

# A Reflection on Tests of AI-Search Tools in the Academic Search Process at the Royal Library, Denmark: A Case Study

**Lorna Wildgaard**

Royal Library, Copenhagen University Library, Copenhagen, Denmark,  
[lowi@kb.dk](mailto:lowi@kb.dk), [orcid.org/0000-0002-3900-5058](https://orcid.org/0000-0002-3900-5058)

**Anne Vils**

The Royal Library Denmark, Copenhagen, Denmark and Aarhus University Library, Aarhus, Denmark, [anvm@kb.dk](mailto:anvm@kb.dk), [orcid.org/0000-0003-3949-7045](https://orcid.org/0000-0003-3949-7045)

**Solveig Sandal Johnsen**

The Royal Library Denmark, Copenhagen, Denmark and Aarhus University Library, Aarhus, Denmark, [ssjo@kb.dk](mailto:ssjo@kb.dk), [orcid.org/0000-0002-4782-1725](https://orcid.org/0000-0002-4782-1725)

## Abstract

Academic search literacy and searches powered by artificial intelligence are a focus of the Royal Library and affiliated university libraries in Denmark. The ambition is to integrate AI-search tools in teaching and services that support literature seeking and hence improve the efficiency of the academic search process. However, before doing so, the library managers needed to learn more about the value AI-powered search tools have for information specialists and library users, and hence make informed decisions regarding investment in such tools. This paper presents a case study of two AI-search tools, which were tested via Think-aloud tests, a hackathon and an expert quality assessment at the Royal Library, Denmark. The results point to both opportunities and barriers for the implementation of AI-search tools at the library and we explore the consequences the results of the tests can have for library users and library services. In conclusion, there is a need

for more research on the value of AI-search tools for information specialists and library users. AI-search tools are continuously being developed and improved. The library needs to provide a critical approach to where in the search process the tools add value. Accordingly, the library needs to develop guidance on how to use AI-search tools as a supplement to more traditional approaches, how to report the use of the tools as part of an academic study and address the limitations of the tools.

**Keywords:** Academic search; Artificial intelligence; Search literacy

## 1. Introduction

Academic search literacy and Artificial Intelligence (AI) search are a focus of future researcher support strategies in the Royal Library and its affiliated university libraries in Denmark. The ambition is to integrate AI-search tools in teaching and services that support literature seeking and improve the efficiency of the academic search process. In this paper, we define an academic literature search as a considered, systematic and thorough search to find key literature, of good quality, across multiple databases and relevant to a specific topic. Conducting an academic search is a comprehensive and detail-oriented task. It is performed as part of a research activity or academic study where parameters such as quality, efficiency, reliability, documentation and transparency are integral. However, before investing time, money and resources in realising the ambition of AI-powered search support the first step for the library is to learn more about the value AI-search tools bring to the academic search.

This paper presents a case study of selected AI powered search tools, (referred to as "AI-search tools" throughout the remainder of this paper). The tests of the AI-search tools were conducted over a two-year period, between 2020–2022. In this paper, we reflect on the value AI-search tools bring to the search and the practice of professional information seeking. We look at the creation of new roles for the library in the development of AI-search tools and services that help library users search in a more effective and perhaps, granular ways. Further, central observations to the usefulness of AI-search tools in an academic search are discussed. In doing so, we move beyond the presentation of our methods, results and discussion of results, which are presented in detailed reports published in Zenodo (Johnsen et al., 2022; Wildgaard et al., 2020, 2021) and instead consider the consequences of the results for future practices at the library.

## 2. Background

AI technology is already used to support all or certain parts of an academic search and review process. For example, in deduplication (Arno et al., 2022), risk of bias assessments (Arno et al., 2022; Khalil et al., 2022; Marshall & Wallace, 2019; Zhang et al., 2022), and data extraction (Hofman-Apitius et al., 2009; Khalil et al., 2022). Technologies which expedite screening processes, such as Rayyan (Ouzzani et al., 2016), EPPI Reviewer (Thomas et al., 2010) and Abstrackr (Wallace et al., 2012) amongst others, are mature enough to be established tools. Yet studies concerning the maturity of AI-search tools, that is tools which automate the literature search processes, suggest a need for improvement (Marshall & Wallace, 2019). There are barriers that need to be overcome before the practical implementation of AI-search tools in conducting academic searches becomes viable, not least of which is the prohibitive cost (Khalil et al., 2022). Another barrier is the tool's inability to fulfill methodological standards and protocols required by review writers. These issues concern principally the accuracy and trust in the tool's methodology (Arno et al., 2022). Other barriers include the inability of the AI tool to replicate the nuance of human judgement and expert opinion in assessments of relevance (Arno et al., 2022), and the continued development of and access to the AI tool particularly after the grant period funding the software development has ended as the grant period is singular (Khalil et al., 2022). These barriers are difficult to solve, and on top of that, expectations regarding what AI-search tools can do, are perhaps unrealistic.

The vast and growing amount of literature renders a manual approach to search a time-intensive and impractical approach, so surely AI-search tools must be able to bring some sort of value to the search process? They provide the technology needed to be able to search across vast amounts of text data quickly. But they can do more than just search quickly. AI-search tools combine elements of semantic search, natural language processing, clustering and classification techniques. They can recommend literature based on weighting the searcher's and similar searchers' search behaviour. The results of each search query are tailored to fit, based on semantic and contextual similarities in the text corpus, in the metadata and sometimes on the previous search activities of the user. The tools aim to deliver personalised results that can improve the relevance of the search results and make the search more efficient and more selective, thus, reducing the number of abstracts that need to be screened by the searcher and save the searcher time and effort (Orgeolet et al., 2020;

Polonioli, 2020). The tools make use of natural language processing and unlike a conventional search engine, map to synonyms and related concepts, and link these synonyms and concepts to literature (Kricka et al., 2020; Polonioli, 2020). They harvest the power of neural networks, which are invaluable in identifying trends and novel associations between the concepts in a search query, which can both contribute to innovation and facilitate human understanding (Gozzo et al., 2022; Kricka et al., 2020). Time is something AI-search tool developers and searchers assign great value to, as discussed in Zhang et al. (2022). By saving time, the AI-search tool can enhance the efficiency and speed of the search, and time can be reclaimed to focus on other activities in the research project (Beller et al., 2018; Marshall & Wallace, 2019).

In summary, AI-search tools have the potential to add value to academic searches, they can analyse a great amount of text and metadata in a short amount of time. Using mathematical formulas, pattern recognition, and machine learning, they can potentially make decisions and suggest results more quickly and precisely than the human searcher can do.

We therefore designed our tests to investigate whether using AI-search tools adds or subtracts value from an academic search based on these parameters.

## **2.1. Objectives**

The project aimed to:

- Acquire knowledge about specific AI-search tools and their application in an academic search.
- Focus on the role of the information specialist.
- Acquire knowledge of the use of AI-search tools in the context of responsible conduct of research and research integrity.

## **2.2. Research Questions**

Accordingly, we pose the following research questions:

- How do AI-search tools support an academic search, where values such as efficiency, trustworthiness, quality, reliability, documentation, and transparency of a system are paramount?

- What value do AI-search tools bring to the academic search process for information specialists?
- What value do AI-search tools bring to the academic search process for library users?

### 2.3. Project Team

The project team at the Royal Library, Denmark, was brought together after a visit to Finland in December 2019. The library at the University of Helsinki had recent experience in implementing an AI-search tool in academic search practices and developing services around the tool. As the leadership at the Royal Danish Library wished to develop support in systematic search and review, we wanted to learn more from the University of Helsinki about the strengths and weaknesses of the AI-search tool, their services, users' needs, use-cases and licence negotiations. Two members of the project team were part of this "innovation-visit" but neither had specific competencies or previous experience in AI before the visit to Helsinki. Two members joined the team after the trip, chosen for their experience in systematic reviews and systematic searching. All four had a background in Library and Information Science, but with no specific skills in the technical aspects of AI. The project group represented different departments at the Royal Library and different disciplinary expertise: Business and Law, Health, Social Science and Humanities and Researcher Support.

The project team reported to a "reference-group" to whom midway reports, milestones and deliverables were presented and discussed. The reference group consisted of individuals with relevant search methodological expertise and technical skills in data analysis, including technical knowledge of search algorithms and information retrieval.

### 2.4. Dissemination and Communication

Over the course of the project, the team has published and communicated the methodology, results and implications of the study. The collected output is available on the Zenodo site: *Artificial Intelligence and Literature Seeking*<sup>1</sup>. Content published on Zenodo includes reports of:

- the selection of the AI-search tools included in our tests.

- the method and framework for the Think-aloud tests and for the Hackathon.
- pilot tests of the Think-aloud tests, protocol of the Hackathon, and
- the results of the tests and other materials produced in the communication of our study.

In the reports we address the applied methods, their limitations and consequences in detail. These methods and all results are therefore not covered in detail in this paper. We refer instead the reader to the Zenodo site.

Results from the project have been communicated at conferences such as: INCONECSS, Virtual Conference (Spring 2022), LIBER Conference in Odense, Denmark (Summer 2022) and the Business Library Association (Spring 2023). The slides are published on our Zenodo site. Further, two articles were published at the very start of the project. The first describing our visit to Helsinki and the second concerning the impact of AI on literature search services at the library. Both articles were published in the Danish Journal "REVV" (Kjær et al., 2020; Lyngsfeldt et al., 2022).

## **2.5. Research Design**

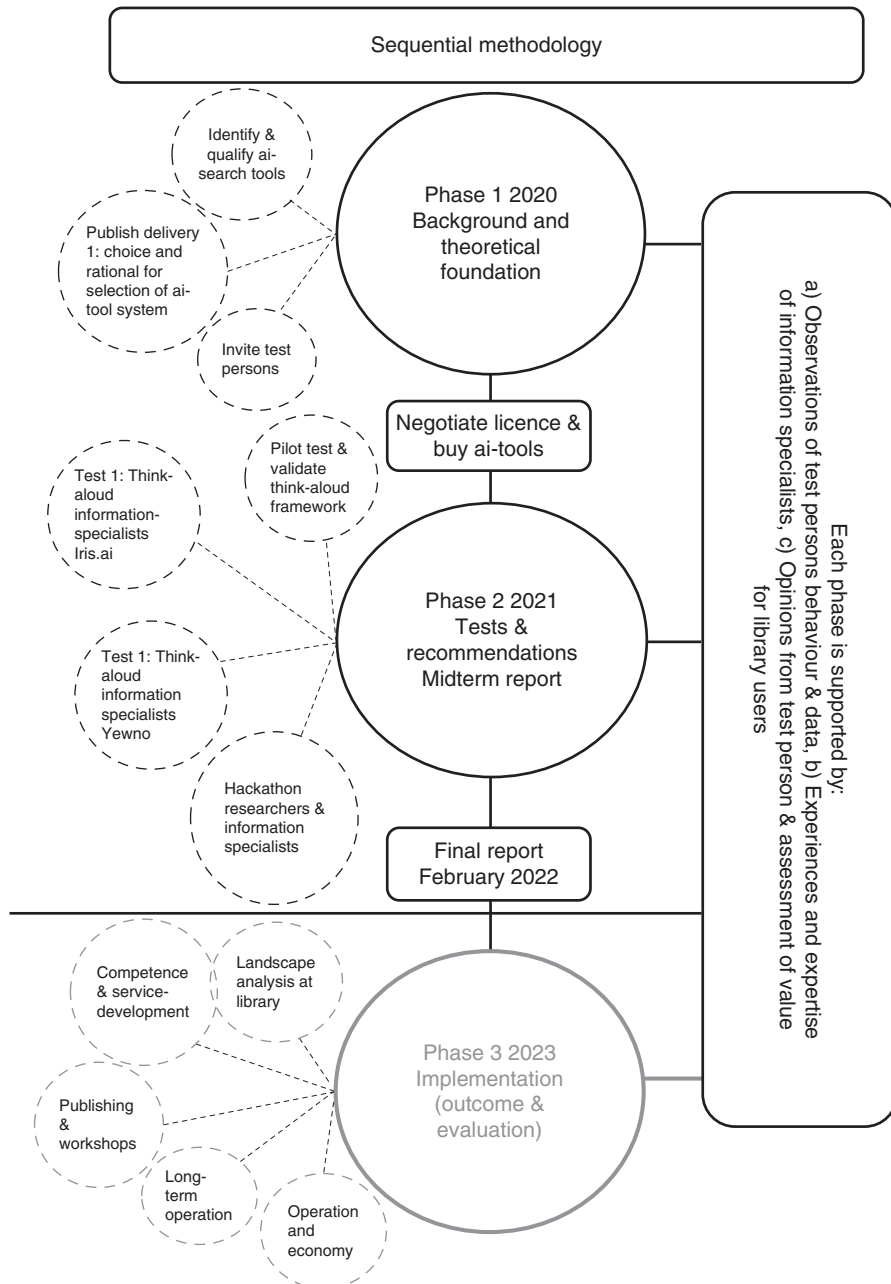
In the following sections, we provide a summary of the applied methods, Section 3, and major findings from our investigation, Section 4. We provide a critical reflection of our observations and consequences for academic searching and librarianship in Section 5. Section 6 provides the conclusions and Section 7 the limitations of our study.

## **3. Methodology**

A sequential method for testing and implementing the AI-search tools was designed and comprised of the following phases (Figure 1):

- 1) Identification and selection of AI-search tools that fulfilled several requirements for academic literature search, research integrity and data security.
- 2) Think-aloud tests of the functionality of selected AI-search tools with information specialists. Search behaviour and observations from

Fig. 1: A sequential approach to identifying, testing and implementing AI-search tools.



the Think-aloud tests informed the design of a hackathon, where researchers and information specialists worked together in the two AI-search tools to solve a common case. Papers identified in the hackathon, were assessed for scientific quality through both qualitative and quantitative assessments. A final report was presented to library managers.

- 3) Based on the final report, the design and implementation of search services at the library. An organisational landscape analysis identifies units and partners in the library who will contribute to the provision of these services and become responsible for maintaining and developing them.

### **3.1. Phase 1**

#### *3.1.1. Selection of AI-Search Tools*

There are many AI-search tools on the market, some free and open source, while others use subscription, 'freemium', or premium models. The latter model comes with the benefit of a product tailored to fit the needs of the user or a specific library catalogue. We assessed both free and paid AI-powered search tools, focusing on the functionalities, declared permissions, third party data collection behaviours and privacy practices of these tools. It is essential that the tools promoted by the library can be used to support academic methods and the responsible conduct of research. This means that any AI tool supporting search, free or paid, must also support methodological transparency, validity, reliability, reproducibility of the search and support good citation practice. Furthermore, in accordance with European law (GDPR), no unnecessary data about the user of the software or any other tracking data may be collected without consent. Thus, the data flow was also a consideration in our evaluation of the various tools.

The requirements described above were formalised into the following list.

- 1) Use one or more of the building blocks of AI-powered search (aggregated behaviour, recommendations, semantic search, and clustering and classification techniques (Wildgaard et al., 2020, p. 1).
- 2) Be designed to support academic literature search.
- 3) Be designed to support discovery of related literature/concepts.



- 4) Be available for testing over a 2–3 year period.
- 5) Be suitable for application in full text resources databases of references and abstracts in the health science disciplines.
- 6) Have clear policies and permissions regarding data collection behaviours and privacy practices.
- 7) Support the responsible conduct of research, such as providing functionalities supporting methodological, transparency, reproducibility and good citation practice including documentation of the applied search and ranking/clustering algorithms (Wildgaard et al., 2020, p. 3).

The search for AI-powered search products was undertaken between May 2020 and October 2020. Software was found by searching the internet and through contact with other libraries that were either in the process of investigating or had investigated the use of AI-search tools. Further, a request for information about other AI-search products not identified in our search was posted on the mailing list “Expert Searching” in September 2020. The search was completed on October 1st, 2020.

## 3.2. Phase 2

### 3.2.1. *Think-Aloud Tests*

Think-aloud tests are a dominant method in usability testing. In a “Think aloud test”, you ask test-takers to use the system while thinking aloud all the time – that is, ...“they simply verbalise their thoughts as they navigate through the user interface”(Nielsen, 1993, p. 195).

Users are asked to say everything they see, think, do, and feel at any given moment. Think-aloud gives us a window into the mind of the system user. There is a concern raised in the literature on the validity of Think-aloud as a methodological approach, specifically concerning the benefits of concurrent and retrospective testing in capturing usability problems in a system and in participant experiences (van den Haak et al., 2003). According to Nielsen (1993), however, the strength of this method is “the wealth of data it can collect from a small number of users” and “show what the users are doing and why they are doing it while they are doing it in order to avoid later rationalizations”(Nielsen, 1993, p. 195). van den Haak et al. (2003) concur:

Think-aloud testing helps us find out what users think about systems and requirements they have to the system in context of real-world experience. Hence, we consider insights into the value of AI-search tools can be gained and we can at the same time observe the participants search behaviour.

Our Think-aloud tests were designed to test the extent an academic search could be conducted in the identified AI-search tools. As the tools identified in Phase 1 were designed with different functionalities and purposes, we developed a set of tasks to fit each tool.

Pilot tests of the Think-aloud tasks were conducted in each AI-search tool in April and May 2021. Two information specialists took part in the pilots, where we assessed the order of the tasks, task formulation and presentation, and how best we could record and respond to the participants verbalisations during the tasks. After each pilot, the tasks were validated by two independent experts. The expert validation included assessment of linguistic validity, content validity and construct validity. The design of the Think-aloud tests was adjusted accordingly.<sup>2</sup>

Ten information specialists were invited to take part in the Think-aloud tests. Before the think aloud tests, the information specialists were taught the basics of how to search in the selected AI-search tools. Training was provided as webinars hosted by developers and consultants from the AI-search tools.

Before each Think-aloud test, the information specialists answered a pre-test survey, enquiring into their demographics, academic search skills and knowledge of AI-search.

The tests were held at the university libraries in Aarhus and Copenhagen. Two testers ran each test as moderator and notetaker. The moderator led the dialogue and guided the testtaker through the set of tasks described in the Think-aloud framework. The notetaker observed search behaviour and reactions, and recorded each testtaker's verbalisations. Further, the testtakers' audio and screen were recorded using Zoom.

The test-takers searched the AI-search tool using the same case. The case was presented at the start of the Think-aloud test. No preparation or prior knowledge of the case was required. Each test-taker completed a series of tasks that took them through all the functions of the tool and were asked to consider the

transparency and documentation of the search. They were asked to verbalise continuously what they were doing, why and their understanding of the search process in each tool.

After the test, the test-takers answered a post-test survey, rating the functionalities and their satisfaction with the tool. They were asked to provide use scenarios and where they thought the tool added value to their work.

### 3.2.2. *Hackathon*

A Scientific Hackathon is a design-sprint like event in which teams of people are exposed to a problem or challenge that requires a collaborative way of problem solving (“Scientific Hackathon”, 2023). We adopt Wu et al.’s (2018) approach to a Hackathon as the challenge “to “hack” a given scientific problem.” Accordingly, we provided a problem, in the form of a search case that was flexible and provided a baseline for the comparison of the search tools. The hackathon test setup is described in (Schoeb et al., 2020; Wu et al., 2018).

Prior to the Hackathon all participants were invited to an introductory webinar in each of the AI-search tools. Supplementary learning material was provided.

On the day of the Hackathon, 8<sup>th</sup> November 2021, the participants were introduced to the project and the main objectives. They were invited to complete a short introductory survey about their knowledge and skills in information seeking, their knowledge/experience in the test tools and databases, and other demographic information (institutional affiliation, status as researcher or information specialist).

The participants were divided into predefined groups to ensure compatibility within the groups and homogeneity across the groups. Three groups consisted of researchers and information specialists and represented a research team construction. A fourth group consisted of information specialists alone. Two of the research team groups searched with one of the selected AI-search tools in each group, respectively Group 1 and Group 2. The information specialist group (Group 3) searched in both AI-search tools. The aim of Group 3 was to investigate the extent information specialists could independently identify relevant results to a complex research question using the AI-search

tools without reference to a researcher. Group 4, the third research team group, searched in a set of traditional resources, and thus functioned as a control group.

The four groups worked on the same problem statement and were required to develop their own research question and consequently design their own search in the search tool or databases they were assigned to. They were tasked with finding up to 10 publications answering their research question. Throughout the Hackathon four observers noted the interaction within the groups using a pre-set observation template. The Hackathon ran 4,5 hours over one day. All four groups worked independently but in proximity to each other. Throughout the day, the groups were provided with technical and practical support and importantly refreshments.

After the Hackathon, a debriefing survey was sent to the participants asking them to rate the systems and comment on the search experience.

### *3.2.3. Qualitative and Quantitative Assessment of Papers Identified in the Hackathon*

On January 2022, the papers each Hackathon group identified as relevant were sent to a qualitative expert assessment. A committee of fourteen independent experts were assigned 2–3 articles from the Hackathon each. They did not know if the articles were retrieved by AI-search tools or by the control group, by the research team groups or by the group of information specialists. The experts had experience in science communication, research integrity and knowledge of the methodologies and analyses relevant to assess the identified articles. They were tasked with evaluating the scientific rigour of each work. They ranked each article on a scale of 1–10 regarding their trust in the results and conclusions. Further, the experts indicated whether the article answered and/or was relevant to the posed research question (Table 6 in Johnsen et al. (2022)).

Finally, we undertook a quantitative assessment of the papers identified as relevant. We described each paper noting the following information: Language, study type (such as opinion paper, primary study, meeting report, etc.), applied method, source publication, publisher, type of peer review and documentation of reviewer support/ethics.

Further, we noted the article impact factor, JIF quartile, number of citations, if the paper was published open access and if the source publication was included in the Danish bibliometric research indicator<sup>3</sup>.

In combination with the qualitative expert review, we hypothesized that the quantitative analysis described above could inform our quality assessment of the found papers.

A detailed report of the Hackathon, the protocol, search case, pre and post survey, instructions, observational framework and expert assessment template is published in Zenodo (Johnsen et al., 2022).

### **3.3. Phase 3**

#### ***3.3.1. Service Design and Implementation***

The intention with Phase 3 was to use the findings from Phase 1 and Phase 2 to inform the acquisition of one or more AI-search tools at the library and design a service structure around these tools.

However, as the results of our tests pointed to the immaturity of the tested AI-search tools the project was closed after Phase 2. A service infrastructure was not developed.

## **4. Results**

### **4.1. Selection of AI-Search Tools**

We identified 16 AI-search tools marketed for academic application. Each tool was examined according to our requirements that is to which extent they indeed supported academic literature search and research integrity, including transparency, reproducibility and data security. Two products, Iris.ai and Yewno.discover, met the posed requirements. The remaining AI-search tools failed on aspects such as data security, academic searching, maturity, content that could be searched and the possibility of documentation. A detailed analysis of the AI-search tools considered for inclusion in our project are presented in Wildgaard et al. (2020).

## **4.2. Main Results of the Think-Aloud Tests**

A report of the results of the Think-aloud tests, methodological framework and set of tasks is published in Zenodo (Wildgaard et al., 2021). In the following, we provide a summary of the main results.

Ten information specialists were invited to take part in the Think-aloud tests in May and June 2021. Nine completed the tests, resulting in five tests in Iris.ai and four in Yewno.discover. The participants were from the natural and technical sciences, business and social sciences, and arts and humanities. They described themselves as good (n1), intermediate (n3) and expert (n5) in a systematic review and systematic search. All were new to the tested AI-search tools.

The participants agreed that Iris.ai and Yewno.discover have the potential to be a supplement to the databases and search systems that the university libraries already offer. They considered the AI-search tools to be especially useful at the start of a project for both students and PhD students, as the tools can be used to explore and investigate concepts, find cross-disciplinary topics and thus generate ideas. Further, they pointed out that AI-search tools may contribute to innovation at the universities, as they “break new ground” in the way they search and present results.

The AI-search tools present search results as a graphical presentation similar to a sociogram, where the searcher can move through hierarchical nodes and links between nodes to explore topics and find literature. This form of display is very different from the bibliographic list of results that is produced in databases employed in academic searches. The graphical display was assessed as valuable by the information specialists. They suggested that the AI-search tools could in this regard be useful in the pre-award stages of a funding application, as the user gains a visual insight into evidence clusters and evidence deserts, trends, disciplinary perspectives and interdisciplinary connections in relation to a research question.

However, the AI-search tools challenged the information specialists' approach to the search. In traditional bibliographic databases information specialists use Boolean, proximity and field search operators. Whereas the search features in the AI-search tools encourage the searcher to explore

connections between results and to investigate the metadata describing the results to identify concepts and related works. As a result, the information specialists found that using the AI-search tools required in-depth knowledge of the functionality of the tool and in-depth knowledge of the specific domain of the search query to be able to assess the success of the search. They commented that in traditional databases, they typically carry out all the search technicalities on behalf of the researcher (i.e., identification of search terms and search string setup). However, in the AI-search tools, they need to be able to make qualified relevance assessments at an early stage in the search process, which in turn form thoughtful exploration of the concepts and literature discovery together with the researcher.

In the AI-search tools, a semantic algorithm assesses the “aboutness” of the texts and links the texts to a list of automatically generated concepts. Dependent on the tool, these concepts can be grouped together to define broader topics and can be labelled with either a name and short definition or just by a number. At the time of testing, both tools worked with concepts, meaning that search terms were not highlighted in the retrieved text, but rather text snippets that are “about” the concept in a specific context were shown to the searcher. The information specialists were sometimes confused by the suggested concepts and how they represented their query statement. They were unable to remove irrelevant concepts from the search. They disagreed with the tools’ “aboutness” assessments, and they were critical of how concepts were defined in the tool.

It was emphasized that both Iris.ai and Yewno.discover could be useful tools for navigating the available Open Access literature as both tools searched open access resources. However, the information specialists were not satisfied with the academic quality of the identified papers in these resources. Overall, the information specialists were not satisfied with the transparency and relevance of the search and the extent to which the search could be documented.

#### **4.3. Observations and Results from the Hackathon, and the Qualitative and Quantitative Assessments**

The assessment of the data collected during the Hackathon and the qualitative and quantitative analyses were completed in March 2022. Hackathon

*Table 1: Participant demographics and group assignment.*

Group	System	Title*	Specialty	Experience SR**	Experience Iris.ai/Yewno
1	Iris.ai	I	Health	Good	Beginner
1	Iris.ai	R	Health	Good	Beginner
1	Iris.ai	I	Health	Expert	Beginner
1	Iris.ai	R	Health	Good	Beginner
2	Yewno.discover	R	Business and social science	Intermediate	Beginner
2	Yewno.discover	I	Natural Science & Technology	Expert	Beginner
2	Yewno.discover	R	Business and social science	Intermediate	Beginner
2	Yewno.discover	I	Business and social science	Expert	Beginner
3	Iris.ai & Yewno. discover	I	Business and social science	Expert	Beginner
3	Iris.ai & Yewno. discover	I	Natural Science & Technology	Good	Beginner
4	Control	I	Health	Expert	Beginner
4	Control	I	Health	Expert	Beginner
4	Control	R	Health	Good	Beginner
4	Control	R	Health	Good	Beginner
4	Control	R	Health	Good	Beginner

\*R = researcher, I = information specialist.

\*\*Participants' subjective evaluation of skills in systematic search and review (SR).

The control group, Group 4, searched in PubMed, Web of Science and Google Scholar.

group demographics are presented in Table 1. The main observations from the Hackathon are presented in Section 4.3.1. The results of the qualitative and quantitative analysis are presented in Section 4.3.2, and Section 4.3.3.

Two of the research team groups searched using the AI-search tool, respectively Iris.ai (Group 1), and Yewno.discover (Group 2). The information specialist group (Group 3) searched in both Iris.ai and Yewno.discover. Group 4, the third research team group, searched in a set of traditional resources, and thus functioned as a control group. They searched in PubMed, Google Scholar and Web of Science.

### ***4.3.1. Main Observations from the Hackathon***

Groups 1, 2 and 3 were challenged by the completely new approach to searching in the AI-search tools. They missed a comprehensive search history and



struggled to document the search. They had a lack of understanding of the applied algorithms. These observations are in alignment with our findings from the Think-aloud tests.

Groups 1 and 3 understood Iris.ai as a tool that provided [quote] “a different way to find relevant literature”, which [quote] “...promotes the encounter with literature that one is not necessarily aware one is seeking”. Iris.ai was also noted to encourage a creative process which as one Hackathon participant describes[quote] “contributes to you as a literature seeker becoming sharper on what you are actually looking for” (Johnsen et al., 2022, p. 10).

Equally positive, most of the participants in Groups 2 and 3 would use Yewno. discover again, either in teaching or in their own research practice. They discussed the added-value of Yewno.discover in supporting the exploration of a subject for students or researchers and its intuitive approach to the search. This conclusion about value supports the considerations made by the information specialists in the Think-aloud tests who agreed that one of the great advantages of Yewno.discover is that it allows a user to explore a concept and the relationship between several concepts. However, we observed during the Hackathon that some of the participants got lost in the cross-references and links between concepts in Yewno.discover. As such, the linking between concepts and papers appeared to confuse the searcher more than it benefited them. Many different avenues of exploration were opened-up to the searcher and thus, we need to be aware of when and how Yewno.discover is used in the search process. Exploration, as discussed further below, was seen to detract from the efficiency of a search.

The participants did not consider either Iris.ai or Yewno.discover as efficient ways to search. The AI-search tools are designed to encourage discovery and even though the participants acknowledged that the tools could and should not be compared to conventional databases such as PubMed and Google Scholar, they continued to return to these known systems. They did this to verify that their search could retrieve relevant results, that they had defined and applied terms correctly in the context of their research question, and that the rationale of the search made sense. Groups 1, 2 and 3 had difficulty deciding when to stop exploring. As the AI-search tools do not support a systematic or stepwise approach to searching, there seemed to be no natural end to the discovery process.

### 4.3.2. The Qualitative Assessment

The complete results from the qualitative assessment are found in Table 6 in Johnsen et al. (2022).

In total, the four Hackathon groups identified 19 potentially relevant publications that were sent to 14 experts for quality assessment, Table 2. The publications were grouped into seven subject packages that we judged would appeal to the individual experts' specialist knowledge. Each expert was invited to read two or three articles and each package of publications was sent to two experts for independent assessment. Eleven out of the fourteen identified experts took part in the quality assessment, giving a 79% participation rate.

*Table 2: Quality assessment of articles identified as relevant during the Hackathon.*

Group	Trust in results* <i>Expert 1, Expert 2</i>	Trust in conclusion* <i>Expert 1, Expert 2</i>	Relevance score** <i>Expert 1, Expert 2</i>	Relevance
Group 1 (Iris.ai)				
Article1	8, –	7, –	1, –	Not relevant
Group 2 (Yewno.discover)				
Article1	7, –	1, –	1, –	Not relevant
Article2	7, –	1, –	1, –	Not relevant
Article3	2, 5	2, 5	1, 1	Not relevant
Article4	6, 10	7, 10	1, 3	Not relevant/ relevant
Article5	10, 5	4, 5	1, 1	Not relevant
Article6	6, 10	6, 10	1, 1	Not relevant
Article7	5, 6	5, 7	1, 2	Not relevant/ partially
Article8	8, 2	8, 2	1, 1	Not relevant
Group 3 (Iris.ai/Yewno.discover)				
Article1	5, 7	5, 6	2, 1	Partially/not relevant
Group 4 (Control)				
Article1	9, –	9, –	3, –	Relevant
Article2	9, 10	9, 10	3, 3	Relevant
Article3	2, 9	2, 9	1, 2	Not relevant/ partially
Article4	8, 10	8, 6	3, 3	Relevant
Article5	8, 3	6, 2	3, 2	Relevant/partially
Article6	5, 9	5, 9	2, 3	Partially /relevant

\*The experts rated their trust in the results and conclusions in the identified papers on a scale of 1–10, where 1 is no trust and 10 is complete trust.

\*\*Relevance was rated on a scale of 1–3, where 1 is not relevant and 3 is relevant.

Sixteen out of nineteen articles were assessed, resulting in a completion rate of 84%.

The experts rated the articles found in an Iris.ai or Yewno.discover as less relevant than the articles found in the control group. The scientific rigour of the articles and trust in the results were also rated lower in articles found using the AI-search tools than in the control group. The expert assessors did not rate the evidence, the credibility of the method and structure of the content in the articles identified in the AI-search as trustworthy.

#### *4.3.3. The Quantitative Assessment*

The results of the quantitative assessment are found in Table 7 & 8 in Johnsen et al. (2022). Articles identified by the control group were primary studies and even though the publications were very recent, they already had a high number of citations. All articles from the control group were published in the first or second JIF quartile and the majority was cited more than average. The majority of articles identified in Yewno.discover were also in the first JIF quartile and had article impact factors indicating that articles published in these journals are cited more than average for journals in the same subject category. However, the articles identified in Yewno.discover were primarily cases, letters to editors and meeting reports which do not require external peer review. Neither do they contribute to the calculation of impact factors and are as such “hitching a ride” on the prestige created by original papers published in the journals.

Only one relevant article was identified in Iris.ai and it was from a journal in the lowest JIF quartile and cited below average for the discipline. Since only one article was identified, it is not appropriate to draw any conclusions. However, we do need to revisit Iris.ai and look even more closely at which databases/sources the AI-search tools search in.

The quantitative analysis also investigated whether the articles were published Open Access and which form of peer review each article had undergone. The articles identified in the AI-search tools were all Open Access and the publishing sources advocated new forms of peer review, including fast-track, open and with “honorarium”. This is in contrast to the articles found in the control group, which supported single and double blinded peer review.

## **5. Strengths, Challenges and Opportunities for the Library**

The following is a reflection on the outcomes of our study. We explore the possible consequences our results and observations can have for academic searching, the role of the library and the role of the information specialist.

### **5.1. Reflection 1: About the Quality of the Search**

The results of the expert analysis and quantitative assessment indicate that the AI-search tools provided papers of lower scientific quality. Without any knowledge of which system, the retrieved publications came from, the expert group unanimously gave the publications identified in the AI-search tools a lower quality rating. The expert group came from a wide range of disciplines and academic profiles. Consequently, we do not consider the expert group biased toward one form of communication or research methodology over another.

The control group retrieved primary studies, whereas the AI-search tools retrieved meeting reports, case studies, reports, letter to the editor and opinion papers. The expert review of these papers questioned the rigor of the academic approach (Tables 6, 7, 8 in Johnsen et al. (2022)). However, this wide range of publication types is not necessarily a weakness of the AI-search tools. Yewno.discover includes grey literature and non-primary works that can be very accessible and easier to digest during the discovery phase of the search, especially to a novice in a research area. Further, as pointed out by the Institute for Work and Health (IWH); *"...because they [grey literature] aren't tied to a conventional structure, they can be longer and provide more detail."* (Institute for Work & Health, 2008). Specifically, once grey literature and studies lower down the pyramid of evidence are properly understood, it is possible to appreciate how these can further the goal of literature discoverability (Polonioli, 2020).

The quantitative assessment also showed that the AI-search tools support the visibility of Open Access publications, which is another area of strategic interest in our University Library. Furthermore, the publications found by the AI-search tools experimented with other forms of peer review. These forms of peer-review mean that papers can potentially be published more quickly, the peer review report can be accessed, and the peer review process is more

dialogue-based and transparent, thus bringing new research to the forefront in the AI-search tools. However, it is important to determine whether or not the innovations in peer review and the speed with which papers are published can affect their quality and content (Barroga, 2020).

### *5.1.1. Consequences*

If studies of lower methodological quality are included in a literature review in an uncritical manner, then this will affect the integrity of the literature synthesis (NHS Centre for Reviews and Dissemination, 2009). The AI-search tools provided the user with very different publication types and different kinds of peer review. These types of publications require the user to be competent in quality assessment and source assessment.

Searching using the tested AI-search tools, points to the increased need for library support in source criticism and scientific integrity (good conduct of research) so that the found literature can be evaluated to determine which publications are most appropriate for the study. It is the responsibility of the human-user not the machine to assess the scientific quality, bias and integrity of an individual study.

## **5.2. Reflection 2: About Algorithmic Literacy and Research Integrity**

Results from our Think-aloud tests, Hackathon and surveys show that it is difficult for both information specialists and researchers to understand how the AI-search tools “think”. It is not possible for the everyday searcher to understand the level of precision of the search, recall, how and why the AI-search tool clusters the results together or labels the clusters in the way that they do. This lack of understanding may lead to distrust, and questions about the reliability and usefulness of the tools. Our combined results show that transparency of the search tool as well as trust in the scientific content in the retrieved articles are significant factors for the test-takers continued use of and engagement in AI-search tools. Trust is further discussed in Beller et al. (2018) and Gasparini and Kautonen (2022).

An academic search should always be conducted with integrity and honesty. The search needs to be assessed as an appropriate method to find literature in

the selected sources and to answer the research question. The search and the sources should be assessed for bias. Bias assessments in an academic search include assessing the risk of publication bias, database bias and selection bias and the effect bias can have for the retrieval of results. Attention should be paid to bias in the search as bias in the search may result in bias in the results and summary of data in the review synthesis. In an AI-powered search, an assessment should also be undertaken of the extent bias in the applied algorithms effect the search. We did not have the technical skills to assess bias in the search tools in our study.

Understanding the system architecture and intention of the AI-search tool is essential if information specialists aim to provide teaching, support, and high-quality professional search services that support research. Increasing our algorithmic literacy throughout the project extended our awareness of what an academic search could be. We began to think beyond systematic search methods, which are developed to support the responsible conduct of research. At the start of the tests our assumption was, that the search methods should be as transparent as possible and documented in a way that enables the search to be validated, evaluated and to some extent, reproduced. We learned however, through working with the AI-search tools that in an academic search there can be other goals apart from methodological transparency, accountability and reproducibility. An academic search can be enriched, particularly in its early phases, by discoverability and serendipity. Discoverability and serendipity are perhaps restricted in traditional Boolean searches and there can be multiple goals and trade-offs between the AI-search tools and the traditional databases involved in the process of the search.

### *5.2.1. Consequences*

AI-search tools require the user to have a critical and technical understanding of the tool and the requirements for an academic search. They require algorithmic literacy, as proposed by Bakke (2020). The aforementioned applies to all users of the tool (the student, the researcher, and the information specialist) and the library providing services that support the use of the AI-search tools. Hence, as information specialists we need to embrace the different

ways of searching that AI tools bring to the table. We need to understand how they can supplement an academic search, the benefits, and challenges, and how best to document the search in a transparent and honest way.

Approaches to responsible searching in AI-search tools is an important discussion to have and a complicated task to solve.

### **5.3. Reflection 3: About Innovation, Serendipity and Preconceptions**

Both Iris.ai and Yewno.discover support the discovery of different semantic connections between concepts in the text corpus. By generating different perspectives to which literature and concepts may be related, idea generation and creativity may be enhanced, divergent thinking increased, and innovative solutions to a research problem may become more perceptible. In addition, through utilising graphical interfaces and node-graphs rather than lists of index terms or bibliographic references, the tools provide a visual entrance to the literature that invites the user to discover “unintended knowledge”. Unintended knowledge is when unexpected connections to the searched topic are identified during the search (Hofman-Apitius et al., 2009). The graphical interfaces in the AI-search tools have the potential to challenge our preconceptions of a concept or topic, reducing our cognitive bias, and may offer a form of serendipity that is lacking in traditional databases. The interdisciplinarity of the search may be increased.

The AI-search tools enabled the searchers in our Hackathon to put their cognitive biases and pre-understanding of a search query to one side. In Iris.ai and Yewno.discover the searchers had to write a research statement which in turn fed words and phrases to the tools’ algorithms and generated concepts and pathways through the literature. Whereas, in our control group which searched traditional search databases, they formulated their queries using the keywords and Boolean logic (Azzopardi, 2021). When searching for literature in the traditional databases, the searcher may be influenced by the conformity of how to search and the desire to confirm the research question (Azzopardi, 2021) rather than discover new connections, as displayed in the AI-search tools, that may identify new perspectives and papers that refute their question and/or provide alternative solutions.

Our participants in both the Think-aloud tests and in the Hackathon could see a potential in using the AI-search tools in the exploratory phase of a literature search. Yet to adhere to the good conduct of research, the explorative phase of the search needs to be documented and rationalised. How to document an explorative search in the project documentation or in the search protocol is at the current time unclear. A case in point is the systematic search and review process. When referring to the PRISMA-S guidelines for reporting literature searched for systematic reviews (Rethlefsen et al., 2021) it is evident that a reproducible and transparent search strategy is mandatory for documenting searches in bibliographic databases and internet search engines. But there are in PRISMA –S (Rethlefsen et al., 2021) so far, no requirements for reporting the initial searches and exploratory methods. In protocols for the JBI evidence synthesis (Aromataris & Munn, 2020) it is recommended to document that initial searches were performed to find articles on the topic and use them to find relevant free text words and index terms for the final search strategy. When reporting on a Cochrane Systematic Review of Interventions any specific methods used to develop the search strategy should be noted (Higgins et al., 2022) and in the Technical Supplement to Chapter 3, by Lefebvre et al. (2022), it is suggested to use text mining tools to objectively identify terms.

### *5.3.1. Consequences*

By providing services and support in AI-search during the exploratory phase of a research project, the library can contribute to innovation and creativity. The library can be a partner at the very start of the project, right where ideas begin to take shape and gaps in knowledge within and across research domains are identified. Such gaps in the literature can in turn be used in arguments for the necessity of the research and innovation (Beller et al., 2018) and investment from funders. Searching professionally in and providing guidance on AI-search tools will extend the portfolio of current services at our library in Denmark. Inspiration could be taken from the PRISMA-S guidelines section “other methods”. Here, examples of non-reproducible searches in personal files or “similar articles” in PubMed are given. Even though AI-search tools are not mentioned, they could be equated to these methods and thereby supplement the literature search.

However, we first need to upskill our own search expertise in AI-search tools before we confidently can begin to provide support to library users.



#### 5.4. Reflection 4: About the Role of the Library, Services and New Collaborations

After completing our study, we were approached by several AI-search tool developers, eager to engage in feedback on their tool from the perspective of the library. Through meeting with developers, we found that typically, information specialists are not part of the software development team, yet the products are marketed libraries. This means that knowledge of reference management, search processes, search and review methodologies, review workflows and requirements for research integrity may be underdeveloped in tool functionalities.

A collaboration between developers and libraries is also discussed in the literature surrounding AI tools and academic searching. For example, developers are keen for libraries to assess the applicability of the AI-search tool to search different fields, across different types of literature and different languages; semantic and topic development in a search (Clark et al., 2021; Khalil et al., 2022), interface development (Marshall & Wallace, 2019), and user-support, particularly the role of the librarian/library in providing support and expertise (Gasparini & Kautonen, 2022).

Beller et al. (2018) address the barriers and facilitators AI-search tools bring to the search process. They define facilitators as places where libraries and AI software developers have the potential to work together with benefits for both parties (Beller et al., 2018). Gasparini and Kautonen (2022) also argue that AI-search tools provide libraries with the voice to act as a facilitator. Acting as a facilitator can give libraries a new or more proactive role, providing not only a collaborative space for AI-search tool development but also a space to provide technically literate search support, to critically assess the quality of the search AI-search tools produce, and to create awareness about the implications AI-assisted searches can have on the responsible conduct of research. Other topics the library could address with developer and research communities are: commerciality vs. trustfulness and the academic credit of the tool (Arno et al., 2022; Gasparini & Kautonen, 2022), bias and transparency in the search (Beller et al., 2018; Clark et al., 2021), the quality of the metadata (Gasparini & Kautonen, 2022), sorting and filtering algorithms (ACRL Research Planning and Review Committee, 2020; Henry, 2019, pp. 47–65), and the compatibility of the AI-search tool with other tools/workflows commonly used by review writers such as reference management, screening and data extraction programmes (Beller et al., 2018).

#### **5.4.1. Consequences**

The application of AI-search tools at the library opens-up for new opportunities for collaboration with AI-search developers and with research communities. As AI-search tools both entail and counter issues of bias and algorithmic literacy, libraries can take a leading role in experimenting with AI-search. Libraries could work together with AI software developers on the design and functionalities of the AI-search tool. Libraries, as suggested by Gasparini and Kautonen (2022) could even host AI laboratories in which library users and staff learn how to deal with new AI-search technology. The established role of libraries as trusted partners at universities and in research communities provides the library with the opportunity and responsibility to act as a “facilitator” and critically assess AI-search tools and discuss requirements for the tools.

At the library, we need to critically appraise the technology behind an AI-search tool to learn more about the extent the tool supports academic methods and documentation of a search. We must also consider when and how to appropriately use the tool in the search process. Only then can we confidently develop new skills and effective services devoted to supporting quality AI-search tools in an academic context.

#### **5.5. Reflection 5: About Reshaping the Role of the Information Specialist**

The AI-search tools we tested challenged conventional approaches to academic searching and the searcher expectations to search methodologies. However, the tools also encouraged new forms of collaboration and dialogue around the search activity. The control group in our Hackathon, searched in PubMed, Google Scholar and Web of Science, and succumbed to the conformity of how to search as discussed in the Reflection 3. They divided responsibilities according to group member profiles: information specialists conducted the search, and researchers judged the relevance of search terms and results. However, the researchers and information specialists who searched using the AI-search tools explored topic areas, discussed literature and search strategies together. This novel form of collaboration could of course be caused by the “newness” of the AI-search tool, which the participants were curious to explore and evaluate. Yet it is worth considering if discovery systems such as Iris.ai and Yewno.discover do require an increased interaction, discussion and collaboration within a research “team”.

Combining AI-search tools and traditional databases in different phases of the search process may mitigate and manipulate cognitive biases to improve the impartiality of the search, the search behaviour of the searcher and outcomes of the search. By mediating the value and limitations of different approaches to literature retrieval in traditional and AI-search tools, the information specialist can equip searchers with the skills they need to use them in combination with traditional approaches to the search. We need to adjust our expectations to what a search on an AI-search tool can deliver; it does not provide a single search solution even though the search is conducted across thousands of multidisciplinary documents.

The role of the information specialist in identifying and clarifying core aspects of the search does not cease with introduction of AI-search tools at the library. If anything, the role of the librarian is rejuvenated, and we are given the opportunity to develop methods for searching in AI-search tools and develop protocols for documenting AI supported approaches to academic search. Asemi et al. (2021), Bethard et al. (2009), and Ewing and Hauptman (1995) predicted the extinction of information specialists in the scholarly ecosystem when professional expertise and behaviour are transferred to automated systems (Gasparini & Kautonen, 2022). On the contrary, our results show the increased need for information specialist involvement in the AI-search process and in AI-search tool development. Gasparini and Kautonen (2022, p. 8) go so far as to suggest, that information specialists need to. *"abandon their old paradigms, practices, and workflows."* Abandonment of existing professional approaches to the search is perhaps extreme. Knowledge and skills in AI-search tools enable information specialists to define new professional roles for themselves and nurture new roles and collaborations in the search and review process with students, researchers, IT and enterprise units, (ACRL Research Planning and Review Committee, 2020; Nolin, 2013) as cited in Gasparini and Kautonen (2022).

### 5.5.1. Consequences

AI-search tools give information specialists the opportunity to reshape the academic literature search and their role during the search and review process. As information specialists we have the unique position to engage in the interaction between the librarian – searcher – AI-search tool (Jakeway et al., 2020; Kennedy, 2019). Rather than rejecting traditional support roles, AI-search

tools provide information specialists with a new way of searching and the potential to engage in new collaborations with searchers as part of a “research team” and, as discussed in Reflection 4, with AI software developers.

## **6. Conclusion**

Iris.ai and Yewno.discover did not support an academic approach to the search sufficiently for the library to renew licences to the tools and develop services around them. Thus, phase three of our project was not implemented. The library continues to investigate the application of AI technology in search tools such as ChatGPT (<https://openai.com/blog/chatgpt>) and monitor the consequences such tools may have on the academic search behaviour of researchers and students. The topics we are investigating are, but not limited to, the use of ChatGPT in answering exam questions, writing assignments, and making literature summaries. We have recently established a national AI and systematic review working group, tasked with monitoring the development of open source AI-search tools and their usefulness in systematic and academic searching.

The results of our Think-aloud tests, Hackathon, quantitative and qualitative assessments point to the immaturity of Iris.ai and Yewno.discover in supporting an academic search, where values such as efficiency, trustworthiness, quality, reliability, documentation and transparency of the search are paramount. Both AI-search tools provided limited possibilities to document the search. The search was identified as unreliable and trust in the validity of the search was low. Our tests indicate that Iris.ai and Yewno.discover are less effective than traditional search resources such as PubMed, Web of Science and Google Scholar in identifying relevant research of a high scientific quality. The scientific quality of the papers retrieved in the AI-search tools scored low on the qualitative assessment. Searching using AI-search tools require an increased knowledge of and technical skills in system architecture, source criticism and research conduct.

We found AI-search tools may have an increased value for researchers and students at the start of the search process. The tools enhance the innovation and discovery. They have the potential to challenge ones preconceptions and reduce cognitive bias in the search process. They can promote hypothesis generation and objective exploration of the literature. With a strategic focus at

our affiliated universities on interdisciplinarity research, this is an interesting capability. Further, both tested AI-search tools searched Open Access publications and hereby “open” research was made more visible in the search. With the Open Science movement also a national strategy, AI-search tools can have a powerful role in promoting Open Access literature and data.

For the information specialist, we identified new, collaborative roles that can bring value to the quality of the search process and can reshape the role of the information specialist. The information specialist is an important partner for AI-search tools developers. The information specialist can facilitate the development and support of AI-search tools in the future. There is still much work to do to improve the documentation of the search conducted in the AI-search tools, both within the tool itself and in project and search protocols.

In summary, the tested AI-search tools force the user to let go of traditional approaches to the search and they broaden our perception of how an academic search can be conducted. Such a reflection may cause a paradigm shift in information seeking practice that demands new terminology, understanding, standards and expectations to the search and to the competencies of the searchers. More research is needed to show the usefulness of AI-search tools in an academic search and guidance needs to be written on how to report the use of them.

## **7. Limitations**

The AI-search tools identified in phase one of our study were found and examined in Autumn 2019. During our two-year study, functions of these tools have improved, and their interfaces have been developed further. During the tests, the stability of Iris.ai and Yewno.discover may have had an effect on their perceived value and technical problems could have biased the participants against the tools. The development of AI-search tools throughout the project is not a factor we have considered, and we have not revisited the tools after the testing period was complete. However, during the testing period, we were in contact with the development teams from Iris.ai and Yewno.discover and these teams were quick to implement our suggestions for improvements and discuss our requirements to an academic search regarding documentation and transparency.

Irias.ai and Yewno.discover are not directly comparable to each other or with the control group. Both Iris and Yewno.discover are discovery tools. Yewno.

discover is designed to develop questions in research projects by exploring semantic links between documents. Iris.ai also explores semantic links but additionally, it includes processes to list retrieved articles. These lists can be used to systematically narrow down the search results over three screening phases using concepts identified in the text corpus to filter the results.

Our study was a small-scale study. Nine information specialists were involved in the Think-aloud tests and seven researchers and eight information specialists took part in the Hackathon. This relatively small number of participants gave us an in depth, rich, but narrow view of the value of the tested AI-search tools. Our findings would be more generalisable if we involved more people in our tests. Further, no students were involved in tests at any time. Therefore, conclusions on the value of the tools for bachelor- and master students are assumptions based on our observations and based on statements from the test participants.

The pre-made cases used in both the Think-aloud tests and the Hackathon were fictive. The participants were not domain-experts in the field of our chosen case. We attempted to account for differences in disciplinary knowledge by making a generic case that could be queried from different disciplinary perspectives. In future studies, we could consider “real” test-cases where we follow a cohort of searchers over a longer time. This approach would improve the robustness of our assessments of the value of the tools and relevance of the found literature.

Finally, we recognise that our results could have been influenced by the design of the Hackathon. The participants had a very limited time to formulate a search query, search with the tools, and assess the relevance of the results.

## **Acknowledgements**

We thank all the researchers and information specialists for their enthusiasm and contribution to the Think-aloud tests and the Hackathon. Thank you to the experts who screened the publications and supplied in depth evaluation of the content. Thank you to the developers of Iris.ai and Yewno.discover who promptly answered our constant stream of questions.

The authors declare no conflicts of interest.

## References

- ACRL Research Planning and Review Committee. (2020). 2020 top trends in academic libraries. *Library Faculty Presentations & Publications*, 21(6), 270. <https://doi.org/10.5860/crln.81.6.270>
- Arno, A., Thomas, J., Wallace, B., Marshall, I. J., McKenzie, J. E., & Elliott, J. H. (2022). Accuracy and efficiency of machine learning-assisted risk-of-bias assessments in “real-world” systematic reviews a noninferiority randomized controlled trial. *Annals of Internal Medicine*, 175(7), 1001–1009. <https://doi.org/10.7326/m22-0092>
- Aromataris, E., & Munn, Z. (2020). *JBIM Manual for Evidence Synthesis*. JBI. <https://doi.org/10.46658/JBIMES-20-01>
- Asemi, A., Ko, A., & Nowkarizi, M. (2021). Intelligent libraries: a review on expert systems, artificial intelligence, and robot. *Library Hi Tech*, 39(2), 412–434. <https://doi.org/10.1108/LHT-02-2020-0038>
- Azzopardi, L. (2021). *Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval*. SIGIR’ 21: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 27–37. <https://doi.org/10.1145/3406522.3446023>
- Bakke, A. (2020). Everyday googling: Results of an observational study and applications for teaching algorithmic literacy. *Computers and Composition*, 57, 1–16. <https://doi.org/10.1016/j.compcom.2020.102577>
- Barroga, E. (2020). Innovative strategies for peer review. *Journal of Korean Medical Science*, 35(20), Article e138. <https://doi.org/10.3346/jkms.2020.35.e138>
- Beller, E., Clark, J., Tsafnat, G., Adams, C., Diehl, H., Lund, H., Ouzzani, M., Thayer, K., Thomas, J., Turner, T., Xia, J., Robinson, K., & Glasziou, P. (2018). Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR). *Systematic Reviews*, 7, Article 77. <https://doi.org/10.1186/s13643-018-0740-7>
- Bethard, S., Ghosh, S., Martin, J., & Sumner, T. (2009). Topic model methods for automatically identifying out-of-scope resources. *JCDL’ 09 International Conference on Digital Libraries (2009)*, 19–28.
- Clark, J., McFarlane, C., Cleo, G., Ishikawa Ramos, C., & Marshall, S. (2021). The impact of systematic review automation tools on methodological quality and time taken to complete systematic review tasks: Case study. *JMIR Medical Education*, 7(2), Article e24418. <https://doi.org/10.2196/24418>
- Ewing, K., & Hauptman, R. (1995). Is traditional reference service obsolete? *The Journal of Academic Librarianship*, 21(1), 3–6. [https://doi.org/10.1016/0099-1333\(95\)90144-2](https://doi.org/10.1016/0099-1333(95)90144-2)

Gasparini, A., & Kautonen, H. (2022). Understanding artificial intelligence in research libraries – Extensive literature review. *LIBER Quarterly*, 32(1), 1–36. <https://doi.org/10.53377/lq.10934>

Gozzo, M., Woldendorp, M. K., & De Rooij, A. (2022). Creative collaboration with the “brain” of a search engine: Effects on cognitive stimulation and evaluation apprehension. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, 209–223. [https://doi.org/10.1007/978-3-030-95531-1\\_15](https://doi.org/10.1007/978-3-030-95531-1_15)

Henry, G. (2019). Research librarians as guides and navigators for AI policies at Universities. *Research Library Issues*, 299, 47–65. <https://doi.org/10.29242/rli.299.4>

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2022). *Cochrane handbook for systematic reviews of interventions version 6.3*. The Cochrane Collaboration. <https://training.cochrane.org/handbook>

Hofman-Apitius, M., Younesi, E., & Kasam, V. (2009). Direct use of information extraction from scientific text for modeling and simulation in the life sciences. *Library Hi Tech*, 27(4), 505–519. <https://doi.org/10.1108/07378830911007637>

Institute for Work & Health (2008, April). *What researchers mean by...Grey literature*. <https://www.iwh.on.ca/what-researchers-mean-by/grey-literature>

Jakeway, E., Algee, L., Allen, L., Ferriter, M., Mears, J., Potter, A., & Zwaard, K. (2020). *Machine learning + libraries summit event summary*. Library of Congress. <https://labs.loc.gov/static/labs/meta/ML-Event-Summary-Final-2020-02-13.pdf>

Johnsen, S. S., Lyngsfeldt, J., Vils, A., & Wildgaard, L. (2022). Exploring Iris.ai and Yewno.discover with a Hackathon and expert quality assessment [Report]. *Zenodo*. <https://doi.org/10.5281/ZENODO.6221505>

Kennedy, M. L. (2019). What do artificial intelligence (AI) and ethics of AI mean in the context of research libraries? *Research Library Issues*, 299, 3–13. <https://doi.org/10.29242/rli.299.1>

Khalil, H., Ameen, D., & Zarnegar, A. (2022). Tools to support the automation of systematic reviews: a scoping review. *Journal of Clinical Epidemiology*, 144, 22–42. <https://doi.org/10.1016/j.jclinepi.2021.12.005>

Kjær, L., Tang, H., & Richter, N. H. (2020). Finnerne satser stort på kunstig intelligens. *REVVY*, 43(1), 18–21. <https://doi.org/10.22439/revvy.v43i1.5939>

Kricka, L. J., Polevikov, S., Park, J. Y., Fortina, P., Bernardini, S., Satchkov, D., Kolesov, V., & Grishkov, M. (2020). Artificial intelligence-powered search tools and resources in the fight against COVID-19. *EJIFCC*, 31(2), 106–116.

Lefebvre, C., Glanville, J., Briscoe, S., Featherstone, R., Littlewood, A., Marshall, C., Metzendorf, M.-I., Noel-Storr, A., Paynter, R., Rader, T., Thomas, J., & Wieland, L.



(2022). Technical Supplement to Chapter 4: Searching for and selecting studies. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions Version 6.3 (updated February 2022)* (2nd ed.). The Cochrane Collaboration. <https://training.cochrane.org/handbook/current/chapter-04>

Lyngsfeldt, J. K., Wildgaard, L., Møller, A. V., & Johnsen, S. S. (2022). Artificial Intelligence og litteratursøgning i biblioteksregi. *REVY*, 45(2), 7–9. <https://doi.org/10.22439/revy.v45i2.6629>

Marshall, I. J., & Wallace, B. C. (2019). Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Systematic Reviews*, 8(1), 163. <https://doi.org/10.1186/s13643-019-1074-9>

NHS Centre for Reviews and Dissemination (2009). *Systematic reviews: CRD's guidance for undertaking reviews in health care*. Centre for Reviews and Dissemination. [https://www.york.ac.uk/media/crd/Systematic\\_Reviews.pdf](https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf)

Nielsen, J. (1993). *Usability engineering*. Academic Press.

Nolin, J. M. (2013). The special librarian and personalized meta-services: Strategies for reconnecting librarians and researchers. *Library Review (Glasgow)*, 62(8–9), 508–524. <https://doi.org/10.1108/LR-02-2013-0015>

Orgeolet, L., Foulquier, N., Misery, L., Redou, P., Pers, J.-O., Devauchelle-Pensec, V., & Saraux, A. (2020). Can artificial intelligence replace manual search for systematic literature? Review on cutaneous manifestations in primary Sjögren's syndrome. *Rheumatology*, 59(4), 811–819. <https://doi.org/10.1093/rheumatology/kez370>

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), Article 210. <https://doi.org/10.1186/s13643-016-0384-4>

Polonioli, A. (2020). In search of better science: on the epistemic costs of systematic reviews and the need for a pluralistic stance to literature search. *Scientometrics*, 122(2), 1267–1274. <https://doi.org/10.1007/s11192-019-03333-3>

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews*, 10(1), 39. <https://doi.org/10.1186/s13643-020-01542-z>

Scientific Hackathon (2023, 15 February 2023). In *Wikiversity*. [https://en.wikiversity.org/wiki/Scientific\\_Hackathon](https://en.wikiversity.org/wiki/Scientific_Hackathon)

Schoeb, D., Suarez-Ibarrola, R., Hein, S., Dressler, F. F., Adams, F., Schlager, D., & Miernik, A. (2020). Use of artificial intelligence for medical literature search: Randomized controlled trial using the hackathon format. *Interactive Journal of Medical Research*, 9(1), Article e16606. <https://doi.org/10.2196/16606>

Thomas, J., Brunton, J., & Graziosi, S. (2010). *EPPI-Reviewer 4.0: Software for research synthesis*. EPPI-Centre Software [Computer software]. EPPI-Centre.

van den Haak, M., De Jong, M., & Schellens, P. J. (2003). Retrospective vs. concurrent Think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & Information Technology*, 22(5), 339–351. <https://doi.org/10.1080/0044929031000>

Wallace, B. C., Small, K., Brodley, C. E., Lau, J., & Trikalinos, T. A. (2012). Deploying an interactive machine learning system in an evidence-based practice center: Abstract. *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, USA*, 819–824. <https://doi.org/10.1145/2110363.2110464>

Wildgaard, L., Johnsen, S. S., & Kiersgaard, J. (2020). Delivery 1: Selection of AI software [Project deliverable]. *Zenodo*. <https://doi.org/10.5281/ZENODO.4279009>

Wildgaard, L., Møller, A. V., Kiersgaard, J., & Johnsen, S. S. (2021). Delivery 2: Exploring Iris.ai and Yewno with Think-Aloud tests – a mid-term perspective [Project deliverable]. *Zenodo*. <https://doi.org/10.5281/ZENODO.5350927>

Wu, R., Stauber, V., Botev, V., Elosua, J., Brede, A., Ritola, M., & Marinov, K. (2018). Scithon™ – An evaluation framework for assessing research productivity tools. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, T. Tokunaga (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. [http://lrec-conf.org/workshops/lrec2018/W24/pdf/7\\_W24.pdf](http://lrec-conf.org/workshops/lrec2018/W24/pdf/7_W24.pdf)

Zhang, Y., Liang, S., Feng, Y., Wang, Q., Sun, F., Chen, S., Yang, Y., He, X., Zhu, H., & Pan, H. (2022). Automation of literature screening using machine learning in medical evidence synthesis: a diagnostic test accuracy systematic review protocol. *Systematic Reviews*, 11, Article 11. <https://doi.org/10.1186/s13643-021-01881-5>

## Notes

---

<sup>1</sup> [https://zenodo.org/communities/ai\\_search\\_royaldanishlibrary/?page=1&size=20](https://zenodo.org/communities/ai_search_royaldanishlibrary/?page=1&size=20).

<sup>2</sup> The pilot tests and validation exercise are available in our Zenodo library: [https://zenodo.org/communities/ai\\_search\\_royaldanishlibrary/search?page=1&size=20](https://zenodo.org/communities/ai_search_royaldanishlibrary/search?page=1&size=20)].

<sup>3</sup> The Danish Research indicator is currently being phased out. Please read the description of the indicator at the time of our tests: <https://journals.lub.lu.se/sciecominfo/article/view/4757/4318>.