Literature Review

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

Abstract

This literature review seeks to answer the question of what has been written on how artificial intelligence (AI) and machine learning (ML) are being used for retention and disposition in Digital Information and Recordkeeping Systems. It provides a brief review of what artificial intelligence, machine learning, and algorithms are and how they work. The literature on AI and retention and disposition is then explored, utilizing ARMA International's Generally Accepted Recordkeeping Principles® (2019) and the Records and Information Management (RIM) Lifecycle as an organizational outline. Articles that are related to the Principles® revealed that AI models created or implemented to make decisions regarding records retention and disposition should be able to make decisions that are explainable and unbiased, demonstrate integrity, and be capable of complete compliance. Articles reviewed that related to the RIM Lifecycle demonstrated that appraisal and classification tools could be used or adapted to determine how long items or collections should be retained and alert recordkeepers to the end of that retention period. A model need not have been created specifically to complete retention and disposition tasks to be utilized for the purpose. From custom-built tools to commercial e-discovery and software-as-a-service tools, various artificial intelligence tools are being used or could be explored to aid in retention and disposition in Digital Information and Recordkeeping Systems.

Introduction

Artificial intelligence (AI) is increasingly present in daily life. From personalized advertisements in our browsers to self-driving cars, AI is growing in complexity, capability, and ubiquity. Machine learning (ML) is enabling AI models to learn and change based on experience, further advancing the systems and expanding their uses. This has led records and information management (RIM) professionals to explore how AI can be implemented in recordkeeping systems. Across the globe, RIM professionals, researchers, and commercial vendors are experimenting with AI and records management, learning what works and what doesn't, and sharing their findings. InterPARES seeks to understand and leverage AI to "support the ongoing availability and accessibility of trustworthy public records" (InterPARES Trust AI, 2021) through research, training, and institutional partnership. This study seeks to examine one aspect of InterPARES' larger query, and that is the question of AI and retention and disposition.

This paper is a review of the literature on AI and retention and disposition. First, we briefly explain what artificial intelligence and machine learning are and how they work and define the terminology utilized in this paper. Then, we explore the literature on the topic of AI and retention and disposition that relate to the Generally Accepted Recordkeeping Principles® (ARMA International, 2019). The Principles® are utilized here as an organizational mechanism through which the resources that discuss topics that apply to the entire lifecycle of records management and could therefore not be included in the final section of the paper are presented. We conclude by investigating how AI intersects with the Records and Information Management Lifecycle, mainly in the areas of Creation, Distribution and Use, and Retention and Disposition (Franks, 2018, p. 36).

Research Question

How are artificial intelligence (AI) and machine learning (ML) being used for retention and disposition in information and recordkeeping systems?

Artificial Intelligence and Machine Learning

Artificial intelligence is constantly evolving, and as a result, there is little consensus on what constitutes “artificial intelligence.” OECD (2019) stated that “there is no universally accepted definition of AI” (p. 3) before presenting the definition they believed to be the most accurate. They said that “an AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments” (OECD, 2019, p. 4). Thomas’ (2019) definition was similar, stating that “AI is an umbrella term for a family of techniques that allow machines to learn from data and to act on what they have learned rather than simply following rote instructions created by a programmer” (p. 2). Lepak (2021) provided a simpler definition, saying that “AI is simply teaching machines to learn so they can make decisions” (para. 5).

Machine learning is a less ambiguous topic. OECD (2019) defined machine learning as systems that “leverage statistical approaches to learn from historical data and make predictions in new situations” (p. 1) or “a set of techniques to allow machines to learn in an automated manner through patterns and inferences rather than through explicit instructions from a human” (p. 7). Goodfellow et al. (2016) described ML as AI systems with “the ability to acquire their own knowledge, by extracting patterns from raw data” (p. 2). Machine learning is a type of AI technology. Not all AI systems have ML components, but all ML systems are a type of AI.

Lepak (2021) also presented a very basic conceptualization of how AI systems typically work. Most AI models utilize algorithms, which are “step-by-step procedures, or set of instructions, that AI use to perform analysis against criteria before making its YES or NO decision” (Lepak, 2021, para. 10). A machine or a program receives data, analyzes it against criteria provided to it by humans (algorithms), then determines if that data fits the criteria or not and proceeds to complete a task as previously directed (Lepak, 2021, para. 6). OECD (2019) provided a more in-depth but still very high-level conceptualization. “The core of an AI system is the AI model” (OECD, 2019, p. 5) which represents the system’s environment, is guided by objectives, and measures performance. The model gathers information from the environment (either real or virtual) via sensors, then uses operational logic (algorithms) to provide an output based on the gathered information. It then uses actuators to make changes to the environment based on the decisions made by the operational logic (OECD, 2019, p. 3).

This is by no means an exhaustive analysis of artificial intelligence, machine learning, algorithms, or how they function. It is meant to provide a brief overview of definitions and functionality to provide background knowledge for the remainder of the research presented in this literature review.

The Principles®

The literature on artificial intelligence and retention and disposition included several resources that focused on aspects of records management that are present through the entire records lifecycle. These elements happened to coincide with three of the Generally Accepted Recordkeeping Principles® (ARMA International, 2019), Transparency, Integrity, and Compliance.

Transparency

The principle of Transparency states that “an organization’s business processes and activities, including its information governance program, shall be documented in an open and verifiable manner, and that documentation shall be available to all personnel and appropriate, interested parties” (ARMA International, 2019). One of the greatest challenges to the application of AI to recordkeeping processes, especially retention and disposition, is the black box issue, or “machine’s current inability to explain their decisions and actions to human users” (Turek, n.d.). AI models often make decisions in a manner that is undocumented and unexplainable. Another challenge faced by the AI and recordkeeping community is that of bias in AI. The decisions made by a biased AI model are affected by the system’s bias, which may be disguised if the model lacks transparency. Several articles have been written regarding the black box issue and bias in AI models, including the following.

Turek (n.d.) presented a research project to create AI solutions that can 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.

Bunn (2020) also examined 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, explaining that

> Here we knew that there was more below than the surface than we could immediately see, but we also had agency in choosing to look above or below the water line. In some cases, we might not care what was below the surface, but in others, it could be very important. (Bunn, 2020, p. 147)

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.

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 (2021) 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 upon 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 a person's 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.

Mehrabi et al. (2021) also 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.

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).

Jo and Gebru (2020) examined the issues of fairness, accountability, transparency, and ethics related to the collection of datasets used to train machine learning 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.

Decisions regarding the retention and disposition of records must be made in a manner that supports transparency, yet current AI processes do not support this. Turek (n.d.) and Bunn (2020) both explored the issue of opaque decision-making by AI systems and advocated for explainable artificial intelligence as a solution to the black box problem. Bunn (2020) specifically linked this issue to the records management field and proposed a solution. However, neither examined the impact that XAI may have on retention and disposition practices explicitly nor discussed how XAI could be implemented to improve retention and disposition decision-making practices. Judge Dixon Jr. (Ret.) (2021) evaluated AI and its uses in the criminal justice system, advocating that models should be more carefully evaluated to ensure they are appropriate to use for a given task before implementation. Mehrabi et al. (2021) and Schwartz et al. (2022) assessed bias in AI systems in general, explored how bias causes harm, and proposed solutions to mitigate this. Jo and Gebru (2020) specifically addressed bias in training datasets as an issue in creating biased models. Any models implemented to make decisions regarding records retention and disposition should consider both explainable and biased artificial intelligence.

Integrity

The principle of Integrity states that “an information governance program shall be constructed so the information assets generated by or managed for the organization have a reasonable guarantee of authenticity and reliability” (ARMA International, 2019). Much has been written on the integrity of digital records, but less has been written on how AI affects integrity. InterPARES’ Authenticity Task Force (2002) sets forth the requirements that must be met to establish the authenticity of electronic records when they are being transferred from creator to preserver and describes requirements that must be met to maintain the authenticity of electronic records once that authenticity has been established. These requirements apply to records that are managed or affected by AI models, yet nothing in the document speaks to how AI could be used to assist the process of establishing and maintaining record authenticity. Likewise, Katuu's (2021b) analysis of an enterprise resource planning (ERP) implementation project explores issues related to the management of records in an AI system. The project Katuu explored suffered from a lack of records management practices and resulted in unreliable, inaccurate, and untrustworthy records. Both resources underline how records management processes and practices can be applied to AI models to improve or maintain the integrity of retained records, but neither examines how AI models can be applied to recordkeeping processes and practices to do the same.

Compliance

The principle of Compliance states that "an information governance program shall be constructed to comply with applicable laws, other binding authorities, and the organization's policies" (ARMA International, 2019). Compliance is an important aspect of disposition as many records management laws and regulations provide requirements for how long records are to be retained or under what circumstances they may be (or must be) disposed of. 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.

Records and Information Management Lifecycle

The Records and Information Management Lifecycle is comprised of five stages, Creation, Distribution & Use, Storage & Maintenance, Retention & Disposition, and Archival Preservation (Franks, 2018, p. 36). The majority of the literature that was reviewed on artificial intelligence and retention and disposition fell into one of three stages, Creation, Distribution & Use, or Retention & Disposition.

Colavizza et al. (2022) presented a similar 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, and how the transformation of archives from physical to digital spaces affects traditional appraisal processes. The authors reviewed several publications surrounding the automation of recordkeeping processes and decisions, including appraisal, metadata, and the handling of sensitive information. Additional articles concerned 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. The article 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. Colavizza et al. 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” (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 having archivists contribute to the development of AI to inform its development.

Creation

One-quarter of the articles reviewed for this paper discussed AI and records creation, appraisal, and classification. While it may initially seem counterintuitive to include these resources in a review of AI and retention and disposition, it is important to remember that the records creation process has a direct effect on retention and disposition practices. Some AI technologies or processes have the potential capability to assign retention periods and disposition conditions to an item upon its ingest into a system. This means that, when discussing AI and retention and disposition, creation needs to be part of the conversation. In this context, the creation phase of the lifecycle includes the appraisal and ingestion of new items into a system or collection and the task of classifying those items.

Appraisal

Records appraisal is particularly important in the digital environment as keeping every item or collection forever is unsustainable. Belovari (2017) expounded that “appraisal means to evaluate the value and quality of content and, beyond identification, it actually involves selecting what should be preserved permanently and what should not be retained" (pp. 56-57). Appraisal enables records managers and archivists to direct their energy and expertise to the care of records with value. The question of how artificial intelligence can assist in this process is a fairly common one.

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 more effectively 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 little to 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 to make the proposed framework work, 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 model.

Belovari's (2017) article described the author's experiment in digital archival records appraisal. The author argued that the ease and speed of processing digital collections after they are appraised makes "certain traditional arrangement tasks unnecessary just as digital search functionalities may render many traditional descriptions redundant" (Belovari, 2017, p. 57). They tested ten different types of software and ultimately selected TreeSize Professional (TSP) as the most effective for their organization and purposes. They created a workflow for digital collections appraisal that utilized software and manual methods and performed the workflow on a collection at the State Archives Ludwigsburg. They began with a quick inspection, assessing the collection for risks. Then, they moved on to do a broad appraisal, manually deleting duplicate directories, using software to delete duplicate files, empty folders, and temporary or technical files. This process took a little more than an hour and resulted in the deletion of 70% of the collection's files. Finally, the author did an in-depth qualitative appraisal where they decided at what level of depth to appraise the collection, outlined criteria for what to keep, and viewed file thumbnails or previews to decide what to keep or delete. The detailed workflow reduced a collection of 677 GB to one-tenth of that size in the space of four days.

Lee (2018) also 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 (NLP) 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.

Makhlouf Shabou et al. (2020) undertook a research project intending to create an archival appraisal tool that could 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 researchers evaluated a root folder for metadata completeness. The file contained 13,179 files, and the appraisal tool identified that 66.1% of the collection had complete metadata, 30.8% had somewhat complete metadata, and 3.1% had no metadata, resulting in the root folder's overall metadata completeness score of 81% (Makhlouf Shabou et al., 2020, pp. 192-193). Archivists could use the information gathered and scores generated by the appraisal tool to make appraisal, retention, and disposition decisions.

The literature published on the intersection of records management or archives, artificial intelligence, and appraisal agrees that the appraisal process is becoming increasingly difficult as digital collections increase in size, complexity, and ubiquity. They also agree that AI could be leveraged to make this process easier. Harvey and Thompson (2010) proposed a high-level framework to automate the appraisal and re-appraisal process for digital objects and explored its potential benefits and limitations. Belovari's (2017) experiment with devising a workflow to utilize manual and AI methods to appraise a digital collection demonstrated the efficacy of such a solution. Lee (2018) reviewed specific technologies (digital forensics, natural language processing, and machine learning) and detailed how they could be leveraged to help with the appraisal process. Makhlouf Shabou et al. (2020) devised a tool that could extract data from collections and assign scores to individual files or entire collections. Their tool could save archivists time and assist in quick decision-making during the appraisal process. While improving the appraisal process is a significant objective, only Harvey and Thompson (2010) examined how AI appraisal tools could be used through the lifetime of collections, and even they did not consider how these tools could assist with retention and disposition workflows. If an appraisal tool can determine if an item or collection is ingested into a system, is it not in the realm of possibility that it could also determine how long that item or collection be retained and alert recordkeepers to the end of that retention period?

Classification

Classification can mean a couple of different things, however, for this paper classification is defined as “the process of assigning some thing to a specific class within a hierarchy, based on the thing's characteristics” (InterPARES Trust AI, 2018). The literature on classification, artificial intelligence, and records management or archives consists of case studies. These tend to focus on one solution or experiment and report on its efficacy or development.

The National Archives of the UK (TNA) (2021) examined how AI could be applied to records selection and evaluated five products for use as tools to help process government records. They detailed how a classification model with machine learning could be trained on data that had been manually classified by an expert, then apply what it learned at scale. They explained that the “most useful tools will report a measure of their confidence alongside the results of the classification” (The National Archives [TNA], 2021, p. 5) so that records managers can better understand where they need to intervene in the process before decisions are made based on the model’s output. The article presented some lessons TNA learned from the evaluation process, as well as general guidance for any other government agency to use when evaluating if they should implement AI. Appendix A outlined the steps of developing an AI classifier: data collection, exploratory analysis, feature engineering, model training and tuning, and production and deployment (TNA, 2021, p. 13). The article then presented TNA’s evaluation of Adlib Elevate, Amazon Web Services, Microsoft Azure, Iron Mountain’s InSight, and Records365 by RecordPoint. They reviewed each product’s data collection, pre-processing, and analysis methods, what features (metadata) the solution could extract from items, the models’ training and tuning process, and each graphic user interface. The article concluded that "although the project endeavoured to standardise evaluation it became clear that direct comparison across different products and approaches is difficult" (TNA, 2021, p. 15). The lessons learned regarding AI evaluation and selection could be utilized to guide an organization through the process so that they could select the best tool to fit their needs. The classification model development process could likewise be replicated by other institutions seeking to create their own classification AI model.

More recently, Franks (2021) wrote about a study they recently performed to determine what kind of natural language processing 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, 2021, p. 15). Based on the experiments described, the author concluded that it is reasonable to expect text classification tools to demonstrate around 88% accuracy and 0.77 F1 (Franks, 2021, 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, 2021, p. 2).

An article by 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 manual 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 (OCR) tool and sent through a text vectorizer (they used Doc2Vec). Finally, the packet was 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 resulting workflow produced predictions quickly, accurately, and with high confidence levels.

Vellino & Alberts (2016) performed a study that examined the email classification practices and decision-making process of eight information management professionals, then developed an AI model to replicate their processes. The authors interviewed the information management professionals and reviewed examples of email triage decision-making. From there, they created a “Model of Value Categories” (Vellino & Alberts, 2016, p. 300) that visualized the categories that the professionals apply to email collections. They then gathered two donated email collections to use as training data for the model they would create. The researchers manually labeled emails as either “Emails of Business Value” or “Emails of NO Business Value”

(Vellino & Alberts, 2016, p. 301) and balanced the sample data so that there were equal amounts of each category. Next, they extracted features from the emails such as the "To/From/Body/Attachment/Importance fields" (Vellino & Alberts, 2016, p. 302) and trained the Support Vector Machine (SVM) classifiers. Four models were created, two for each mailbox, one of each utilized the extracted features while the other did not. The models were "built with the LightSide Labs Researcher's Workbench (2015), an open-source text-mining tool that integrates the Apache OpenNLP" (Vellino & Alberts, 2016, p. 304). They validated the classifiers with "10-fold cross-validation" (Vellino & Alberts, 2016, p. 305) and tested the accuracy of each classifier against the mailbox it had not been trained on. Neither model was "able to detect 'Business Value' to any degree of accuracy on the other data set" (Vellino & Alberts, 2016, p. 306), so they merged the two mailboxes and trained a new model on the resulting collection. They compared the new model against a Spam detecting one and a randomly selected set of emails from Enron, learning that the SVM classifier was "slightly less accurate than the Spam/Ham SVM, they are nevertheless quite precise" (Vellino & Alberts, 2016, p. 306). Additionally, the system successfully replicated the experts' processes with high levels of accuracy. The authors concluded by reiterating the feasibility of their approach but pointing out the highly context-sensitive nature of a model created using this methodology (Vellino & Alberts, 2016, p. 309).

The Industrial Memories Project at the University College in Dublin, Ireland was a digital humanities group that utilized word embedding and text classification to analyze a 2,600-page long report and distill its findings into useable information (Leavy et al., 2017). The 2009 Ryan Report is "the report of the Irish Government's investigation into abuse at Irish Industrial Schools" from 1920 to 1990 (Leavy et al., 2017, p. 1). The report is 2,600 pages long, over

500,000 words, and documents a nine-year-long investigation (Leavy et al., 2017, p. 1). The researchers digitized the report and designed a "web-based exploratory interface" (Leavy et al., 2017, p. 1) with a relational database to enable search and analysis. Segmenting the report into usable data entries, the researchers identified and tagged names using a natural language processing toolkit. They created a set of categories to annotate entries with to further assist with discovery. The categories were created using automated text classification, a rules-based search, a random forest classifier, and manual methods. Researchers attempted a bag-of-words approach, but it "yielded results that were over-fitted due to the small samples of training data" (Leavy et al., 2017, p. 1). Instead, they used a word embedding algorithm on the training data to identify "seed-words" that were in turn used by Word2Vec to generate content-specific, semantic lexicons (Leavy et al., 2017, p. 1) for each knowledge category. The first category, "Movements of Staff and Clergy (Transfer Paragraphs)" (Leavy et al., 2017, p. 1) had to have their seed-words initially identified manually because of the obscuring language that was used in the report. The lexicon for the second category, "Witness Testimony (Witness Paragraphs)" was built on reporting verbs such as "said," "told," and "explained" (Leavy et al., 2017, p. 2). Seed-words for the third category, "Descriptions of Abusive Events (Abuse Paragraphs)" were more difficult to identify, so a support vector machine algorithm was used that ended up creating five lexicons to identify these paragraphs (Leavy et al., 2017, p. 2). Data entries were then classified under the appropriate knowledge category based on the created lexicons. A performance evaluation on random samples returned precision rates from 0.58 to 0.86, recall rates from 0.88 to 1.0, F-score rates from 0.73 to 0.88, and accuracy rates from 92% to 95% (Leavy et al., 2017, p. 3). The project has been made available to the public at https://industrialmemories.ucd.ie/ and demonstrated that machine learning used in conjunction with word embedding and content-specific lexicons can classify large documents, making them more accessible for use.

Most of the literature on classification, artificial intelligence, and records management or archives are case studies that focus on the efficacy or development of a particular solution. The National Archives of the UK (TNA) (2021) iterated the value of automated classification and evaluated five different AI solutions. They also presented guidance on selecting an AI solution for an organization or creating a classifier. Franks (2021) determined that Transformer language models outperform term frequency-inverse document frequency, support vector machines, and neural network architectures in effective natural language processing. Tanvir (2021) created a multi-page document classification solution that identified whether a given page was the first, middle, or last page of a document, resulting in a workflow that circumnavigated a tricky bottleneck in mortgage industry processes. Vellino & Alberts (2016) developed an SVM classifier that could replicate the decision-making process of an information professional when applied to a collection of emails. Leavy et al. (2017) demonstrated that an AI model that utilized machine learning, natural language processing, word embedding, and content-specific lexicons can classify large documents, making them more accessible for use and review. Some of these different methods of classification could be used or adapted to determine retention periods and disposition requirements.

Distribution and Use

Another significant portion of the literature concerning artificial intelligence and AI focuses on the distribution and use phase of the RIM lifecycle. Much like the creation phase, it may seem counterintuitive to include distribution and use in a discussion that is meant to focus on retention and disposition. However, how an item is used and moved during its lifetime can affect its evidential value, changing its retention period or ultimate disposition. Many of the following studies discuss solutions or workflows that could be explored, adapted, or altered to fit retention and disposition needs and workflows.

Obukhov et al. (2020) created a software tool and an algorithm that could be utilized to alter and personalize the interface of an electronic document management system (EDMS). This changed the ways users interacted with the EDMS, and by extension, how they interacted with records. The algorithm formalized different workflow processes, automatically adapted the EDMS interface to the user’s needs, and assessed the system’s capability to change (Obukhov et al., 2020). It automatically collected user preference data and utilized it to increase system flexibility. Obukhov et al. (2020) found that this resulted in users having a better first experience with the EDMS.

Baron (2005) discussed the lack of a benchmark for evaluating electronic record search results during the e-discovery process. The article outlined various search methodologies, including advanced boolean searches, statistical techniques, concept searching, natural language search, and fuzzy logic techniques, describing briefly how each works. They expounded that the problem that prevents many search tools from being efficient is the issue of balancing the number of records that are recalled by a given search with the relevance of those records. “The retrieval of large numbers of false positive unresponsive documents is certainly burdensome and vexatious; however, the failure to find responsive documents can be critical” (Baron, 2005, p. 243). A solution was proposed where a benchmark for search processes is established and software vendors who provide search functionality be tested by an accredited standards body and compared to that benchmark.

Conrad (2010) defined and explored e-discovery, intending to make 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 to explore trends among e-discovery service providers and their customers. This revealed 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 (2010) 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 substantially improve retrieval effectiveness. The author 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). They 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 were 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 (2010) recommended to be investigated was anticipatory e-discovery, which is a method that prepares an enterprise for the possibility of legal action and legal holds.

The National Archives of the UK (TNA) (2016) 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 TNA to conclude that e-discovery tools can "support government departments during appraisal, selection and sensitivity review" (The National Archives [TNA], 2016, p. 5). They learned 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 "extract meaning from a large collection of born-digital records" (TNA, 2016, p. 17) through categorization, clustering, and email visualization processes. These solutions can also help locate and redact sensitive information. Researchers "found a mature eDiscovery market" (TNA, 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" (TNA, 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 as well as increased acceptance of the legality of e-discovery tool use in the legal process.

The literature on how AI intersects with records distribution and use focuses primarily on document recovery and user interaction with records. Obukhov et al. (2020) wrote about a tool they created to alter and personalize the interface of an electronic document management system (EDMS) and found that its use resulted in users having a better first experience with the EDMS. Baron (2005) examined electronic records search methodology and proposed the creation of a benchmark for search processes. Conrad (2010) discussed e-discovery from a technological perspective and evaluated a model that they found to accurately and precisely identify more responsive documents. The National Archives of the UK (TNA) (2016) conducted trials of e-discovery software and reported that there are increasing levels of confidence in the accuracy and legality of e-discovery solutions. The lessons learned and models or algorithms developed in these case studies could be explored, adapted, or altered to fit the needs of retention and disposition workflows.

Retention and Disposition

There has been little written specifically on retention and disposition and artificial intelligence. Rolan et al.'s (2019) "More Human than Human? Artificial Intelligence in the Archive" is the exception. The authors presented several case studies where AI either has been or will be implemented in recordkeeping environments.

The first case study reviewed the Australian Public Record Office Victoria's (PROV) recently created a proof of concept that used AI to address the problem of email appraisal. They used a commercial e-discovery tool called Nuix on 1.5 TB of data to perform a technical appraisal of the composition of the collection used in the study. They then had Nuix identify and remove duplicate records, finding that "roughly 40% of the 4.6 million emails in the dataset" (Rolan et al., 2019, p. 189) were duplicates. Finally, PROV had Nuix assess and apply metadata elements for items in the dataset and evaluate if items should be retained or not. The evaluation was based on email addresses and domains, "partner agencies' role definitions; action verbs/objects; and function/activity terms" (Rolan et al., 2019, p. 189). PROV concluded that "the Nuix eDiscovery tool could effectively be used to reduce the volume of email needed to be analysed by PROV for appraisal" (Rolan et al., 2019, p. 190). This model's ability to dispose of duplicates is a capability shared by only one other model discussed above, Belovari's (2017) workflow. PROV's model was also unlike those listed in the "Appraisal" section of this article in that it applied rudimentary retention requirements.

Another case study explored by Rolan et al. (2019) was that of the New South Wales State Archives (NSWSAR) study conducted in 2017. The goal of the study was to explore the use of software to apply retention and disposal periods to a set of unstructured records and check its accuracy. The researchers evaluated software solutions to use for this experiment and were restricted to low-cost or free solutions due to the resource limitations of the project. They ultimately selected "scikit-learn, a free and open-source machine-learning library for the Python programming language" (Rolan et al., 2019, p. 191). The collection used for the experiment was "30 GB of data, in 7,561 folders, containing 42,653 files" (Rolan et al., 2019, p. 191) and had been manually assigned at the folder level to be retained by the State Archives. The researchers used only those files that text could be extracted from, such as PDF, DOC, and DOCX files for their study. They ran a text extractor and cleaned the data by removing formatting, stop words, unnecessary documents, and converting all text to lowercase (Rolan et al., 2019, p. 192). Next, researchers ran a text vectorizer and feature extractor, utilizing a bag-of-words approach (Rolan et al., 2019, p. 192). Finally, the researchers ran two classification algorithms to find out which would be more effective. Each algorithm was run on two copies of the data, one that had been cleaned and one that had not. Of the two algorithms, Multinomial Naïve Bayes and Multi-Layer Perceptron, it was discovered that “the Multi-Layer Perceptron algorithm with cleaned data was the most successful, with a maximum of 84% success rate” (Rolan et al., 2019, p. 192). The researchers concluded that, while an 84% accuracy rate may not be what they would want to see in regular operations, it was promising in the light of the fact that the study was relatively short and limited in its scope and resources. This case study revealed a workflow and specific algorithm that could be useful for assigning retention periods and may be worth further exploration.

Rolan et al. (2019) also discussed two case studies of projects that had not been completed yet at the time of the article’s publication. The National Archives of Australia was working on a research project to explore the creation and issuance of disposal and retention authorizations automatically or with minimum human involvement. The project was expected to evaluate various machine learning technologies, including auto-classification, clustering, and indexing tools (Rolan et al., 2019, p. 194). This project is different from many other case studies in its specificity on a single task that could be automated to save time for information managers, rather than a model or solution that could automate the entire process or a proof of concept that could be extended to include other tasks.

The other project was a proof of concept that was undertaken by the Australian Government Department of Finance to “test the application of microservices architecture and linked data technologies for automating records management” (Rolan et al., 2019, p. 195). The idea was to evaluate user records management needs, then create microservice tools to meet those needs and deploy them on Amazon Web Services. The proposed proof of concept was a model that would automatically assess the business value of captured emails and classify them according to a retention schedule. The study had not been completed at the time of Rolan et al.'s (2019) writing, but further research into the Australian Government Department of Finance's website revealed additional details on the study. The researchers ultimately concluded that

> While the Government is best placed to describe its functions, [the] industry is working towards automation and would be best placed to provide the digital records management systems that would be compatible with the government-developed Australian Government Records Interoperability Framework. (Australian Government Department of Finance, 2021)

The Australian Government Department of Finance ultimately selected a software-as-a-service product, "Records365 from Australian company RecordPoint" (Birmingham, 2021) to fill its needs. This implied that for this organization, the task of automating records management functions was complex enough that they decided to seek assistance.

The *Emergency Medicine Australasia* journal published an article about a study that was performed to predict where an Emergency Department patient would need care based on their presenting problem (Rendell et al., 2019). While this does not seem at first glance to apply to records management, the techniques and technologies used by the researchers could be adapted from predicting the disposition needs of a patient to predicting the disposition needs of a record. The study analyzed six classification algorithms and five feature selection techniques and evaluated each model based on the specificity of the presenting problem, comparing them to the existing model, the Sydney Triage to Admission Risk Tool (START). When the presenting problem was broad, the model with the "nearest neighbour algorithm with manual feature selection had the best area under the curve (AUC) of 0.8206 (95% confidence indicator [CI] ±0.0006)" while the model with the "decision tree with no feature selection had the best accuracy of 74.83% (95% CI ±0.065)" (Rendell et al., 2019). When the presenting problem was narrow, the model with the "nearest neighbour with information gain feature selection had the best AUC of 0.8267 (95% CI ±0.0006)" and the model with the "decision tree with wrapper or no feature selection had the best accuracy of 75.24% (95% CI ±0.064)" (Rendell et al., 2019). It is interesting to note that the algorithm that provided the best AUC in both cases was the nearest neighbor algorithm and the one that provided the most accuracy was the decision tree with no feature selection. A closer look into the researchers' methodology could reveal model elements that information professionals could explore when considering the retention and disposition of records and information.

Challen et al. (2019) also 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 their accuracy (Challen et al., 2019, p. 234). Other issues included 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. These questions and the issues uncovered in the medical field could be examined for applicability in records management and archives contexts.

While not as much has been written on the intersection of AI and retention and disposition as the intersection of AI and other aspects of the RIM lifecycle, several case studies demonstrate that the topic has been explored. Rolan et al. (2019) provided a snapshot of several Australian AI and recordkeeping initiatives. The Australian Public Record Office Victoria's (PROV) case study utilized an e-discovery tool to appraise emails and apply rudimentary retention requirements. 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 apply 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. Additionally, the Australian Government Department of Finance explored options for creating its own AI system for managing records and ultimately selected RecordPoint to fill its needs. Rendell et al. (2019) and Challen et al. (2019) both wrote about AI and the medical disposition process. Rendell et al.'s (2019) exploration of algorithms that could be utilized to determine a patient's long-term needs based on their presenting issues revealed nearest neighbor and decision tree algorithms as the best options. Challen et al.'s (2019) discussion of some of the issues encountered when AI is put into action resulted in several questions that could be asked to facilitate the assessment and quality control of AI systems. From e-discovery tools to specific algorithms and software-as-a-service, various artificial intelligence solutions are being used to aid in retention and disposition workflows in records management.

Conclusion

This literature review seeks to answer the question of what has been written on how artificial intelligence (AI) and machine learning (ML) are being used for retention and disposition in information and recordkeeping systems. We briefly reviewed what artificial intelligence, machine learning, and algorithms are and how they work, with explanations provided by Goodfellow et al. (2016), Lepak (2021), OECD (2019), and Thomas (2019). From there, we explored literature on artificial intelligence and retention and disposition, utilizing ARMA International's Principles® (2019) and the Records and Information Management Lifecycle as an organizational outline.

Several articles focused on aspects of records management that are present through the entire records lifecycle, coinciding with three of the Generally Accepted Recordkeeping

Principles® (ARMA International, 2019), Transparency, Integrity, and Compliance. It was revealed that, while decisions regarding the retention and disposition of records must be made in a manner that supports transparency, current AI processes do not support this. Turek (n.d.) and Bunn (2020) both explored the issue of opaque decision-making by AI systems and advocated for explainable artificial intelligence as a solution to the black box problem. Judge Dixon Jr. (Ret.) (2021), Mehrabi et al. (2021), Schwartz et al. (2022), and Jo and Gebru (2020) all reviewed the issue of bias in AI models and explored solutions that could be leveraged to mitigate the problem. Together, these articles demonstrated that any AI models implemented to make decisions regarding records retention and disposition should consider both explainable and biased artificial intelligence. InterPARES' Authenticity Task Force (2002) and Katuu (2021b) explored the problem of the integrity of AI models and concluded that records management processes and practices could be applied to AI models to improve their integrity and that of the records they manage. The ability to demonstrate a model's integrity is particularly important for models that perform retention and disposition tasks, as the choice to retain or dispose of an item is important and often irreversible. Fosch Villaronga et al. (2017) explored the issue of compliance and AI, raising the concern that it may not be possible for AI models to comply with particular privacy laws and theorizing that the disconnect between legal requirements and technical reality extends to other areas of compliance. This is particularly concerning with regard to retention and disposition as choices made to retain or dispose of items are often made to facilitate compliance with one or more laws, regulations, or policies.

The remainder of the articles reviewed each fell into one of three stages of the Records and Information Management Lifecycle: Creation, Distribution & Use, or Retention & Disposition. Those that discussed records creation focused on either records appraisal or classification. Harvey and Thompson (2010) discussed a framework to utilize to automate the appraisal process, while Belovari (2017), Lee (2018), and Makhlouf Shabou et al. (2020) each explored specific technologies to do the same. Belovari (2017) tested ten different types of software and ultimately selected TreeSize Professional (TSP) as the most effective for their organization and purposes, Lee (2018) advocated for the use of digital forensics, natural language processing (NLP), and machine learning technology, and Makhlouf Shabou et al. (2020) created their own tool that assigned a score to each item in a collection. The articles that reviewed records classification were case studies that focused on the efficacy or development of a particular solution. Some focused on commercial solutions. The National Archives of the UK (TNA) (2021) evaluated Adlib Elevate, Amazon Web Services, Microsoft Azure, Iron Mountain's InSight, and Records365 by RecordPoint, ultimately concluding that their differences precluded a standardized evaluation. Vellino & Alberts (2016) adopted LightSide Labs Researcher's Workbench (2015), a Support Vector Machine (SVM) that utilized Apache OpenNLP. Other articles reported on the authors' experiments in creating their own classifiers. Franks (2021) advocated for the use of Transformer language models in natural language processing, and Tanvir (2021) created their own classification model using optical character recognition (OCR) and Doc2Vec (an NLP tool). Leavy et al. (2017) created their own classifier using content-specific lexicons formulated by automated text classification, a rules-based search, a random forest classifier, manual methods, a word embedding algorithm, and Word2Vec (an NLP tool). It is possible that these appraisal and classification tools could be used or adapted to determine how long items or collections should be retained and alert recordkeepers to the end of that retention period.

The literature on how AI intersects with records distribution and use focused primarily on user interactions with records and document recovery. Obukhov et al. (2020) discussed alterations to a model's graphic user interface (GUI) and Baron (2005) advocated for a standardized electronic records search methodology. Conrad (2010) discussed how e-discovery tools could be improved by using intelligent relevance feedback, social network analysis, and NLP that leveraged morphological analysis, ontologies, and named entity recognition (NER). They also examined Xerox's e-discovery tool, CategoriX, finding that it accurately and precisely identifies more responsive documents than manual review. The National Archives of the UK (TNA) (2016) conducted trials of e-discovery software and reported that more effective e-discovery tools use categorization, clustering, and email visualization processes. The lessons learned and models or algorithms developed in these case studies could also be explored, adapted, or altered to fit the needs of retention and disposition workflows.

Articles that focus on the intersection of AI and retention and disposition are also typically case studies that focus on the implementation or evaluation of a particular tool or solution. Rolan et al. (2019) provided a snapshot of several Australian AI and recordkeeping initiatives. The Australian Public Record Office Victoria (PROV) selected a commercial e-discovery tool called Nuix and used it for email appraisal, including the disposition of extraneous emails. The New South Wales State Archives (NSWSAR) selected an open-source ML library called scikit-learn that utilized a text extractor, text vectorizer, a feature extractor with a bag-of-words approach, and a Multi-Layer Perceptron classification algorithm to apply retention and disposal periods to a set of unstructured records. The Australian Government Department of Finance tested Amazon Web Services' microservice tools for effectiveness in the automatic creation and issuance of disposal and retention authorizations but ultimately chose to select a commercial solution, Records365 from RecordPoint. Other articles on disposition and AI were written not from a recordkeeping perspective, but from the point of view of medical triage. Rendell et al. (2019) evaluated algorithms to utilize to determine long-term disposition needs and selected nearest neighbour and decision tree algorithms as the most effective. Challen et al. (2019) discussed the prevalence of rules-based systems, supervised learning, and reinforcement learning in medical disposition AI models and explored some of the benefits and drawbacks of their use.

There has been little written specifically on retention and disposition and artificial intelligence, and much on AI and other areas of recordkeeping. However, exploration of the literature on AI and records management has revealed that a model need not be created specifically to complete retention and disposition tasks to be utilized for the purpose. From custom-built tools to commercial e-discovery and software-as-a-service tools, various artificial intelligence tools are being used or could be explored to aid in retention and disposition in Digital Information and Recordkeeping Systems.

References

ARMA International. (2017). *The Principles®*. https://www.arma.org/page/principles

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

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

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

Birmingham, S. (2021). *Australian company raises the bar for digital data management*. Senator the Hon. Simon Birmingham Minister for Finance. https://www.financeminister.gov.au/media-release/2021/06/25/australian-company-raises-bar-digital-data-management

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

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

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

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

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

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

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

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

Franks, P. C. (2018). *Records and information management* (2nd ed.). Chicago, IL: ALA Neal-Schuman.

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&d

q=related:MHq4MMenr-gJ:scholar.google.com/&ots=MNQ-cnqHPZ&sig=r6rDVtNwWSC45emOUC4VMKI2NDY

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

InterPARES Trust AI. (2018). *Classification [English].* https://interparestrustai.org/terminology/term/classification

InterPARES Trust AI. (2021). *InterPARES Trust AI.* https://interparestrustai.org/trust

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

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

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

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

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

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

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

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

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

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

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

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

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

The National Archives [TNA]. (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 [TNA]. (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

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

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

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

Bibliography

Resources that were reviewed during this project that were not included in the final literature review, these articles can be referenced for more information on managing AI projects, Enterprise Resource Planning, and archives and social justice issues.

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

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

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

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

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

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

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