The Journal of Academic Librarianship 50 (2024) 102943

Contents lists available at ScienceDirect

The Journal of Academic Librarianship

journal homepage: www.elsevier.com/locate/jacalib

"What could go wrong?": An evaluation of ethical foresight analysis as a tool to identify problems of AI in libraries

Helen Bubinger *, Jesse David Dinneen

Berlin School of Library and Information Science, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

ARTICLE INFO

Keywords:
Academic libraries
AI ethics
AI auditing
Ethical foresight analysis
Delphi method

ABSTRACT

Artificial intelligence (AI) has entered libraries in various ways and raised concern about its potential ethical consequences therein. A number of approaches have been developed to encourage ethical AI and audit the ethics of specific AI applications, but very few approaches have been applied or tested, especially in a library setting, and so it remains unclear which, if any approaches are suitable or useful for encouraging ethical AI in libraries. We applied *Ethical Foresight Analysis* as an approach to identify possible ethical risks of an AI project for (semi-) automated subject indexing in a large research library. Specifically, to identify risks we conducted a two-round ethical Delphi study wherein experts on AI development, library practices, and AI ethics sought consensus on potential risks and their relative importance. The experts' post-test reflections on the procedure were then collected to inform an evaluation of the approach's feasibility. A variety of ethical risks of the specific project and of general AI indexing were indeed identified, most notably discrimination and under-representation stemming from attributes of the bibliographic training data provided by the library (e.g. varied historical contexts and gaps left by unindexed items). However, we identified some drawbacks of the approach tested: (1) it is time-consuming, which is likely prohibitive for many libraries, and (2) the identified risks were mainly well-known issues of AI and its training data rather than the subtle, application-specific, and human-centred issues that *ethical foresight analysis* might be employed to identify. Thus, although libraries should continue to model ethical AI through careful planning and auditing, alternative development and auditing approaches may be more practical to undertake and more effective at identifying novel or application-specific issues.

Introduction

Various applications of artificial intelligence (AI) are now available to libraries, such as humanoid greeter robots, software to handle patron requests (Vecera, 2020, p. 51), and automated indexing (Vecera, 2020, p. 49). The deployment of such AI applications has increased significantly (Lund, 2021, p. 2), particularly in academic libraries.

In the course of AI's breakthrough in other domains it has also brought major problems, especially with regard to ethical aspects, bias, transparency, explainability, privacy, data protection, and copyright (Kazim & Koshiyama, 2021). Outside the library sector, the ethics of AI are an active area of interest and research, for example in the case of facial recognition technology, where there is considerable room for abuse and discrimination (Buolamwini & Gebru, 2018). The variety of AI-related ethical concerns is large and shows that AI applications should always be viewed from a critical perspective and their risks taken into account. AI audits are one way of identifying such problems and risks and are currently attracting so much attention in the scientific debate that a topical collection of *Digital Society* was recently dedicated to them (Mökander, 2023, p. 19). Libraries are a particularly promising context for ethical AI research because of libraries' pro-social values and human-centred technological skills (V. Singh, Bilal, Cox, Chidziwisano, & Dinneen, 2023); yet while AI ethics has quickly become an integral part of research in fields like healthcare (Char, Abràmoff, & Feudtner, 2020), research on ethical AI in libraries is still coalescing. In particular there is very little research on practical methods for working towards ethical AI in libraries.

Bubinger and Dinneen (2021) previously outlined the need for AI ethics research in libraries, identified approaches to AI audits intended for other domains that appear promising for use in libraries, and noted a particular need to test and evaluate such approaches. In this study we take up that suggestion by examining Ethical Foresight Analysis (EFA) as a tool for identifying potential ethical issues with AI applications in libraries. In particular we explore its efficacy and feasibility for

* Corresponding author.
E-mail addresses: helen.bubinger@alumni.hu-berlin.de (H. Bubinger), jesse.dinneen@hu-berlin.de (J.D. Dinneen).

https://doi.org/10.1016/j.acalib.2024.102943
Received 15 December 2023; Received in revised form 14 August 2024; Accepted 19 August 2024
Available online 24 August 2024

identifying ethical issues of AI in academic libraries by applying it via an ethical Delphi study of a pilot project using AI for semi-automated content indexing at a large research library in Germany.

Literature review

AI in libraries

AI is being increasingly deployed in varied library settings and applications with both great promise and considerable risks for patrons and staff. Researchers have so far responded to the adoption of AI by libraries by considering the long-term effects or examining specific application areas.

AI is widely considered one of a few main factors that will change future knowledge management and library work (Bartlett, 2021), even prompting consideration of the future of the library profession within a fourth industrial revolution characterised by "*the evolution of information technology towards greater automation and interconnectedness*" (Lund, 2021, p. 1). While some predictions are dystopian, others point to the effect of past technological innovations (e.g. computers and the Internet) and suggest rather that librarians' jobs will simply change in nature and that they must upskill accordingly (e.g. Cox, 2022a), though inadvertent overall deskilling is also a risk (Østerlund et al., 2021).

Focusing on the shorter term, numerous application areas have been identified from general managerial uses, like AI-powered human resource allocation, to library specific applications. AI-powered chatbots for example have been investigated regarding their potential benefits for libraries (Panda & Chakravarty, 2021) and the challenges they may pose (Sanji, Behzadi, & Gomroki, 2022), resulting in the recommendation that librarians might collaborate with chatbot designers to get a better understanding of patrons' and librarians' expectations and therefore to improve the chatbots' performance. Further tasks have been taken over by *embodied* AI – robots – in several libraries, for example in user services and makerspaces, but also as support for librarians in tasks like stock-taking, locating missing books, or speeding up digitisation (De Sarkar, 2023; Früh, 2018; Vecera, 2020, pp. 50–51). AI may also improve the efficiency of library services like literature analysis via text mining and automatic metadata generation (Underwood, 2019), and with further development of applications may improve stacks management and information retrieval (Li & Wang, 2022).

Furthermore, AI can be helpful in management and retrieval of digital collections, for example to simplify the retrieval in large collections like digitised historical newspapers in an academic context (B. C. G. Lee, Berson, & Berson, 2021). Additionally, several methodologies like machine learning (ML) can classify images in provincial research (Krickl, Mayer, & Zangger, 2022). *Named Entity Recognition* is able to identify entities like persons or locations in digitised historical documents (Ehrmann, Hamdi, Pontes, Romanello, & Doucet, 2021) and the natural language processing model BERT was developed to improve Optical Character Recognition (OCR), enable text classification, and thus handle large volumes of incoming textual data (Haffenden, Fano, Malmsten, & Börjeson, 2022). Therefore data curation has often been an important point of reference for the use of AI in libraries and general information management. Libraries that could be seen as "*data stores*" (Korošec, 2020, p. 55) are an important player in data mining and should use specific tools, preferably AI, to find new relations between data and its meaning (Korošec, 2020).

AI-assisted content indexing is particularly promising; for example, automated metadata creation may save librarians time, but is at a very early stage of development in which librarian participation is crucial to ensure an effective system (Corrado, 2021). This application therefore currently exists often in a prototype or project phase, such as the one in the present study, but the National Library of Finland has developed a leading and open-source toolkit and Web-based API service (Souminen, Inkinen, & Lehtinen, 2022).

For further AI applications in libraries, see Das and Islam (2021), Lippincott (2021), Oyelude (2021), and Seeliger et al. (2021).

AI ethics

There are numerous ethical concerns surrounding AI, which are generally discussed under the common topics of fairness, bias and discrimination (Strasser & Niedermayer, 2021), privacy and safety (Kazim & Koshiyama, 2021, pp. 8–9), explainability and transparency, and accountability (Kroll, 2020). However there are also intersectional concerns around the training data such as stolen labour and lack of consent, and further still, global equity issues derived from AI development, its benefits, and its considerable environmental impacts (Bender, Gebru, McMillan-Major, & Shmitchell, 2021).

Numerous examples illustrate the varied ethical problems of AI systems (Dubber, Pasquale, & Das, 2020), but perhaps the most widely known issue with AI today – also one very illustrative of its pernicious risks – is bias. Owing variously to bias in the real world that is accurately captured by the data used to train AI and to bias manifested as incompleteness in the contents of the data sets (e.g. insufficiently diverse phenotypical data; Jo & Gebru, 2020), minorities along various demographic properties (e.g. origin, gender, skin colour) are frequently excluded, misclassified, or otherwise discriminated against by an AI system (Borenstein & Howard, 2021). For example, people of colour have been notoriously misclassified by facial recognition technology (Gebru, 2020, p. 266) and popular automated hiring software has significantly disadvantaged women (Gebru, 2020, p. 255). However, bias also manifests in system design decisions, such as metadata rules that narrowly implement gender and thus exclude transgender or non-binary people (Gebru, 2020, pp. 259–260).

Many initiatives have sought to address such ethical concerns, for example by preventing or reducing known issues (Borenstein & Howard, 2021, p. 63). Broad guidelines have been developed such as the *Montréal Declaration for Responsible Development of Artificial Intelligence*, which provides 10 abstractly formulated principles to be considered in the context of AI development like "*respect for autonomy, protection of privacy and intimacy, solidarity, equity, diversity inclusion, and responsibility*" (Université de Montréal, 2017). The research and non-governance organisation AlgorithmWatch promotes ethical concerns of AI in several projects, like the design of a map including published AI ethics guidelines worldwide (AlgorithmWatch, 2022a; AlgorithmWatch, 2022b). A similar map shows 84 national and international AI ethics guidelines and a distillation of 11 ethical values (Jobin, Ienca, & Vayena, 2019), while others have clarified the normative implications of many AI ethics guidelines for developers and users (Ryan & Stahl, 2021).

Although these guidelines provide a general orientation for AI software development, they have so far been largely abstract and non-binding (Brundage et al., 2020; Hagendorff, 2020; Munn, 2023). There are, however, dozens of more practically, legally, and ethically oriented approaches, tools, and frameworks for developing ethical AI, auditing AI systems, managing AI risks (e.g. NIST's AI Risk Management Framework), and identifying potential problems in advance (Mökander, 2023; Mökander & Floridi, 2021). These range from tools to identify bias during AI system development (Lee & Singh, 2021) to domain-specific tools like ethics auditing procedures for AI in healthcare (Char et al., 2020) or psychiatric diagnosis (Burr, Morley, Taddeo, & Floridi, 2020). The approaches are similarly diverse in their acuteness. For example, ambitious end-to-end auditing processes may monitor and report on all steps of an AI application's life cycle from development to deployment with the aim of closing accountability gaps (Raji et al., 2020). Another framework, *Z-Inspection*, has been developed to evaluate the trustworthiness of AI applications across different stages of an AI life cycle, and involves achieving consensus among experts through several phases (Vetter et al., 2023). Experts are similarly drawn upon in a collection of six different methods for forecasting potential risks of AI applications, such as *Ethical Foresight Analysis* (EFA; Floridi & Strait, 2020). No tool is without limitation (Mökander, Morley, Taddeo, & Floridi, 2021), but the

reality of the tools may be worse than that: in most cases these frameworks have not been implemented so far and are still proposals (Brundage et al., 2020; Bubinger & Dinneen, 2021), which means that there is no guarantee they help mitigate harm nor indication of which domain-agnostic approaches are practical in any particular context.

AI ethics in libraries

Although many Webinars in the past few years have acknowledged that the general ethical risks of AI may also apply in libraries, and some libraries have even produced guidelines for their use of AI (Van Wessel, 2020), a relatively small number of published works propose how to address AI-ethical issues in libraries in particular. Valuable works have identified potential adoption strategies, ethical challenges, and possible problematic outcomes of AI in libraries (Cox, 2022b; Nayyer & Rodriguez, 2022; Smith, 2021). As above, bias and the future of the profession are issues in the forefront of such discourse. Principles that contravene bias, like inclusion, equity, and fairness, have long been present in discussions of librarians' professional values (Zaiane, 2011) but are now renewed; for example, the *explainability* of AI has been emphasised to aid in establishing if a system in a library warrants "*significant concerns about bias, unfairness, and veracity*" (Ridley, 2022, p. 1). Regarding the future development of the library profession and the long-standing fear of librarians that they will lose their tasks, the term *human in the loop* reflects that AI applications in libraries should not replace humans nor make independent decisions, but rather should support a librarian's activity and operate under their oversight (Van Wessel, 2020).

Even rarer than principles and possibilities are initial actionable approaches to promote ethical AI in libraries. Some auditing methodologies have been proposed for libraries, primarily via adaptation from other settings (Bubinger & Dinneen, 2021), but as noted above, hardly any of such approaches have been applied, let alone tested, and to our knowledge none have been applied or tested in libraries in particular. It therefore remains unclear which approach, if any, is effective at identifying, preventing, or addressing the ethical issues of AI in libraries, and further if any approach is sufficiently practical to implement (e.g. detailed and doable).

Method

Aim of the study and research questions

In light of the gaps identified above, the aim of this study was to establish if *Ethical Foresight Analysis* (EFA) – which is intended primarily as an *ex ante* auditing tool to identify broad ethical problems (as opposed to narrow technical issues) prior to deployment (Mökander, 2023, p. 12) – is an effective and feasible tool for encouraging ethical AI in libraries. A study using EFA was thus conducted, which followed the aim to identify potential ethical risks and implications as they pertain to AI in the context of a particular AI application in an academic library. The idea of testing EFA as an approach followed the most promising among possible approaches to ethical AI in libraries identified by Bubinger and Dinneen (2021); libraries may have more capacity to prevent ethical issues before they arise than to address them afterwards, and there is no clear reason it could not be used in the context of library AI (Bubinger & Dinneen, 2021). Although there are six methods for implementing EFA (Floridi & Strait, 2020), the method *Crowd-sourced single prediction framework* was chosen, specifically a Delphi study, as its resource requirements are relatively modest (i.e. it does not require a system-wide rollout or sustained contact with developers across an application life cycle) and is thus suitable for a singular AI project and accessible to a greater number of libraries. Further, the Delphi method is an established method for predicting issues (Lund, 2020, p. 936) for which there are considerable resources that can be consulted.

While in the planning stages of the study and seeking partner libraries with AI applications, a project to use AI for (semi-)automated indexing was identified and was deemed suitable as it was in the early stages of deployment and in a large research library. Further details of the study setting are provided below.

Considering the gaps identified in the previous section and the goals and conditions of the study presented here, our particular research questions were:

RQ1: What are the foreseeable, ethical risks/problems of AI for (semi-)automated content indexing in an academic library?

RQ2: What are the advantages and disadvantages of the Delphi-powered *Ethical Foresight Analysis* for libraries?

Study setting

This Delphi study was realised in cooperation with the Berlin State Library (Staatsbibliothek zu Berlin), one of the largest libraries in Germany with a main collection of 32.5 million items and 31.000 patrons that takes on important transregional tasks (Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, n.d.; Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, 2020). It is financed by both federal and state governments and operates under the trusteeship of the Prussian Cultural Heritage Foundation (Stiftung Preußischer Kulturbesitz) (Gantert, 2016, pp. 16–17).

The library launched a new large-scale project on AI on July 01, 2022, called Mensch. Maschine. Kultur – Künstliche Intelligenz für das digitale Kulturelle Erbe (Man. Machine. Culture - AI for Digital Cultural Heritage; Neudecker, 2022), which consists of four subprojects aiming to enable new AI-powered services for librarians and library patrons. With the project in development and yet to fully launch, the library had not yet had any concrete experiences and thus potential issues could still be identified.

Since this study was intended to examine a specific AI application in the library for ethical concerns, it was decided to focus on one of the four subprojects. The selected subproject deals with AI methods for (semi-) automated indexing procedures to support librarians and information professionals in their daily tasks as well as to semantically enrich discovery systems (Neudecker, 2022). This project was chosen as the library did not have any previous experience with automation in this area of work, especially content indexing, and so the applications had not been fully developed or tested by beginning of the study (September 2022), making it suitable for an *ex ante* auditing approach.

The chosen subproject comprised two work packages taking different approaches to automating content indexing. The first work package entailed developing semiautomated AI applications to support librarians in subject indexing by suggesting suitable keywords, meaning that AI or ML models would be trained on catalogue data and from which finally the proposed keywords would be pulled from existing bibliographic databases and digitised collections. The second work package was to develop a fully automated AI application for discovery systems, wherein digitised materials available as full texts would be made searchable for/in discovery systems. Specifically, information on entities that occur in the digitised materials were integrated into search processes and visual results via Named Entity Recognition (NER) and Named Entity Linking (NEL). Therefore specific names or locations could be automatically recognised in text documents and passed forward to discovery layers (Neudecker et al., 2021, p. 153). Since these two work packages were designed for varying degrees of automation, we use *(semi-)automated* to refer to both.

Research design, participants, and data collection

This study implemented the Delphi method (as a form of crowd-sourced single prediction for EFA). The Delphi method was first implemented by the RAND Corporation (Häder, 2014, p. 15) in 1964 to forecast long-term technical and scientific developments. It has since used in many scholarly studies, particularly in medicine (Lam, Iqbal, Purkayastha, & Kinross, 2021, p. 2), politics (Häder, 2014, p. 75), and

technology (Gallego & Bueno, 2014). It is characterised by the iteration of questions and discussion in rounds, controlled feedback, anonymity, and (statistical) aggregation (Millar, Thorstensen, Tomkins, Mepham, & Kaiser, 2007, p. 58). It also typically involves "*highly specialised experts for individual disciplines*" (Häder, 2014, p. 16), though many variations of the Delphi method exist for pursuing different goals (Gallego & Bueno, 2014, p. 991).

Though it has a general procedure, the Delphi method is highly adaptable (Jones et al., 2015, p. 3514). In many cases it can pursue for example the aim of idea aggregation and consensus (Häder, 2014, pp. 31–32), of collecting *different* opinions (Baruchson-Arbib & Bronstein, 2002, p. 399), or to foresee an uncertain future situation (Häder, 2014, pp. 32–33). In information science and technology, Delphi has mostly been used as a tool to forecast outcomes of new technologies (Gallego & Bueno, 2014, p. 991); in LIS, forecasting is the second most common goal of Delphi studies (Lund, 2020, p. 936). Yet a Delphi study aimed at aggregating ideas may pose open-ended qualitative questions (Puig & Adams, 2018) and indeed even quantitative questions although it does not aim to provide a statistical evaluation or claim to representation since it typically only involves a relatively small interdisciplinary group of experts (Häder, 2014, p. 32). Especially the *ethical Delphi* study variation as described by Millar et al. (2007, p. 56)"*does not look for consensus on future actions/developments as its target.*" In particular, it is recommended to use the ethical Delphi to collect ethical risks that may arise in the course of an emerging technology (Millar et al., 2007, p. 56).

Thus this study utilised a two-round ethical Delphi design with qualitative and quantitative data (discussion and importance scores, respectively) to answer RQ1 (i.e. identify risks of the project) and participants' post-test statements and researcher observations to answer RQ2 (i.e. reflect on the feasibility of the EFA approach). An overview on the data collection and analyses is presented in Table 1.

Seven experts from different fields around AI and libraries as well as one active library patron were involved: two AI project managers from libraries, two software developers working in libraries, one representative from the software industry (i.e. library software vendor), one librarian, one university professor with expertise in AI ethics, and one active user of an academic library. The small number of representatives of each group of experts follows the aim of a Delphi study that should aggregate ideas, where only one or two representatives are adequate for each group of expertise, as this study should only bring up ideas of possible ethical risks rather than establish a consensus representative of a singular group (Häder, 2014, p. 32).

The procedure of the Delphi method conducted here followed a combination of different Delphi approaches. The main focus in the first round was the iterative dialog between the participants to collect their different views and to aggregate ideas (Mankoff, Rode, & Faste, 2013, p. 1634). Each participant received a Web-based text document that served as a combined protocol and live editing environment for inputting their responses. Participants were instructed to answer each question and were also able to see the anonymised answers of the other participants so that responses would engage existing inputs and a discussion would emerge in the text document. It was decided to use the collaborative open-source tool *CryptPad* to conduct the study, as it can be accessed without any registration and therefore guarantees anonymity (barring self-identification in responses, which did not occur). This configuration of the Delphi method is called "*technological, real-time online Delphi*" (Lund, 2020, p. 931) and is similar to a written focus group but conducted online (Floridi & Strait, 2020, p. 81). The online implementation of the study was primarily due to organisational reasons, as this significantly expanded the group of experts that could participate and allowed including from countries beyond where the AI library project took place (Lam et al., 2021, p. 2).

Table 1
Overview of data collection and analysis.

Research question	Data collection	Analyses
RQ1 (foreseeable risks)	Delphi study ($n = 7$): discussion (rounds 1 & 2), importance scores and justifications (round 2)	Content analysis, tabular examination
RQ2 (evaluation of method)	Post-test statements (7), researcher observations (rounds 1 + 2)	Content analysis

The second round of the Delphi study continued structured discussion and added a quantitative component informed by the answers from round one. Named ethical issues and arguments collected in round one were mapped to three categories (procedure described below) and a questionnaire was produced for each category with items that summarised the risks as stated in participants' answers from round one. The three short questionnaires were present in the second round protocol and experts were asked to agree or disagree with each item on a Likert scale from 1 to 5. The answers thus reflected the participants' perceptions of the seriousness of the risks and enabled broadly ranking them by their relative importance. Participants were additionally asked to provide reasons for every rating provided for each named risk.

Finally, to generate data for RQ2 (reflection on the feasibility of the EFA approach), at the end of round two every participant was asked for a short statement on their experience of the study and their perceived advantages and disadvantages of the Delphi-powered EFA. For the same reason researcher observations of the procedure were collected throughout both rounds.

Data analysis

Participants' text-based responses from round 1 were coded immediately upon completion of the round using the qualitative data analysis software MAXQDA. The coding followed an inductive approach (i.e. codes were derived from the data; O'Reilly, 2009, p. 105) and three categories of risks were identified (example risks identified are presented below):

1. Data and bias
2. The AI system's effects on humans
3. Future job profiles for LIS professionals

Upon completion of round two, the experts' provided Likert scores were presented in tables and summarised (e.g. the clearest trends and rankings of agreement or risk seriousness were identified). The assessments via Likert scale only serve as a rough tendency of the experts to verify which risks were classified as particularly serious (i.e. realistic, severe, pressing) and which of the risks from round one appear to be lower priority (i.e. nominal risk, a remote possibility, or otherwise less important for immediate consideration); in other words, as a Delphi study aiming at idea aggregation, exact risk scores did not need to be calculated and no claim is made about the data being representative of the participant groups (Häder, 2014, p. 32). The experts' provided justifications of the given scores, also text that other participants could review and comment on, were then analysed in MAXQDA together with the data from the first round the. The frequency of the topics mentioned by the experts was established (i.e. frequency analysis; Mayring, 2015, pp. 13–15).

Results

Identified ethical risks

An immediately clear result from round one was that the discussion and risks identified did not distinguish between nor apply differently to the two work packages nor their respective technologies, but rather discussed their common application area of (semi)-automated indexing. For this reason, further differentiation between the two work packages and their possible scenarios was omitted (e.g. in the protocol for round

two). There were 21 issues identified in round one: eleven in the category *Data and bias*, six in *The AI system's effects on humans*, and four in *Future job profiles for LIS professionals*. Those categories' identified risks are presented next.

Risks that were not regarded as serious are presented in the tables for completeness but are not discussed unless particularly relevant to the most notable risks.

Category 1: data and bias

Most of the mentioned risks could be assigned to the category *Data and bias* (Table 2) and such risks mainly reflected problems with regards to the data used to generate the output of an AI system for (semi-) automated indexing. The bias and values reflected in data created in different epochs, regions, political and ideological systems, cultures, was named a "*core reason*" for significant problems of discrimination against marginalised (e.g. on the basis of "*origin, appearance, gender/sex, political or social groups*"). Discrimination may also stem from homogeneous or imbalanced training data that over- or under-represents certain languages, subjects, gender or cultures; for example, non-western cultures may be under-represented in otherwise relevant generated keywords. The majority of experts agreed on both discrimination-related risks.

The unavoidable bias present in decision-making in the course of library practice was also an identified and agreed-upon risk. Past choices, like decisions made by librarians on what to preserve and what not, will have an impact on libraries' bibliographic data and thus also on the AI training data. Participants agreed that this concern applies also to indexing and classification systems determined by human choice about (e.g.) which terms should exist or be replaced. It was further noted that such systems carry the historical perspectives of their time of creation, which like the collection data, would influence the hierarchies and terms that would be generated or suggested during (semi-)automated content indexing. In particular, a participant urged in the second round to see such systems as "*legacy codes*".

Finally technical factors influencing biased data sets were classified as a moderate problem. Among such technical problems were bibliographic data not always being machine-readable (e.g. if it comes from old card catalogues) and the difficulty of using external data following different standards (e.g. Wikidata). It was noted that both challenges could lead to certain time periods not being represented in the training data. Institutional circumstances and challenges of digitisation (e.g. funding) were also acknowledged for their role in determining data quality and diversity. However, data gaps can also be traced back to automatically imported cataloguing data, e-book packages, non-indexed library stocks, and several coexisting indexing systems. The experts only partially agreed on these issues, which are perhaps more easily solved, and some noted that completeness is not absolutely necessary. Similarly, factual inaccuracies in the training data were identified for their risk of producing incorrect keywords, but participants were divided about the seriousness of the issue.

Table 2
Identified risks category 1: Data and bias.

Category 1: data and bias	Agree–Not agree				
	1	2	3	4	5
Collections are biased by human choices.	4	2	–	1	–
Bias can be caused by digitisation factors.	4	2	–	1	–
Cataloguing and indexing systems are biased depending on the time of creation, this has an influence of hierarchies and terminologies suggested.	3	3	1	–	–
Biased data from different epochs, regions, political and ideological systems or cultures do not reflect contemporary values.	3	2	2	–	–
Cataloguing and indexing systems are biased by human choices.	3	3	–	1	–
Over- or under-representation of certain languages, subjects, people or cultures has an impact on keywords.	4	1	1	1	–
Technical issue: Many bibliographic data are not machine-readable and external data do not reach library standards. This leads to data gaps.	2	2	3	–	–
Information loss and gaps due to bibliographic data delivered by e-book packages.	2	3	–	2	–
There are opaque rules the legacy data are based on, so it is difficult to retrace where data come from.	–	3	3	1	–
Misinformation in documents could generate incorrect keywords.	–	3	2	2	–
Incomplete bibliographic data due to the fact that only a small part of library stock is subject indexed.	–	3	2	3	–

Category 2: the AI system's effects on humans

The risks mentioned in this category (Table 3) were assessed as generally less serious than those of data and bias. The possible influence of AI-generated keywords on users' search behaviour was the most agreed-upon risk; for example, users may search for, find, and understand materials differently with the AI-generated keywords than they would have otherwise (and existing biases may then be further perpetuated). Detail about why *exactly* this was a risk was lacking (e.g. what negative consequences might occur), but it may have been readily identified since it is already a known issue in Web search (identified in round two as "*confirmation bias*") for example. One participant noted that in the event of an unethical outcome caused by the AI, such as long-term negative influences on users, the library may be held accountable; there was however no overall agreement on that risk. It was also mentioned, but only partly agreed, that that the use of AI should be made clear to users and an "*active-learning*" loop should be established to integrate their feedback, but that such knowledge of how the system is worked is also likely to change some users' behaviour.

The potential influence of AI on *librarians* was also identified (and mostly agreed upon), though once again the exact consequences did not emerge in the discussion: as AI-generated suggestions are the goal of the indexing-related applications, users will undoubtedly be influenced by those suggestions, with an extreme possible consequence being an uncritical acceptance of all suggestions. Some participants reported having such experiences before, but others stated that the benefit of an accelerated workflow prevails. Moreover, one participant with library expertise noted it is still possible within many systems to assign additional keywords manually, so the influence of the system can be resisted.

Category 3: future job profiles for LIS profession-als

A few speculative and eventual risks of the AI application to the future of librarianship were identified (Table 4). What had the most agreement was that rather than causing redundancy, the implementation of AI applications in libraries is very staff-intensive and generally outstrips libraries' capacity, particularly since the implemented applications need to be monitored and corrected. It was generally agreed that time will be saved by using AI for content indexing according, but that jobs will not be in danger since the tasks will tend to shift to either "*post-*

Table 3
Identified risks category 2: The AI system's influence on humans.

Category 2: the AI system's influence on humans	Agree–Not agree				
	1	2	3	4	5
Patrons may adapt their search habits to AI-generated content.	2	2	1	2	–
The search behaviour of a patron can be influenced if the user realises that a machine is responsible for the keywords.	1	2	2	1	–
Suggested keywords could influence the librarian in its decisionmaking	1	1	3	2	–
The institution could be responsible for unethical outcomes caused by AI.	1	1	3	1	1
The enriched discovery system based on NER/NEL will make critical (i.e. sensitive) publications easier to find	–	1	4	1	1
Risk of surveillance of the librarians, when choosing the keywords, by the AI model.	–	–	4	2	1

Table 4
Identified risks category 3: Future job profiles for LIS professionals.

Category 3: future job profiles for LIS professionals	Agree–Not agree				
	1	2	3	4	5
AI applications are very staff intensive	3	3	1	–	–
Change of librarians role could bring uncertainty among librarian as they might find it difficult to redefine their tasks	1	2	1	3	–
The rise of AI could lead to a certain insensitivity to errors in subject indexing, a loss of importance of subject indexing as a task for specialists as well as a loss of the know-how necessary to accompany AI / ML.	1	–	4	2	–
Semi-automated content indexing will replace (only a few) jobs over time	–	1	5	–	1

correction" or other content management tasks. A potential risk of uncertainty was identified regarding the change of librarians' roles and difficulty redefining their tasks. This was assessed very differently by the experts, however: on the one hand, change brings uncertainty, but on the other hand, adaptation of job profiles is so slow that new generations are slowly coming up and older generations have forewarning before they must master new AI skills. Two experts in particular emphasised that most librarians are aware that only machine-based solutions can adequately handle the volume of relevant literature, so the advantages were seen as greater than the uncertainty of not being able to cope with new tasks.

A risk identified as less likely or pressing was a lowering of the quality of library work over time as a result of ongoing use of AI, such as building a tolerance or indifference to system errors, diminished expertise and skills as AI takes the lead on certain tasks, and thus a diminished emphasis or value placed on that expertise (specifically, content indexing). However, the importance and goal of such quality was also questioned, particularly whether it is important to fulfilling patron needs. Regardless, the importance of these risks were also doubted so long as a "*human in the loop*" remains.

Reported experience of using ethical foresight analysis

As noted above, participants were asked about their reflections and experiences of the implemented Delphi-driven EFA and its usefulness in libraries, and researcher observations about the study procedure and results were also collected. We summarise both here.

Regarding researcher observations, round one appeared helpful for generating ideas, but a notable disadvantage was that discussion was sometimes limited to agreement with the first response to a question. Thus the first participant to respond might strongly steer discussion and influence the opinions of the others. Additionally, answers became significantly shorter towards the end of the protocol, which may have reflected participant fatigue, but either way meant less detail was provided than could have been. Round two saw a renewal of efforts, however, as participants mostly used the opportunity to add comments to their Likert scale responses, even though this was not mandatory, and several thoughts from round one were returned to and expanded upon.

Six of seven participants identified both advantages and disadvantages (one participant responded that they felt unable to provide reflections). Perceived advantages were (a) the opportunity to clarify issues and questions across rounds, (b) the opportunity to prioritise risks in the second round, and (c) the integration of experts with different backgrounds to share and compare opinions. The latter advantage in particular was perceived as increasing the diversity and quality of the results. One expert was positive about the implemented method and reported finding it helpful for libraries, but did not provide further detail.

Several disadvantages and aspects that could be improved were identified by participants. Especially the first round was criticised by most of the experts for the same reason described above in the researcher observations: the first to respond could determine the direction of discussion and "*bias all other comments*". It was also criticised that this first person had to put the most effort into the answer (i.e. nothing to build upon) and that many participants have agreed more promptly with the previous speaker than writing their own answer, which is indeed observable in the raw data of the first round. This resulted in some redundant answers, which participants also mentioned as a disadvantage. Regarding the questions, it was lamented (a) that the questions were rather abstract and so more in-depth questions would have been helpful and (b) that there is not yet enough known about the "*background and connections of libraries and AI*", so answering the questions was challenging. Of course, that lack of knowledge also demonstrates the need for attempting to foresee ethical issues.

Discussion

What are the foreseeable, ethical risks/problems of AI for (semi-) automated content indexing in an academic library?

Here we discuss and interpret the results presented above, first along the three categories of risks identified (*data and bias*, *the AI system's influence on humans*, *future job profiles for LIS professionals*) and then with regards to evaluating EFA.

Finally, a short synthesis and then study limitations follow.

Data and bias

Problems with data and bias are fundamental to AI, applicable to most applications, and by now commonly known issues in AI ethics (Tredinnick & Laybats, 2020, p. 8). Such issues were also identified in the library-centred applications considered. Many of the foreseeable risks identified in the course of this study relate primarily to the bibliographic data on which the test AI application for (semi-)automated content indexing is based. This applies to the fully automated procedures for semantic enrichment of the discovery, too, which is also a part of the third subproject: the question of data origins and problems that affect the performance of an AI application are the same and were thus assessed as even more serious in the second case.

The core of many risks in this context consisted in discrimination or exclusion, which has been identified as one of six central risk areas in relation to AI-driven language models (Weidinger et al., 2022). Some issues named, like the technical problem of non-machine-readable data, may not initially seem like ethical problems but can nonetheless lead to such, such as under-representation of certain historical periods due to data gaps persisting from bibliographic data through to AI training data. So, under-representation as well as discrimination (e.g. of languages, social groups, cultures or gender that may occur when data come from historical contexts transporting values that are no longer justifiable today) are common risks associated with AI applications. Such risks also conflict with (a) the IFLA code of ethics, which states that social, ethnic or religious exclusion should be avoided and that libraries are committed to neutrality (Garcia-Febo, Hustad, Rösch, Sturges, & Vallotton, 2012) and (b) the libraries' claim to representation in their collections (Rösch, 2020, p. 304). Library AI is not novel in its potential for bias; for example, popular large language models are a case of AI where different kinds of bias are present (Stahl & Eke, 2024) and AI voice assistants use and reinforce stereotypical gender roles (Gebru, 2020, p. 260). Many of the risks of bias identified in the study are also well known in other library and information contexts (c.f. Weidinger et al., 2022).

Beyond existing bias in the world, it was also noted human decisions also influence bibliographic data and thus can lead to biased AI data and performance (Nayyer & Rodriguez, 2022, p. 167). Especially the question of who chooses what will be acquired, preserved, or discarded determines the resulting library collections, and thus it has been the subject of past discussion (Corrado, 2021, p. 397). Human decisions that have eventual influence on data curation are thus a notable risk: library collections used for AI data (e.g. subject indexing terms that may be

suggested) are based on historically shaped views by a society and its ethical values, which often change over time (Donath, 2020, p. 61). In some regions (e.g. German-speaking countries) such problems are acknowledged and attended to by regularly updated rules for cataloguing and indexing (Scheven, 2021, p. 93), but nonetheless a risk of negative effects of bias cannot be eliminated.

Indexing keywords change over the years though not only through unavoidable bias but also through new developments and deliberate review. If that procedure is automated, i.e. if the changes will be transferred to and influenced by automated indexing systems, is a human decision. Without human oversight of the reuse of terms, problematic terms can be replicated and thus persist (e.g. pejoratives used decades ago but now regarded as discriminatory; Lo, 2019).

There are a variety of technical problems that may get in the way of successfully training AI for indexing. A considerable volume of consistently formatted data is needed, but different indexing systems co-exist, and these individual systems have only indexed about 10 % of all publications (Albrecht, Block, Kratzer, & Thiessen, 2021, pp. 362–364). Modern acquisition practices such as the automated import of e-book packages and the accompanying automated import of associated metadata are susceptible to errors and contribute to the gaps in bibliographic data that serve as the basis for the keywords suggested by AI. According to the results in the second round, however, the aim is less to achieve data completeness and to train AI/ML models and more to have a small amount of digitised bibliographic data for certain time periods as examples. Due to the mass of data and the huge workload for librarians, external data transfer and the so-called re-use of subject indexing data can no longer be avoided in subject indexing and keyword allocation. This means that the issue of *data loss* and therefore the transfer of incomplete or incorrect data will continue to arise. And as mentioned earlier, the bias associated with such data gaps can eventually affect proposed keywords or discovery systems results (Albrecht et al., 2021, pp. 369–371).

The AI system's effects on humans

Further noteworthy risks of AI for (semi-)automated subject indexing and enriched discovery search identified in this study concern (a) possible negative effects of AI on people and who should be responsible and (b) influence an AI system might have on the work and actions of the person operating or interacting with it, such as the librarian in their decision-making or a patron in their search behaviour.

The long-standing information-ethical issues of privacy and accountability are reflected in the foreseeable risks identified (Kroll, 2020; Tredinnick & Laybats, 2020). AI complicates and makes more important the question of responsibility and prompts a need for *explainability*, as many different parties are involved in the technology's design, deployment, and use. In our study there was disagreement regarding the question of who is responsible if an AI leads to an unethical outcome: is it the institution who applies the AI system or the parties involved in producing or assembling the training data? The fact the disagreement is not easily settled reflects the state of understanding on such topics – even experts are unsure where the problem originates and who should be contacted if something goes wrong – as well as the still-developing state of regulation (Ridley, 2022, p. 7). Although the responsibilities of organisations or institutions should be articulated precisely (Kroll, 2020, p. 195), the results show that this is not always the case in practice (at least, not yet), since, on the one hand, precise documentation of responsibilities was already emphasised in the context of *Data and bias* and this apparently has not yet been enforced and, on the other hand, according to the experts the responsibility for ethical problems does not lie exclusively with the institution. The lack of explainability regarding influences on a system's performance can therefore be interpreted as producing uncertainty and thus potential ethical puzzles for its deployers (i.e. the academic library).

Furthermore the search strategies and behaviour of users may be influenced by the search terms, index keywords, and search results produced or suggested by an AI system: if users repeatedly use common keywords because they produce the desired or successful outcome (e.g. search results), one could argue that their search behaviour has been influenced. This influence may then be further perpetuated in systems designed to self-learn or *"adapt [their] behaviour based on external inputs and interactions with their environment"* (Mökander et al., 2021, pp. 6–7): influenced user inputs that optimise for the AI-generated keywords are then used to further promote those keywords, thus in practice further exaggerating initial biases that were present in the training data.

With regard to the librarian and their interaction with the AI application, by seeing keyword suggestions, librarian's decision making can also be significantly influenced, e.g. in an extreme case causing them to blindly accept keywords as suitable regardless of their suitability. Though it seems unlikely, the scenario is nonetheless worrisome as two key ethical principles of information work would be transgressed: professional integrity and information access (i.e. poor quality keywords do not facilitate access; Garcia-Febo et al., 2012).

Category 3: future job profiles for LIS profession-als

With regard to the question of whether AI poses a serious threat of radical change to the work of librarians, the answer visible in the discussion in the Delphi study would be *No*. First, the core tasks of librarians will not change directly as a result of AI deployment but rather there will be a chance in the kinds of technical knowledge required. A specific risk in this regard was not clear, but generally it seems the training of librarians will adapt to help them use new technologies to fulfil their core duties. Thus AI literacy and algorithmic literacy should be a central aim for librarians as well as library patrons (Archambault, 2024; Ng, Leung, Chu, & Qiao, 2021; Ridley & Pawlick-Potts, 2021). Integrating management of AI in information professionals' education is already demanded from other perspectives, for example by introducing librarians in interacting with AI and robotics, which is still rarely taught in current curricula (Tait & Pierson, 2022). Accordingly, the future job profile is more of a change than a threat, but as an academic library it is important to adjust accordingly (and cannot be taken for granted; Cox, 2023). Consequently, it is necessary to have a large number of employees who are familiar with artificial intelligence and can control and observe AI systems to deal with the many problems that were identified. A corresponding risk could therefore arise with regard to the fact that AI may require more work effort and working time of the librarians, although it is to be used for the reason of saving time. To go beyond the experts' views, AI ethics should also play a central role in the training of information professionals and librarians, and should familiarise future librarians with emerging ethical concerns (Borenstein & Howard, 2021).

What are the advantages and disadvantages of the delphi-powered Ethical Foresight Analysis for libraries?

After conducting the Delphi-powered EFA, some advantages and disadvantages emerged that could help determine to what extent this kind of method is appropriate for ethical audits in an academic library and indicate if it is effective at achieving ethical AI.

Regarding the efficacy of the Delphi-powered EFA, a central advantage of this approach was that various facets of risks and problems could be mapped from experts with different backgrounds and experiences, who could share different opinions as mentioned by one expert. This variety of views shows the range of risks that AI used in the library can bring, as many different research and professional fields are affected by the use of such a technology in this setting, and thus the experts from all these fields are aware of and consider different risks relevant. Developers and vendors in particular need to understand that the technologies they build are not free of ethical risks, which is why it is important to involve this group of participants from the field of AI, without a direct library connection (Borenstein & Howard, 2021, p. 63). Risks were indeed identified, and the possible risks that were identified in round one could be reviewed in detail, prioritised or clarified in round

two. Further, the experts' opinions on the individual problems mentioned in the panel were revealed separately, which was highlighted by the experts as well as by Floridi and Strait (2020, p. 80), who emphasised the importance of a prioritisation of new technology's "*harmful impacts*" in order to better situate them in the socio-political environment. This made it possible to classify or filter problems mentioned in round one, such as human decisions that influence the collection and thus also the AI data, or lack of transparency regarding the generation of AI data, as important problems for the library involved and to address them first. In contrast, other problems mentioned in round one, such as the possible danger of librarians being surveilled by the AI system, were classified as rather unrealistic and therefore less pressing.

Several possible upcoming risks and approaches to mitigate these risks were identified, as most of the experts from both AI development and project management already had experience from other AI projects or are part of professional associations and expert committees. The topic of AI ethics is also not new in libraries, as shown by the *Seven Principles for AI in Libraries* (Van Wessel, 2020) or "*eight ethics scenarios*" (Cox, 2022b). Thus a participant's statement that there is not enough knowledge about AI and libraries is partly true: most of the risks mentioned are known in other domains where AI is used. Thus the inclusion of various experts who contribute information that goes beyond the area of libraries and can combine this knowledge with the area of libraries is an advantage for the method and its use in the library environment for finding particular risks, but the approach also risks re-identifying known ethical issues rather than discovering novel or particular ones.

Some disadvantages limited the efficacy of the approach as implemented. As noted by Floridi and Strait (2020, p. 82), a Delphi study has the possible disadvantage that one expert has a completely different understanding of the technology investigated than another expert, which can undermine the "*effectiveness of any kind of insight from a crowdsourced analysis*". Very different views of the experts were indeed observed, especially with regard to the possible solutions. To put it one way, although EFA is designed to identify risks and this is strengthened by including diverse experts, doing so can produce disagreement and the method provides little guidance on how one should regard or interpret risks that (e.g.) only one expert finds serious.

Finally, although the implementation of a multi-round Delphi study does not entail notable financial costs (e.g. free tools can be used), participation is very time-consuming for the experts and managing the procedure and reviewing the results is time-consuming for the organisers. This ultimately requires working hours for everyone involved, which may be especially scarce for participating librarians who, as noted before, have long been struggling with an increased publication volume and potential backlogs (Corrado, 2021, p. 396). And until technologies like automated indexing truly relieve such time constraints (i.e. by freeing up time from manual indexing), it seems rare that library staff will have time to participate in a study like the Delphi-based EFA.

Synthesis

With the research questions answered (i.e. ethical risks and methodological advantages and disadvantages identified) we now consider if EFA is suitable as a method for ethical audits of AI in libraries. In short: it depends. Although a positive result was in principle possible (i.e. the method might have worked great overall), and although a number of risks were indeed identified, most risks were already wellknown issues of AI (or information or library practice), some expert disagreement is left unresolved, and the exercise was time-consuming for all involved. The use of established ethical issues of some technology to find the potential issues of some new application is a known ethical review approach (Stahl & Eke, 2024) and the content produced by the method describes how those issues might manifest in a library in particular, which may be useful (e.g. for convincing project leaders that biased training data is not a concern only for high-risk AI but also for libraries). But the extent to which *novel* issues are *foreseen* is difficult to judge in this case; data bias, harm, and an uncertain future are somewhat obvious risks of AI today (Dubber et al., 2020).

Yet despite the lack of a clearly positive result, it is difficult to say a Delphipowered EFA can *never* be suitable. The configurability of Delphi makes it difficult to infer with certainty if modified approaches would be more suitable; this depends very much on the implementation (Lund, 2020). A Delphi can be carried out in various ways that could attempt to account for the disadvantages, for example with more participants as in a classic Delphi, but that would optimise for consensus rather than exploration and risk identification.

Further, Floridi and Strait (2020) provide EFA in a way that leaves a lot of room for implementation. For this reason, the identified advantages and disadvantages are indicative but not conclusive for all possible AI implementations. Similarly to Delphi, EFA could be carried out in other ways, as there were also five other approaches for Ethical Foresight, like long-term studies such as the *Technical Assessment* (Floridi & Strait, 2020). Thus, this study only tested one way of conducting EFA, which, as already mentioned, is only suitable to a limited extent and not necessarily transferable to ethical audits in general in libraries. If risk identification is a priority, EFA as implemented may suffice. A less costly approach, such as a literature review, might also identify widely known and possibly applicable risks of AI. However, at least at this relatively early stage of deployment of AI in libraries, we believe it too would be unlikely to identify detailed, application-specific risks.

Lastly it should be considered if other approaches that were suggested by Bubinger and Dinneen (2021) might be more suitable to identify ethical issues in AI library. This includes, for example, the tool of a questionnaire for the developers of the AI application in order to prevent possible problems in the process of development (Lee & Singh, 2021). Another promising approach worth testing is the SMACTR framework (Raji et al., 2020) for conducting an end-to-end algorithmic audit, which might be more time-consuming but already provides necessary materials like principles, cases, and check lists. And since those methods were put forward several new methods have emerged but have not yet been applied to AI in libraries (Vetter et al., 2023); these too should be considered for their suitability.

Limitations

This study has a number of limitations, which were identified during the auditing process and are explained in the following. Notably, the data collected could have been more exhaustive; for example, participants sometimes only briefly agreed with each other rather than expanding, they did not justify every answer, and responses became shorter over time, indicating potential participant *fatigue* (Gallego & Bueno, 2014, p. 989). It is unclear if this was due to the exact questions asked or if any such implementation of EFA would face similar problems, which is in turn an obstacle to unconditionally recommending for or against similar uses of EFA; our conditional takeaways (above) stand nonetheless and suggestions for future research follow in the conclusion.

Second, it should be noted that the study did not seek to be representative and, following the Delphi approach to idea aggregation, only encouraged a gathering of ideas and the exchange among experts. Thus, the importance of the risks is may not be accurate for all library institutions, though this was not the aim of the study. Nonetheless, the risks identified should be considered by future similar evaluations or audits of AI in libraries and commonly identified risks should be identified in summarising research. Finally, as only ethical issues were sought, possible ethical *benefits* were not considered, but should nonetheless be included in an overall ethical evaluation of a technology.

Conclusion

This study followed a suggestion of the usefulness of applying EFA to

AI applications (Floridi & Strait, 2020) and took up a later proposal to then apply and evaluate EFA in a library setting (Bubinger & Dinneen, 2021). A Delphi-powered study was conducted to (1) identify foreseeable ethical risks of AI for (semi-)automated indexing in an academic library and (2) identify advantages and disadvantages of the approach. A modified two-round Delphi study was conducted, which is mainly based on the form of the Delphi survey for idea aggregation that followed a qualitative approach. Furthermore a quantitative element was used in the second round so that experts could assess the identified risks (i.e. severity or likelihood) of the AI application.

In round one of the study 21 potential risks were identified, which were either of direct ethical concern or else might nonetheless contribute or lead to ethical risks. These issues were grouped in three categories: *data and bias*, *the AI system's effects on humans* and *future job profiles for LIS professionals*. Since the data for the planned AI application is intended to be generated from bibliographic data, the study showed that the main concern is with biased data that could negatively affect the outputs of the AI. Various sources of potential bias were noted, including the origin of the various indexing systems and keywords that reflect potentially problematic values due to historical contexts and human decisions. It was noted that these risks are worsened by technical problems that lead to gaps in the data. The main risks of AI applications for (semi-)automated subject indexing thus lie primarily in the origin and complexity of the data on which an AI system is trained. Fewer problems were found with regard to the effects and influence of the system on humans, and the significance of these risks than in category one. Finally, risks regarding the impact of the AI application on librarians' professional profiles was similarly named but not regarded as particularly serious or pressing: the technology will not replace librarians in a significant way, rather more staff may be needed to control the systems considering the numerous issues previously mentioned.

A goal of the study was to consider if EFA is an appropriate auditing tool for AI in libraries, and several advantages and disadvantages were noted. Among the advantages were that EFA is inexpensive, does not involve a long-term study, and perhaps thanks to the varied expert perspectives does indeed identify library relevant ethical risks. The main disadvantages were (1) the time expenses for all parties involved and (2) that several of the identified risks were already known issues of AI in other contexts and so it arguably re-identified general risks rather than uncover or predict unknown ones. Therefore although the method seems better than *no* analysis or consideration of ethical issues of AI projects, which may be laden with ethical issues and thus deserve consideration, it is nonetheless a limited auditing tool. Alternative implementations of EFA *may* prove more efficient or effective, but from our results it seems lighter-weight approaches like a literature review might be more feasible and similarly effective.

The identification of risks is an important part of implementing new technologies in libraries and elsewhere, particularly considering how modern AI seems to entail varied, pervasive, and insipid ethical issues. Regardless of the method used, a project team dedicated to accounting for the ethical challenges of each AI project is therefore highly important (e.g. the team created for the Berlin State Library AI projects; Neudecker, 2022). However, considering the results of this study and the fact that most methods or tools that teams would use remain untested (Brundage et al., 2020; Bubinger & Dinneen, 2021), it is clear that further work is required to develop, identify, and achieve feasible methods and tools to support those teams.

CRediT authorship contribution statement

Helen Bubinger: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Jesse David Dinneen:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

None.

References

Albrecht, R., Block, B., Kratzer, M., & Thiessen, P. (2021). Quantität als Qualität – Was die Verbünde zur Verbesserung der Inhaltserschließung beitragen können. In M. Franke-Maier, A. Kasprzik, A. Ledl, & H. Schürmann (Eds.), *Qualität in der Inhaltserschließung (pp. 361–386)*. https://doi.org/10.1515/9783110691597-018. De Gruyter Saur.

AlgorithmWatch. (2022a). Leitbild. Retrieved 30th November 2022, from https://algorithmwatch.org/de/leitbild/.

AlgorithmWatch. (2022b). Projekte. Retrieved 30th November 2022, from https://algorithmwatch.org/de/projekte/.

Archambault, S. G. (2024). Toward a new framework for teaching algorithmic literacy. *Information and Learning Science, 125*(1/2), 44–67.

Bartlett, J. A. (2021). *Knowledge management: A practical guide for librarians*. Rowman & Littlefield.

Baruchson-Arbib, S., & Bronstein, J. (2002). A view to the future of the library and information science profession: A Delphi study. *Journal of the American Society for Information Science and Technology, 53*(5), 397–408. https://doi.org/10.1002/asi.10051

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big?. Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623).

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics, 1*, 61–65. https://doi.org/10.1007/s43681-02000002-7

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., Khlaaf, H., Yang, J., Toner, H., Fong, R., Maharaj, T., Koh, P. W., Hooker, S., Leung, J., Trask, A., Bluemke, E., Lebensold, J., O'Keefe, C., Koren, M., … Anderljung, M. (2020). Toward trustworthy AI development: Mechanisms for supporting verifiable claims. arXiv. http://arxiv.org/abs/2004.07213.

Bubinger, H., & Dinneen, J. D. (2021). Actionable approaches to promote ethical AI in libraries. *Proceedings of the Association for Information Science and Technology, USA, 58*, 682–684. https://doi.org/10.1002/pra2.528

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the 1st conference on fairness, accountability and transparency, USA, 81, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html.

Burr, C., Morley, J., Taddeo, M., & Floridi, L. (2020). Digital psychiatry: Risks and opportunities for public health and wellbeing. *IEEE Transactions on Technology and Society, 1*(1), 21–33. https://doi.org/10.1109/TTS.2020.2977059

Char, D. S., Abràmoff, M. D., & Feudtner, C. (2020). Identifying ethical considerations for machine learning healthcare applications. *The American Journal of Bioethics, 20*(11), 7–17. https://doi.org/10.1080/15265161.2020.1819469

Corrado, E. M. (2021). Artificial intelligence: The possibilities for metadata creation. *Technical Services Quarterly, 38*(4), 395–405. https://doi.org/10.1080/07317131.2021.1973797

Cox, A. (2022a). How artificial intelligence might change academic library work: Applying the competencies literature and the theory of the professions. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.24635

Cox, A. (2022b). The ethics of AI for information professionals: Eight scenarios. *Journal of the Australian Library and Information Association, 71*(3), 201–214. https://doi.org/10.5445/IR/1000082215

Cox, A. (2023). How artificial intelligence might change academic library work: Applying the competencies literature and the theory of the professions. *Journal of the Association for Information Science and Technology, 74*(3), 367–380.

Das, R. K., & Islam, M. S. U. (2021). Application of artificial intelligence and machine learning in libraries. *A Systematic Review. arXiv.*. https://doi.org/10.48550/arXiv.2112.04573

De Sarkar, T. (2023). Implementing robotics in library services. *Library Hi Tech News ahead-of-print.*. https://doi.org/10.1108/LHTN-11-2022-0123

Donath, J. (2020). Ethical issues in our relationship with artificial entities. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 51–73). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.3.

Dubber, M. D., Pasquale, F., & Das, S. (Eds.). (2020). *The Oxford handbook of ethics of AI. Oxford University Press*. https://doi.org/10.1093/oxfordhb/9780190067397.001.0001

Ehrmann, M., Hamdi, A., Pontes, E. L., Romanello, M., & Doucet, A. (2021). *Named entity recognition and classification on historical documents: A survey. arXiv*. https://doi.org/10.48550/arXiv.2109.11406

Floridi, L., & Strait, A. (2020). Ethical foresight analysis: What it is and why it is needed? *Minds and Machines, 30*(1), 77–97. https://doi.org/10.1007/s11023020-09521-y

Früh, R. (2018). Roboter in Bibliotheken und die flexible Ordnung von Sammlungen. *ABI Technik, 38*(1), 2–7. https://doi.org/10.1515/abitech-2018-0002

Gallego, D., & Bueno, S. (2014). Exploring the application of the Delphi method as a forecasting tool in information systems and technologies research. *Technology Analysis & Strategic Management, 26*(9), 987–999. https://doi.org/10.1080/09537325.2014.941348

Gantert, K. (2016). *Bibliothekarisches Grundwissen (9., vollständig aktualisierte und erweiterte Auflage)*. De Gruyter. https://doi.org/10.1515/9783110321500

Garcia-Febo, L., Hustad, A., Rösch, H., Sturges, P., & Vallotton, A. (August 2012). *IFLA code of ethics for librarians and other information workers*. International Federation of Library Associations. Retrieved 19th November 2022, from https://www.ifla.org/publications/ifla-code-of-ethics-for-librarians-and-otherinformation-workers-full-version/.

Gebru, T. (2020). Race and gender. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 252–269). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.16.

Häder, M. (2014). Delphi-Befragungen: Ein Arbeitsbuch (3). *Springer VS Wiesbaden.*. https://doi.org/10.1007/978-3-658-01928-0

Haffenden, C., Fano, E., Malmsten, M., & Börjeson, L. (2022). Making and using AI in the library: Creating a BERT model at the National Library of Sweden. *SocArXiv*. https://doi.org/10.31235/osf.io/k9duq

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

Jo, E. S., & Gebru, T. (2020). *Lessons from archives: Strategies for collecting sociocultural data in machine learning. Proceedings of the 2020 conference on fairness, accountability, and transparency, USA, 306–316*. https://doi.org/10.1145/3351095.3372829

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256019-0088-2

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns, 2*(9), Article 100314. https://doi.org/10.1016/j.patter.2021.100314

Korošec, M. R. (2020). Libraries and challenges for the future: Data mining and its purpose, benefits and meaning. Proceedings Western Balkan information and media literacy conference 2020, BosniaHerzegovina, 53–58. https://www.wbimlc.org/_files/ugd/3a3c2d_ea89dcfcf74745cb8ddbebd0e3e00eb1.pdf.

Krickl, M., Mayer, S., & Zangger, E. (2022). Mit Machine Learning auf der Suche nach Provenienzen – ein Use Case der Bildklassifikation an der Österreichischen Nationalbibliothek. *Bibliothek Forschung und Praxis, 46*(1), 227–238. https://doi.org/10.1515/bfp-2021-0090

Kroll, J. A. (2020). *Accountability in computer systems. In M. D. Dubber, F. Pasquale & S. Das (Eds.), The Oxford handbook of ethics of AI (pp. 180–196)*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.10

Lam, K., Iqbal, F. M., Purkayastha, S., & Kinross, J. M. (2021). Investigating the ethical and data governance issues of artificial intelligence in surgery: Protocol for a Delphi study. *JMIR Research Protocols, 10*(2), Article e26552. https://doi.org/10.2196/26552

Lee, B. C. G., Berson, I. R., & Berson, M. J. (2021). Machine Leanring and the social studies. *Social Education, 85*(2), 88–92.

Lee, M. S. A., & Singh, J. (2021). *Risk Identification Questionnaire for Detecting Unintended Bias in the Machine Learning Development Lifecycle. Proceedings of the 2021 AAAI/ACM conference on AI, Ethics, and Society, USA* (pp. 704–714). https://doi.org/10.1145/3461702.3462572

Li, J., & Wang, H. (2022). Application of artificial intelligence in libraries. In *2021 3rd international conference on artificial intelligence and advanced manufacture (AIAM), United Kingdom* (pp. 323–329). https://doi.org/10.1109/AIAM54119.2021. 00072

Jones, W., Capra, R., Diekema, A., Teevan, J., Pérez-Quiñones, M., Dinneen, J., & Hemminger, B. (2015). "For Telling" the present: using the delphi method to understand personal information management practices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, USA* (pp. 3513–3522). https://doi.org/10.1145/2702123.2702523

Lippincott, S. (2021). Mapping the current landscape of research library engagement with emerging Technologies in Research and Learning. Association of Research Libraries. https://www.arl.org/resources/mapping-the-currentlandscape-of-research-library-engagement-with-emerging-technologies-inresearch-and-learning/.

Lo, G. (2019). "Aliens" vs. catalogers: Bias in the Library of Congress subject heading. *Legal Reference Services Quarterly, 38*(4), 170–196.

Lund, B. D. (2020). Review of the Delphi method in library and information science research. *Journal of Documentation, 76*(4), 929–960. https://doi.org/10.1108/JD09-2019-0178

Lund, B. D. (2021). The fourth industrial revolution. *Information Technology and Libraries, 40*(1), 1–4. https://doi.org/10.6017/ital.v40i1.13193

Mankoff, J., Rode, J. A., & Faste, H. (2013). Looking past yesterday's tomorrow: Using futures studies methods to extend the research horizon. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, France, 1629–1638*. https://doi.org/10.1145/2470654.2466216

Mayring, P. (2015). *Qualitative Inhaltsanalyse (12., vollständig überarbeitete und aktualisierte Aufl.)*. Beltz Verlagsgruppe. Retrieved 3rd February 2023, from https://content-select.com/de/portal/media/view/552557d1-12fc-4367-a17f4cc3b0dd2d03?forceauth=1.

Millar, K., Thorstensen, E., Tomkins, S., Mepham, B., & Kaiser, M. (2007). Developing the ethical Delphi. *Journal of Agricultural and Environmental Ethics, 20*(1), 53–63. https://doi.org/10.1007/s10806-006-9022-9

Mökander, J. (2023). Auditing of AI: Legal, ethical and technical approaches. *Digital Society, 2*(49). https://doi.org/10.1007/s44206-023-00074-y

Mökander, J., & Floridi, L. (2021). Ethics-based auditing to develop trustworthy AI. *Minds and Machines, 31*(2), 323–327. https://doi.org/10.1007/s11023-021-09557-8

Mökander, J., Morley, J., Taddeo, M., & Floridi, L. (2021). Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. *Science and Engineering Ethics, 27*(44), 1–30. https://doi.org/10.1007/s11948-02100319-4

Munn, L. (2023). The uselessness of AI ethics. *AI and Ethics, 3*(3), 869–877.

Nayyer, K., & Rodriguez, M. (2022). Ethical implications of implicit Bias in AI: Impact for academic libraries. In S. Hervieux, & A. Wheatley (Eds.), *The rise of AI: Implications and applications of artificial intelligence in academic libraries (pp. 165–174)*. ACRL https://hdl.handle.net/1813/111207.

Neudecker, C. (2022). "Mensch.Maschine.Kultur" – Neues Projekt zu Künstlicher Intelligenz für das digitale Kulturelle Erbe. *SBB aktuell.* Retrieved 30th March 2022, from https://blog.sbb.berlin/mensch-maschine-kultur-neues-projekt-zurkuenstlichen-intelligenz/.

Neudecker, C., Zaczynska, K., Baierer, K., Rehm, G., Gerber, M., & Schneider, J. M. (2021). Methoden und Metriken zur Messung von OCR-Qualität für die Kuratierung von Daten und Metadaten. In M. Franke-Maier, A. Kasprzik, A. Ledl, & H. Schürmann (Eds.), *Qualität in der Inhaltserschließung* (pp. 137–165). De Gruyter.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, Article 100041.

O'Reilly, K. (2009). Key concepts in ethnography. London. 10.4135/9781446268308. Retrieved 2nd December 2022, from. https://sk.sagepub.com/books/keyconcepts-in-ethnography.

Østerlund, C., Jarrahi, M. H., Willis, M., Boyd, K., Wolf, T., & C.. (2021). Artificial intelligence and the world of work, a co-constitutive relationship. *Journal of the Association for Information Science and Technology, 72*(1), 128–135.

Oyelude, A. A. (2021). AI and libraries: Trends and projections. *Library Hi Tech News, 38*(10), 1–4. https://doi.org/10.1108/LHTN-10-2021-0079

Panda, S., & Chakravarty, P. R. (2021). *Implementing conversational AI in libraries: A practical approach. Impact of COVID-19 in academic institutions, ed. by S. Bhattacharjee. India, 124–145*. https://doi.org/10.5281/zenodo.4755976

Puig, A., & Adams, C. M. (2018). Delphi Technique. In *The SAGE encyclopedia of educational research, measurement, and evaluation* (p. 480). SAGE Publications, Inc.. https://doi.org/10.4135/9781506326139.n190

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., SmithLoud, J., Theron, D. & Barnes, P. (2020). Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on fairness, Accountability, and Transparency, USA, 33–44. doi: https://doi.org/10.1145/3351095.3372873.

Ridley, M. (2022). Explainable Artificial Intelligence (XAI). *Information Technology and Libraries, 41*(2). https://doi.org/10.6017/ital.v41i2.14683

Ridley, M., & Pawlick-Potts, D. (2021). Algorithmic literacy and the role for libraries. *Information Technology and Libraries, 40*(2), 1–15. https://doi.org/10.6017/ital.v40i2.12963

Rösch, H. (2020). Informationsethik und Bibliotheksethik: Grundlagen und praxis. *De Gruyter Saur.*. https://doi.org/10.1515/9783110522396

Ryan, M., & Stahl, B. C. (2021). Artificial intelligence ethics guidelines for developers and users: Clarifying their content and normative implications. *Journal of Information, Communication and Ethics in Society, 19*(1), 61–86. https://doi.org/10.1108/JICES-12-2019-0138

Sanji, M., Behzadi, H., & Gomroki, G. (2022). Chatbot: An intelligent tool for libraries. *Library Hi Tech News, 39*(3), 17–20. https://doi.org/10.1108/LHTN-01-2021-0002

Scheven, E. (2021). *Qualitätssicherung in der GND. In Qualität in der Inhaltserschließung (pp. 93–112). De Gruyter Saur*. https://doi.org/10.1515/9783110691597-006

Seeliger, F., Puppe, F., Ewerth, R., Koch, T., Kasprzik, A., Maas, J. F., … Greifeneder, E. (2021). Zum erfolgversprechenden Einsatz von KI in Bibliotheken: Diskussionsstand eines White Papers in progress – Teil 1. *BIT online, 24*(2), 173–178. https://www.b-i-t-online.de/heft/2021-02fachbeitraege.

Singh, V., Bilal, D., Cox, A., Chidziwisano, G. H., & Dinneen, J. D. (2023). Global AI initiatives: From theory to practice. *Proceedings of the Association for Information Science and Technology, 60*(1), 836–840.

Smith, C. (2021). *Automating intellectual freedom: Artificial intelligence, bias, and the information landscape. IFLA journal special issue: 20th anniversary of the IFLA statement on libraries and intellectual freedom*. https://doi.org/10.1177/03400352211057145

Souminen, O., Inkinen, J., & Lehtinen, M. (2022). Annif and Finto AI: Developing and implementing automated subject indexing. In G. Bergamin, & M. Guerrini (Eds.), *Bibliographic control in digital ecosystems* (pp. 265–282). Firenze University Press. https://doi.org/10.36253/978-88-5518-544-8.

Staatsbibliothek zu Berlin – Preußischer Kulturbesitz. Porträt. Retrieved 15th January 2023, from https://staatsbibliothek-berlin.de/die-staatsbibliothek/portraet.

Staatsbibliothek zu Berlin – Preußischer Kulturbesitz. (2020). *Bericht der Generaldirektorin für den Zeitraum November 2019 - Oktober 2020 (Auswahl an Ereignissen, Projekten und Themen)*. Retrieved 15th January 2023 from https://staatsbibliothek-berlin.de/diestaatsbibliothek/portrait/jahresberichte.

Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT–exploring the ethical issues of an emerging technology. *International Journal of Information Management, 74*, Article 102700.

Strasser, K., & Niedermayer, B. (2021). Unvoreingenommenheit von KünstlicheIntelligenz-Systemen. Die Rolle von Datenqualität und Bias für den verantwortungsvollen Einsatz von künstlicher Intelligenz. In R. Altenburger, & R. Schmidpeter (Eds.), *CSR und Künstliche Intelligenz* (pp. 121–135). Springer. https://doi.org/10.1007/978-3-662-63223-9_6.

Tait, E., & Pierson, C. M. (2022). Artificial intelligence and robots in libraries: Opportunities in LIS curriculum for preparing the librarians of tomorrow. *Journal of the Australian Library and Information Association, 71*(3), 256–274. https://doi.org/10.1080/24750158.2022.2081111

Tredinnick, L., & Laybats, C. (2020). Applied information ethics. *Business Information Review, 37*(1), 6–9. https://doi.org/10.1177/0266382120911260

Underwood, T. (2019). Distant horizons: Digital evidence and literary change. *University of Chicago Press.*. https://doi.org/10.7208/9780226612973

Université de Montréal. (2017). The Declaration - Montreal Responsible AI. Montreal Declaration for a Responsible Development of AI. Retrieved 30th November 2022, from. https://www.montrealdeclaration-responsibleai.com/the-declaration.

Van Wessel, J. W. (2020). AI in libraries: Seven principles. *National Library of the Netherlands.*. https://doi.org/10.5281/zenodo.3865344

Vecera, E. (2020). Künstliche Intelligenz in Bibliotheken. *Information - Wissenschaft & Praxis, 71*(1), 49–52. https://doi.org/10.1515/iwp-2019-2053

Vetter, D., Amann, J., Bruneault, F., Coffee, M., Düdder, B., Gallucci, A., Krendl Gilbert, T., Hagendorff, T., Halem, I. van, Hickman, E., Hildt, E., Holm, S., Kararigas, G., Kringen, P., Madai, V. I., Wiinblad Mathez, E., Tithi, J. J., Westerlund, M., Wurth, R., Zicari, R. V. & Initiative, Z.-I. (2023). Lessons learned from assessing trustworthy AI in practice. DISO 2(35). doi:https://doi.org/10.1007/s44206-02300063-1.

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., … Gabriel, I. (2022). *Taxonomy of Risks posed by Language Models. FAccT '22: 2022 ACM conference on fairness, Accountability, and Transparency, South Korea* (pp. 214–229). https://doi.org/10.1145/3531146.3533088

Zaiane, J. R. (2011). Global information ethics in LIS: An examination of select national library association English-language codes of ethics. *Journal of information ethics, 20* (2), 25.