Annotated Bibliography

Prepared for

Employing AI for Retention & Disposition in Digital Information and Recordkeeping Systems (AA01)

An InterPARES Trust AI Project

Patricia C. Franks, Researcher (SJSU)

Souvick Ghosh, Researcher (BSISDA & SJSU)

Alicia Butler, Graduate Research Assistant (SJSU)

May 20, 2022

Australian Government Department of Finance. (2021, July 6). *More about the digital records transformation initiative*. https://www.finance.gov.au/government/digital-records-transformation-initiative/more-digital-records-transformation-initiative

A timeline of the Australian Government Department of Finance's digital records efforts. Provides a follow-up to the study discussed in Rolan et al., 2019.

Authenticity Task Force. (2002). *Requirements for assessing and maintaining the authenticity of electronic records*. InterPARES. http://www.interpares.org/display_file.cfm?doc=ip1_authenticity_requirements.pdf

This document set forth the requirements that must be met to establish the authenticity of electronic records when they are being transferred from creator to preserver. It also described requirements that must be met to maintain the authenticity of electronic records once that authenticity has been established.

Baron, J. R. (2005). Toward a federal benchmarking standard for evaluating information retrieval products used in e-discovery. *Sedona Conference Journal*, *6*, 237–246. https://thesedonaconference.org/sites/default/files/publications/237-246%20Baron_237-246%20Baron.qxd__0.pdf

This article discussed the lack of a benchmark for evaluating electronic record search results during the e-discovery process. Baron (2005) outlined various search methodologies and described the variance between record recall and search precision rates during the search process. The article proposed the establishment of a benchmark for search processes and suggested that different software vendors be tested by an accredited standards body and compared to that benchmark.

Belovari, S. (2017). Expedited digital appraisal for regular archivists: an MPLP-type approach. *Journal of Archival Organization*, *14*(1–2), 55–77. https://doi.org/https://doi.org/10.1080/15332748.2018.1503014

This article described the author's experiment in digital archival records appraisal. The author utilized software and manual de-duplication methods, then manually previewed files and compared their contents to previously determined selection criteria. The detailed workflow reduced a collection from 677 GB to one-tenth of that size in the space of four days. The software used provided file analysis and de-duplication.

Bunn, J. (2020). Working in contexts for which transparency is important: A recordkeeping view of explainable artificial intelligence (XAI). *Records Management Journal*, *30*(2), 143–153. https://doi-org.libaccess.sjlibrary.org/10.1108/RMJ-08-2019-0038

Bunn (2020) examined explainable artificial intelligence (XAI) and how recordkeeping professionals can engage with it. The article pointed out that "the increasing use of more opaque AI techniques is generally framed as disruptive for recordkeeping" (Bunn, 2020, p. 144) and recommended that recordkeeping professionals are uniquely suited to help develop XAI models. Bunn (2020) reported on an interdisciplinary workshop organized by the author that focused on human-centered explainable AI and explored the human need for explanation. Workshop attendees expressed a desire for better public understanding of AI and proposed that the implementation of XAI could change the common metaphor of the black box to that of an iceberg. The article advocated strongly for interdisciplinary, exploratory conversations about AI and explainability, and recommended that recordkeepers help with XAI development by learning about AI and joining these conversations.

Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. *BMJ Quality & Safety*, *28*(3), 231–237. https://doi.org/https://doi.org/10.1136/bmjqs-2018-008370

Challen et al. (2019) explored artificial intelligence in the medical field. They discovered that “the bulk of research into medical applications of ML has focused on diagnostic decision support” (Challen et al., 2019, p. 231). Diagnostic decisions are decisions made to identify a patient’s ailment and make a decision on what to do for the patient. This process parallels the archival appraisal, retention, and disposition process, meaning that issues in medical AI are issues that may arise during the development and use of AI in archives. The article discussed how rules-based systems, supervised learning, and reinforcement learning are the most common forms of AI used and researched in the medical setting, and that research trends are evolving from reactive systems to more proactive autonomous systems (Challen et al., 2019, p. 232). They discussed issues that have arisen during the use of AI in healthcare, such as distributional shifts, a system’s insensitivity to the impact of decisions it makes, “black box” decision making, and predictions produced without confidence in accuracy (Challen et al., 2019, p. 234). Other issues include practitioners becoming complacent in their use of AI and giving more weight to the system’s predictions than their own, systems reinforcing outdated practices through an inability to adapt to new changes, and system implementation that "reinforces the outcome it is designed to detect" (Challen et al., 2019, p. 234). The authors then explored some theoretical issues with AI quality and safety that had been observed in test environments (Challen et al., 2019, p. 234). These included unintended negative side effects that resulted from a system performing a task without accounting for wider contextual information, "reward hacking" (Challen et al., 2019, p. 234), or the system finding an alternate method to achieve its reward without actually fulfilling its goal, exploration of new strategies in a manner that is not safe for patients, and implementation of or changes to a system that are not scalable (Challen et al., 2019, p. 234). The article then listed several questions to ask to facilitate the assessment and quality control of AI systems.

Colavizza, G., Blanke, T., Jeurgens, C., & Noordegraaf, J. (2022). Archives and AI: An overview of current debates and future perspectives. *Journal on Computing and Cultural Heritage*, *15*(1), 1–15. https://doi.org/https://doi.org/10.1145/3479010

Colavizza et al. (2022) presented a survey of recent literature concerning the intersection of Artificial Intelligence and archival theory and practice through the lens of the Records Continuum Model (Colavizza et al., 2022, p. 1). They explored the theoretical and professional considerations of archives and AI, including how AI affects archival theory, the transformation of archives from physical to digital spaces, and how that affects traditional appraisal processes and the profession at large. The article discussed how "the digital transformation has put pressure on archival concepts such as provenance and original order" (Colavizza et al., 2022, p. 5) and how archivists can leverage their expertise to inform AI development. The authors reviewed a number of publications surrounding the automation of recordkeeping processes and decisions, including appraisal, metadata, and the handling of sensitive information. More articles concern methods for organizing and accessing archives, automatic content extraction and indexation, alternative ways to read archival records, and tactics to improve search and retrieval. They explored novel forms of digital archives, and reviewed trends in the literature concerning the ethical use of AI and how it might be utilized to create a more inclusive and diverse archival record. Colavizza et al. discussed how AI is pushing archival principles to their limits, introducing a new dimension to the recordkeeping world, and noted the lack of discussion there appears to be regarding the limits and consequences of AI implementation. They also commented on how "there is ample room to design and develop AI-powered solutions to improve and enrich the way scholars can use archives" (Colavizza et al., 2022, p. 10). They noted that much of the literature on this topic focuses on the "organize" and "pluralize" dimensions of the Records Continuum Model, while there is little written on topics connected to "capture" and less for "create" (Colavizza et al., 2022, p. 10). They concluded by exploring areas where further work would benefit the archives and AI community, such as the creation of literature on transforming case studies and projects into long term practice, working on the ethical framework of AI to improve trust in AI systems, updating archival theory to be informed by AI developments, and archivists contributing to the development of AI to inform its creation with the principles of "provenance, appraisal, contextualisation, transparency, and accountability" (Colavizza et al., 2022, p. 11).

Conrad, J. G. (2010). E-discovery revisited: The need for artificial intelligence beyond information retrieval. *Artificial Intelligence and Law*, *18*, 321–345. https://doi.org/https://doi.org/10.1007/s10506-010-9096-6

This article defined and explored e-discovery with the goal of making the e-discovery field more available to AI and law researchers. The author explored the e-discovery process and provided several different examples of e-discovery in practice. The U.S. National Institute of Standards and Technology (NIST)'s Text REtrieval Conference (TREC) activities over the preceding four years were summarized, assessed, and critiqued. The author expounded upon the multidisciplinary nature of e-discovery and provided an e-discovery model designed to frame the process from a "technological perspective" (Conrad, 2010, p. 334). They continued on to explore trends among e-discovery service providers and their customers, revealing that customers have been tending to try to handle the e-discovery process on their own, and enterprises that manage the entire process from beginning to end sell better than those that handle only one aspect of e-discovery. Conrad went on to discuss several new technologies that they believed would benefit the e-discovery process. Intelligent relevance feedback, or "a partial release of relevant documents, followed by a second ''consultation,''' (Conrad, 2010, p. 337-338) could potentially substantially improve retrieval effectiveness. Conrad asserted that having computers respond to a query and then employing humans to review that output would be more effective than entrusting the entire inquiry to either humans or computers (Conrad, 2010, p. 338). Conrad also advocated for more effective email management, as, at the time of writing, "at least 50% of the material in today's E-Discovery environment is in the form of e-mail" (Conrad, 2010, p. 338). Natural language processing that includes “morphological analysis, ontologies, and named entity resolution” (Conrad, 2010, p. 339) could greatly simplify the email e-discovery process. The author also discussed the impact that social network analysis could have on the e-discovery process by enabling researchers to filter out “extraneous electronic content” (Conrad, 2010, p. 339) early on in the workflow, decreasing the amount of time spent analyzing content that is not relevant to the case. Machine learning techniques are also discussed, with Xerox’s CategoriX program as an example. CategoriX uses two ML models, one that learns from a set of data that has been “manually categorized by Subject Matter Experts (SMEs) using a pre-defined taxonomy” (Conrad, 2010, p. 339) then another predictive model that classifies a set of similar documents. An evaluation of CategoriX demonstrated that the system accurately identified more responsive documents and had a precision rate that was similar to human reviewers. The final technology Conrad recommended to be investigated was anticipatory e-discovery or methods that prepare an enterprise for the possibility of legal action and legal holds.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world: Don’t start with moon shots. *Harvard Business Review*, *January-February*, 2–10. https://www.kungfu.ai/wp-content/uploads/2019/01/R1801H-PDF-ENG.pdf

Davenport and Ronanki’s (2018) article explored the reasons behind the setbacks and failures of large-scale, ambitious AI projects and suggested a framework to implement that could help organizations successfully integrate AI into business processes. A study performed on 152 projects revealed that “highly ambitious moon shots are less likely to be successful than ‘low-hanging fruit’ projects that enhance business processes”

(Davenport & Ronanki, 2018, p. 4). For example, a cancer center's project to use AI to diagnose and treat patients was more costly and less successful than their project to use AI to help staff address IT problems. The article argued that starting small, taking an incremental approach, and focusing on augmenting human work rather than trying to replace it will yield better results (Davenport & Ronanki, 2018, p. 4). Davenport and Ronanki suggested a framework to follow when implementing AI solutions. First, it is important to understand the different technologies that exist, what each one does, and their strengths and weaknesses. Then, an organization should create a portfolio of AI-related projects they need or want to implement. They should identify areas of the business that could benefit from AI implementation, including bottlenecks in information flow, challenges in scaling information use, and areas where more computing power is needed to process gathered data. Once they've identified these areas of opportunity, they should evaluate cases where process improvement would "generate substantial value and contribute to business success" (Davenport & Ronanki, 2018, p. 8). Then, the organization should evaluate available technology to see if there's anything available that can complete the task needed. Once the organization has decided what project to implement, it should begin with a proof of concept pilot to test the project's actual efficacy and perform a business process redesign. Finally, the organization can scale up the project, spreading its use to the entire organization. The authors emphasized the importance of change management in this step, as employees may resist the project or feel threatened by AI, fearing displacement. Davenport & Ronanki provided guidance for any organization looking to implement AI, advocating for caution during project selection and suggesting a framework anyone can use to more effectively implement their AI solution.

Dixon Jr. (Ret.), J. H. B. (2021). Artificial intelligence: Benefits and unknown risks. *Judges' Journal*, *60*(1), 41–43. https://search-ebscohost-com.libaccess.sjlibrary.org/login.aspx?direct=true&db=a9h&AN=148239554&site=ehost-live&scope=site

Judge Dixon Jr. (Ret.) (2021) evaluated AI and its uses in the criminal justice system. The article discussed how AI is being used for e-discovery, predictive policing, solving crimes, and risk assessment. Judge Dixon examined the risks of AI bias in predictive policing and assessing the risk of recidivism (the likelihood that a person will commit a crime again once being released from custody). The article provided examples where AI models used for these purposes made incorrect and obviously biased decisions, especially in instances where race was a variable. The author concluded by calling for more carefully evaluating AI, its capabilities, and its appropriateness to a given task before model implementation.

Fosch Villaronga, E., Kieseberg, P., & Li, T. (2017). Humans forget, machines remember: Artificial intelligence and the right to be forgotten. *Computer Law & Security Review*, *34*, 304–313. https://doi.org/https://doi.org/10.1016/j.clsr.2017.08.007

Fosch Villaronga et al. (2017) examined how AI and the Right to Be Forgotten intersect. The authors performed a legal analysis of the Right to Be Forgotten, its history, and relevant definitions. They discussed legal controversies over the law and examined the technical details of deletion to determine if the Right to Be Forgotten works with AI. They concluded that "it may be impossible to fulfill the legal aims of the Right to Be Forgotten in artificial intelligence environments" (Fosch Villaronga et al., 2017, p. 304) and theorized that the disconnect between legal requirements and technical reality extends to other areas of privacy compliance and AI.

Franks, J. (2021). Text classification for records management. *Journal on Computing and Cultural Heritage*, *Just Accepted*. https://doi.org/https://doi.org/10.1145/3485846

This article described a study recently performed to determine what kind of natural language processing (NLP) technology is most effective to assist in the automatic classification of records. Experiments were conducted on authentic records data, each using a different text classification model. One model used term frequency-inverse document frequency (TF-IDF) and a support vector machine (SVM), three used different neural network architectures, and three others used different Transformer language models. The experiments found that "Transformer language models outperform both neural networks with no pre-training and statistical techniques on text classification tasks when tested against authentic records data" (Franks, 2022, p. 15). Based on the experiments described, the author concluded that it is reasonable to expect text classification tools to demonstrate skill of around 88% accuracy and 0.77 F1 (Franks, 2022, p. 16). The author iterated that classification is used in records management software most often to determine retention periods and disposition requirements or to identify sensitive information in records and that using AI and ML techniques can help records managers complete these tasks more efficiently (Franks, 2022, p. 2).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press. https://books.google.com/books?hl=en&lr=&id=omivDQAAQBAJ&oi=fnd&pg=PR5&dq=related:MHq4MMenr-gJ:scholar.google.com/&ots=MNQ-cnqHPZ&sig=r6rDVtNwWSC45emOUC4VMKI2NDY

This book is an introduction to machine learning concepts. It reviewed basic mathematical tools used in developing machine learning software, described several different deep learning algorithms, and explored some areas for further study and development. It was written with the assumption that readers "come from a computer science background" (Goodfellow et al., 2016, p. 12).

Harvey, R., & Thompson, D. (2010). Automating the appraisal of digital information. *Library Hi Tech*, *28*(2), 313–322. https://doi.org/https://doi-org.libaccess.sjlibrary.org/10.1108/07378831011047703

Harvey and Thompson (2010) investigated requirements for the automation of the appraisal and re-appraisal process for digital objects. They articulated that the main problems behind the inability to automate the appraisal process are the sheer volume of born-digital materials and the technical experience needed to manage them (Harvey & Thompson, 2010, p. 314). They approached appraisal as "part of the ongoing process of life-cycle management" (Harvey & Thompson, 2010, p. 314) and an essential aspect of responsible long-term collections management. Once an item is assigned a retention period or determined to be part of permanent holdings, the repository is responsible for ensuring its survival and accessibility. Re-appraisal enables the recordkeepers to evaluate an item’s risk for technological obsolescence (a significant threat to the survival of digital items) and act to prevent it. The article suggested “that the re-appraisal of technical aspects on an ongoing basis is a prime contender for some level of automation” (Harvey & Thompson, 2010, p. 317) and outlined a high-level framework for an automated re-appraisal process. The AI solution would first validate the file format of an object, then identify the version of the format. It would identify the application(s) needed to render the file, and (optionally) validate the file against a hash key (Harvey & Thompson, 2010, p. 318). If any of those steps failed, a technical failure is likely to have occurred and the program would alert the recordkeeper or another system to the issue. Advantages to automated re-appraisal include increased efficiency, the ability to notice issues sooner (providing increased time to respond to the issues), reliable processes (if the system was designed well), the ability to plan ahead, and increased capacity to properly manage larger collections. This approach is limited in that it cannot work entirely without human input, it only works with technical metadata, metadata created by the process may only be machine-readable, other systems need to be created to act on the information discovered by the re-appraisal process, and it has no value for short-term collections. Additionally, the authors raised the question “can an automated process that runs unattended be fully trusted?” (Harvey & Thompson, 2010, p. 319) They explored some requirements needed for automated technical appraisal, namely sufficient quantity and quality of metadata and additional systems or processes to act on the findings of the re-appraisal tool. They also raised the point that the cost and complexity of creating and implementing an AI re-appraisal tool are unknown and could provide a significant barrier to implementation.

They concluded by calling for more research into the practical application of their conceptual modeling.

Jimerson, R. C. (2007). Archives for all: Professional responsibility and social justice. *The American Archivist*, *70*(2), 252–281. https://doi.org/https://doi.org/10.17723/aarc.70.2.5n20760751v643m7

Jimerson (2007) discussed the power of information, archives, and archivists and explored how archivists can use that power responsibly to promote accountability and social justice. The article detailed how archives protect the rights of citizens, preserve cultural heritage, and have been used against marginalized communities in the past. Jimerson advocated that archivists need to embrace the power of information rather than deny its existence (2007, p. 254) and use that power for the public interest through promoting accountability, open government, social justice, and diversity. Archivists can do this by being objective (not neutral), being willing to take a stand against those who would abuse power, and examining personal and professional "assumptions, methods, and practices in light of the desired outcomes of justice and diversity" (Jimerson, 2007, p. 270). Additionally, archivists can use the power of archives to be a public advocates, resist pressures to alter systems or practices, draw attention to injustices, and speak out in defense of archival values and the rights of citizens. By committing to accountability and social justice, archivists can help create a more just society.

Jo, E. S., & Gebru, T. (2020). Lessons from archives: Strategies for collecting sociocultural data in machine learning. *FAT 2020 - Proceedings of the 2020 Conference on Fairness,*

Accountability, and Transparency, 306–316.
https://doi.org/https://doi.org/10.1145/3351095.3372829

Jo and Gebru (2020) examined the issues of fairness, accountability, transparency, and ethics related to the collection of datasets used to train machine learning (ML) systems and argued that this process should be informed by archival and library policies and practices (Jo & Gebru, 2020, p. 306). They advocated that since archivists and librarians have been managing collections for longer than ML professionals, ML processes could be improved upon by approaching data collection through an archival or library lens. The article explored the concepts of consent, inclusivity, power, transparency, ethics, and privacy. It then listed examples of actions that ML professionals can take to collect better quality datasets in a more ethical manner.

Katuu, S. (2020). Enterprise resource planning: Past, present, and future. *New Review of Information Networking*, *25*(1), 37–46.
https://doi.org/https://doi.org/10.1080/13614576.2020.1742770

This article by Katuu (2020) provided a general overview of enterprise resource planning (ERP) systems and analyzed current trends in ERP evolution. Katuu explored how ERPs can be both a concept (changes to an institution when a system is implemented) and a technology (the system itself) (Katuu, 2020, p. 39). ERPs began in the 1960s as inventory control (IC) systems, which, as the name implies, simply tracked inventory stocks and monitored usage. IC systems evolved into material requirements planning (MRP) systems in the 1970s, which had the added capacity to plan production utilizing a master schedule. MRPs evolved into manufacturing resource planning II (MRP II) systems in the 1980s, with a focus "on optimizing manufacturing processes by synchronizing material and production requirements" (Katuu, 2020, p. 40). In the 1990s, ERPs were developed to integrate different business processes (Katuu, 2020, p. 40). In the 2000s, ERPs evolved into a three-tiered system, and some moved to become cloud-based. These were referred to as extended ERPs, and in the mid-2010s they evolved into postmodern ERPs, which were "seen as more agile and outward-facing" (Katuu, 2020, p. 42), embracing RPA and AI technologies.

Katuu, S. (2021a). Trends in the enterprise resource planning market landscape. *Journal of Information & Organizational Sciences*, *45*(1), 55–75. https://doi.org/10.31341/jios.45.1.4

This article discussed enterprise resource planning (ERP) systems, the marketplace, and the impacts different technologies have had on their development. Katuu defined ERPs as "the integrated management of institutional activities mediated by technology" (2021a, p. 55) that are "designed to support and leverage the capabilities of the tools and processes used by an organization" (2021a, p. 56). The article explored existing literature on the ERP marketplace and concluded that market analyses are quickly outdated because of how quickly ERP software changes and how infrequently such analyses are made (Katuu, 2021a, p. 58). It then proceeded to evaluate four different technology trends and their impact on ERP software and the ERP market. The Fourth Industrial Revolution, or the increased automation of manufacturing and use of smart technology, is expected to rely heavily on the use of ERPs to continue to grow. ERPs are utilizing artificial narrow intelligence by integrating predictive inventory management, data analysis and processing, virtual assistants, chatbots, and predictive analysis models into their systems

(Katuu, 2021a, p. 63). They are shifting to be more cloud-based, and working on developing blockchain infrastructure (Katuu, 2021a, p. 65).

Katuu, S. (2021b). Managing records in enterprise resource planning systems. *IEEE International Conference on Big Data (Big Data)*, 2240–2245. https://doi.org/10.1109/BigData52589.2021.9672034

In this paper, Katuu explored "a multi-year ERP implementation project by the United Nations" (Katuu, 2021b, p. 2240) known as Umoja and highlighted some recordkeeping challenges and implications faced by the project. The project was launched in 2006 with the purpose of optimizing the U.N. Secretariat's workflows, methods for conducting business, and resource management (Katuu, 2021b, p. 2241). Katuu's analysis of external audit reports on the project revealed two main challenges faced by the project. First, employee master data (name, date of birth, beneficiary information) was often incomplete or incorrect. Second, users and past employees had access to information and power over processes they don't need. These two issues revealed an underlying problem of poor data quality, resulting in unreliable, inaccurate, and ultimately untrustworthy records. Katuu concluded by advocating for increased consideration of records management practices when making changes to ERP system management, stating that proper records management practices could help address the "challenges of project governance and management [and] issues related to the trustworthiness of records" (Katuu, 2021b, p. 2242).

Leavy, S., Pine, E., & Keane, M. T. (2017, August). *Mining the cultural memory of Irish industrial schools using word embedding and text classification*. Digital Humanities 2017 Conference, Montreal, Canada. https://dh2017.adho.org/abstracts/098/098.pdf

A research group utilized word embedding and text classification to analyze a 2,600-page long report and distill its findings into useable information. Segmenting the report into usable data entries, they created lexicons based on sets of "seed-words" (Leavy et al., 2017, p. 1). The researchers then ran an algorithm that utilized the lexicons to classify each data entry into one of three categories. The algorithm also identified and tagged names. This enabled researchers to better understand the lengthy report.

Lee, C. A. (2018). Computer-assisted appraisal and selection of archival materials. *IEEE International Conference on Big Data (Big Data)*, 2721–2724. https://doi.org/10.1109/BigData.2018.8622267

Lee (2018) discussed the appraisal of archival materials and how computers can be leveraged to assist archivists in the appraisal process. The article explored how the section and appraisal of digital materials differs from that of analog materials as "digital materials exist at multiple levels of representation" (Lee, 2018, p. 2721) and their inherent machine-readable nature makes it easier for users to identify patterns. Lee reviewed three types of technology that can be utilized to assist in archival appraisal. Digital forensics can be used to extract metadata from diverse collections and construct timelines from the extracted information. Natural language processing can be used to "capture and provide access to contextual information" (Lee, 2018, p. 2723), especially through named entity recognition. Machine learning tools can be utilized to automate classification and reduce the amount of time it takes to process a collection. Lee listed a few projects or publications that have explored each technology and concluded with a call to further research and develop technologies to enhance archival selection and appraisal.

Lepak, N. (2021, June 24). *What is artificial intelligence & why is it valuable for information management?* [Vendor]. Collabware. https://blog.collabware.com/what-is-artificial-intelligence-4-ways-to-take-advantage-of-ai-in-records-management

Lepak (2021) explained that AI is the process and result of teaching machines how to learn and make decisions. A machine or a program receives data, analyzes it against criteria provided to it by humans, then determines if that data fits the criteria or not, and proceeds to complete a task as directed. The article also identified different types of algorithms.

Luca, M., Kleinberg, J., & Mullainathan, S. (2016). Algorithms need managers, too. *Harvard Business Review*, *January-February*, 96–101. https://hbr.org/2016/01/algorithms-need-managers-too

In a Harvard Business Review article, “Algorithms Need Managers, Too” (Luca et al., 2016), the authors asserted that managers dealing with algorithms need to understand them better to make them more effective. They postulated that management requires making predictions and that “algorithms make predictions more accurate” (Luca et al., 2016, p. 97), advocating for the benefits algorithm use could provide to managers. They went on to caution that algorithms come with risks, as they don’t evolve automatically as people or circumstances change and can be too focused on one outcome to the exclusion of other priorities. To mitigate those risks, the authors said that "managers need to understand what algorithms do well—what questions they answer and what questions they do not" (Luca et al., 2016, p. 97). The article outlined core elements of algorithms that managers should understand. First, algorithms behave differently from humans. They are extremely literal and often provide predictions without being able to demonstrate the rationale behind those predictions (Luca et al., 2016, p. 98). To work around these differences, the authors said that managers should "be explicit about all your goals" (Luca et al., 2016, p. 99), include long-term outcomes in algorithm design alongside short-term goals, and carefully select input data. The article argued that by more closely understanding algorithms, managers in any field can utilize them more effectively.

Makhlouf Shabou, B., Tièche, J., Knafou, J., & Gaudinat, A. (2020). Algorithmic methods to explore the automation of the appraisal of structured and unstructured digital data. *Records Management Journal*, *30*(2), 175–200. https://doi.org/https://doi.org/10.1108/RMJ-09-2019-0049

This article detailed a research project with the goal of creating an archival appraisal tool that can identify and extract relevant data from a collection full of diverse formats and contents, then assist in decision-making based on the extracted data. The researchers created a list of variable data attributes and programmed software to assign a score to each item in a collection for each category of variable. The scores then provided a numerical value to the archivist representing that attribute's presence in a set of documents. For example, the root folder is being evaluated for metadata completeness and contains 13,179 files, 66.1% of which are complete, 30.8% are somewhat complete, and 3.1% have no metadata. That root folder has a metadata completeness score of 81% (Makhlouf Shabou et al., 2020, pp. 192-193). Archivists can then use the information gathered in the scores to make decisions.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6), 1–35. https://doi.org/https://doi.org/10.1145/3457607

Mehrabi et al. (2021) examined the issue of bias in machine learning. They explored examples of algorithmic unfairness in systems that demonstrate discrimination and analyzed types of bias in data, algorithms, and user experiences. The article then presented a cycle of bias in ML models. If a model's training data is biased, then the algorithm that trained on that data will be biased. That algorithm then produces a biased outcome, which influences user interactions with the model and creates more biased data. The article explored several definitions of fairness and concluded that "no universal definition of fairness exists" (Mehrabi et al., 2021, p. 11) but that "broadly, fairness is the absence of any prejudice or favoritism towards an individual or a group based on their intrinsic or acquired traits in the context of decision-making" (Mehrabi et al., 2021, p. 11). They then surveyed the literature on methods to utilize to make algorithms and machine learning operate more fairly.

Obukhov, A., Krasnyanskiy, M., & Nikolyukin, M. (2020). Algorithm of adaptation of electronic document management system based on machine learning technology [Abstract]. *Progress in Artificial Intelligence*, *9*, 287–303. https://doi.org/https://doi.org/10.1007/s13748-020-00214-2

Obukhov et al. (2020) created a software tool and related algorithm that could be utilized to alter and personalize the interface of electronic document management systems (EDMS). The algorithm formalized workflow processes, automatically adapted the EDMS interface to the user's needs, and assessed the system's capability to adapt (Obukhov, 2020). It automatically collected user preference data and utilized it to increase system flexibility. This resulted in users having a better first experience with the EDMS.

OECD. (2019). *Artificial intelligence in society*. OECD Publishing. https://doi.org/10.1787/eedfee77-en

This document presented an overview of AI and ML basics, including their history and the AI system lifecycle. It also proposed a taxonomy of topics for future study.

OECD. (2022). *OECD framework for the classification of AI systems*. OECD Publishing. https://doi.org/10.1787/20716826

This document presented a framework developed to be used to assess and characterize AI systems in order to promote understanding of how AI works, inform on its use, support industry-specific solutions, and facilitate risk assessment and management. It evaluated the impact of AI on five dimensions; people and planet, economic context, data and input, the model itself, and the system's tasks and output. The framework was tested by a number of stakeholders via a survey and was found to be most effective when applied to a specific solution, rather than a general type of technology.

Rendell, K., Koprinska, I., Kyme, A., Ebker-White, A. A., & Dinh, M. M. (2019). The Sydney Triage to Admission Risk Tool (START2) using machine learning techniques to support disposition decision-making [Abstract]. *Emergency Medicine Australasia*, *31*(3), 429–435. https://doi.org/10.1111/1742-6723.13199

A study was performed where several different types of prediction models were created and tested for accuracy in predicting where an Emergency Department patient would need care based on their presenting problem. This could translate to records management, as similar techniques might be able to determine a record's retention period based on its contents.

Rolan, G., Humphries, G., Jeffrey, L., Samaras, E., Antsoupova, T., & Stuart, K. (2019). More human than human? Artificial intelligence in the archive. *Archives & Manuscripts*, *47*(2), 179–203. https://doi.org/https://doi.org/10.1080/01576895.2018.1502088

Rolan et al. (2019) provided a snapshot of several Australian AI and recordkeeping initiatives. The Australian Public Record Office Victoria's (PROV) case study focused on appraisal and classification and revealed that e-discovery tools can be helpful in processing emails. The New South Wales State Archives (NSWSAR) case study explored a workflow using a Multi-Layer Perceptron algorithm that classified documents according to retention schedules, revealing a methodology that could be refined to help enforce retention periods for digital records. The National Archives of Australia's unfinished (in 2019) study focused AI implementation on the task of automatic disposal and retention authorizations to help humans to be more efficient, rather than trying to overhaul an entire program. Finally, the Australian Government Department of Finance explored options for creating its own AI system for managing records and ultimately selected a software-as-a-service product to fill its needs.

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1270

The National Institute of Standards and Technology (NIST) recently published a document that explored biases in artificial intelligence technology and provided guidance for addressing these biases with the goal of beginning a discussion that will lead to the creation of a NIST standard to help in this area (Schwartz et al., 2022). The authors explored the context and categories of AI biases, discussed how biases in AI can cause harm, and proposed the adoption of a socio-technological approach to AI creation and an updated AI lifecycle. The challenges to bias mitigation in AI they identified included features of datasets, testing and evaluation issues, and human factors (Schwartz et al., 2022, p. ii). The paper concluded with NIST's commitment to continue collaborating with the research community and other stakeholders to provide further socio-technical guidance on addressing bias in AI models (Schwartz et al., 2022, p. 48).

SOS Archivi. (2022, February 15). *What does AI look like when archival concepts inform its development?* [Webinar]. LinkedIn. https://www.linkedin.com/video/event/urn:li:ugcPost:6897112667836833792/

This webinar was a discussion of various AI projects and research related to archives. It provided terminology to utilize for the literature review.

Tanvir, Q. (2021, August 7). *Multi page document classification using machine learning and NLP*. Towards Data Science. https://towardsdatascience.com/multi-page-document-classification-using-machine-learning-and-nlp-ba6151405c03

An article by Qaisar Tanvir (2021) explored a multi-page document classification solution that could be utilized to circumnavigate bottlenecks in the mortgage industry. When mortgage companies perform mortgage loan audits they must analyze a loan package, which is a set of scanned pages that can be anywhere from around 100 to 400 pages long, containing sub-components that may range from one to around 30 pages (Tanvir, 2021). Analyzing these documents is generally outsourced and completed through a mixture of manual labor and semi-automation, generating questionably accurate results and taking a significant amount of time (Tanvir, 2021). This study was developed with the intent to create a document classification solution that would reduce the amount of human effort that goes into this process while increasing the accuracy of document analysis. The researchers focused on creating a solution that would identify the distinctions between different documents in the packet. First, the packet was split into individual pages, which were then processed through an optical character recognition tool and sent through a text vectorizer (they used Doc2Vec). Finally, the packet is run through a logistic regression classifier, where each page was tagged as the first page in a document, the last page in a document, or other (representing the middle pages) and assigned a confidence score for the selected category (Tanvir, 2021).

The National Archives. (2016). *The application of technology-assisted review to born-digital records transfer, inquires and beyond* (pp. 1–27). https://www.nationalarchives.gov.uk/documents/technology-assisted-review-to-born-digital-records-transfer.pdf

The National Archives of the UK conducted trials of e-discovery software and looked at additional research to test how the tools and processes could meet the challenges of born digital records. The research led the National Archives to conclude that e-discovery tools can “support government departments during appraisal, selection and sensitivity review” (The National Archives, 2016, p. 5). Lessons learned included that e-discovery tools can give a high-level understanding of an organization’s digital information, reduce the amount of information needed to be manually reviewed during the e-discovery process, and “to extract meaning from a large collection of born-digital records” (The National Archives, 2016, p. 17) through categorization, clustering, and email visualization processes. These solutions are also helpful in locating and redacting sensitive information. Researchers “found a mature eDiscovery market” (The National Archives, 2016, p. 21) with both well-established products and less-developed solutions with potential. They also learned that a solution’s “user interface is as important as the quality of the algorithm” (The National Archives, 2016, p. 22), and that coordination with information technology colleagues is vital to successful solution deployment. They concluded that there are increasing levels of confidence in the accuracy of e-discovery solutions and increased acceptance of the legality of e-discovery tool use.

The National Archives. (2021). *Using AI for digital records selection for government: Guidance for records managers based on an evaluation of current marketplace solutions*. https://cdn.nationalarchives.gov.uk/documents/using-ai-digital-selection-in-government.pdf

The National Archives evaluated five products for use as tools to help process government records. They outlined their findings and lessons learned, as well as general guidance for any other government agency to use when evaluating if they should implement AI.

Thomas, R. (2019). *The AI ladder*. O'Reilly. https://www.oreilly.com/online-learning/report/The-AI-Ladder.pdf

Thomas (2019) outlined the main challenges that prevent artificial intelligence implementation and presented a framework for the application of artificial intelligence solutions in any organization. The outlined challenges include a lack of understanding of AI technology, difficulty getting control of an organization's data, the lack of relevant skills in the workforce to administer AI, lack of trust in AI processes, and the difficulty of changing workplace culture and business models to include AI (Thomas, 2019, pp. 3-5). The AI Ladder is a framework that businesses can follow to successfully integrate AI into business processes. The first step is to collect the organization's data of all data types, and make it simple and accessible. Second, organize and catalog the data, evaluating its quality and making it accessible only to authorized users. Third, analyze the data by building, running, and managing transparent AI models. Fourth, infuse AI into operations across the entire enterprise. Through this entire process, modernize the organization by "building an information architecture for AI that provides choice and flexibility across the organization" (Thomas, 2019, p. 7). Implementation of this framework can help an organization to understand where they are with their AI initiatives and move forward to "a governed, efficient, agile, and future-proof" (Thomas, 2019, p. 7) use of AI technologies.

Turek, M. (n.d.). *Explainable artificial intelligence (XAI)*. Defense Advanced Research Projects Agency. https://www.darpa.mil/program/explainable-artificial-intelligence

Turek (n.d.) presented a research project to create AI solutions that are able to explain their decision-making rationale to users through a user interface. The article explored the inability of AI models to explain their output values to users and commented that this limits their effectiveness. Their project aimed to explore the psychology of explanation and develop AI solutions that would "have the ability to explain their rationale, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future" (Turek, n.d.). Turek advocated that explainable artificial intelligence (XAI) models will be more trustworthy and effective than existing models.

Vellino, A., & Alberts, I. (2016). Assisting the appraisal of e-mail records with automatic classification. *Records Management Journal*, *26*(3), 293–313. https://doi.org/https://doi.org/10.1108/RMJ-02-2016-0006

This article reported on a study that examined the methodology and decision-making process of eight information management professionals and then applied their processes to an AI system that included ML technology. The system successfully replicated the experts' processes with high levels of accuracy.

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, *July-August*, 2–11. https://hometownhealthonline.com/wp-content/uploads/2019/02/ai2-R1804J-PDF-ENG.pdf

Wilson and Daugherty's (2018) article discussed how AI can be utilized to improve business processes and explored the benefits of pairing AI with skilled workers. The authors observed that "artificial intelligence is transforming business—and having the most significant impact when it augments human workers instead of replacing them" (Wilson & Daugherty, 2018, p. 4). Humans assist machines by training them how to perform tasks, explaining machine output to other humans, and ensuring AI sustains safe and responsible functionality (Wilson & Daugherty, 2018, pp. 5-6). Machines help humans by amplifying our abilities by providing information, enabling us to interact with other humans in more effective ways, and augmenting human workers' abilities (Wilson & Daugherty, 2018, pp. 6-7). The article argued that "in order to get the most value from AI, operations need to be redesigned" (Wilson & Daugherty, 2018, p. 8). First, the organization determines an operation to improve. Wilson & Daugherty recommended looking for processes where the organization wants to improve flexibility, speed, scale, decision-making capabilities, or increase personalization (2018, p. 9). Then the organization works with stakeholders to develop a solution, implement, scale, and sustain it (Wilson & Daugherty, 2018, p. 8). The article also recommended five principles to follow to make the most of the human-machine dynamic in the workplace. The principles are: “reimagine business processes; embrace experimentation/employee involvement; actively direct AI strategy; responsibly collect data; and redesign work to incorporate AI and cultivate related employee skills” (Wilson & Daugherty, 2018, p. 5). The authors mentioned a survey conducted that found that the more of the principles an organization followed, the more effective their AI initiatives were, but gave no further details on the principles or the study (Wilson & Daugherty, 2018, p. 5). This article excellently explained how humans and AI complement each other, provided several demonstrations, and advocated for careful business process redesign to take advantage of the benefits of AI and human cooperation.