

How Data Scientists Review the Scholarly Literature

Sheshera Mysore
smysore@cs.umass.edu
University of Massachusetts, Amherst
USA

Mahmood Jasim
mjasim@cs.umass.edu
University of Massachusetts, Amherst
USA

Haoru Song
hsong@umass.edu
University of Massachusetts, Amherst
USA

Sarah Akbar
sakbar@umass.edu
University of Massachusetts, Amherst
USA

Andre Kenneth Chase Randall
andrekeneth@umass.edu
University of Massachusetts, Amherst
USA

Narges Mahyar
nmahyar@cs.umass.edu
University of Massachusetts, Amherst
USA

ABSTRACT

Keeping up with the research literature plays an important role in the workflow of scientists – allowing them to understand a field, formulate the problems they focus on, and develop the solutions that they contribute, which in turn shape the nature of the discipline. In this paper, we examine the literature review practices of data scientists. Data science represents a field seeing an exponential rise in papers, and increasingly drawing on and being applied in numerous diverse disciplines. Recent efforts have seen the development of several tools intended to help data scientists cope with a deluge of research and coordinated efforts to develop AI tools intended to uncover the research frontier. Despite these trends indicative of the information overload faced by data scientists, no prior work has examined the specific practices and challenges faced by these scientists in an interdisciplinary field with evolving scholarly norms. In this paper, we close this gap through a set of semi-structured interviews and think-aloud protocols of industry and academic data scientists ($N = 20$). Our results while corroborating other knowledge workers’ practices uncover several novel findings: individuals (1) are challenged in seeking and sensemaking of papers beyond their disciplinary bubbles, (2) struggle to understand papers in the face of missing details and mathematical content, (3) grapple with the deluge by leveraging the knowledge context in code, blogs, and talks, and (4) lean on their peers online and in-person. Furthermore, we outline future directions likely to help data scientists cope with the burgeoning research literature.

CCS CONCEPTS

• **Information systems** → **Users and interactive retrieval; Digital libraries and archives; Collaborative and social computing systems and tools**; • **Human-centered computing** → **HCI design and evaluation methods**.

ACM Reference Format:

Sheshera Mysore, Mahmood Jasim, Haoru Song, Sarah Akbar, Andre Kenneth Chase Randall, and Narges Mahyar. 2023. How Data Scientists Review

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).
CHIIR '23, March 19–23, 2023, Austin, TX, USA

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0035-4/23/03...\$15.00
<https://doi.org/10.1145/3576840.3578309>

the Scholarly Literature. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '23)*, March 19–23, 2023, Austin, TX, USA. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3576840.3578309>

1 INTRODUCTION

Literature reviews play an important role in the workflow of various scientists - with researchers spending upwards of 11 hours a week reading the research literature [91]. They seek this literature with a variety of intents ranging from exploring the frontier of problems and solutions, keeping up their understanding, to quick look-up of facts, etc [46, 94]. A challenge in this space is presented by an increasing volume of papers with studies estimating that the past few decades have seen a doubling of published research every 9 years [71, 129]. This deluge of information, as Chu and Evans [23] notes, presents a situation in which scientific progress as a whole sees a slowdown with individual scientists and peer-reviewers struggling to recognize and understand novel ideas. In this work, we examine the literature review practices of a currently emerging group of knowledge workers: data scientists.

Mirroring broader trends in the sciences, recent years have seen a surge in data science publications with a doubling every 2 years [11, 14, 66], the development of several tools intended specifically to aid data scientists¹, and coordinated efforts to develop AI tools to identify the data science research frontier [66]. This hints at the information overload faced by this group. Further, data science as a discipline has had a significant impact in business [103], government [98], and science [13], among others. Despite these trends, to the best of our knowledge, no prior work has examined the practices or challenges of data scientists as they review the research literature. Our goal is to address this gap.

Here, we view *data scientists* as individuals engaged in data work, applied engineering, and research work and with training in computer science and statistics as well as disciplines of specific application domains such as economics and biology [24]. We view *literature reviews* as a knowledge-building process involving the gathering of research literature, development of relationships between gathered data, and the emergence of synthesized information – spanning the tasks of *information seeking*, *sensemaking*, and *composition* [122, 140]. However, in this work, we omit examination of *composition* given its narrower focus on publication and the difficulty of clearly demarcating these stages [126, Sec 3].

Specifically, we examine the practices and challenges in the iterative process of information seeking and sensemaking, where

¹<https://search.zeta-alpha.com/>, <https://papers.labml.ai/papers>, <https://alphasignal.ai/>

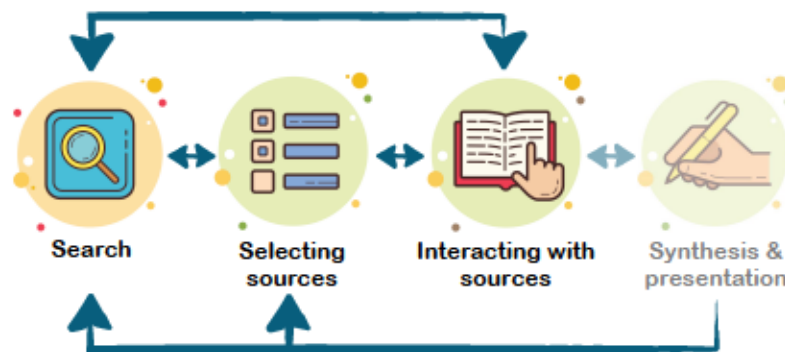


Figure 1: In this work, we examine the literature review practices of data scientists. We examine practices and challenges spanning information seeking and sensemaking: search and discovery, selection of literature, skimming and reading, and formation of a synthesized understanding of a collection of papers. §4 presents the results of our thematic analysis anchored to these stages. However, we omit the study of longer-term synthesis and composition.

research literature is searched for, selected, read, and then synthesized into a mental model to achieve task-specific goals [110, 133]. In studying both information seeking and sensemaking as closely related activities, we draw on a line of work in *search as learning* that views search as a learning process closely tied with sensemaking, and advocates support for users in both gathering information and use of the found results [69, 78, 117, 126]. Our study differs from the body of prior work which has examined practices of information seeking and sensemaking in *isolation* – focusing on search behaviors [7], use of academic social networks [4], use of information management systems [92], and note taking to capture ideas [52]. In cases when information seeking and sensemaking have been examined in conjunction it is in the context of specific literature-review systems [80, 122] or through the lens of abstract cognitive processes [6, 125] with limited focus on application grounded examination of literature review by expert knowledge workers.

Furthermore, in examining the literature review practices of data scientists our work also extends a recent body of work shedding light on the work practices of data scientists ranging from: teaching practices [68], code and data workflows [79, 88, 120], to documentation practices [130], and others. Our special focus on examining the practices of data scientists is also necessitated by prior work noting differences in literature review practices of scientists in different disciplines due to differences in research methods, norms of communication, and disciplinary values [4, 12, 17, 49].

To conduct this study we recruited 20 data scientists from industry and academic institutions and employed two different methods of gathering data for analysis: semi-structured interviews and think-aloud observations of open-ended search tasks conducted by participants. Our analysis used multiple rounds of coding and thematic analysis. In part, our study corroborates past findings, but also unearths several new and under-studied findings across information seeking and sensemaking: 1) Participants struggled in seeking literature, skimming, and establishing the credibility of literature outside their domains of expertise. 2) They struggled to understand papers in the face of missing details and mathematical content. 3) They faced an onslaught of seemingly similar papers and sought to understand them through their differences. 4) In coping

with these problems they leveraged their peers and a range of on-line resources forming the *knowledge context* of papers: discussion forums and social media, code, talks, and blogs. Given our findings, we highlight under-explored areas providing a rich canvas for future work and meaningful solutions for data scientists reviewing the literature.

2 RELATED WORK

Here we overview four relevant lines of work. The first line of work overviews *information seeking* systems and practices for various knowledge workers. Alongside this, we overview the *sensemaking* systems and practices for the scholarly literature. In examining the practices of knowledge workers we ground our findings of data scientists’ practices in prior work. Similarly, our examination of existing systems combined with the challenges of data scientists (§4) allows us to discuss meaningful future work (§5). Following this work, we outline work on *search as learning* which has examined the use of search systems in the context of learning tasks – much like literature reviews. Complementary to work on information seeking and sensemaking, this line of work also contributes important findings on the effects of tasks such as search, reading, and note-taking on learning-oriented goals. While these three lines of work have been explored in somewhat disjoint communities they may be seen as closely related, with each information seeking, sensemaking, and learning following one other [78]. Finally, we overview the recent work examining *data scientists* practices – our work most directly extends this literature.

2.1 Info-seeking & Sensemaking - Practices

2.1.1 Information seeking. A large body of work has studied the information-seeking practices of a range of different knowledge workers, ranging from medical researchers conducting systematic literature reviews [63], their use of search interfaces [76], design students’ use of search for ideation [95], social science and education researchers practices [51], astrophysicists practices [112], and engineering professors perceptions of institutional library services [29]. Niu et al. [91] and Alhoori et al. [4] present extensive reviews

of early work in this area. Given our focus on data scientists, we review studies that have examined information-seeking practices of populations in related disciplines [4, 7, 53, 81, 118].

Athukorala et al. [7] examine computer scientists for their goals in conducting literature searches and their tools. Their primary findings indicate that respondents used search most frequently to stay up to date on topics and find the exploration of unknown areas most challenging. They find that the most common form of navigating papers involves following citations to and from a known seed paper, often obtained via an initial keyword search. Alhoori et al. [4] examine STEM researchers' information-seeking behavior in conjunction with social media use and reference management tools. Hoeber et al. [46] examine the differences in practices of STEM researchers across levels of seniority. Finally, recent work of Soufan et al. [118] presents a large-scale survey of STEM researchers intended to reveal the alignment between theoretical models of exploratory search and actual practices followed by researchers. A few studies have also examined behaviors grounded in specific systems: Ishita et al. [53] investigate which sections of a paper returned in search results researchers examined, while McCay-Peet et al. [81] examine the difference between social scientists and computer scientists in using control features of digital libraries.

2.1.2 Sensemaking. While a sizable body of work has examined information-seeking behaviors, a smaller body of work has examined practices in sensemaking tasks such as managing gathered literature, reading, and note-taking. Nosheen et al. [92] examines the challenges of personal information management in engineering researchers, finding fragmentation of data across different systems and difficulty determining the future value of information to be a challenge. Similarly, Inie et al. [52] explore researchers' tool use patterns in managing ideas, finding interoperability between tools to be important. Finally, Morabito and Chan [85] studies researchers' use of tools to support the capture of context and metadata for documents to facilitate the effective reuse of found information.

In contrast with this body of work which has examined information-seeking and sensemaking tasks in isolation, our work examines the two in conjunction through a series of interviews and think-aloud observations. This combined examination more faithfully examines the iterative processes of information seeking and sensemaking [110, 133], especially in the early stages of sensemaking. This allows our work to shed light on the practices and challenges at the intersection of these activities informing the development of systems for both activities.

2.2 Info-seeking & Sensemaking - Systems

2.2.1 Information seeking. Prior work building information-seeking systems for the scientific literature has explored a range of strategies: discovery through seed papers and citation chaining [21, 40, 74, 101], through metadata such as authors or keyword [27, 28, 60, 102], and systems exploring more specialized querying methods [19, 22, 30, 61, 106].

AgroScholar [74] and PaperQuest [101] facilitate visualization and interaction with the citation network starting from a seed set of papers - allowing incremental addition of papers to a core set of papers based on which others may be recommended. Work in Apolo [21] and PaperPoles [40], further supports user interactions such as

grouping of papers and specification of preferences over keywords to influence the set of recommended papers. While this line of work has focused on citation-based discovery others leverage paper metadata more heavily. PivotPaths [28] presents a visualization system to allow researchers to explore paper collections through author and keyword metadata with a special emphasis on exploring the relationships between metadata elements and facilitating serendipitous discovery. Similarly, Portenoy et al. [102] and Kang et al. [60] enable the discovery of relevant authors and papers from disciplines different than that of the seed author or paper. Finally, a range of work has explored specialized querying methods going beyond papers or their metadata, more heavily leveraging the paper content. Work of Chan et al. [19] and Kang et al. [61] allow searching for papers based on the “challenges”, “solutions”, or “results” they tackle and work in DataHunter [31] allows search of datasets based on research problem queries.

2.2.2 Sensemaking. We organize prior systems for sensemaking into aids for a *collection* of documents [45, 80, 113], and systems to facilitate reading [32, 42, 131, 139] and note-taking [38, 58, 121, 142] of *individual* documents.

Since sensemaking often consists of iterative stages of information seeking and construction of mental models of gathered data [100], systems for sensemaking from collections of documents also contain components to aid information seeking - allowing users to examine and construct relationships while also allowing discovery of additional work [21, 28]. Other work has sought to help users determine salient time-aligned trends in the scientific literature. Shahaf et al. [113] construct “metro maps” from collections of scientific papers to depict the evolution of multiple lines of related work, similarly Heimerl et al. [45] leverage a streamgraph metaphor to allow users to examine patterns in citations, topics, and research communities over time. Finally, PaperForager [80] targets a different problem - minimizing the context switches in document filtering, skimming, and reading. It represents paper collections entirely as persistently visible document images.

In augmenting reading experiences prior work has sought to help readers relate documents to other concepts and papers or have sought to enrich the reading experience of individual documents. Zhang et al. [139] leverage bubble-tree maps to help visualize and explore hierarchies of concepts in a document. Similarly, Wang et al. [131] helps users better understand the related work sections of individual papers. On the other hand, reading aids for individual papers have examined aids for skimming through highlights over salient text [32], while others have explored augmentations for reading equations and terms defined in a paper [42]. Besides reading, work has also explored various forms of note-taking support to allow the synthesis and reorganization of gathered data. Work in Passages [38] and Threddy [58] help users to generate and organize text snippets and notes from reading papers that persist across applications and documents. While Passages emphasize persistent metadata and cross-application movement to retain the provenance of information, Threddy emphasizes the discovery of additional papers based on note collections. Finally, a line of work has also explored integrated systems for a literature review, spanning search and discovery, organization, synthesis, and composition [122, 140]. Compared to work developing systems that study practices in the

context of a specific system we conduct a more exploratory study examining scientists in their natural workflows aiming to contribute broader directions for future systems.

2.3 Search as Learning

The body of work on search as learning has studied several different aspects of how learning occurs in the context of the search process and provides several findings relevant to information-seeking and sensemaking. A bulk of the focus has been directed at studying learning outcomes while varying factors such as 1) user characteristics 2) search system characteristics, and 3) search behaviors [126]. Here, learning has most generally been considered the formation of mental models of knowledge, their retention over time, and their application [78]. This work also provides a perspective on search that goes beyond the finding and gathering of results. Instead, viewing search as a series of tasks in a constructive process - necessitating support for search as well as the use of search results [69, 78, 117, 126]. It is worth noting, however, that the bulk of work in SAL has focused on classroom students or drawn inferences with controlled experiments with crowd workers rather than knowledge workers such as data scientists.

2.3.1 User characteristics. Work on user characteristics has often examined users' domain knowledge. Willoughby et al. [134] examine the outcome of domain knowledge alongside the use of search engines in an essay writing task. They find searching to only benefit essays where participants had greater domain knowledge, with participants noting domain knowledge to be beneficial for searching. Roy et al. [108] examine learning with vocabulary assessments in the course of a search session as participants explore a topic, finding participants with greater prior topic knowledge to gain knowledge toward the end of their search session, while those without gaining more at the start of the session. While work examining user characteristics has often focused on domain knowledge Vakkari et al. [128] also find domain knowledge to influence searching only for users with knowledge of the search system.

2.3.2 System characteristics. In examining system characteristics, prior work has explored search features [18, 25, 104] as well as features relating to reading and note-taking [33, 109, 123]. Here, di Sciascio et al. [25] find a transparent and controllable search system to deliver better learning outcomes than the widely used PubMed, and Qiu et al. [104] find web-search interfaces to lead to better knowledge gain compared to conversational search interfaces though the latter lead to improved knowledge retention. In studying reading experiences, Freund et al. [33] find simpler text environments to lead to improved comprehension, Roy et al. [109] find highlights in reading to improve topic coverage in essay writing tasks and note taking to lead to the inclusion of more facts, and Syed et al. [123] finding automatically generated questions embedded in the text to improve learning outcomes.

2.3.3 Search behaviors. Work examining search behaviors has contributed several findings. Vakkari et al. [128] find psychology students to use more specific queries as their vocabulary improved though their search strategies employed remained consistent. Dosso et al. [26] find domain knowledge to not influence the number or length of queries across medicine and computer science students,

though they find medicine students to use more domain-specific vocabulary in queries. Vakkari and Huuskonen [127] find increased effort in examining documents to be associated with improved essays even in the face of lower precision search results, finding students to compensate for bad search results – a result also mirrored elsewhere [76, 116]. Work of Moraes et al. [86] finds search paired with instructor lectures to have improved learning outcomes than the lectures alone. Finally, Urgo and Arguello [125] abstract away from specific interactions and characterizes the “pathway” toward learning objectives in terms of learning-oriented sub-goals for tasks varying in cognitive complexity. This also ties to a line of work examining task complexity and learning outcomes.

In our work, rather than focusing on specific learning outcomes and implementing interventions to facilitate learning, we contribute an exploratory needs-finding study for highlighting the practices and challenges throughout the process of search and the use of search results by data scientists in a learning-oriented task. In this respect, we draw on SAL by viewing search as a constructive process necessitating a focus on both information-seeking and sensemaking rather than the gathering of search results alone [69, 126].

2.4 Practices of Data Scientists

Besides the lines of work covered above, a sizable body of recent work has examined the broader practices of data scientists – a currently emerging body of knowledge workers [24], we briefly discuss this work. Work of Crisan et al. [24] helps paint a picture of who data scientists are by conducting a review of prior work examining data scientists. A few papers have also examined the dataset-related information-seeking needs of data scientists. Kross and Guo [68] interview individuals engaged in training data scientists in industry and academia and identify that instructors find searching for pedagogically-relevant datasets to be one of their challenges. Koesten et al. [64] investigate data scientists' challenges in searching for structured data repositories. A range of work also examines data science workflows via codebooks [79, 120], their documentation practices [130], and how they arrive at analysis decisions [62]. Other recent work examines the challenges and needs of data scientists in developing fair machine learning systems [48], and their adoption of explainable machine learning systems [67].

While this body of work has examined aspects of how data science is conducted, how data scientists seek and make sense of expanding research literature remains unknown – as we have noted, the recent proliferation of tools aimed at data scientists and the expanding scholarly literature indicates that this is an important challenging activity. Therefore, in conducting our needs-finding exercise, our work serves to extend the understanding of the working practices of data scientists with a specific focus on information-seeking and sensemaking of the scholarly literature.

3 STUDY METHODS

3.1 Study Design

Our study primarily consisted of two components, a semi-structured interview followed by a think-aloud observation in the working environment of the participants' choice. While our semi-structured interviews probed participants' self-described practices and challenges, the think-aloud aimed to observe more tacit behaviors,

Table 1: Description of our participants. Research Areas were self-described in our semi-structured interview and our presentation respects participant requests for maintaining confidentiality about their research areas. Of 20 participants 13 were enrolled in Ph.D. programs in Universities across the USA and EU, and 7 participants held Data Scientist, ML Engineer, or Statistician designations at for-profit industry or non-profit organizations.

Participant	Organization	Designation	Research Areas
P1	Industry	Data Scientist	Conversational Information Retrieval
P2	Non-profit	Statistician	Statistics
P3	Industry	Data Scientist	Computer Vision
P4	Industry	Data Scientist	NLP for Legal Text
P7	Industry	Data Scientist	Speech Processing
P5	University	Ph.D. Student	ML for Public Health
P6	University	Ph.D. Student	Video Processing
P8	University	Ph.D. Student	Search as Learning
P9	University	Ph.D. Student	Robustness in NLP
P10	University	Ph.D. Student	Question Answering and Reasoning
P11	University	Ph.D. Student	Robustness in NLP and Hate Speech
P12	University	Ph.D. Student	NLP for Journalism
P13	University	Ph.D. Student	Variational Inference
P14	Industry	ML Engineer	NLP
P15	University	Ph.D. Student	NLP for Scientific Documents
P16	University	Ph.D. Student	Differential Privacy
P17	University	Ph.D. Student	CS Theory
P18	Non-profit	Data Scientist	Cryptography
P19	University	Ph.D. Student	Reinforcement Learning and Causal Inference
P20	University	Ph.D. Student	Data Management

workflows, and practices in a realistic task scenario. In our think-aloud participants conducted a literature review of their choice in response to one of three open-ended task prompts. The design and content of our study were developed in a pilot study with 15 participants run from February-May 2022. In our pilot study, we sought to probe information-seeking behaviors involved in literature review, however, in initial results we observed substantial sense-making activities closely tied to information-seeking. In the second iteration of our study with 20 new participants, run from July-August 2022, we updated our study to also probe broader sensemaking activities in literature reviews - our paper reports on this second iteration. Appendix A includes our questions and think-aloud prompts.

3.2 Study Participants

We invited participants for our study using social media posts on Twitter and LinkedIn, and email invitations sent to university mailing lists. In recruitment forms, a broad definition of "data scientist" was displayed and respondents were asked if they identified as data scientists. Further, participants were also asked to submit links/titles of 3 research papers they found useful or enjoyable to read in the past year. Of the respondents, 20 were selected as participants for our study (P1-P20) on a first-come-first-serve basis while ensuring that they identified as data scientists and had read papers in the past year. Participants were compensated for their time with a \$25 gift card and all study procedures and materials were approved by the university IRB. Appendix A includes our recruitment form.

Of 20 participants 11 noted gender pronouns he/him, 9 noted she/hers, and 2 noted they/them, with some noting multiple pronouns. 14 participants were enrolled in or had completed Ph.D. degrees and 6 had completed master's degrees. 13 participants worked in universities and 7 worked in non-profit or for-profit

industry labs. On average participants had published 4 research papers. Table 1 describes participants' research areas and the type of organization where they conducted work.

3.3 Study Procedure

Each study session was conducted by the authors and lasted about 1 hour with an equal split between the semi-structured interview and the think-aloud. The whole session was conducted over Zoom. We began each session by explaining the study procedure, obtaining participants' consent for participation, and having them fill out a short demographic survey. This was followed by a semi-structured interview. Here participants were asked about their research focus and their goals, practices, and challenges for conducting literature reviews. Participants were encouraged to think of literature reviews as broader than a single directed search and discuss all interactions with the scientific literature. This was followed by a think-aloud where participants conducted a literature review over Zoom screen-share. Participants were shown 3 broad task scenarios and selected one as a prompt for their think-aloud. One prompt asked them to recall and demonstrate a previous literature review, another asked them to enhance their knowledge on a known topic, and the final asked them to research the literature for a future project of their interest, an area they had lesser experience in. The latter two scenarios were drawn from the work of Hoeber et al. [46]. Our analysis revealed that participants distributed uniformly between the 3 scenarios. Note that simple look-up searches were not used as task scenarios since we sought to understand more complex exploratory searches. If any papers were relevant enough to require in-depth reading participants were invited to move on and continue their search or terminate the study session. Audio and video of the study were recorded for analysis.

3.4 Data Collection and Analysis

All our analyses used transcribed audio of the interview with the video examined for our think-aloud session. Automatic transcription of the audio was obtained from Zoom and was corrected substantially for errors. This was followed by 3 rounds of coding by 3 authors of this paper and a thematic analysis of the interview and think-aloud transcripts – this process was conducted from Aug–October 2022. In our first round of coding 2 authors independently examined 5 transcripts and generated a set of codes using open and axial coding, then, these codes were consolidated into a unified set of codes. In consolidation, both authors examined each other's codes independently, then met to discuss any differences and consolidated them into a unified set of codes. This was followed by application of the unified codes to the 5 transcripts by both coders. We observed an agreement of 0.92 in terms of nominal Krippendorff's alpha. This unified set contained 167 codes of which 51 noted names of tools used by participants or logistic aspects of the study. This was followed by discussion and resolution of disagreements in codes and application of the resultant codes to the remaining 15 transcripts while also adding any new codes to our initial codebook. This process added 59 new codes of which 32 noted names of tools or logistic aspects. The coded transcripts were then used for thematic analysis – which resulted in themes corresponding in part with the stages of the Information Search Process [126, Sec 3], we present our themes next. Our codes and extended quotes per theme may be examined online.²

4 RESULTS

In presenting the results of our thematic analysis we broadly organize our themes by the Information Search Process as consisting of the formulation of an information need (§4.1), query formulation and search (§4.2), assessment of search results through skimming and reading (§4.4, §4.3), and synthesis of a collection of documents (§4.4, §4.3). Besides these, our final theme elaborates on how participants relied on social ties at various stages of the ISP (§4.5). Importantly, while these stages provide an organizing model to aid understanding, they are iterative processes with poorly defined boundaries. In presenting our results we present each theme followed by specific relevant prior work where appropriate.

4.1 Why do data scientists access the scientific literature?

4.1.1 Actively understanding disciplinary norms. Participants most sought the literature actively as a means to understand disciplinary norms where they lacked this familiarity (18/20). Here, they sought to understand disciplinary vocabulary, form a mental model of the discipline, understand the contexts and community preferences of problems, solutions, and evaluation practices:

“The first question that comes to mind for any researcher is what has been already done, are there similar problems which have already been tackled or related problems whose corresponding methods might be used in my problem.” - P16

“When I'm starting work in a problem ... I'm not sufficiently familiar with to work to know what the typical approaches are, how is this evaluated, what kinds of approaches are falling out of favor versus becoming more accepted by the community.” - P15

Participants focus on understanding problems and solutions reflect prior understanding across scientific and creative problem-solving disciplines with individuals often thinking in terms of “problem” and “solution” [44, 95]. A focus on evaluations and metrics also matches understanding of a large focus on quantitative performance in data science [12].

4.1.2 Passively following a discipline. Besides actively seeking to understand a discipline, participants also noted accessing the literature more passively (10/20) – keeping up with trends in the community or keeping updated on the work of peers: “Where the community is going, or what people that I have previously followed the works of are up to right now” - P10, and remaining on the lookout for future projects: “Relevant reading for my own research what I had in mind ... at some point, I thought I would do future research in. [But] now I think that that's been on the back burner” - P17. This motivation to keep up is also mirrored in related work [7].

4.1.3 Brainstorming solutions. A majority of participants also leveraged existing scientific literature as a resource to aid brainstorming (18/20). Here participants leveraged the literature to find open problems (12/20): “Making sure I'm not redoing what has been done before, that's one major aspect – figuring out like open questions and existing work that I could work on” - P10. Others examined the space of solutions in prior work (13/20): “I am looking for existing methods which do what I am doing. What are other people doing to solve this? Is it even solved? Are there methods that already solve this problem?” - P16. While others established the novelty of their own ideas by seeking similar solutions (10/20): “After you figure out [the problem], it's like I have an idea for what you could do better, and then it's seeing if others have done something similar before” - P5.

This process of seeking problems, developing solutions, and establishing the novelty of their ideas formed an iterative brainstorming loop. The use of search for creative brainstorming has been noted elsewhere [95], and the desire for finding problems and solutions in the scientific literature has seen the development methods and resources intended to facilitate this [19, 61, 89] - our findings further support this effort. Focus on novelty and building on prior work also matches understanding their importance in data science research [12] and creative disciplines more broadly [52].

4.1.4 Seeking solutions for application. Besides seeking the literature for developing solutions, several participants also sought solutions to problems (17/20). Here, participants often sought solutions for direct application: “The other thing I will look this how other people have applied [this method], to see if other people have applied it to similar data or use cases to mine” - P2, or as baseline systems to aid research: “What are the general approaches that people have taken to solve a given task... That helps when we are trying to publish a paper... So this forms what are the existing baseline methods to compare against for quantitative research” - P6.

Seeking solutions as baselines for comparison or as solutions for application matches findings from citation analysis indicating that

²Codes and extended quotes: <https://github.com/MSheshera/dslitreview-study>

solutions either assume the role of vetted methods which are widely adopted or ones on the frontier of research knowledge and seeing current development, necessitating comparisons [57, Sec 8]. Further, the acts of “creating” and “applying” also correspond to cognitive processes often used to precisely define learning objectives [125].

4.2 How do data scientists access the scientific literature?

4.2.1 Data scientists seeking the literature. Expectedly, most participants sought the literature through keyword-based web searches (17/20), following citations to and from source articles (15/20) obtained via search or recommendations, or by examining papers of authors of relevant publications (13/20). This largely matches prior understanding [7] and has seen the development of several tools intended to aid citation chaining behaviors [21, 101, 102]. Despite these methods being well established, participants often lamented the challenge of coming up with appropriate keywords to describe their information needs (13/20), sometimes requiring weeks to stumble into the right keywords. This was further exacerbated in unfamiliar disciplines:

“I had an idea in my head, but I was not sure how to map this into normative terms used by communities or even necessarily what communities are interested in this ... Ultimately it just took trial and error like finding some papers and coming back to it over several weeks, and eventually I kind of started to find things that actually matched.” - P15

To tackle this challenge in complex searches, prior work has explored query recommendation strategies [22, 83, 96], with large-scale systems also implementing them [115]. More generally, challenges in formulating queries in corroborated in prior work noting literature searches to be exploratory with evolving, ill-defined, and open-ended searches – where users often learn terms in the course of search [118]. While many systems have been developed for exploratory search (see §2.2), large-scale systems for literature search lack support for it [90].

Finally, in the face of not knowing where to begin a search, participants also noted soliciting recommendations from expert peers (4/20): *“If you don’t know where to start and you just reach out to someone, especially in a company like [redacted] it’s a lot easier, there are a lot of subject matter experts and they help us.”* - P7. As we note next support for leveraging social ties remains under-explored.

4.2.2 The literature finding data scientists. Besides seeking the literature through search, participants also discovered scientific literature more passively (20/20). Their methods for obtaining automated recommendations primarily relied on following individuals on social media (13/20), subscribing to email alerts from arXiv or Google Scholar (7/20), following the work of specific authors (13/20), or newsletters intended to curate the literature (4/20). Here, however, participants noted being overwhelmed with the alerts which were supposed to inform them (5/20): *“It would be nice to keep up with people’s work ... Maybe at least you open up 10% of this stuff instead of being like this is just junk mail.”* - P1. Further, participants also noted tools for discovery often trapped them in a disciplinary bubble (4/20): *“I’m probably heavily in my own bubble of papers ... if*

I’m working on hate speech, most of my recommendations will be very computer science based but maybe there’s relevant stuff in social science that I’m probably never going to come across.” - P11.

Besides automated methods for discovering literature, a majority of participants noted receiving recommendations from their peers (14/20) – albeit less frequently than automated methods. This had several advantages – peers had a close understanding of participant interests, had vetted the paper, and offered deeper engagement:

“There are some of my peers from both industry and academia, who if they come across something interesting and they know [name] is working in these problems or she finds them interesting they just send it over. Sometimes it’s, not even a paper it’s just a blog post, which has links to papers.” - P9

“I guess the benefit [of recommendations from peers is] its gone past one set of eyes. So there’s some added incentive to read it, somebody said it was interesting and that the claims make sense.” - P10

Recent work of Kang et al. [59] finds even the addition of automatic social network-based explanations in scholarly email alerts to improve engagement and retention while allowing scientists to see themselves as part of the community.

4.3 How do data scientists select papers?

4.3.1 An onslaught of papers. Having obtained literature through search and discovery participants are now faced with filtering and organizing literature. Here they often reported being faced with a large volume of similar papers (16/20) and credited it to heavily crowded disciplines of data science. This made it hard to distinguish between many similar or incremental steps and seminal papers:

“If the field is very crowded - sometimes I find RL, and the problems I am focusing on to be crowded, then it becomes frustrating and you’re always finding papers [that do the same thing].” - P19

“There are a lot of papers that are incremental updates so to find that one seminal paper that actually started it all is going to be painful and unless someone helps you out it’s actually very hard.” - P7

To cope with this deluge participants turned to surveys or good reviews of the literature to find salient papers (7/20). Some choose to focus on a handful of papers, minimizing the overlap between papers they examined, and pairing the examination of a few papers with citation chaining to balance the breadth and depth of exploring results (9/20). Others also leveraged repeated references to specific concepts or papers as a sign of having found the papers worthy of examination (6/20): *“Usually there will be a collection of papers cited in all of [the papers] and for me that is a good proxy of here is the core papers I should be looking at.”* - P16. In the presence of many similar variants, work in information foraging notes that users understand variants in terms of their differences or as time-aligned stories [119]. While this is explored in the context of seeking code, we note a similar desire in the more challenging case of text documents, further elaborated on in §4.4.3.

4.3.2 Establishing the credibility of papers. An important challenge of coping with this deluge was establishing the credibility of papers

(13/20). Here participants relied upon expected indicators: relying on known authors, affiliations, publication venues, and citation counts as markers of credibility. As we noted in §4.2.2 some relied on the vetting provided in peer recommendations, others sought forum discussions: *“One thing is that its hard to figure the credibility of a paper, so it’s sort of trying to figure out the credibility based on discussions by online forums like Twitter, Reddit or Openreview. Even if this is highly reviewed what do other people who have worked in similar domains think about it”* - P14.

Challenges with credibility however did not end with the selection of papers, participants also noted the content of papers sometimes betray their information scent (12/20). While this could in part be attributed to disciplinary writing norms [50], participants also noted the challenge of papers with unclear or exaggerated contributions, re-branding of ideas, and papers not structured with information needs of a reader:

“People do a lot of re-branding, sometimes a lot of ideas are not very new but the motivation section is like poetry and when you read the details you feel [its] not what they are claiming they do. ... [or] exaggerating their contribution and not meeting the expectation in their experiments. So identifying those trends from papers is very important.” - P19

The use of publication metadata is well implemented in popular search engines, and its use for determining credibility has seen critical examination [10, 36]. However, methods and systems to facilitate the use of social discourse around papers to augment skimming and filtering have been under-explored. Further, while challenges like exaggerated contributions have been examined in science communication [132] it remains understudied in literature searches by experts - we see evidence for this phenomenon here. These challenges arising from exaggerated/re-branded claims and needing to sift through many similar papers (§4.3.1) may be a result of the crowded and highly incentivized disciplines of data science. While we present, to the best of our knowledge, the first report of these challenges they warrant deeper examination in data science and other rapidly expanding disciplines.

4.3.3 Everyone skims papers. Finally, in the selection of results all participants skimmed individual papers (20/20) often to make quick decisions of correctness or glean contributions of a paper. They often relied on knowing the discipline to know where to look for specific information and being challenged otherwise:

“I try to jump to wherever they say “in this paper” because I’m fairly familiar with the space I don’t need to actually look at the introduction.” - P7

“One challenge was that [this discipline] was very active in the 90s. ... They were published in different venues, and the approaches that they used were not familiar to me. That made it much harder to skim a paper and get the essence of its contribution.” - P15

Skimming is often interspersed with information seeking with frequent context changes between the two, however, few prior lines of work have examined close integration of information seeking and skimming [80], or tools to aid skimming of papers [32].

4.4 What challenges do data scientists face in reading papers?

4.4.1 Understanding the hidden details. In reading and understanding papers, participants also noted the challenge born from missing details in a paper (13/20). Here they noted that papers were written to be accepted with little incentive for authors to include the details which made a solution effective: *“Often the things that are in the paper or the things that make a good story, but the things that actually matter in the paper, for example, like how you set the hyperparameters are missing.”* - P5. Some also noted a tension between including a lot of detail which would interfere with readers hoping to get a high-level idea of a paper: *And then of course there’s always a thing that I have the time to read a paper and you are getting into it and you face a lot of detail. Sometimes, thanks to the length and sometimes these certain details have to be omitted and, at times, those are the very details that would cause certain confusion to a reader.* - P9. In a bid to cope with missing details, participants noted the value of augmentations provided by code (7/20).

[I ask authors if] there is any publicly available code for what you’re doing. Because many of these papers look well on paper but then its unclear how to implement them. Or its unclear which specific hyperparameter choices they made. - P16

While platforms such as paperswithcode.com, pair papers and code, use of code to augment reading experiences for scientific papers presents a future venue for improvement. More broadly, the challenges arising from unclear documentation for models and data have led the data science community to pursue initiatives intended to document these artifacts [34, 84], sometimes incentivizing them at publication.³ However, even availability of code alongside papers remains at 25% as of Sep 2022⁴ with an understanding of incentives for documentation currently emerging [20].

4.4.2 Understanding the math on display. While participants noted the absence of details the flipside was the inclusion of extensive math which also presented challenges (10/20). Here participants noted disciplinary subcultures which often rewarded papers with extensive math, equating hardness to understand with quality:

In writing for niche audiences it requires having to show that [an idea] is important or useful and often that means that they will add equations or theorems [for an idea] that really are not as complicated ... if there’s a lot of math or if it’s hard to understand it must be impressive. - P5

In coping with this, participants noted the importance of alternative sources such as code, blogs, and talks, or the inclusion of examples in papers to aid understanding (9/20). Also noting that talks and blogs presented better incentives to ensure understanding in audiences: *Blogs help because sometimes the writing [in a paper] is not easy to understand compared to an informal way of writing [I: So maybe papers are more mathematical but blogs have the high-level ideas?] Exactly* - P6.

Aids for understanding math in reading papers have seen some precedent in recent work in the form of reading tools intended to

³<https://naacl2022-reproducibility-track.github.io/>

⁴<https://paperswithcode.com/trends>

better define mathematical symbols in papers [42, 43]. However, this presents an open problem, with no work having explored augmentations from code, blogs, or talks to aid math understanding. It is also worth noting that challenges arising from missing details in papers and of understanding math, and the coping strategies of leveraging code and the knowledge context surrounding papers represent challenges that are likely specific to interdisciplinary and computational disciplines such as data science – our work presents an initial report of this.

4.4.3 Struggling to understand the increments of progress. In reading papers, participants often noted the importance of understanding the specific “delta” that papers offered in comparison to other work that was influential or known papers (9/20). This was important to forming an understanding of the discipline and contextualizing their contributions:

“Once you’ve read 10, 20, 30 papers it becomes a lot clearer what the core idea is and what the delta from the core is. ... For things you’re not as well versed in, it’s often a nightmare when you don’t know the context in which the paper is being written. Understanding what exists in the literature and what doesn’t is hard, then what the contribution is and why the contribution matters is hard.” - P5

To understand the increments participants noted the importance of gaining familiarity with the norms of a sub-discipline and the challenge in its absence: *In grad school a professor synthesizes these things and says hey, this is the main theme of all these papers. When that information is there for you and you start reading the paper it tells you what to expect otherwise you’re spending a lot of time and don’t understand how different it is from previous papers*” - P7.

However, it was a challenge to establish if a paper was poorly written or if the participants were missing necessary context: *“The lack of clarity on their contribution - it might come from how they wrote that paper or it might come from my complete lack of understanding of what they might be doing”* - P19.

As we note in §4.3.1 in the presence of many variants of an item, prior work has noted users seeking to understand the differences between items [119] - we see this here. Methods and systems to explore free text differences in documents to aid filtering, skimming and sense-making of collections remain under-explored. However, recent work in NLP has begun to examine this problem - Luu et al. [77] explore methods for explaining relationships between papers, while Rajagopal et al. [107] and Kuznetsov et al. [70] have studied textual edits in other application contexts.

4.5 How do data scientists lean on social ties?

Besides leveraging social ties for the discovery of papers in §4.2.2, participants also leaned on their peers in other ways.

4.5.1 Collaboratively brainstorming and making sense of papers. All participants noted leveraging social ties in making sense of the literature (20/20). Here they noted the value of group discussions centered on papers to keep up with the literature, help brainstorm ideas, or spark new research directions (12/20):

“In independent research at the very least, I talk about my idea with someone else just to see if I’m in the

right direction. On the other hand, we have bi-weekly brain-storming sessions in which we discuss at least one paper that is very relevant to our work and so we get a lot of inputs ... this might be useful, this might be a limitation, it might not work.” - P7

Others noted the value of discussions with collaborators to understand the details of specific important papers (11/20): *“If I’m having one on one meetings then we dig into why people made certain decisions in their paper, and if we should be following the same”* - P10. Finally, a few participants (5/20) noted the value of sharing notes and literature with collaborators to establish the provenance of ideas: *I usually share these notes with collaborators to give them a sense of where I am getting this idea from or where this hypothesis is coming from* - P19 or correctness of information: *It also helps with collaborators double checking my writing.* - P6.

Besides close peers, participants also sought weaker social ties and online discussions (14/20). Here, they noted many of the same benefits that interaction with peers provided - seeking recommendations from experts on forums: *You will find a Quora or Yahoo answer which is dominated by exceptional mathematicians. Someone will have asked a similar question [to yours] and someone will have pointed them to some relevant work* - P17, establishing the credibility of papers as in §4.3.2, and engaged in discussion to understand the specifics of papers: *[The ML Collective] discord channel has a lot of volunteers and contributors if you can figure out someone interested in the same kind of work ... you can be talking about how the algorithm goes from this step to another or what is this variable* - P14.

A body of work has examined collaborative information seeking [87], with some examining social reading [97, 138], and sensemaking with collaborative annotations [114, 143]. However, examinations of collaborative reading and sensemaking in specific task contexts are understudied avenues that are likely to present specific challenges, as noted in cases of information re-use in software teams [75] or collaboratively verifying the credibility of claims [41].

4.5.2 Leveraging authors. While peers were useful in-person and online, participants also found value in interacting with authors (11/20). While some contacted them actively (5/20), others noted more passively seeking them in recorded talks and forums such as Twitter or Reddit (6/20). Direct communication ensured a fuller understanding of work, helped develop ties with other researchers, and sometimes turned into fruitful collaborations (2/20): *A little while ago a new paper got released that was very similar to my work I emailed the authors. This is something I do in this type of situation to create a conversation and also connect with researchers who are doing similar work to me. I had a multi-day email chain with them where we discussed things and I wanted to confirm the differences [of my method] with them, which was extremely helpful* - P18. Others also found these useful to broach under-discussed aspects: *I send them a message congratulating them about the work they’ve released and understand some of the secrets behind their work* - P4.

Here, prior work has explored recommending expert peers [102] and “people recommendation” more broadly [37]. However, important challenges remain in how incentive structures must be set up to facilitate these interactions e.g. Breiting et al. [15] note a trade-off in the discovery of ongoing research and keeping it confidential.

While an active engagement was important for a deep understanding, passively interactions with authors through talks or forum posts made work easier to understand than their papers. Participants noted the value of visual communication and an incentive for authors to communicate their idea for understanding: *Sometimes getting a very good intuition of what their motivation helps understanding. It might be not that clear in the paper but when they explain it to you in the video it's much more exciting. I think visual communication [is useful], their talk used a lot of good design and slide animations ... that cannot be communicated in a paper* - P19. However, some noted that talks were only recorded for “famous people”: *But [use of talks] isn't always applicable because it's only the more famous people's projects that get this much coverage. But I'm presuming I'm going into a brand new area and starting with the most famous people around* - P9.

Examination of authors' engagement through social media has only recently begun to be studied [35, 65] and incorporation of this into aids for sensemaking remains under-explored. The usefulness of author interactions beyond their papers also indicates the value of alternative publication formats [1, 47], once again, challenges remain in setting up incentives for publishers and authors [124].

5 DISCUSSION

In our Results, we examined the practices and challenges of data scientists reviewing the scientific literature, while anchoring our results along the formulation of an information need, query formulation and search, assessment of search results through skimming and reading, synthesis of a collection of documents, and leveraging social ties to accomplish a number of these tasks. Next, we note challenges that spanned across these themes and speculate future work likely to present solutions.

5.1 Support cross-disciplinary access

Data science is seeing exponential growth in the number of papers – a doubling every two years [66]. As fields develop several sub-disciplines emerge around specific problems and methods each with their own disciplinary vocabulary and norms [57]. Individual scientists are now tasked with conducting work in increasingly fragmented disciplines only some of which are familiar to them while knowing that significant innovation and progress is to be had from drawing across multiple disciplines [111]. The challenges stemming from fragmented knowledge are echoed in our findings at each examined stage of the information search process. Future work necessary to support information seeking and sensemaking of cross-disciplinary research can take several forms.

To tackle the challenge of search in unknown disciplines aids such as query recommendation have been explored (§4.2.1), however querying strategies intended to allow richer specification of user context remain under-explored. These may take the form of verbose queries [2, 89], conversational and interactive searches [5, 39] paired with mechanisms for cross-domain retrieval – Kang et al. [60] present a preliminary example. Further, the exploration of papers in unfamiliar disciplines is likely to benefit from the presentation of “explanations” to aid in understanding their credibility and relevance. While explanations have been extensively studied in recommender systems [141], examination of explanations for

cross-disciplinary exploration remains an open question. Aids may also be developed for skimming. Since skimming relies on knowing norms of writing in a discipline, these aids may take the form of adaptive document layouts [54] or automatically generated “FAQs” for a paper [8], personalized to the discipline of a reader. Similarly, reading aids may take the form of paraphrasing text for different disciplinary audiences or presenting “pre-requisite” concepts or papers necessary to understand a given paper [73] – thereby aiding readers to judge the quality of a paper independent from their own knowledge gaps. However, since scientists see better learning outcomes in tasks perceived as challenging the trade-offs involved in easing this process remains to be seen [76, 127].

5.2 Facilitate reliance on close peers

At several stages in reviewing the literature, participants relied on close peers – teammates, lab members, collaborators, mentees, and mentors. From peers, they received recommendations, engaged in brainstorming, established the credibility of papers, and understood the details of papers. The collaborative practices of data scientists have been examined in the context of code and data work [137] – we extend this to include the creation and understanding of ideas. Careful understanding and support for these practices are likely to be fruitful – especially with the move toward remote work [135].

The dominant line of work in information-seeking has taken an egocentric perspective and exploration of methods and systems to leverage communities has been examined largely in early work in collaborative search [87] and group recommendations [55]. This line of work may fruitfully be re-explored in our present circumstance – evidence for this is provided in recent work of Piao et al. [99] who find bringing “friends-into-the-loop” of recommender systems to result in more accurate and diverse recommendations. Re-incarnations of early work melding sensemaking, reading, and discovery in a collaborative feed-reader [3] also promise to leverage the trust scientists have in their peers for selection and consumption of research. A different line of work is suggested by Morris [87] where users preferred to interleave egocentric search with lightweight communication. Templates for this work are provided in aids to summarize group chats [136] or collaborative conversational agents [9] – as aids in brainstorming. Implementing these tools requires careful design, however – Brucks and Levav [16] note virtual communication through video conferences to curb creative ideation. Similarly, the work of Inie et al. [52] notes users' preferences for their own scholarly tool chains, making interoperability of new tools important for uptake.

5.3 Leverage the knowledge context of papers

While a close community of expert peers was important this was not always available to participants. As an expanding discipline with many new entrants, this is also likely in data science [56]. In this scenario, a variety of other resources were sought to augment papers – forum and social media discussions, recorded talks and videos, blogs, and code. These resources provided starting points for exploration, helped establish the credibility of papers, and aided in understanding missing details and math in papers. Traditionally this information (e.g. entity cards, videos) referred to as the *knowledge context* has seen use in web search engine result pages (SERP) and

aids users in making information literate decisions. Smith and Rieh [117] advocate for greater use of this knowledge context instead of their reduction and search engines “getting out of the way” of users. Here, we echo this push and observe its values for data scientists searching the literature. Furthermore, we also note its value in augmenting reading and skimming aids for found documents.

In the presentation of the knowledge context alongside academic search numerous questions remain. Prior work in the TREC Blog Track has examined blog retrieval with an emphasis on their subjective contents – similar efforts are necessary for the retrieval of knowledge contexts for academic publications [93]. The presentation of this information is also likely to require further investigation, e.g. Levi et al. [72] report that only some queries benefit from the presentation of a cluster of results from a Community Question Answering site. Further, questions remain about the fairness implications of augmenting academic search results with their knowledge context [82]. Besides use in information seeking, this knowledge context is likely to benefit skimming and reading aids. Recent work of Rachatasumrit et al. [105] presents an early example and places discussions from follow-on work into the margins of a paper. There is room however for more complex augmentations – This may take the form of using blogs, and videos for aiding math understanding, using code to infer missing details, or use of social media and blogs to aid skimming. This space remains under-explored with each type of information presenting challenges to retrieval, presentation, and the subsequent implications of use.

6 FUTURE WORK AND LIMITATIONS

We note three limitations to our study, each of which points toward future work. These stem from our study design, our choice of participants, and the nature of the challenges probed. First, given that our study spanned a single session we did not probe longer-running activities such as longer-term synthesis and composition, it is likely that examining this aspect is also likely to shed more light on the organizing and note-taking strategies of scientists. This will require longer-term observation and introspection from participants. Second, in the recruitment of participants we noted that all participants were based in the USA or EU with ample infrastructural access, it is likely that a more varied participant pool will result in different findings. Further larger-scale studies are also likely to be beneficial. Third, in several cases we noted challenges arising from incentive structures surrounding disseminating research, while reasonable in hindsight, these were under-examined in our study and warrant future work [20]. Finally, while unlikely to be a limitation, we also note that our study was conducted in the aftermath (February to October 2022) of the COVID-19 pandemic and its influence on work practices [135] – this likely influenced our study and results.

7 CONCLUSIONS

An exponential rise in the number of scholarly publications, its expanding influence, a large number of new entrants, and its likely impact on data science drove us to examine information seeking and sensemaking practices of data scientists reviewing the literature. In our examination, we ran an exploratory interview study with 20 data scientists recruited from both industry and academic

institutions. Here, we established their goals for accessing the literature, then we examined their practices and challenges in search and discovery, selection of search results, skimming and reading, and their reliance on peers in a number of these tasks. A number of our results corroborate those seen in prior work – here, our work offers a synthesis of disparate prior work. Next, our work also uncovers specific results stemming from our focus on data scientists and our joint examination of information-seeking and sensemaking. In our findings, we highlighted challenges arising from fragmented scientific disciplines influencing all stages of the information search process – we believe this represents a special challenge of data science given its inter-disciplinary nature. Besides, we highlighted the challenges of missing detail and mathematical content in reading papers – likely, a feature of computational disciplines such as data science. Our joint examination of information seeking and sensemaking revealed the nature of information sought at the intersection of these two activities: a desire to establish the credibility of papers owing to exaggerated/re-branded claims or unfamiliarity with the discipline, and a desire to understand the incremental differences between many seemingly similar papers – this likely represents a consequence of crowded and incentivized sub-disciplines of data-science. To cope with these challenges, we found our participants to leverage the knowledge context surrounding scientific papers in the form of code, blogs, talks, and forums and by leaning on their peers – these practices were also examined in our work. Finally, examination of our participants’ challenges and coping mechanisms lead us to carefully speculate on future work likely to present meaningful solutions to these challenges while also presenting meaningful scientific questions for future work in the IR, NLP, HCI, and CSCW communities.

ACKNOWLEDGMENTS

We thank our study participants for their valuable insights and time. We also thank Alyx Burns of the HCI-VIS Lab for extensive guidance in his role as Teaching Assistant for the Advanced HCI class in the formative stages of this project. We acknowledge partial funding from the National Science Foundation under Grant Number IIS-1922090, the Chan Zuckerberg Initiative under the project Scientific Knowledge Base Construction, the National GEM Consortium, and the Intel Scholars Program.

REFERENCES

- [1] ICLR 2022. 2022. *ICLR 2022 Blog Track*. <https://iclr-blog-track.github.io/> Accessed: 15 October 2022.
- [2] Jafar Afzali, Aleksander Mark Drzewiecki, and Krisztian Balog. 2021. POINTREC: A Test Collection for Narrative-Driven Point of Interest Recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (Virtual Event, Canada) (SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 2478–2484. <https://doi.org/10.1145/3404835.3463243>
- [3] Netta Aizenbud-Reshef, Ido Guy, and Michal Jacovi. 2009. Collaborative Feed Reading in a Community. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work (Sanibel Island, Florida, USA) (GROUP '09)*. Association for Computing Machinery, New York, NY, USA, 277–280. <https://doi.org/10.1145/1531674.1531716>
- [4] Hamed Alhoori, Mohammed Samaka, Richard Furuta, and Edward A Fox. 2019. Anatomy of scholarly information behavior patterns in the wake of academic social media platforms. *International Journal on Digital Libraries* 20, 4 (2019), 369–389. <https://doi.org/10.1007/s00799-018-0255-9>
- [5] Mohammad Aliannejadi, Leif Azzopardi, Hamed Zamani, Evangelos Kanoulas, Paul Thomas, and Nick Craswell. 2021. Analysing Mixed Initiatives and Search

- [6] Lorin W Anderson and David R Krathwohl. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- [7] Kumaripaba Athukorala, Eve Hoggan, Anu Lehtiö, Tuukka Ruotsalo, and Giulio Jacucci. 2013. Information-seeking behaviors of computer scientists: Challenges for electronic literature search tools. *Proceedings of the American Society for Information Science and Technology* 50, 1 (2013), 1–11. <https://doi.org/10.1002/meet.14505001041> arXiv:[https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/meet.14505001041](https://arxiv.org/abs/https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/meet.14505001041)
- [8] Tal August, Lucy Lu Wang, Jonathan Bragg, Marti A Hearst, Andrew Head, and Kyle Lo. 2022. Paper Plain: Making Medical Research Papers Approachable to Healthcare Consumers with Natural Language Processing. *arXiv preprint arXiv:2203.00130* (2022). <https://doi.org/10.48550/arXiv.2203.00130>
- [9] Sandeep Avula, Gordon Chadwick, Jaime Arguello, and Robert Capra. 2018. SearchBots: User Engagement with ChatBots during Collaborative Search. In *Proceedings of the 2018 Conference on Human Information Interaction Retrieval* (New Brunswick, NJ, USA) (*CHIIR '18*). Association for Computing Machinery, New York, NY, USA, 52–61. <https://doi.org/10.1145/3176349.3176380>
- [10] Leif Azopardi. 2021. Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia) (*CHIIR '21*). Association for Computing Machinery, New York, NY, USA, 27–37. <https://doi.org/10.1145/3406522.3446023>
- [11] Alina Beygelzimer, Emily Fox, Florence d'Alché Buc, and Hugo Larochelle. 2019. What we learned from NeurIPS 2019 data. <https://neuripsconf.medium.com/what-we-learned-from-neurips-2019-data-11lab996462c>
- [12] Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2022. The Values Encoded in Machine Learning Research. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) (*FACCT '22*). Association for Computing Machinery, New York, NY, USA, 173–184. <https://doi.org/10.1145/3531146.3533083>
- [13] David M. Blei and Padhraic Smyth. 2017. Science and data science. *Proceedings of the National Academy of Sciences* 114, 33 (2017), 8689–8692. <https://doi.org/10.1073/pnas.1702076114> arXiv:<https://www.pnas.org/doi/pdf/10.1073/pnas.1702076114>
- [14] Marcel Bollmann and Desmond Elliott. 2020. On Forgetting to Cite Older Papers: An Analysis of the ACL Anthology. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 7819–7827. <https://doi.org/10.18653/v1/2020.acl-main.699>
- [15] Corinna Breittinger, Patrick Wortner, Bela Gipp, and Harald Reiterer. 2019. "Too Late to Collaborate": Challenges to the Discovery of in-Progress Research. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. 134–137. <https://doi.org/10.1109/JCDL.2019.00028>
- [16] Melanie S Brucks and Jonathan Levav. 2022. Virtual communication curbs creative idea generation. *Nature* 605, 7908 (2022), 108–112. <https://doi.org/10.1038/s41586-022-04643-y>
- [17] Hilary Bussell, Jennifer Schnabel, and Amanda K. Rinehart. 2020. Meeting graduate student needs: an exploration of disciplinary differences. *Public Services Quarterly* 16, 4 (2020), 213–233. <https://doi.org/10.1080/15228959.2020.1818663> arXiv:<https://doi.org/10.1080/15228959.2020.1818663>
- [18] Arthur Câmara, Nirmal Roy, David Maxwell, and Claudia Hauff. 2021. Searching to Learn with Instructional Scaffolding. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia) (*CHIIR '21*). Association for Computing Machinery, New York, NY, USA, 209–218. <https://doi.org/10.1145/3406522.3446012>
- [19] Joel Chan, Joseph Che Chang, Tom Hope, Dafna Shahaf, and Aniket Kittur. 2018. SOLVENT: A Mixed Initiative System for Finding Analogies between Research Papers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 31 (Nov. 2018), 21 pages. <https://doi.org/10.1145/3274300>
- [20] Jiyouo Chang and Christine Custis. 2022. Understanding Implementation Challenges in Machine Learning Documentation. In *Equity and Access in Algorithms, Mechanisms, and Optimization* (Arlington, VA, USA) (*EAAAM '22*). Association for Computing Machinery, New York, NY, USA, Article 16, 8 pages. <https://doi.org/10.1145/3551624.3555301>
- [21] Duen Horng Chau, Aniket Kittur, Jason I. Hong, and Christos Faloutsos. 2011. Apolo: Making Sense of Large Network Data by Combining Rich User Interaction and Machine Learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (*CHI '11*). Association for Computing Machinery, New York, NY, USA, 167–176. <https://doi.org/10.1145/1978942.1978967>
- [22] Kiyoung Choe, Seokwon Jung, Seokhyeon Park, Hwajung Hong, and Jinwook Seo. 2021. Papers101: Supporting the Discovery Process in the Literature Review Workflow for Novice Researchers. In *2021 IEEE 14th Pacific Visualization Symposium (PacificVis)*. 176–180. <https://doi.org/10.1109/PacificVis52677.2021.00037>
- [23] John S. G. Chu and James A. Evans. 2021. Slowed canonical progress in large fields of science. *Proceedings of the National Academy of Sciences* 118, 41 (2021). <https://doi.org/10.1073/pnas.2021636118>
- [24] Anamaria Crisan, Brittany Fiore-Gartland, and Melanie Tory. 2020. Passing the data baton: A retrospective analysis on data science work and workers. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 1860–1870. <https://doi.org/10.1109/TVCG.2020.3030340>
- [25] Cecilia di Sciascio, Eduardo Veas, Jordan Barria-Pineda, and Colleen Culley. 2020. Understanding the Effects of Control and Transparency in Searching as Learning. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) (*IUI '20*). Association for Computing Machinery, New York, NY, USA, 498–509. <https://doi.org/10.1145/3377325.3377524>
- [26] Cheyenne Dosso, Lynda Tammine, Pierre-Vincent Paubel, and Aline Chevalier. 2022. The Impact of Expertise on Query Formulation Strategies During Complex Learning Task Solving: A Study with Students in Medicine and Computer Science. In *Proceedings of the 21st Congress of the International Ergonomics Association* (IEA 2021), Nancy L. Black, W. Patrick Neumann, and Ian Noy (Eds.). Springer International Publishing, Cham, 621–627. https://doi.org/10.1007/978-3-030-74614-8_77
- [27] Marcel Dunaiski, Gillian J Greene, and Bernd Fischer. 2017. Exploratory search of academic publication and citation data using interactive tag cloud visualizations. *Scientometrics* 110, 3 (2017), 1539–1571. <https://doi.org/10.1007/s11192-016-2236-3>
- [28] Marian Dörk, Nathalie Henry Riche, Gonzalo Ramos, and Susan Dumais. 2012. PivotPaths: Strolling through Faceted Information Spaces. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2709–2718. <https://doi.org/10.1109/TVCG.2012.252>
- [29] Debra Engel, Sarah Robbins, and Christina Kulp. 2011. The Information-Seeking Habits of Engineering Faculty. *College & Research Libraries* 72, 6 (2011), 548–567. <https://doi.org/10.5860/crl-155>
- [30] Michael Färber and Ann-Kathrin Leisinger. 2021. *DataHunter: A System for Finding Datasets Based on Scientific Problem Descriptions*. Association for Computing Machinery, New York, NY, USA, 749–752. <https://doi.org/10.1145/3460231.3478882>
- [31] Michael Färber and Ann-Kathrin Leisinger. 2021. DataHunter: A System for Finding Datasets Based on Scientific Problem Descriptions. In *Proceedings of the 15th ACM Conference on Recommender Systems* (Amsterdam, Netherlands) (*RecSys '21*). Association for Computing Machinery, New York, NY, USA, 749–752. <https://doi.org/10.1145/3460231.3478882>
- [32] Raymond Fok, Andrew Head, Jonathan Bragg, Kyle Lo, Marti A Hearst, and Daniel S Weld. 2022. Scim: Intelligent Faceted Highlights for Interactive, Multi-Pass Skimming of Scientific Papers. (

- [41] Lu He and Changyang He. 2022. Help Me #DebunkThis: Unpacking Individual and Community's Collaborative Work in Information Credibility Assessment. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 413 (nov 2022), 31 pages. <https://doi.org/10.1145/3555138>
- [42] Andrew Head, Kyle Lo, Dongyeop Kang, Raymond Fok, Sam Skjonsberg, Daniel S. Weld, and Marti A. Hearst. 2021. Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 413, 18 pages. <https://doi.org/10.1145/3411764.3445648>
- [43] Andrew Head, Amber Xie, and Marti A. Hearst. 2022. Math Augmentation: How Authors Enhance the Readability of Formulas Using Novel Visual Design Practices. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 491, 18 pages. <https://doi.org/10.1145/3491102.3501932>
- [44] Kevin Heffernan and Simone Teufel. 2018. Identifying problems and solutions in scientific text. *Scientometrics* 116, 2 (2018), 1367–1382. <https://doi.org/10.1007/s11192-018-2718-6>
- [45] Florian Heimerl, Qi Han, and Steffen Koch. 2016. CiteRivers: Visual Analytics of Citation Patterns. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (jan 2016), 190–199. <https://doi.org/10.1109/TVCG.2015.2467621>
- [46] Orland Hoerber, Dolinkumar Patel, and Dale Storie. 2019. A Study of Academic Search Scenarios and Information Seeking Behaviour. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (Glasgow, Scotland UK) (CHIIR '19). Association for Computing Machinery, New York, NY, USA, 231–235. <https://doi.org/10.1145/3295750.3298943>
- [47] Fred Hohman, Matthew Conlen, Jeffrey Heer, and Duen Horng (Polo) Chau. 2020. Communicating with Interactive Articles. *Distill* (2020). <https://doi.org/10.23915/distill.00028> <https://distill.pub/2020/communicating-with-interactive-articles>
- [48] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé, Miro Dudik, and Hanna Wallach. 2019. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3290605.3300830>
- [49] Chien-yu Huang, Arlene Casey, Dorota Glowacka, and Alan Medlar. 2019. Holes in the Outline: Subject-Dependent Abstract Quality and Its Implications for Scientific Literature Search (CHIIR '19). Association for Computing Machinery, New York, NY, USA, 289–293. <https://doi.org/10.1145/3295750.3298953>
- [50] Chien-yu Huang, Arlene Casey, Dorota Glowacka, and Alan Medlar. 2019. Holes in the Outline: Subject-Dependent Abstract Quality and Its Implications for Scientific Literature Search. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (Glasgow, Scotland UK) (CHIIR '19). Association for Computing Machinery, New York, NY, USA, 289–293. <https://doi.org/10.1145/3295750.3298953>
- [51] Sharon Favaro Ince, Christopher Hoadley, and Paul A. Kirschner. 2018. A Study of Search Practices in Doctoral Student Scholarly Workflows. In *Proceedings of the 2018 Conference on Human Information Interaction; Retrieval* (New Brunswick, NJ, USA) (CHIIR '18). Association for Computing Machinery, New York, NY, USA, 245–248. <https://doi.org/10.1145/3176349.3176877>
- [52] Nanna Inie, Jonas Frich, and Peter Dalsgaard. 2022. How Researchers Manage Ideas. In *Creativity and Cognition* (Venice, Italy). Association for Computing Machinery, New York, NY, USA, 83–96. <https://doi.org/10.1145/3527927.3532813>
- [53] Emi Ishita, Yasuko Hagiwara, Yukiko Watanabe, and Yoichi Tomiura. 2018. Which Parts of Search Results Do Researchers Check When Selecting Academic Documents?. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries* (Fort Worth, Texas, USA) (JCDL '18). Association for Computing Machinery, New York, NY, USA, 345–346. <https://doi.org/10.1145/3197026.3203867>
- [54] Charles Jacobs, Wil Li, Evan Schrier, David Barger, and David Salesin. 2004. Adaptive Document Layout. *Commun. ACM* 47, 8 (aug 2004), 60–66. <https://doi.org/10.1145/1012037.1012063>
- [55] Anthony Jameson and Barry Smyth. 2007. Recommendation to groups. In *The adaptive web*. Springer, 596–627. https://doi.org/10.1007/978-3-540-72079-9_20
- [56] Nicole Janssen. 2022. *The Data Science Talent Gap: Why It Exists And What Businesses Can Do About It*. <https://www.forbes.com/sites/forbestechcouncil/2022/10/11/the-data-science-talent-gap-why-it-exists-and-what-businesses-can-do-about-it/>
- [57] David Jurgens, Srijan Kumar, Raine Hoover, Dan McFarland, and Dan Jurafsky. 2018. Measuring the Evolution of a Scientific Field through Citation Frames. *Transactions of the Association for Computational Linguistics* 6 (2018), 391–406. https://doi.org/10.1162/tacl_a_00028
- [58] Hyeonsu Kang, Joseph Chee Chang, Yongsung Kim, and Aniket Kittur. 2022. Threddy: An Interactive System for Personalized Thread-Based Exploration and Organization of Scientific Literature. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) (UIST '22). Association for Computing Machinery, New York, NY, USA, Article 94, 15 pages. <https://doi.org/10.1145/3526113.3545660>
- [59] Hyeonsu B Kang, Rafal Kocielnik, Andrew Head, Jiangjiang Yang, Matt Latzke, Aniket Kittur, Daniel S Weld, Doug Downey, and Jonathan Bragg. 2022. From Who You Know to What You Read: Augmenting Scientific Recommendations with Implicit Social Networks. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 302, 23 pages. <https://doi.org/10.1145/3491102.3517470>
- [60] Hyeonsu B Kang, Sheshera Mysore, Kevin J Huang, Haw-Shiuan Chang, Thorben Prein, Andrew McCallum, Aniket Kittur, and Elsa Olivetti. 2022. Augmenting Scientific Creativity with Retrieval across Knowledge Domains. In *Second Workshop on Bridging Human-Computer Interaction and Natural Language Processing at NAACL 2022*. <https://doi.org/10.48550/arXiv.2206.01328>
- [61] Hyeonsu B. Kang, Xin Qian, Tom Hope, Dafna Shahaf, Joel Chan, and Aniket Kittur. 2022. Augmenting Scientific Creativity with an Analogical Search Engine. *ACM Trans. Comput.-Hum. Interact.* (mar 2022). <https://doi.org/10.1145/3530013> Just Accepted.
- [62] Mary Beth Kery, Bonnie E. John, Patrick O'Flaherty, Amber Horvath, and Brad A. Myers. 2019. Towards Effective Foraging by Data Scientists to Find Past Analysis Choices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300322>
- [63] Ian A. Knight, Max L. Wilson, David F. Brailsford, and Natasa Milic-Frayling. 2019. Enslaved to the Trapped Data: A Cognitive Work Analysis of Medical Systematic Reviews. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (Glasgow, Scotland UK) (CHIIR '19). Association for Computing Machinery, New York, NY, USA, 203–212. <https://doi.org/10.1145/3295750.3298937>
- [64] Laura M. Koesten, Emilia Kacprzak, Jennifer F. A. Tension, and Elena Simperl. 2017. The Trials and Tribulations of Working with Structured Data: A Study on Information Seeking Behaviour. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 1277–1289. <https://doi.org/10.1145/3025453.3025838>
- [65] Kaisu Koivumäki, Timo Koivumäki, and Erkki Karvonen. 2020. "On Social Media Science Seems to Be More Human": Exploring Researchers as Digital Science Communicators. *Media and Communication* 8, 2 (2020), 425–439. <https://doi.org/10.17645/mac.v8i2.2812>
- [66] Mario Krenn, Lorenzo Buffoni, Bruno Coutinho, Sagi Eppel, Jacob Gates Foster, Andrew Gritsevskiy, Harlin Lee, Yichao Lu, Joao P Moutinho, Nima Sanjabi, et al. 2022. Predicting the Future of AI with AI: High-quality link prediction in an exponentially growing knowledge network. (2022). <https://doi.org/10.48550/arXiv.2210.00881>
- [67] Satyapriya Krishna, Tessa Han, Alex Gu, Javin Pombra, Shahin Jabbari, Steven Wu, and Himabindu Lakkaraju. 2022. The Disagreement Problem in Explainable Machine Learning: A Practitioner's Perspective. *arXiv preprint arXiv:2202.01602* (2022). <https://doi.org/10.48550/arXiv.2202.01602>
- [68] Sean Kross and Philip J. Guo. 2019. Practitioners Teaching Data Science in Industry and Academia: Expectations, Workflows, and Challenges. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300493>
- [69] Carol Kuhlthau. 1993. Seeking Meaning: a process approach to library and information services" Ablex Publishing. (01 1993).
- [70] Ilia Kuznetsov, Jan Buchmann, Max Eichler, and Iryna Gurevych. 2022. Revise and Resubmit: An Intertextual Model of Text-based Collaboration in Peer Review. *arXiv preprint arXiv:2204.10805* (2022). <https://doi.org/10.48550/arXiv.2204.10805>
- [71] Esther Landhuis. 2016. Scientific literature: Information overload. *Nature* 535, 7612 (2016), 457–458. <https://doi.org/10.1038/nj7612-457a>
- [72] Or Levi, Ido Guy, Fiana Raiber, and Oren Kurland. 2018. Selective Cluster Presentation on the Search Results Page. *ACM Trans. Inf. Syst.* 36, 3, Article 28 (feb 2018), 42 pages. <https://doi.org/10.1145/3158672>
- [73] Irene Li, Alexander R Fabbri, Robert R Tung, and Dragomir R Radev. 2019. What should i learn first: Introducing lecturebank for nlp education and prerequisite chain learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 6674–6681. <https://doi.org/10.1609/aaai.v33i01.33016674>
- [74] Kevin Li, Haoyang Yang, Anish Upadhyay, Zhiyan Zhou, Jon Saad-Falcon, and Duen Horng Chau. 2021. Argo Scholar: Interactive Visual Exploration of Literature in Browsers. (2021). <https://doi.org/10.48550/arXiv.2110.14060>
- [75] Michael Xieyang Liu, Aniket Kittur, and Brad A. Myers. 2021. To Reuse or Not To Reuse? A Framework and System for Evaluating Summarized Knowledge. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 166 (apr 2021), 35 pages. <https://doi.org/10.1145/3449240>
- [76] Ying-Hsang Liu, Paul Thomas, Tom Gedeon, and Nicolay Rusnachenko. 2022. Search Interfaces for Biomedical Searching: How Do Gaze, User Perception, Search Behaviour and Search Performance Relate?. In *ACM SIGIR Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) (CHIIR

- '22). Association for Computing Machinery, New York, NY, USA, 78–89. <https://doi.org/10.1145/3498366.3505769>
- [77] Kelvin Luu, Xinyi Wu, Rik Koncel-Kedziorski, Kyle Lo, Isabel Cachola, and Noah A. Smith. 2021. Explaining Relationships Between Scientific Documents. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online, 2130–2144. <https://doi.org/10.18653/v1/2021.acl-long.166>
- [78] Gary Marchionini. 2018. Search, sense making and learning: closing gaps. *Information and Learning Sciences* (2018). <https://doi.org/10.1108/ILS-06-2018-0049>
- [79] Moshe Mash, Stephanie Rosenthal, and Reid Simmons. 2021. *DSWorkflow: A Framework for Capturing Data Scientists' Workflows*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411763.3451683>
- [80] Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2021. Paper Forager: Supporting the Rapid Exploration of Research Document Collections. In *Graphics Interface 2021*. <https://openreview.net/forum?id=QhN4tUZd8r>
- [81] Lori McCay-Peet, Anabel Quan-Haase, and Dagmar Kern. 2015. Exploratory search in digital libraries: a preliminary examination of the use and role of interface features. *Proceedings of the Association for Information Science and Technology* 52, 1 (2015), 1–4. <https://doi.org/10.1002/pr2.2015.145052010070>
- [82] Graham McDonald, Craig Macdonald, and Iadh Ounis. 2022. Search results diversification for effective fair ranking in academic search. *Information Retrieval Journal* 25, 1 (2022), 1–26. <https://doi.org/10.1007/s10791-021-09399-z>
- [83] Alan Medlar, Jing Li, and Dorota Glowacka. 2021. Query Suggestions as Summarization in Exploratory Search. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia) (CHIIR '21). Association for Computing Machinery, New York, NY, USA, 119–128. <https://doi.org/10.1145/3406522.3446020>
- [84] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) (FAT* '19). Association for Computing Machinery, New York, NY, USA, 220–229. <https://doi.org/10.1145/3287560.3287596>
- [85] John S Morabito and Joel Chan. 2021. Managing Context during Scholarly Knowledge Synthesis: Process Patterns and System Mechanics. In *Creativity and Cognition* (Virtual Event, Italy). Association for Computing Machinery, New York, NY, USA, Article 39, 5 pages. <https://doi.org/10.1145/3450741.3465244>
- [86] Felipe Moraes, Sindunuraga Rikarno Putra, and Claudia Hauff. 2018. Contrasting Search as a Learning Activity with Instructor-Designed Learning. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (Torino, Italy) (CIKM '18). Association for Computing Machinery, New York, NY, USA, 167–176. <https://doi.org/10.1145/3269206.3271676>
- [87] Meredith Ringel Morris. 2013. Collaborative Search Revisited. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) (CSCW '13). Association for Computing Machinery, New York, NY, USA, 1181–1192. <https://doi.org/10.1145/2441776.2441910>
- [88] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3290605.3300356>
- [89] Sheshera Mysore, Tim O'Gorman, Andrew McCallum, and Hamed Zamani. 2021. CSFCube - A Test Collection of Computer Science Research Articles for Faceted Query by Example. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*. <https://doi.org/10.48550/arXiv.2103.12906>
- [90] Ya R Nedumov and Sergei D Kuznetsov. 2019. Exploratory search for scientific articles. *Programming and Computer Software* 45, 7 (2019), 405–416. <https://doi.org/10.1134/S0361768819070089>
- [91] Xi Niu, Bradley M Hemminger, Cory Lown, Stephanie Adams, Cecelia Brown, Allison Level, Merinda McLure, Audrey Powers, Michele R Tennant, and Tara Cataldo. 2010. National study of information seeking behavior of academic researchers in the United States. *Journal of the American Society for Information Science and Technology* 61, 5 (2010), 869–890. <https://doi.org/10.1002/asi.21307>
- [92] Fatima W. Nosheen, Irfan Ali, and Shazia Yasmeen. 2018. Keeping found things found. *Information and Learning Science* 119, 12 (2018), 712–720. <https://doi.org/10.1108/ILS-07-2018-0064>
- [93] Iadh Ounis, Craig MacDonald, and Ian Soboroff. 2021. On the TREC Blog Track. *Proceedings of the International AAAI Conference on Web and Social Media* 2, 1 (Sep. 2021), 93–101. <https://ojs.aaai.org/index.php/ICWSM/article/view/18622>
- [94] Elisabeth Pain. 2016. How to keep up with the scientific literature. *Science Careers* 30 (2016). <https://www.science.org/content/article/how-keep-scientific-literature-rev2>
- [95] Srishti Palani, Zijian Ding, Stephen MacNeil, and Steven P. Dow. 2021. *The "Active Search" Hypothesis: How Search Strategies Relate to Creative Learning*. Association for Computing Machinery, New York, NY, USA, 325–329. <https://doi.org/10.1145/3406522.3446046>
- [96] Srishti Palani, Zijian Ding, Austin Nguyen, Andrew Chuang, Stephen MacNeil, and Steven P. Dow. 2021. CoNate: Suggesting Queries Based on Notes Promotes Knowledge Discovery. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 726, 14 pages. <https://doi.org/10.1145/3411764.3445618>
- [97] Jennifer Pearson, Tom Owen, Harold Thimbleby, and George R. Buchanan. 2012. Co-Reading: Investigating Collaborative Group Reading. In *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries* (Washington, DC, USA) (JCDL '12). Association for Computing Machinery, New York, NY, USA, 325–334. <https://doi.org/10.1145/2232817.2232876>
- [98] Timothy Persons. 2016. Data and Analytics Innovation: Emerging Opportunities and Challenges. (2016).
- [99] Jinghua Piao, Guozhen Zhang, Fengli Xu, Zhilong Chen, Yu Zheng, Chen Gao, and Yong Li. 2021. Bringing Friends into the Loop of Recommender Systems: An Exploratory Study. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 439 (oct 2021), 26 pages. <https://doi.org/10.1145/3479583>
- [100] Peter Piroli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, Vol. 5. McLean, VA, USA, 2–4.
- [101] Antoine Ponsard, Francisco Escalona, and Tamara Munzner. 2016. PaperQuest: A Visualization Tool to Support Literature Review. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) (CHI EA '16). Association for Computing Machinery, New York, NY, USA, 2264–2271. <https://doi.org/10.1145/2851581.2892334>
- [102] Jason Portenoy, Marissa Radensky, Jevin D West, Eric Horvitz, Daniel S Weld, and Tom Hope. 2022. Bursting Scientific Filter Bubbles: Boosting Innovation via Novel Author Discovery. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 309, 13 pages. <https://doi.org/10.1145/3491102.3501905>
- [103] Gil Press. 2013. *A Very Short History of Data Science*. <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/?sh=310bcc2a55cf>
- [104] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. 2020. Towards Memorable Information Retrieval. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval* (Virtual Event, Norway) (ICTIR '20). Association for Computing Machinery, New York, NY, USA, 69–76. <https://doi.org/10.1145/3409256.3409830>
- [105] Napol Rachatasumrit, Jonathan Bragg, Amy X. Zhang, and Daniel S Weld. 2022. CiteRead: Integrating Localized Citation Contexts into Scientific Paper Reading. In *27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) (IUI '22). Association for Computing Machinery, New York, NY, USA, 707–719. <https://doi.org/10.1145/3490099.3511162>
- [106] Behnam Rahnadari and Peter Brusilovsky. 2021. PaperExplorer: Personalized Exploratory Search for Conference Proceedings. In *IUI Workshops*.
- [107] Dheeraj Rajagopal, Xuchao Zhang, Michael Gamon, Sujay Kumar Jauhar, Diyi Yang, and Eduard Hovy. 2022. One Document, Many Revisions: A Dataset for Classification and Description of Edit Intents. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 5517–5524. <https://aclanthology.org/2022.lrec-1.591>
- [108] Nirmal Roy, Felipe Moraes, and Claudia Hauff. 2020. Exploring Users' Learning Gains within Search Sessions. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (Vancouver BC, Canada) (CHIIR '20). Association for Computing Machinery, New York, NY, USA, 432–436. <https://doi.org/10.1145/3343413.3378012>
- [109] Nirmal Roy, Manuel Valle Torre, Ujwal Gadiraju, David Maxwell, and Claudia Hauff. 2021. Note the Highlight: Incorporating Active Reading Tools in a Search as Learning Environment. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia) (CHIIR '21). Association for Computing Machinery, New York, NY, USA, 229–238. <https://doi.org/10.1145/3406522.3446025>
- [110] Daniel M. Russell, Mark J. Stefik, Peter Piroli, and Stuart K. Card. 1993. The Cost Structure of Sensemaking. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) (CHI '93). Association for Computing Machinery, New York, NY, USA, 269–276. <https://doi.org/10.1145/169059.169209>
- [111] Andrey Rzhetsky, Jacob G. Foster, Ian T. Foster, and James A. Evans. 2015. Choosing experiments to accelerate collective discovery. *Proceedings of the National Academy of Sciences* 112, 47 (2015), 14569–14574. <https://doi.org/10.1073/pnas.1509757112>
- [112] Hemant Kumar Sahu and Surya Nath Singh. 2013. Information seeking behaviour of astronomy/astrophysics scientists. In *Aslib Proceedings*. Emerald Group Publishing Limited. <https://doi.org/10.1108/00012531311313961>

- [113] Dafna Shahaf, Carlos Guestrin, and Eric Horvitz. 2012. Metro Maps of Science. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Beijing, China) (KDD '12). Association for Computing Machinery, New York, NY, USA, 1122–1130. <https://doi.org/10.1145/2339530.2339706>
- [114] Rina Shaikh-Lesko. 2019. Web annotation tool Hypothesis hits a milestone. *Nature* 569, 7756 (2019), 295–296.
- [115] Namit Shetty. 2017. *Query Suggestions for Detailed Queries*. <https://scholar.googleblog.com/2017/08/query-suggestions-for-detailed-queries.html>
- [116] Catherine L. Smith and Paul B. Kantor. 2008. User Adaptation: Good Results from Poor Systems. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Singapore, Singapore) (SIGIR '08). Association for Computing Machinery, New York, NY, USA, 147–154. <https://doi.org/10.1145/1390334.1390362>
- [117] Catherine L. Smith and Soo Young Rieh. 2019. Knowledge-Context in Search Systems: Toward Information-Literate Actions. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (Glasgow, Scotland UK) (CHIIR '19). Association for Computing Machinery, New York, NY, USA, 55–62. <https://doi.org/10.1145/3295750.3298940>
- [118] Ayah Soufan, Ian Ruthven, and Leif Azzopardi. 2022. Searching the Literature: An Analysis of an Exploratory Search Task. In *ACM SIGIR Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) (CHIIR '22). Association for Computing Machinery, New York, NY, USA, 146–157. <https://doi.org/10.1145/3498366.3505818>
- [119] Sruti Srinivasa Ragavan, Sandeep Kaur Kuttal, Charles Hill, Anita Sarma, David Piorkowski, and Margaret Burnett. 2016. Foraging Among an Overabundance of Similar Variants. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 3509–3521. <https://doi.org/10.1145/2858036.2858469>
- [120] Krishna Subramanian, Johannes Maas, and Jan Borchers. 2020. *TRACTUS: Understanding and Supporting Source Code Experimentation in Hypothesis-Driven Data Science*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376764>
- [121] Hariharan Subramonyam, Colleen Seifert, Priti Shah, and Eytan Adar. 2020. TexSketch: Active Diagramming through Pen-and-Ink Annotations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376155>
- [122] Nicole Sultanum, Christine Murad, and Daniel Wigdor. 2020. Understanding and Supporting Academic Literature Review Workflows with LitSense. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Salerno, Italy) (AVI '20). Association for Computing Machinery, New York, NY, USA, Article 67, 5 pages. <https://doi.org/10.1145/3399715.3399830>
- [123] Rohail Syed, Kevyn Collins-Thompson, Paul N. Bennett, Mengqiu Teng, Shane Williams, Dr. Wendy W. Tay, and Shamsi Iqbal. 2020. Improving Learning Outcomes with Gaze Tracking and Automatic Question Generation. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) (WWW '20). Association for Computing Machinery, New York, NY, USA, 1693–1703. <https://doi.org/10.1145/3366423.3380240>
- [124] Editorial Team. 2021. Distill Hiatus. *Distill* (2021). <https://doi.org/10.23915/distill.00031> <https://distill.pub/2021/distill-hiatus>.
- [125] Kelsey Urgo and Jaime Arguello. 2022. Understanding the “Pathway” Towards a Searcher’s Learning Objective. *ACM Trans. Inf. Syst.* 40, 4, Article 77 (jan 2022), 42 pages. <https://doi.org/10.1145/3495222>
- [126] Pertti Vakkari. 2016. Searching as learning: A systematization based on literature. *Journal of Information Science* 42, 1 (2016), 7–18. <https://doi.org/10.1177/0165551515615833>
- [127] Pertti Vakkari and Saila Huuskonen. 2012. Search effort degrades search output but improves task outcome. *Journal of the American Society for Information Science and Technology* 63, 4 (2012), 657–670. <https://doi.org/10.1002/asi.21683> <https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.21683>
- [128] Pertti Vakkari, Mikko Pennanen, and Sami Serola. 2003. Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *Information Processing & Management* 39, 3 (2003), 445–463. [https://doi.org/10.1016/S0306-4573\(02\)00031-6](https://doi.org/10.1016/S0306-4573(02)00031-6)
- [129] Richard Van Noorden. 2014. Global scientific output doubles every nine years. *Nature news blog* (2014).
- [130] April Yi Wang, Dakuo Wang, Jaimie Drozdal, Xuye Liu, Soya Park, Steve Oney, and Christopher Brooks. 2021. *What Makes a Well-Documented Notebook? A Case Study of Data Scientists’ Documentation Practices in Kaggle*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411763.3451617>
- [131] Yun Wang, Dongyu Liu, Huamin Qu, Qiong Luo, and Xiaojuan Ma. 2016. A Guided Tour of Literature Review: Facilitating Academic Paper Reading with Narrative Visualization. In *Proceedings of the 9th International Symposium on Visual Information Communication and Interaction* (Dallas, TX, USA) (VINCI '16). Association for Computing Machinery, New York, NY, USA, 17–24. <https://doi.org/10.1145/2968220.2968242>
- [132] Jevin D. West and Carl T. Bergstrom. 2021. Misinformation in and about science. *Proceedings of the National Academy of Sciences* 118, 15 (2021), e1912444117. <https://doi.org/10.1073/pnas.1912444117>
- [133] Ryen W White and Resa A Roth. 2009. Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services* 1, 1 (2009), 1–98. <https://doi.org/10.2200/S00174ED1V01Y200901ICR003>
- [134] Teena Willoughby, S. Alexandria Anderson, Eileen Wood, Julie Mueller, and Craig Ross. 2009. Fast searching for information on the Internet to use in a learning context: The impact of domain knowledge. *Computers & Education* 52, 3 (2009), 640–648. <https://doi.org/10.1016/j.compedu.2008.11.009>
- [135] Longqi Yang, David Holtz, Sonia Jaffe, Siddharth Suri, Shilpi Sinha, Jeffrey Weston, Connor Joyce, Neha Shah, Kevin Sherman, Brent Hecht, et al. 2022. The effects of remote work on collaboration among information workers. *Nature human behaviour* 6, 1 (2022), 43–54. <https://doi.org/10.1038/s41562-021-01196-4>
- [136] Amy X. Zhang and Justin Cranshaw. 2018. Making Sense of Group Chat through Collaborative Tagging and Summarization. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 196 (nov 2018), 27 pages. <https://doi.org/10.1145/3274465>
- [137] Amy X. Zhang, Michael Muller, and Dakuo Wang. 2020. How Do Data Science Workers Collaborate? Roles, Workflows, and Tools. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 22 (may 2020), 23 pages. <https://doi.org/10.1145/3392826>
- [138] Huiwen Zhang, Dana McKay, and George Buchanan. 2021. I’ve Got All My Readers With Me: A Model of Reading as a Social Activity. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia) (CHIIR '21). Association for Computing Machinery, New York, NY, USA, 185–195. <https://doi.org/10.1145/3406522.3446022>
- [139] Xiaoyu Zhang, Senthil Chandrasegaran, and Kwan-Liu Ma. 2021. ConceptScope: Organizing and Visualizing Knowledge in Documents Based on Domain Ontology. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 19, 13 pages. <https://doi.org/10.1145/3411764.3445396>
- [140] Xiaolong Zhang, Yan Qu, C. Lee Giles, and Piyong Song. 2008. CiteSense: Supporting Sensemaking of Research Literature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) (CHI '08). Association for Computing Machinery, New York, NY, USA, 677–680. <https://doi.org/10.1145/1357054.1357161>
- [141] Yongfeng Zhang, Xu Chen, et al. 2020. Explainable recommendation: A survey and new perspectives. *Foundations and Trends® in Information Retrieval* 14, 1 (2020), 1–101. <https://doi.org/10.1561/15000000066>
- [142] Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. 2012. Successful Classroom Deployment of a Social Document Annotation System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) (CHI '12). Association for Computing Machinery, New York, NY, USA, 1883–1892. <https://doi.org/10.1145/2207676.2208326>
- [143] Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. 2012. Successful Classroom Deployment of a Social Document Annotation System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) (CHI '12). Association for Computing Machinery, New York, NY, USA, 1883–1892. <https://doi.org/10.1145/2207676.2208326>

A STUDY MATERIALS

Here we present the recruitment form used for our study, the questions used in our semi-structured interviews, and the prompts used in our think-aloud. Additionally, we report participant demographics in Table 1

A.1 Participant Recruitment

The following contents were displayed in a web form to solicit information for participant recruitment:

- (1) Your name. *Response: Free text*
- (2) Your email. This will help us to contact you for scheduling a study session. *Response: Free text*
- (3) Do you identify as a “data science” practitioner? Data science practitioners may be conceived as being very broad and inclusive of data work, applied engineering work, and research with individuals trained in computer science and statistics as well as disciplines of specific application domains such as economics and biology, among others. *Response: Yes, No*

- (4) Please list 3 research papers you have read in the past year and found enjoyable or otherwise useful. Please interpret "read" very broadly - this can include a full reading of the paper, skimming, or summaries read via blogs or tweets.
Response: Free text

A.2 Stage 1: Questions for Semi-structured Interview

The following questions were asked in our semi-structured interview with follow-up questions (sub-points below) asked when appropriate and responses summarized to participants before the next question.

- (1) Tell me a bit about your role as a data scientist.
 - For example: What kinds of problems do you work on? What is the nature of the work you conduct on a daily basis?
- (2) What are your goals in conducting a lit review?
- (3) What kinds of supporting tools do you use in conducting your reviews? Eg. search engines, any bibliography managers, note-taking apps, etc.
- (4) Are there points where you interact with others (virtually or in person) in the process of conducting reviews?
- (5) Tell me about a time that you found conducting a literature review to be frustrating or tedious.

- (6) Can you mention different ways in which you come across research literature? *asked if time permits*
 - What do you think the benefits and tradeoffs of those different methods are?
- (7) Before we conclude this stage, are there any additional thoughts about interactions with the literature that you would like to share?

A.3 Stage 2: Prompts for Think-Aloud with Task Scenario

In this stage, participants conducted a literature review and shared their screens so they could be observed.

Participants were shown the following prompts and instructed to start their literature review process.

"Recall a literature review you conducted in the past. Imagine you were re-starting this process and show us how you went through the literature review."

"Imagine you are interested in finding and documenting the latest work on a topic of your interest, show us how you go about this process."

"Imagine you are planning future work on a problem you are interested in, conduct the literature review to help plan your future work."