]: wayfinding, timely filtering of irrelevant information, comparison across results, and sense-making. Each item was rated on an 11-point scale ranging from 0 to 10, a format Bandura recommends as both simple and sensitive [53] (Cronbach's $\alpha = 0.94$).

Table 3.4: Survey items for measuring self-efficacy

No.	Survey Items
Q1	I think I'm able to develop search queries that accurately reflect what I'm looking for.
Q2	I think I'm able to distinguish between results that are relevant to my needs and those that are not.
Q3	I think I'm able to find an adequate amount of information.
Q4	I think I'm able to navigate through information spaces to find what I need.
Q5	I think I'm able to quickly decide which results to skip and which to pursue.
Q6	I think I'm able to compare multiple results and draw an integrated conclusion.
Q7	I think I'm able to form and revise my understanding as I discover new information.

Note. Each item was rated on an 11-point scale ranging from 0 to 10, where higher scores indicate greater self-efficacy. Beyond fundamental search skills (Q1–Q3), items Q4–Q7 were adapted from White and Roth's framework for exploratory search, specifically assessing wayfinding ability, selective filtering of irrelevant information, cross-item comparison, and sense-making capabilities.

3.2.6 Topic Familiarity

Lastly, I measured topic familiarity prior to the search session to assess participants' readiness to engage in exploratory search with the given task (see Table 3.5). Following the conditions under which exploratory search occurs as described in [21], I constructed a topic familiarity measurement consisting of three items. According to this work, exploratory search requires that participants (1) be unfamiliar with the domain, (2) lack certainty about their search goals, and (3) be uncertain about how to achieve those goals. Accordingly, I used these items to measure how actively participants were positioned to engage in search; the results showed a Mean of 3.77 and $SD = 1.36$ (Cronbach's $\alpha = 0.79$). Thus, participants in this study exhibited a moderate level of readiness for exploratory

search, suggesting they were open to serendipitous discoveries.

Table 3.5: Survey items for measuring pre-task topic familiarity

No.	Survey Items
Q1	How familiar are you with the {Topic}?
Q2	To what extent are you knowledgeable about keywords and concepts related to {Topic}?
Q3	How clear is your plan for finding interesting information about {Topic}?

Note. All items were rated on a 7-point Likert scale (1: Not at all, 7: Very much), where {Topic} was replaced with the specific assigned search task (e.g., GMOs, nanotechnology, and cultivated meat). These items were administered prior to the search task to control for the baseline knowledge of participants across different topics.

3.3 Analysis Plan

I applied different analysis methods tailored to each RQ and hypothesis to address the goals of this study. For RQ 1, RQ 2, and RQ 3, which aimed to compare outcomes across different search systems, I primarily employed ANOVA or independent t-tests across varying dependent variables. First, RQ 1 examined the level of serendipity among participants across three conditions (WebSearch, RAGSearch, and GenSearch). Accordingly, I employed a one-way ANOVA to examine whether there were significant differences in serendipity among the three systems. Subsequently, I conducted post hoc tests using Tukey's Honestly Significant Difference (HSD) to perform pairwise comparisons of (1) GenSearch vs. RAGSearch, (2) WebSearch vs. RAGSearch, and (3) GenSearch vs. WebSearch. Next, to address RQ 2 and RQ 3, I applied one-way ANOVA and independent samples t-tests to compare search engagement (RQ 2) and the quality of information (RQ 3). Since GenSearch and WebSearch operate on inherently different mechanisms, I conducted pairwise comparisons against RAGSearch for specific fields such as click counts and query length. Following these analyses, H1, H2, RQ 4, and RQ 5 investigated the relationships between variables. Specifically, H1 and H2 tested the effect of their respective independent variables on serendipity, while RQ 4 and RQ 5 treated serendipity as an

independent variable for downstream outcomes.

Prior to the analysis results, descriptive statistics are presented in Table 4.1, followed by the full results listed in Table 4.3. In Table 4.3, I reported the ANOVA results and pairwise comparisons for RQ 1, RQ 2, and RQ 3. I present the independent samples t-test results in the text to compare indicators that could only be compared between specific systems. Lastly, I presented the results of the correlation between serendipity and its psychological outcomes in Table 4.2 for RQ 4 and RQ 5. In conclusion, throughout Chapter 4, I organize the results around three overarching questions, highlighting key patterns alongside these tables.

Chapter 4: Results

4.1 GenSearch and RAGSearch showed numerically but not significantly higher serendipity than WebSearch

First, the analysis (see Table 4.1 and Figure 4.1) for RQ 1 revealed that GenSearch users experienced the highest level of serendipity (M = 5.40, SD = 1.25), followed by RAGSearch (M = 5.37, SD = 1.19), and WebSearch (M = 4.96, SD = 1.45). However, there was no statistically significant difference in the level of serendipity across the three conditions (WebSearch, RAGSearch, GenSearch), F(2, 66) = 2.159, p = .124, η^2 = .06 (see Table 4.3). When applying Tukey's HSD pairwise comparisons, none of the three comparisons reached statistical significance. The difference between WebSearch and RAGSearch showed a medium effect size but did not reach conventional significance (t(43) = -2.080, p = .114, d = -0.620), while the comparisons between GenSearch and WebSearch (t(45) = -1.248, p = .344, d = -0.364) and between GenSearch and RAGSearch (t(44) = 0.749, p = .788, d = 0.221) were also not significant. Thus, RQ 1 results indicated that the search system did not produce significant differences in users' serendipity experiences.

4.2 WebSearch users showed more search engagement than RAGSearch users

Second, for RQ 2, the comparison of search engagement across the two different search mechanisms revealed several cases in which users' search engagement was significantly higher in WebSearch than in RAGSearch (see Table 4.1 and Table 4.3). Specifically, independent samples t-tests further revealed that the number of link clicks (t(43) = 4.700, p < .001, d = 1.402) and the number of queries was statistically significantly higher among WebSearch users than among both RAGSearch users (t(43) = 3.801, p = .001, d = 1.134) and GenSearch users (t(45) = 2.989, p = .003, d = 0.872). In contrast, search

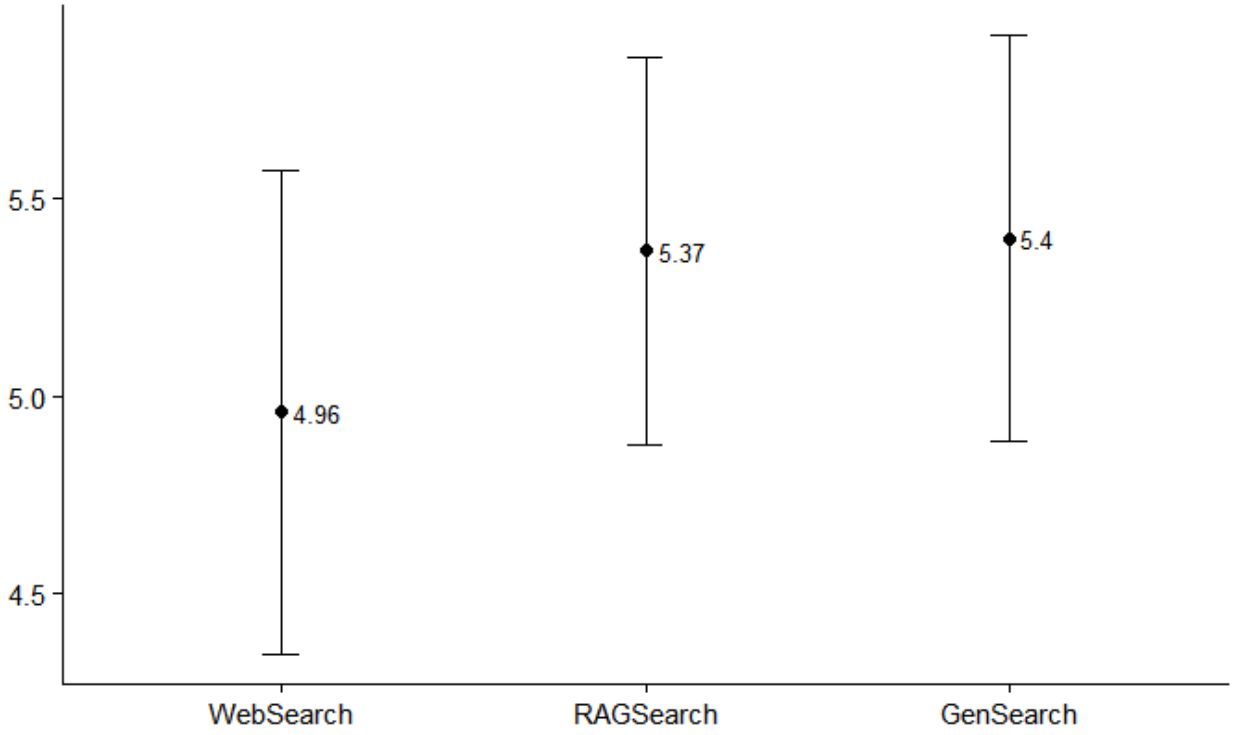


Figure 4.1: Mean plot with 95% confidence intervals for user-perceived serendipity across the three search systems (WebSearch, RAGSearch, and GenSearch). The data indicate the extent to which serendipity levels differ across search systems.

Table 4.1: Descriptive statistics of comparison across search systems

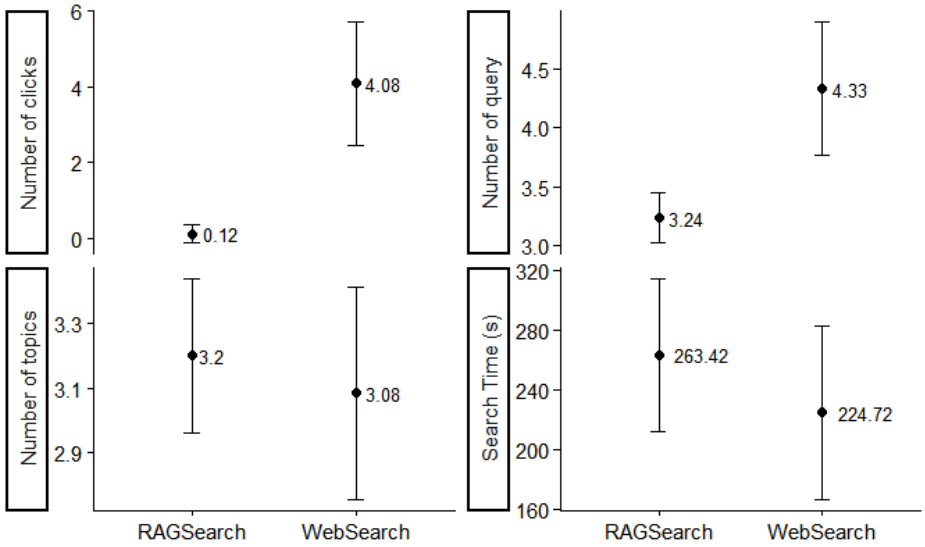
Metric	WebSearch ($n = 24$)		RAGSearch ($n = 25$)		GenSearch ($n = 26$)	
	M	SD	M	SD	M	SD
Serendipity	4.96	1.45	5.37	1.19	5.40	1.25
Number of queries	4.39	1.34	3.23	0.53	3.38	0.97
Number of topics	3.13	0.76	3.18	0.59	3.43	0.95
Search Time (s)	231.0	137.0	265.0	129.0	252.0	138.0
Number of clicks	4.13	3.93	0.14	0.64	N/A	N/A
Length of query (words)	N/A	N/A	11.10	10.60	8.52	3.87
Novelty of Information	0.94	0.11	0.85	0.10	0.82	0.12
User Satisfaction	5.77	0.75	5.69	0.92	6.02	0.80
Self-Efficacy	9.60	1.14	9.26	1.42	9.37	1.48

Note. M = Mean, SD = Standard Deviation. Serendipity, Novelty, and User Satisfaction were measured on a 7-point scale, while Self-Efficacy was measured on an 11-point scale (0–10). "N/A" indicates metrics not applicable to the specific system interface (e.g., query length was not captured in WebSearch, and clicks were not applicable to GenSearch system). Search Time is reported in seconds.

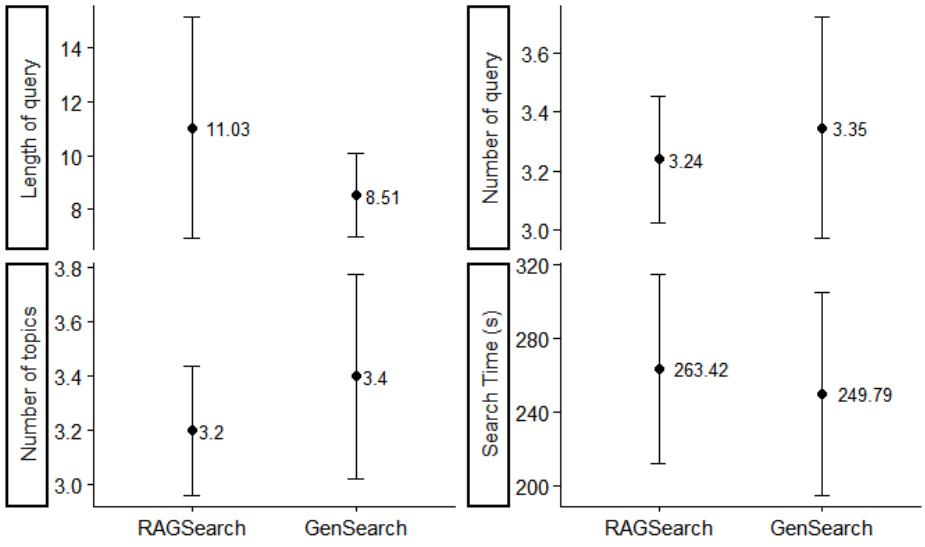
time and the number of topics per query were higher among RAGSearch and GenSearch users than among WebSearch users, although these differences did not reach statistical significance (as shown in Table 4.3). Furthermore, I found no areas of search engagement in which RAGSearch and GenSearch differed significantly. Therefore, I concluded that the navigational characteristic (the key distinction between GenSearch and RAGSearch) did not significantly affect users' search engagement, whereas the conversational characteristic (the key distinction between WebSearch and RAGSearch) had some significantly negative effects on users' search engagement.

4.3 WebSearch exhibited more novel information than RAGSearch and GenSearch

Third, for RQ 3, my analysis (see Table 4.1, Table 4.3, and Figure 4.3) showed that the novelty of the system's output information was significantly highest in the WebSearch condition (M = 0.94, SD = 0.11, $F(2, 59) = 6.046$, $p = .004$, $\eta^2 = .17$), followed by RAGSearch (M = 0.85, SD = 0.10) and



(a) Comparison of WebSearch and RAGSearch



(b) Comparison of GenSearch and RAGSearch

Figure 4.2: Mean with 95% confidence intervals for search engagement variables across system comparisons. (a) Comparison of WebSearch and RAGSearch across number of clicks, number of topics, number of queries, and search time; (b) Comparison of GenSearch and RAGSearch across length of query, number of topics, number of queries, and search time. Points represent group means and vertical lines indicate 95% confidence intervals.

GenSearch (M = 0.82, SD = 0.12). Notably, WebSearch's information novelty was also significantly higher than that of RAGSearch ($t(37)$[1] = 2.525, p = .050, d = 0.816) and GenSearch ($t(37)$ = 3.300, p = .003, d = 1.055) in pairwise comparisons. However, there was no significant difference in information novelty between RAGSearch and GenSearch. Taken together, these results suggested that the retrieved knowledge base (the key distinction between GenSearch and RAGSearch) did not significantly affect information novelty, whereas the involvement of parametric knowledge (the key distinction between WebSearch and RAGSearch) appeared to have a negative effect on information novelty.

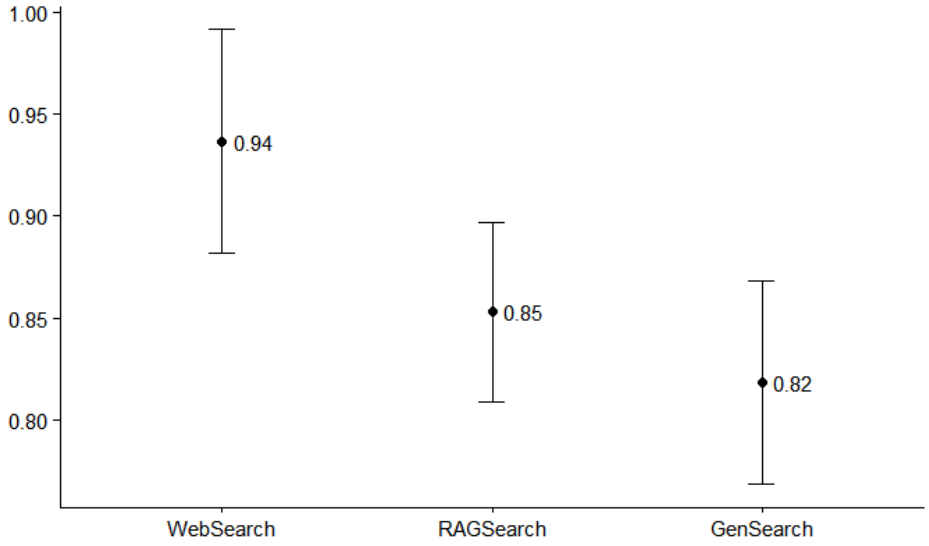


Figure 4.3: Mean plot with 95% confidence intervals for information novelty scores across the three search systems. The data indicate the extent to which information novelty differs across search systems.

4.4 Search engagement and information novelty were not related to serendipity

Next, neither search engagement (H1) nor information novelty (H2) significantly predicted serendipity. For H1, the composite engagement measure did not significantly predict serendipity (B =

[1]Degrees of freedom reflect listwise exclusion of participants with missing α-nDCG scores.

0.361, $SE = 0.222$, $p = .109$), and individual engagement measures showed non-significant correlations with serendipity (see Table 4.2). For H2, information novelty was also not significantly correlated with serendipity (r = -0.064, p = .623), and this non-significant relationship persisted after controlling for the experimental condition (B = -0.130, SE = 1.135, p = .909). Thus, neither H1 nor H2 was supported.

Table 4.2: Pearson correlations between search engagement and serendipity, and between information novelty and serendipity

	Overall		**WebSearch**		**RAGSearch**		**GenSearch**	
Measure	r	p	r	p	r	p	r	p
Search Engagement (H1)								
Number of queries	-0.081	.510	0.124	.572	0.182	.417	-0.140	.514
Number of subtopics	0.000	1.000	0.000	.998	0.230	.302	-0.118	.592
Search time (s)	0.207	.088	0.307	.154	0.220	.326	0.051	.814
Information Novelty (H2)								
Novelty of output information	-0.064	.623	-0.023	.931	-0.129	.568	0.046	.834

Note. r = Pearson's correlation coefficient, p = two-tailed significance. The correlation analysis examines the relationship between serendipity and participants' search engagement/information novelty across different system conditions. Significance is typically denoted at the * $p < .05$ level, though no results reached statistical significance.

4.5 The effect of Serendipity on user satisfaction and self-efficacy

Lastly, I found that users' serendipity experience significantly predicted their satisfaction (B = 0.197, SE = 0.089, p = .027, f^2 = 0.073). However, this relationship varied considerably across conditions: the effect was large and significant for WebSearch users ($r = 0.625$, $p = .001$), but negligible and non-significant for both RAGSearch (r = 0.059, p = .794) and GenSearch users (r = -.003, p = .991). In contrast, serendipity did not significantly predict self-efficacy for future search across all three conditions (RQ 5; B = 0.195, SE = 0.151, p = .200), with negligible correlations observed in WebSearch (r = 0.113, p = .607), RAGSearch (r = 0.257, p = .247), and GenSearch (r = 0.215, p = .313). Therefore, these results suggested that serendipity contributed to user satisfaction primarily

in traditional web search contexts (RQ 4), while its effect on self-efficacy for future search appeared limited regardless of the search system (RQ 5).

Table 4.3: ANOVA and pairwise comparisons across search systems

Metric & Comparisons	**ANOVA Test Result** $F(df_1, df_2), p, \eta^2$	*Est.*	*SE*	*t*	*p*	*d*
		Pairwise Comparison (Tukey's HSD)				
Serendipity	*(Not Supported)*					
RAGSearch vs. WebSearch	$F(2, 66) = 2.159$,	0.606	0.291	-2.080	.114	-0.620
GenSearch vs. WebSearch	$p = .124, \eta^2 = .06$	0.411	0.329	-1.248	.344	-0.364
GenSearch vs. RAGSearch		-0.195	0.260	0.749	.788	0.221
Search Engagement						
1. Number of Queries	*(Supported)*					
RAGSearch vs. WebSearch	$F(2, 66) = 9.051$,	-1.164	0.306	3.801	.001	1.134
GenSearch vs. WebSearch	$p < .001, \eta^2 = .22$	-1.016	0.340	2.989	.003	0.872
GenSearch vs. RAGSearch		0.148	0.233	-0.633	.873	-0.187
2. Number of Topics	*(Not Supported)*					
RAGSearch vs. WebSearch	$F(2, 65) = 1.000$,	0.051	0.203	-0.253	.973	-0.076
GenSearch vs. WebSearch	$p = .373, \eta^2 = .03$	0.304	0.253	-1.205	.387	-0.355
GenSearch vs. RAGSearch		0.253	0.236	-1.072	.525	-0.320
3. Search Time	*(Not Supported)*					
RAGSearch vs. WebSearch	$F(2, 66) = 0.371$,	34.341	39.823	-0.862	.672	-0.257
GenSearch vs. WebSearch	$p = .691, \eta^2 = .01$	21.128	40.209	-0.525	.854	-0.153
GenSearch vs. RAGSearch		-13.213	39.560	0.334	.941	0.099
Information Novelty	*(Supported)*					
RAGSearch vs. WebSearch	$F(2, 59) = 6.046$,	-0.084	0.033	2.525	.050	0.816
GenSearch vs. WebSearch	$p = .004, \eta^2 = .17$	-0.118	0.036	3.300	.003	1.055
GenSearch vs. RAGSearch		0.035	0.032	1.076	.530	0.321

Note. Est. = Estimate, SE = Standard Error, d = Cohen's d (effect size). Significance levels are reported as exact p-values. Bold text for "Supported" indicates statistical significance at the $p < .05$ level.

Chapter 5: Conclusion

5.1 AI-powered search systems enable effort-efficient serendipitous discovery

Although no statistically significant difference in serendipity was found across the three conditions, the descriptive pattern showed GenSearch and RAGSearch users reporting numerically higher serendipity than WebSearch users, with a medium effect size ($d = -0.620$) that did not reach conventional significance ($p = .114$). While this difference should be interpreted with caution given the non-significant test, the observed direction is consistent with the possibility that conversational, AI-mediated search may play a role in facilitating unexpected and meaningful information discovery. Rather than requiring users to actively navigate and filter through a large volume of results, AI powered search system may lower the cognitive threshold for serendipity by contextually synthesizing information in a way that surfaces personally relevant content. Future research with larger samples would be needed to determine whether this trend reflects a reliable effect.

5.2 Higher search engagement, higher novelty in output information, but lower serendipity in WebSearch users

WebSearch users demonstrated significantly higher search engagement, submitting more queries and clicking on more external links, and the system produced objectively higher information novelty compared to both AI-based conditions. Yet despite these advantages, WebSearch users reported lower serendipity than RAGSearch users, revealing an important dissociation: active engagement and quantitative novelty do not, on their own, guarantee a serendipitous experience. One interpretation is that the navigational nature of WebSearch, while expanding the breadth of information exposure, places a greater burden on users to make meaningful connections themselves, such that the cognitive load asso-

ciated with actively seeking, filtering, and evaluating a large number of results may actually impede the spontaneous sense of unexpected discovery that characterizes serendipity. By contrast, RAGSearch, despite eliciting fewer queries and clicks, appears to facilitate serendipity more efficiently by delivering contextually integrated responses that enable users to encounter meaningful information with less deliberate effort. This pattern ultimately challenges a common assumption in the information retrieval literature that novelty is a reliable proxy for serendipity, as our results suggest these are empirically distinct constructs. What matters is not simply whether information is new, but whether it is presented in a way that allows users to recognize its relevance and connect it to their existing knowledge and interests, and in this regard, the contextual relevance facilitated by AI-based systems may be more conducive to serendipity than the sheer volume of novel content surfaced by traditional web search.

5.3 Serendipity leads to user satisfaction but does not impact self-efficacy

Serendipity was a significant predictor of overall user satisfaction, suggesting that unexpected and valuable information discovery contributes meaningfully to how users evaluate their search experience. Notably, however, this effect was driven primarily by WebSearch users, while the relationship between serendipity and satisfaction was less pronounced in the AI-based conditions. This may reflect differences in user expectations across systems: when users interact with a traditional search engine, the discovery of unexpected yet relevant information stands out as a particularly rewarding experience against a backdrop of effortful navigation. In contrast, users of AI-based systems may attribute their satisfaction more broadly to the convenience and coherence of the responses they receive, making it more difficult for any single serendipitous moment to register as distinctively impactful. Serendipity, however, did not significantly predict self-efficacy in any of the three conditions. This null finding

suggests that isolated serendipitous encounters, even when satisfying, may not be sufficient to meaningfully alter users' beliefs in their own search capabilities. Developing self-efficacy likely requires repeated, recognizable success experiences that users can attribute to their own agency — a condition that a single experimental search session may not be able to fulfill. Future research should examine whether cumulative serendipity experiences, sustained over time and across varied search contexts, eventually contribute to the development of search self-efficacy.

5.4 Discussion

Through these findings, I observed a descriptive pattern consistent with RAGSearch enabling effort-efficient serendipitous discovery. Although the difference in serendipity between RAGSearch and WebSearch did not reach statistical significance, RAGSearch users reported numerically higher serendipity while submitting significantly fewer queries and clicks to web sources. This pattern is consistent with the possibility that AI can facilitate unexpected discoveries by contextually connecting relevant information in advance, with minimal active engagement on the part of the user.

Furthermore, given that no significant differences were found between RAGSearch and GenSearch users in terms of either search engagement or the novelty of system's output information across all analyses, I conclude that despite its hybrid nature, RAGSearch tends to align more closely with GenSearch in practice. The behavioral similarity between RAGSearch and GenSearch users in terms of search engagement may be explained by the fact that RAGSearch users rarely utilized the external website feature; notably, only a single RAGSearch user followed a link to an external source.

Additionally, I found that information novelty does not straightforwardly translate into serendipity experience. While prior studies have suggested a potential link between information novelty and

serendipity, these findings reveal a meaningful gap between objectively measured novelty and subjectively experienced serendipitous discovery. Indeed, WebSearch outperformed AI-based systems in information novelty, yet this did not translate into higher serendipity. This suggests that simply presenting previously unseen information is insufficient for serendipity; rather, meaningful contextual relevance matters more than quantitative novelty. Moreover, an excessive focus on delivering a greater volume of novel information may instead give rise to problems such as cognitive overload [55]. In this regard, although RAGSearch may present a smaller quantity of novel information than WebSearch, its ability to deliver personally relevant and contextually appropriate information may be more conducive to serendipitous discovery by keeping cognitive load within a manageable range while fostering meaningful connections for the individual.

Finally, to establish serendipity as an enduring pattern of individual behavior rather than a mere one-time chance encounter, I reasoned that users would need to report both high satisfaction with their search experience and positive outcomes in self-efficacy. Accordingly, I examined these two downstream effects. Among them, I found that serendipity significantly contributed to user satisfaction. Such positive search experiences may foster the conditions for users to become what Erdelez terms "super-encounterers"—individuals who are "serendipity-prone" in that they actively seek to create circumstances favorable to information encountering and are positively reinforced through successful experiences [19]. However, this trajectory is likely shaped by a wide range of individual characteristics [40]. Consistent with this, these findings also suggest that the pathways from serendipity to satisfaction and self-efficacy vary across individuals. For instance, factors such as demographic characteristics, frequency of use of search systems, and familiarity with the search topic may all play a role. Future research should therefore incorporate more diverse individual-level variables into its analyses to better account for this variability.

In this paper, I compared three search systems with a focus on serendipity, examining how each system fosters serendipitous discovery and how such experiences may in turn positively influence users' search behavior and future encounters with serendipity. Serendipity, as a measure of the degree to which users subjectively perceive or experience information as interesting or valuable, is a more conservative yet fundamentally user-centered metric than information novelty, and one that deserves greater consideration in the development of search systems. Furthermore, people frequently do not know what they do not know, a condition commonly referred to as "unknown unknowns" [56], underscoring the importance of systems that can surface meaningful information beyond the boundaries of users' existing awareness.

I therefore argue that serendipity should be more actively incorporated into the design and evaluation of AI-powered search systems. Given that human information behavior is fundamentally nonlinear, dynamic, holistic, and flowing [57], future efforts must grapple more carefully with how to account for individual-level factors—such as social context, cognitive state, emotional condition, and prior knowledge—in shaping the conditions for serendipitous discovery.

Bibliography

[1] Lori McCay-Peet and Elaine G. Toms. Investigating serendipity: How it unfolds and what may influence it. *Journal of the Association for Information Science and Technology*, 66(7):1463–1476, 2015.

[2] Ohid Yaqub. Serendipity: Towards a taxonomy and a theory. *Research Policy*, 47(1):169–179, 2018.

[3] Naresh Kumar Agarwal. Towards a definition of serendipity in information behaviour. *Information research: an international electronic journal*, 20(3):n3, 2015.

[4] Sandra Erdelez. Information encountering: It's more than just bumping into information. *Bulletin of the American Society for Information Science and technology*, 25(3):26–29, 1999.

[5] Yaxi Liu, Chunxiu Qin, Xubu Ma, and Huigang Liang. Serendipity in human information behavior: a systematic review. *Journal of Documentation*, 78(2):435–462, 07 2021.

[6] Sanda Erdelez. Investigation of information encountering in the controlled research environment. *Information Processing Management*, 40(6):1013–1025, 2004.

[7] Sandra Erdelez et al. Towards understanding information encountering on the web. In *Proceedings of the Annual Meeting-American Society for Information Science*, volume 37, pages 363–371. Information Today; 1998, 2000.

[8] Ian G. Kennedy, Dean Whitehead, and Debra Ferdinand-James. Serendipity: A way of stimulating researchers' creativity. *Journal of Creativity*, 32(1):100014, 2022.

[9] Qazi Mohammad Areeb, Mohammad Nadeem, Shahab Saquib Sohail, Raza Imam, Faiyaz Doctor, Yassine Himeur, Amir Hussain, and Abbes Amira. Filter bubbles in recommender systems: Fact or fallacy—a systematic review. *WIREs Data Mining and Knowledge Discovery*, 13(6):e1512, 2023.

[10] Nikhil Sharma, Q Vera Liao, and Ziang Xiao. Generative echo chamber? effect of llm-powered search systems on diverse information seeking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2024.

[11] Liwei Jiang, Yuanjun Chai, Margaret Li, Mickel Liu, Raymond Fok, Nouha Dziri, Yulia Tsvetkov, Maarten Sap, Alon Albalak, and Yejin Choi. Artificial Hivemind: The Open-Ended Homogeneity of Language Models (and Beyond). *arXiv e-prints*, page arXiv:2510.22954, October 2025.

[12] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys '10, page 257–260, New York, NY, USA, 2010. Association for Computing Machinery.

[13] Marius Kaminskas and Derek Bridge. Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Trans. Interact. Intell. Syst.*, 7(1), December 2016.

[14] Pablo Castells, Neil Hurley, and Saul Vargas. Novelty and diversity in recommender systems. In *Recommender systems handbook*, pages 603–646. Springer, 2021.

[15] Denis Kotkov, Alan Medlar, and Dorota Glowacka. Rethinking serendipity in recommender systems. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*, CHIIR '23, page 383–387, New York, NY, USA, 2023. Association for Computing Machinery.

[16] Lori McCay-Peet and Elaine G. Toms. The serendipity quotient. *Proceedings of the American Society for Information Science and Technology*, 48(1):1–4, 2011.

[17] Wanshu Niu, Liqiang Huang, and Mingliang Chen. Spanning from diagnosticity to serendipity: An empirical investigation of consumer responses to product presentation. *International Journal of Information Management*, 60:102362, 2021.

[18] Robert K. Merton. The bearing of empirical research upon the development of social theory. *American Sociological Review*, 13(5):505–515, 1948.

[19] Sanda Erdelez. Information encountering: a conceptual framework for accidental information discovery. In *Proceedings of an International Conference on Information Seeking in Context*, ISIC '96, page 412–421, GBR, 1997. Taylor Graham Publishing.

[20] Gary Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

[21] Ryen W White and Resa A Roth. *Exploratory search: Beyond the query-response paradigm*. Number 3. Morgan & Claypool Publishers, 2009.

[22] Elaine Toms. Browsing digital information: examining the affordances in the interaction of user and text. 1998.

[23] Lennart Björneborn. Serendipity dimensions and users' information behaviour in the physical library interface. *Information Research*, 13(4):13–4, 2008.

[24] Lori McCay-Peet, Elaine G Toms, and E Kevin Kelloway. Examination of relationships among serendipity, the environment, and individual differences. *Information Processing & Management*, 51(4):391–412, 2015.

[25] W Bruce Croft, Donald Metzler, Trevor Strohman, et al. *Search engines: Information retrieval in practice*, volume 520. Addison-Wesley Reading, 2010.

[26] Gary Marchionini. *Information seeking in electronic environments*. Number 9. Cambridge university press, 1995.

[27] Mark Levene. *An introduction to search engines and web navigation*. John Wiley & Sons, 2011.

[28] G. Marchionini and B. Shneiderman. Finding facts vs. browsing knowledge in hypertext systems. *Computer*, 21(1):70–80, 1988.

[29] Yongfeng Zhang, Xu Chen, Qingyao Ai, Liu Yang, and W. Bruce Croft. Towards conversational search and recommendation: System ask, user respond. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, page 177–186, New York, NY, USA, 2018. Association for Computing Machinery.

[30] 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, page 117–126, New York, NY, USA, 2017. Association for Computing Machinery.

[31] Fengran Mo, Kelong Mao, Ziliang Zhao, Hongjin Qian, Haonan Chen, Yiruo Cheng, Xiaoxi Li, Yutao Zhu, Zhicheng Dou, and Jian-Yun Nie. A survey of conversational search. *ACM Trans. Inf. Syst.*, 43(6), September 2025.

[32] Marcelo Mendoza and Nicolás Torres. Evaluating content novelty in recommender systems. *Journal of Intelligent Information Systems*, 54(2):297–316, 2020.

[33] Charles L.A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, page 659–666, New York, NY, USA, 2008. Association for Computing Machinery.

[34] 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 H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 9459–9474. Curran Associates, Inc., 2020.

[35] Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. Language models as knowledge bases? In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 2463–2473. Association for Computational Linguistics, 2019.

[36] Eli Pariser. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin, 2011.

[37] Philipp Schaer, Philipp Mayr, Sebastian Sünkler, and Dirk Lewandowski. How relevant is the long tail? a relevance assessment study on million short. In Norbert Fuhr, Paulo Quaresma, Teresa Gonçalves, Birger Larsen, Krisztian Balog, Craig Macdonald, Linda Cappellato, and Nicola Ferro, editors, *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 227–233, Cham, 2016. Springer International Publishing.

[38] Christoph Lutz, Christian Pieter Hoffmann, and Miriam Meckel. Online serendipity: A contextual differentiation of antecedents and outcomes. *Journal of the Association for Information Science and Technology*, 68(7):1698–1710, 2017.

[39] Albert Bandura. *Self-efficacy: The exercise of control*, volume 11. Freeman, 1997.

[40] Professor Peter Willett, Allen Edward Foster, and David Ellis. Serendipity and its study. *Journal of Documentation*, 70(6):1015–1038, 10 2014.

[41] Jaap Walhout, Paola Oomen, Halszka Jarodzka, and Saskia Brand-Gruwel. Effects of task complexity on online search behavior of adolescents. *Journal of the Association for Information Science and Technology*, 68(6):1449–1461, 2017.

[42] Maida Mobeen, Muhammad Majeed, Samreen Fatima, Akim Oladokoun, Ayesha Akbar, Aftab Umar, Jameel M Al-Khayri, Waseem Akhtar Quresh, Ambreen Kalsoom, Javed Iqbal, et al. Regulations and ethical considerations for nanotechnology in agri-food applications. In *Nanobiosensors for Crop Monitoring and Precision Agriculture*, pages 307–334. Springer, 2025.

[43] Mengxue Ou, Shirley S. Ho, and Stanley A. Wijaya. Harnessing ai to address misinformation on cultivated meat: The impact of chatbot expertise and correction sidedness. *Science Communication*, 47(6):859–896, 2025.

[44] Yedi Herdiana. Halal challenges and health risks in genetically modified organisms (gmos): a critical approach. *Cogent Food & Agriculture*, 11(1):2565716, 2025.

[45] Anita Crescenzi and Lan Li. Assessing realism in simulated work tasks. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*, CHIIR '22, page 266–271, New York, NY, USA, 2022. Association for Computing Machinery.

[46] Pia Borlund. Experimental components for the evaluation of interactive information retrieval systems. *Journal of documentation*, 56(1):71–90, 2000.

[47] Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

[48] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yixin Dai, Jiawei Sun, Haofen Wang, Haofen Wang, et al. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2(1):32, 2023.

[49] Yuyu Yang, Kelsey Urgo, Jaime Arguello, and Robert Capra. Search+chat: Integrating search and genai to support users with learning-oriented search tasks. In *Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval*, CHIIR '25, page 57–70, New York, NY, USA, 2025. Association for Computing Machinery.

[50] Charles L.A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, page 659–666, New York, NY, USA, 2008. Association for Computing Machinery.

[51] James E. Bailey and Sammy W. Pearson. Development of a tool for measuring and analyzing computer user satisfaction. *Management Science*, 29(5):530–545, 1983.

[52] Jack J. Baroudi and Wanda J. Orlikowski. A short-form measure of user information satisfaction: A psychometric evaluation and notes on use. *Journal of Management Information Systems*, 4(4):44–59, 1988.

[53] Albert Bandura. Guide for constructing self-efficacy scales. In *Self-Efficacy Beliefs of Adolescents*. Emerald Publishing Limited, 01 2006.

[54] Kathy Brennan, Diane Kelly, and Yinglong Zhang. Factor analysis of a search self-efficacy scale. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, CHIIR '16, page 241–244, New York, NY, USA, 2016. Association for Computing Machinery.

[55] Torkel Klingberg. *The overflowing brain: Information overload and the limits of working memory*. Oxford University Press, 2009.

[56] Ian G Kennedy, Dean Whitehead, and Debra Ferdinand-James. Serendipity: A way of stimulating researchers' creativity. *Journal of Creativity*, 32(1):100014, 2022.

[57] Allen Foster. A nonlinear model of information-seeking behavior. *Journal of the American society for information science and technology*, 55(3):228–237, 2004.